Document 1:

Chaque espèce d'êtres vivants se distingue des autres espèces par des caractères héréditaires dits caractères spécifiques. Les individus de la même espèce se distinguent les uns des autres par des caractères héréditaires qualifiés de caractères individuels. Certains de ces caractères sont externes, donc facilement observables ; on peut les classer en deux types:

- les caractères ***qualitatifs*** (comme la couleur de l'épiderme chez l'homme, la couleur des fleurs chez plantes, la forme de graine…). On peut facilement suivre la transmission de ses caractères de génération en génération, mais ils ne sont pas mesurables quantitativement.
- les caractères ***quantitatifs*** (comme le poids, la taille, nombre de gaines dans un fruit, nombre d'œufs pondus...). Ces caractères sont mesurables et peuvent être représentés par des nombres, on les qualifie de ***variables.*** On distingue deux types de caractères quantitatifs :
- les caractères quantitatifs à variation discontinue: les variables prennent des valeurs en nombres entiers positifs (nombre fini) , exemples : nombre de sépales d'une fleur, nombre de naissances par portée chez les femelles d'un mammifère …
- les caractères quantitatifs à variation continue: les variables peuvent prendre toutes les valeurs possibles (nombres illimités) à l'intérieur du domaine de variation ; exemples : poids des tubercules de pomme de terre, quantité du lait produit par une vache…

Document 2:

Pour étudier les caractères héréditaires quantitatifs, on utilise des méthodes mathématiques statistiques appliquées en biologie : c'est la biométrie. Pour un caractère donné chez une population donnée, on rassemble des données statistiques pour réaliser des représentations graphiques de ce qu'on appelle ***la distribution des fréquences***.

Dans le cas d'une variation discontinue, la représentation graphique des données statistiques se fait en deux étapes :

1: classement des données statistiques dans un tableau appelé tableau de la distribution des fréquences et qui comprend :
- Les variables : ensemble des valeurs que prend le caractère quantitatif chez les individus de la population étudiée.
- Les fréquences : nombre d'individus de la population qui ont la même valeur du variable.

2: Utilisation des données du tableau de la distribution des fréquences pour réaliser des représentations graphiques : le diagramme en bâtons, le polygone de fréquence et la courbe de fréquence.

a. ***Réalisation du diagramme en bâtons*** :
- tracer un axe de coordonnées et un axe d'abscisses ;
- porter les variables (xi) en abscisse et les fréquences (fi) en ordonnée.
- à chaque valeur xi de la variable, on fait correspondre un segment parallèle à l'axe des ordonnées et dont la longueur est proportionnelle à la fréquence fi correspondante.

b. ***Réalisation du polygone de fréquence***: on trace le diagramme en bâtons, puis on relie les points du sommet des trais verticaux par des segments de droite.

c. ***Réalisation de la courbe de fréquence***: on ajuste les contours du polygone de fréquence sans dépasser son domaine ; mais cet ajustement est difficile à faire de manière rationnelle.

- Si la courbe de fréquence (ou le polygone de fréquence) est unimodale, on déduit que la population étudiée est peut être homogène, ses individus appartiennent peut être à une même race (mais on n'est pas sûr).

En général, une courbe en cloche à symétrie axiale (appelée courbe de Gauss) indique que l'échantillon mesuré est suffisamment grand et représentatif de la population dont il est issu. La courbe de Gauss indique aussi qu'aucun facteur n'intervient dans la réalisation du caractère étudié.

- Si la courbe de fréquence (ou le polygone de fréquence) est plurimodale, on déduit que la population étudiée est obligatoirement hétérogène, elle comprend plusieurs races pour le caractère étudié.

Document 3:

A maturité, le Pavot forme un fruit appelé capsule (figure 1). Ce dernier est divisé en plusieurs loges par des cloisons appelés bandes stigmatiques. Pearson réalisa une étude statistique sur le nombre de bandes stigmatiques par capsule chez un échantillon comprenant 1927 capsules. Le tableau de la figure 2 présente les résultats de cette étude.

1. En justifiant la réponse, déterminez le type de variation étudiée.
2. En utilisant des couleurs différentes, représentez graphiquement la distribution des fréquences sous forme de diagramme en bâtons, de polygone et courbe de fréquence.
3. Que constatez-vous à partir de l'analyse de la courbe de fréquence ?

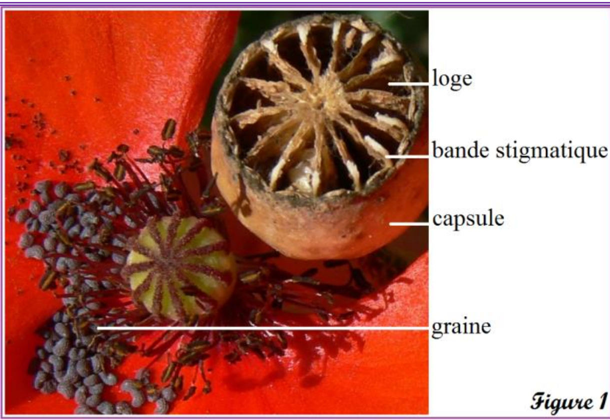


Figure 1

Variables (xi) : nombre de bandes	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Fréquences (fi): nombre de capsules	1	9	35	110	162	236	308	320	304	235	132	51	18	4	2

Figure 2 : distribution des fréquences du nombre de bandes stigmatiques chez 1927 capsules de Pavot

Document 4:

Dans le cas d'une variation continue, la représentation graphique des données statistiques se fait en deux étapes :

1: classement des données statistiques dans le tableau de la distribution des fréquences et qui comprend :

- Les variables : ensemble des valeurs que prend le caractère quantitatif chez les individus de la population étudiée. Ces valeurs sont réparties en classes d'intervalles égales.
- Les fréquences : nombre d'individus appartenant à la même classe.

2: Utilisation des données du tableau de la distribution des fréquences pour réaliser des représentations graphiques : l'histogramme de fréquence, le polygone de fréquence et la courbe de fréquence.

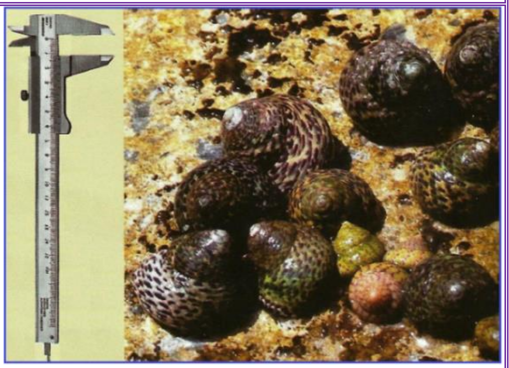
a. ***Réalisation de l'histogramme de fréquence***:
- tracer un axe de coordonnées et un axe d'abscisses ;
- porter en abscisse les classes des valeurs du caractère et en ordonnée les fréquences ;
- dessiner une suite de rectangles juxtaposés qui représente chacun une classe. La largeur de chaque rectangle correspond à l'intervalle de la classe qu'il représente, alors que sa hauteur de correspond à la fréquence de cette classe.

b. ***Réalisation du polygone de fréquence***:
- dessiner l'histogramme de fréquence ;
- relier successivement, par des segments de droite, les points correspondants aux milieux des côtés supérieurs des rectangles.

c. ***Réalisation de la courbe de fréquence***: ajuster les contours du polygone de fréquence sans dépasser son domaine.

Document 5:

Gibbule est un mollusque gastéropode marin à coquille spiralée conique tachée, très commun sur les côtes marocaines. On peut facilement mesurer le diamètre de sa coquille en utilisant un pied à coulisse (figure ci-contre).
Les mesures réalisées sur un échantillon de 375 d'individus ont permis d'établir le tableau ci-dessous. Ces mesures sont réparties en classes ; on a choisi ici des classes d'intervalles de 0,5mm.



1. En justifiant la réponse, déterminez le type de variation étudiée.
2. En utilisant des couleurs différentes, représentez graphiquement la distribution des fréquences sous forme d'histogramme et de polygone.
3. Que constatez-vous à partir de l'analyse des résultats ?

Variables (xi): diamètre des coquilles ($x10^{-1}$mm)	[115,120[	[120,125[	[125,130[	[130,135[	[135,140[	[140,145[	[145,150[	[150,155[	[155,160[	[160,165[
Fréquences (fi): nombre de coquilles	1	8	29	55	107	82	61	26	3	3

Document 6:

Les paramètres de position permettent, en général et de façon absolue, le positionnement des valeurs moyennes de la variable autour de laquelle se répartissent les autres valeurs. Les paramètres de position les plus utilisés sont le ***mode* M** et la ***moyenne arithmétique*** $\overline{\mathbf{X}}$

a. ***Le mode***

Dans le cas d'une variation discontinue, le mode traduit la valeur de la variable qui correspond à la fréquence la plus élevée. Pour une variation continue, le mode représente la valeur de la moyenne de la classe qui correspond à la fréquence la plus élevée.

b. ***La moyenne arithmétique***

La moyenne arithmétique d'une série statistique est la moyenne ordinaire, c'est-à-dire la valeur moyenne de la variable, c'est le rapport de la somme d'une distribution d'un caractère statistique quantitatif par le nombre de valeurs dans la distribution.

Pour calculer la moyenne arithmétique, on utilise la formule suivante:

$$\boxed{\overline{X} = \frac{\sum_1^i (fixi)}{n}}$$

$\overline{X}$: moyenne arithmétique $\sum_1^i$: somme fi : fréquence n : effectif total

xi : valeur des variables dans le cas de la variation discontinue ou des milieux des classes dans le cas de la variation continue

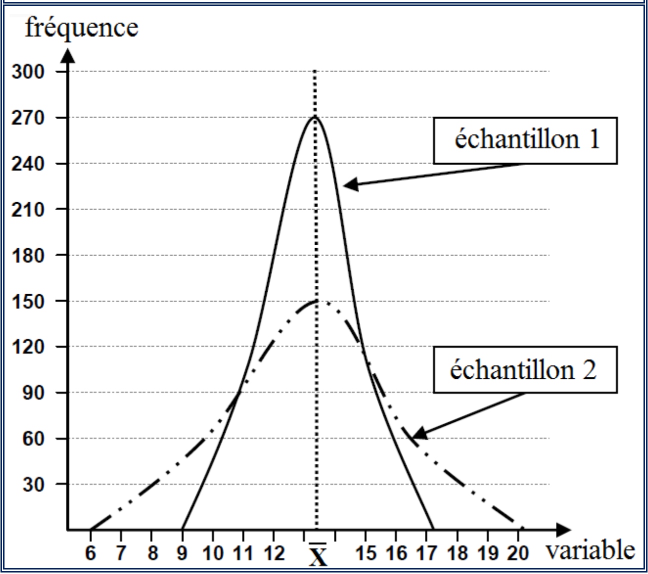
Exercice d'application

Déterminez à partir des données des documents 3 et 5 :

a. le mode et la moyenne arithmétique dans le cas de la variation du nombre de bandes stigmatiques de la capsule du Pavot.

b. le mode et la moyenne arithmétique dans le cas de la variation du diamètre des coquilles du Gibbule.

Document 7:



Document 11:

En 1903, W. Johannsen réalise une étude statistique sur la variation du poids des graines chez le Haricot.

* ***Première étape*** : Johannsen isole 1337 graines d'une population **P** et pèse chaque graine, il constate que le poids des graines prend des valeurs entre 20 et 90cg, il dispose alors les mesures en 14 classes dont l'intervalle est 5cg. Le tableau 1 représente les résultats obtenus.

Tableau 1 : étude statistique de la variation du poids des graines chez la population **P**

Classes	*21-25*	*26-30*	*31-35*	*36-40*	*41-45*	*46-50*	*51-55*	*56-60*	*61-65*	*66-70*	*71-75*	*76-80*	*81-85*	*86-90*
Fréquences	2	14	32	89	182	293	267	209	130	66	26	17	9	1

1. ***Déterminez***, en justifiant la réponse, le type de variation étudiée.
2. a: ***Représentez*** graphiquement la variation étudiée sous forme de polygone de fréquence.
   b: A partir de l'analyse du polygone réalisé, que peut-on supposer en ce qui concerne l'homogénéité de la population **P** ?

* ***Deuxième étape*** : Johannsen réalise une expérience de sélection artificielle, il isole les graines les plus légères de la population **P** (celles de la classe 21cg – 25cg) d'une part et les graines les plus lourdes (celles de la classe 86cg – 90cg) d'autre part. Il cultive ensuite ces deux classes séparément et il obtient deux sous populations : P_1 issue des graines légères et P_2 issue des graines lourdes. Johannsen réalise enfin une étude statistique de la variation du poids des graines chez les deux sous populations ; les résultats de cette étude sont présentés par les tableaux 2 et 3.

Tableau 2 : étude statistique de la variation du poids des graines chez la sous population P_1

Classes	*21-25*	*26-30*	*31-35*	*36-40*	*41-45*	*46-50*	*51-55*	*56-60*	*61-65*
Fréquences	2	7	18	23	20	16	10	5	2

Tableau 3 : étude statistique de la variation du poids des graines chez la sous population P_2

Classes	*36-40*	*41-45*	*46-50*	*51-55*	*56-60*	*61-65*	*66-70*	*71-75*	*76-80*	*81-85*	*86-90*
Fréquences	2	5	9	14	21	22	24	23	17	6	2

3. ***Représentez***, sous forme de polygones de fréquence, la variation du poids des graines chez les deux sous populations P_1 et P_2.
4. a: ***Déterminez*** les modes de **P**, de P_1 et de P_2.
   b: Que ***déduisez-vous*** en ce qui concerne l'homogénéité de la population **P** ?

* ***Troisième étape*** : Pour évaluer l'homogénéité des sous populations P_1 et P_2, Johannsen réalise sur chacune d'elle une opération de sélection artificielle semblable à celle effectuée sur la population **P**, il obtient à chaque fois une descendance ayant la même distribution et le même mode que la sous population d'origine.

5. A partir des résultats de la sélection artificielle effectuée sur les sous populations P_1 et P_2, que ***constatez-vous*** sur l'homogénéité de ces deux sous populations ?

Document 8:

Les paramètres de dispersion permettent de connaitre le degré d'homogénéité d'une population, d'évaluer l'amplitude d'une variation et de donner une idée sur la dispersion des valeurs autour de la moyenne. Ces constantes statistiques sont : l'écart moyen arithmétique **E**, la variance **V**, l'écart-type **σ** et le coefficient de variabilité **k**.

A. ***L'écart moyen arithmétique* E**

L'écart moyen arithmétique est la moyenne des écarts entre la valeur de chaque variable et la moyenne arithmétique $\bar{X}$. Il prend toujours une valeur positive et on le calcule en utilisant la formule suivante:

$$\boxed{E = \frac{\sum_1^i fi\left|xi - \bar{X}\right|}{n}}$$

E : écart moyen arithmétique $\sum_1^i$: somme fi : fréquence
xi : valeur des variables $\bar{X}$: moyenne arithmétique n : effectif total

Pour un caractère quantitatif donné, plus la valeur de l'écart moyen arithmétique est faible, plus les mesures sont groupées.

B. ***La variance* V**

L'écart moyen arithmétique **E** est une constante insuffisante pour expliquer la dispersion à cause des écarts négatifs, c'est pour cela qu'on mesure la moyenne des carrées des écarts appelée ***variance*** ou ***fluctuation***.

La variance **V** est la moyenne des carrées des écarts entre la valeur de chaque variable et la moyenne arithmétique $\bar{X}$. Pour calculer la variance, on utilise la formule suivante:

$$\boxed{V = \frac{\sum_1^i fi\left(xi - \bar{X}\right)^2}{n}}$$

V : variance $\sum_1^i$: somme fi : fréquence
xi : valeur des variables $\bar{X}$: moyenne arithmétique n : effectif total

C. ***L'écart-type* σ**

** L'écart-type **σ** est la racine carré de la variance. On le calcule en utilisant la formule suivante :

$$\boxed{\sigma = \sqrt{V} = \sqrt{\frac{\sum_1^i fi\left(xi - \bar{X}\right)^2}{n}}}$$

σ : écart-type **V** : variance $\sum_1^i$: somme fi : fréquence
xi : valeur des variables $\bar{X}$: moyenne arithmétique n : effectif total

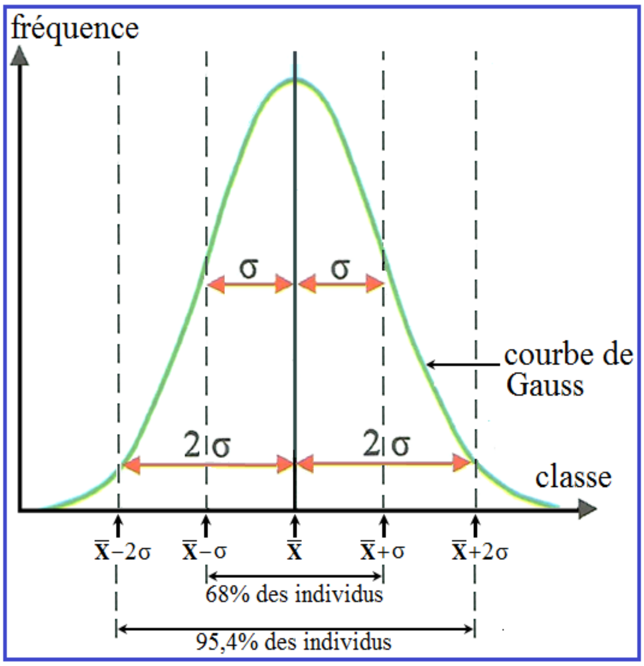
** ***Significations statistiques de l'écart-type*** :

* On utilise l'écart-type **σ** et la moyenne arithmétique $\bar{X}$ pour calculer l'intervalle de confiance qui prend les significations suivantes :

\+ Dans l'intervalle $[\bar{X} - \boldsymbol{\sigma} ; \bar{X} + \boldsymbol{\sigma}]$, on trouve 68% des individus de la population.

\+ Dans l'intervalle $[\bar{X} - 2\boldsymbol{\sigma} ; \bar{X} + 2\boldsymbol{\sigma}]$, on trouve 95,4% des individus de la population.

L'écart-type exprime la distribution réelle du variable surtout s'il s'agit d'une distribution normale qui correspond à la courbe de Gauss (figure ci-contre). Dans ce cas, les valeurs du variable chez 68% des individus de l'échantillon étudié se trouvent dans l'intervalle $[\bar{X} - \boldsymbol{\sigma} ; \bar{X} + \boldsymbol{\sigma}]$ et 95,4% dans l'intervalle $[\bar{X} - 2\boldsymbol{\sigma} ; \bar{X} + 2\boldsymbol{\sigma}]$.



* L'écart-type est une constante essentielle pour la comparaison de la dispersion du variable chez des échantillons de la même population ou appartenant à des populations différentes, plus la valeur de **σ** est faible, plus la dispersion est faible est vice-versa.

D. ***Le coefficient de variabilité* k**

Le coefficient de variabilité **k** représente la relation entre **σ** et $\bar{X}$, il permet de déterminer le degré d'homogénéité d'une population et la nature de la dispersion. On calcule le coefficient de variabilité k en utilisant la formule suivante :

$$\boxed{k = \frac{\sigma . 100}{\bar{X}}}$$

k : coefficient de variabilité **σ** : écart-type $\bar{X}$: moyenne arithmétique

- si **k** ≤ 15%, la population est homogène et la dispersion est faible ;
- si 15% < **k** ≤ 30%, l'homogénéité de la population est moyenne et la dispersion est aussi moyenne ;
- si 30% < **k**, la population est hétérogène et la dispersion est forte.

Document 9:

Les figures 1 et 2 représentent des tableaux d'application pour calculer certains paramètres statistiques de la distribution des fréquences :

- figure 1 : tableau d'application pour la variation du nombre de bandes stigmatiques de la capsule du Pavot (voir document 3).
- figure 2 : tableau d'application pour la variation du diamètre des coquilles du Gibbule (voir document 5).

1. a: ***Calculez*** pour chaque variation : la moyenne arithmétique, la variance, l'écart-type et le coefficient de variabilité.
   b: Que ***constatez-vous*** en ce qui concerne l'homogénéité et la nature de la dispersion des populations étudiées ?
2. Déterminez l'intervalle de confiance pour les populations étudiées.

xi	*fi*	*fi xi*	*xi - X*	$(xi - X)^2$	$fi\ (xi - X)^2$
6	1	6	- 6,77	45,8329	45,8329
7	9	63	- 5,77	33,2929	299,6361
8	35	280	- 4,77	22,7529	796,3515
9	110	990	- 3,77	14,2129	1563,419
10	162	1620	- 2,77	7,6729	1243,0098
11	236	2596	- 1,77	3,1329	739,3644
12	308	3696	- 0,77	0,5929	182,6132
13	320	4160	0,23	0,0529	16,928
14	304	4256	1,23	1,5129	495,9216
15	235	3525	2,23	4,9729	1168,6315
16	132	2112	3,23	10,4329	1377,1428
17	51	867	4,23	17,8929	912,5379
18	18	324	5,23	27,3529	492,3522
19	4	76	6,23	38,8129	155,2516
20	2	40	7,23	52,2729	104,5458
Somme	$n = \sum fi = 1927$	$\sum (fi\ xi) = 24611$	------------	------------	$\sum fi\ (xi - X)^2 = 9593{,}5383$

Figure 1

classe	*xi*	*fi*	*fi xi*	*xi - X*	$(xi - X)^2$	$fi\ (xi - X)^2$
[115 -120[	117,5	1	117,5	- 22,24	494,6176	494,6176
[120 -125[	122,5	8	980	- 17,24	297,2176	2377,7408
[125 -130[	127,5	29	3697,5	- 12,24	149,8176	4344,7104
[130 -135[	132,5	55	7287,5	- 7,24	52,4176	2882,968
[135 -140[	137,5	107	14712,5	- 2,24	5,0176	536,8832
[140 -145[	142,5	82	11685	2,76	7,6176	624,6432
[145 -150[	147,5	61	8997,5	7,76	60,2176	3673,2736
[150 -155[	152,5	26	3965	12,76	162,8176	4233,2576
[155 -160[	157,5	3	472,5	17,76	315,4176	946,2528
[160 -165[	162,5	3	487,5	22,76	518,0176	1554,0528
Somme	-------	$n = \sum fi = 375$	$\sum (fi\ xi) = 52402{,}5$	---------	-----------	$\sum fi\ (xi - X)^2 = 21668{,}4$

Figure 2

Document 10:

Deux échantillons (A et B) de pommes de terre sont récoltés dans deux champs différents (figure 1). On mesure pour chaque échantillon le poids en g des tubercules ; les résultats des mesures sont présentés par les tableaux de la figure 2 (les mesures sont disposées en classe d'intervalles de 10g). La figure 3 représente les histogrammes et les courbes normales de distribution de fréquences des deux échantillons de pomme de terre.

1. ***Déterminez***, en justifiant la réponse, le type de variation étudiée.
2. A partir de la figure 1, ***comparez*** la répartition des fréquences des deux échantillons étudiés.
3. a: ***Calculez*** pour chaque échantillon la moyenne arithmétique, l'écart-type et l'intervalle de confiance $[\bar{X} - \boldsymbol{\sigma}\,;\ \bar{X} + \boldsymbol{\sigma}]$.
   b: ***Comparez*** les résultats obtenus et ***déduisez*** quel est l'échantillon le plus homogène. ***Justifiez*** votre réponse.

Figure 1

Echantillon A

Classes	[35 – 55[	[55 – 75[	[75 – 95[	[95 – 115[	[115 – 135[	[135 – 155[	[155 – 175[
Fréquence	4	14	21	45	64	85	68

Classes	[175 – 195[	[195 – 215[	[215 – 235[	[235 – 255[	[255 – 275[	[275 – 295[
Fréquence	62	49	24	14	8	3

Echantillon B

Classes	[115 – 135[	[135 – 155[	[155 – 175[	[175 – 195[	[195 – 215[	[215 – 235[	[235 – 255[
Fréquence	24	45	73	92	83	38	22

Figure 2

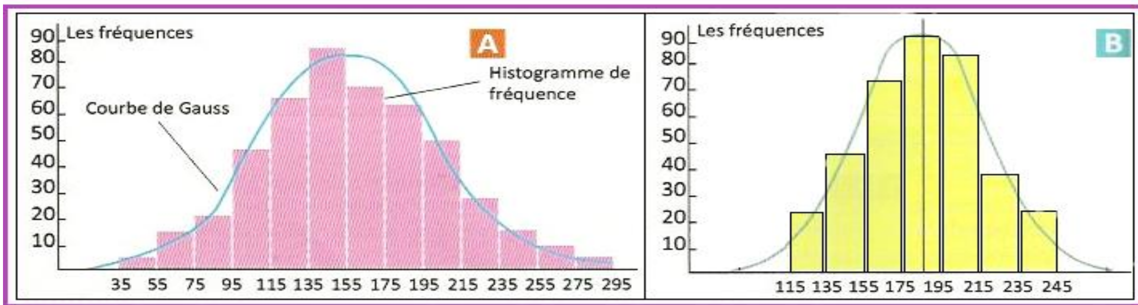


Figure 3