L'étude quantitative de la variation : La biométrie

Introduction

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.

La biométrie est la branche de la biologie qui s'intéresse à l'étude de la variation quantitative en appliquant des méthodes mathématiques et statistiques dans le but d'expliquer la distribution des caractères quantitatifs.

……

I-Notion de variation

1 –Notion des caractères quantitatifs

Pour étudier les caractères héréditaires quantitatifs, on utilise des méthodes mathématiques statistiques appliquées en biologie : c'est la biométrie.
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..

2 – L'étude de la variation discontinue

On parle d'une variation discontinue lorsque la variable prend des valeurs exprimées par des nombres entiers. $\boxed{X \in \mathbb{Z}}$

➢ Représentation graphique de la distribution des fréquences

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.

① Exemple de la variation discontinue: (Voir document 1)

On considère une population P de petits mammifères (souris), d'un élevage, tous de même espèce, ayant des poils noirs, courts et frisés ; on relève soigneusement dans cette population le nombre de jeunes mis au monde par les femelles à chaque portée ; pour 100 portées. Le tableau suivant présente les résultats :

Nombre des nouveau-nés : (x_i)	1	2	3	4	5	6	7	8	9
Nombre des femelles : (n_i)	2	8	12	16	23	18	10	7	1

1) **Dégager** du document la variable étudiée.
2) **Déterminer** le type de variation étudiée. Justifiez votre réponse
3) **Représenter** graphiquement la distribution des fréquences sous forme de diagramme en bâtons et polygone de fréquences.

Correction

② Exercice d'application: (Voir document 2)

A maturité, le Pavot forme un fruit appelé capsule (Figure 1). Ce dernier est divisé en plusieurs loges par des cloisons appelés bandes stigmatiques. On réalise 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 :

Figure 1

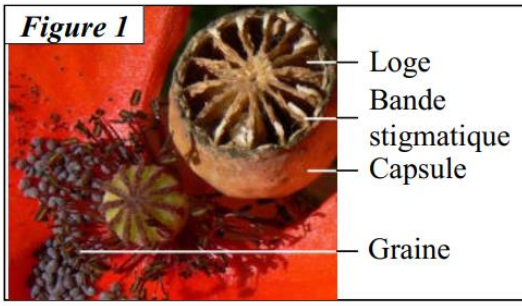


Figure 2

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

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 ?

3– L'étude de la variation continue

On parle d'une variation continue lorsque la variable peut prendre toutes les valeurs possibles dans un intervalle donné. $X \in \mathbb{R}$

➢ Représentation graphique de la distribution des fréquences

➢ 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 représentationgraphiques: 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

① Exemple de la variation continue: (Voir document 3)

Le tableau suivant montre la longueur des gousses chez le petit pois:

Variables (x_i) : classes de longueur en mm	40 - 49	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99	100 - 109
Effectif (n_i) : nombre de gousses	7	47	179	172	79	49	9

1) **Déterminer** le type de variation étudiée. Justifiez votre réponse
2) **Réaliser** l'histogramme et le polygone de fréquence de la distribution de la longueur des gousses.

....

....

....

② Exercice d'application: (Voir document 4)

Le Forficule ou Perce-oreille est un insecte de petite taille très répandu et inoffensif. Il possède un abdomen qui se termine par deux pinces. Chez les mâles, la longueur des pinces est un caractère héréditaire variable (elle varie entre 2mm et 9mm). On a mesuré, chez une population P, la longueur des pinces chez 586 mâles. Le tableau du **document 1** résume les résultats obtenus.

Les classes	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[7-8[	[8-9]
Les fréquences	66	177	19	66	132	112	14

1. Déterminez le type de variation étudiée. Justifiez votre réponse

2.Dressez l'histogramme et le polygone de fréquence de la distribution de la longueur des pinces chez les individus de la population P**.**

(Utilisez 2cm pour chaque classe et 1cm pour une fréquence de 20)

Correction

Remarque

Les représentations graphiques de la distribution des fréquences d'une variation, permettent de rassembler des données numériques; mais elles restent insuffisantes pour réaliser des comparaisons et donnent peu de renseignements sur les caractéristiques de cette variation. C'est pour cela qu'on utilise des constantes mathématiques spécifiques appelés les paramètres de la distribution des fréquences.

II-Les paramètres caractéristiques d'une distribution de fréquence

1- Paramètres de position :

Il s'agit de paramètres qui caractérisent la répartition des mesures par rapport aux valeurs centrales. On distingue deux types de paramètre de position :

a– Le mode (M):

- Dans le cas d'une variation discontinue c'est la valeur de la variable qui correspond à la fréquence la plus élevée (c.-à-d. au plus grand nombre d'individus).
- Dans une variation continue, le mode est la valeur moyenne de la classe ayant la plus grande fréquence.

....
....

⇨Le mode permet de déterminer l'homogénéité de la distribution d'une variable :

- Si le polygone de fréquence est **unimodale**, l'échantillon étudié est **homogène**
- Si le polygone de fréquence est **bimodale**, ou **plurimodale**, l'échantillon étudié est **hétérogène**.

....
....
....
....
....

b– La moyenne arithmétique: (lire X barre)

La moyenne arithmétique ($\bar{X}$) d'une série statistique est 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

$$\bar{X} = \frac{\sum_1^i (n_i \cdot x_i)}{N}$$

$$\bar{X} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_n x_n}{n_1 + n_2 + \ldots + n_n}$$

- ✓ ***xi***: la valeur de la variable (dans une variation discontinue) ou le centre de la classe (dans une variation continue).
- ✓ n_i : l'effectif de la variable.
- ✓ ***N:*** le nombre total d'individus dans la population étudiée.

Elle nous renseigne sur la valeur centrale du variable tenant compte des effectifs.

***Exemple 1**

Calcul de $\bar{\mathbf{X}}$ pour la distribution de nombre de nouveau nés chez les femelles de souris

Variable x_i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
Effectif f_i	2	8	12	16	23	18	10	7	1	$n = \Sigma fi = ...$
xi . fi	….	….	….	….	….	….	….	….	….	$\Sigma\, xi.fi = ...$

$$\bar{\mathbf{X}} = \frac{\sum_1^i (fi\ xi)}{n} = \cdots$$

Exercice d'application

Pour comparer la distribution de la masse des tubercules de pomme de terre dans deux champs différents, on a pris un échantillon de pomme de terre de chaque champ et on a mesuré leur masse. Les résultats sont résumés dans les tableaux suivants :

▶Echantillon du champ 1 :

Masse des tubercules de pomme de terre en g	115-135	135-155	155-175	175-195	195-215	215-235	235-255
Centre des classes (*xi*)	…	…	…	…	…	…	…
Nombre de tubercules = fréquences (*fi*)	34	55	73	92	83	58	22
xi . fi	…	…	…	…	…	…	…

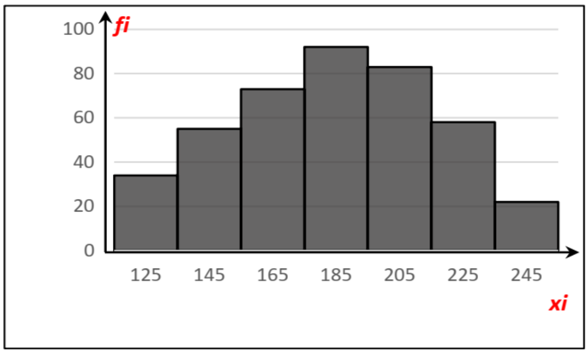
▶Echantillon du champ 2 :

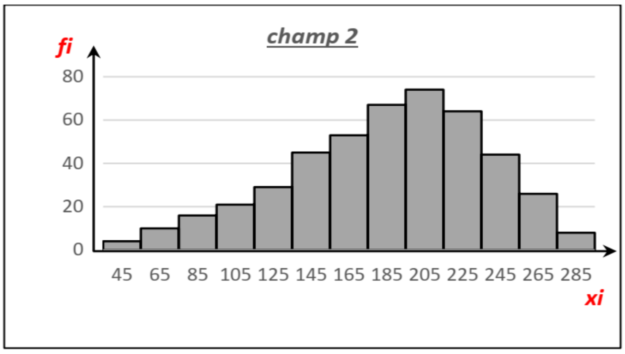
Masse des tubercules de pomme de terre en g	35-55	55-75	75-95	95-115	115-135	135-155	155-175	175-195	195-215	215-235	235-255	255-275	275-295
Centre des classes (*xi*)	…	…	…	…	…	…	…	…	…	…	…	…	…
Nombre de tubercules = fréquences (*fi*)	4	10	16	21	29	45	53	67	74	64	44	26	8
xi . fi	…	…	…	…	…	…	…	…	…	…	…	…	…

1. **Dressez** l'histogramme et le polygone de fréquence pour chaque échantillon.
2. **Déterminez** le mode et **calculez** la moyenne arithmétique $\overline{X}$ dans chaque cas.
3. **Comparez** les deux distributions. Que pouvez-vous **déduire** ?

1.

Champ 1 :





2.

	Champ 1	Champ 2
Mode	M_1= 185	M_2=205
Moyenne arithmétique	$\overline{X}$= 184.04	$\overline{X}$= 184.82

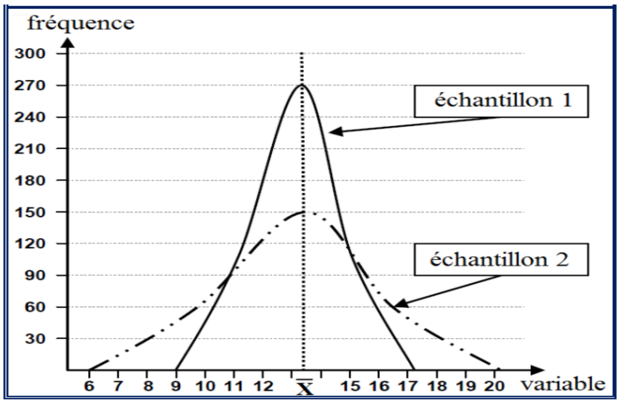
3. les deux distributions ont une moyenne arithmétique identique, alors que leur mode et leur polygone de fréquence sont différents, celui de la deuxième distribution étant plus dispersé.

Remarque :

Il se peut qu'une variable ait la même moyenne dans deux distributions, mais les valeurs se présentent avec des dispersions différentes.

Donc la moyenne arithmétique …………………

Pour décrire la distribution d'un caractère quantitative. On aura recours à des paramètres de dispersion.



2- Paramètres de dispersion :

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. Il prend toujours une valeur positive et on le calcule en utilisant la formule suivante:

$$E = \frac{\sum_1^i |x_i - \bar{X}|\, n_i}{N}$$

- ✓ Xi: valeur de la variable étudiée correspondant à la classe de rang i.
- ✓ $\bar{X}$: moyenne arithmétique.
- ✓ ni: l'effectif correspondant à la classe de rang i.
- ✓ Σ: total.
- ✓ N : nombre total d'individu.

……

……

……

b - La variance (V):

La variance V est la moyenne des carrées des écarts entre la valeur de chaque variable et la moyenne arithmétique. Pour calculer la variance, on utilise la formule suivante:

$$V = \frac{\sum_1^i (x_i - \bar{X})^2\, n_i}{N}$$

- ✓ Xi: valeur de la variable étudiée correspondant à la classe de rang i.
- ✓ $\bar{X}$: moyenne arithmétique.
- ✓ ni: l'effectif correspondant à la classe de rang i.
- ✓ Σ: total.
- ✓ N : nombre total d'individu.

c - L'écart type (σ)

L'écart type est défini comme la racine carrée de la variance et se calcule par la formule suivante: $\boldsymbol{\sigma = \sqrt{V}}$

Plus l'écart type est petit, plus la variation est faible (moins dispersée) et plus la population est homogène.

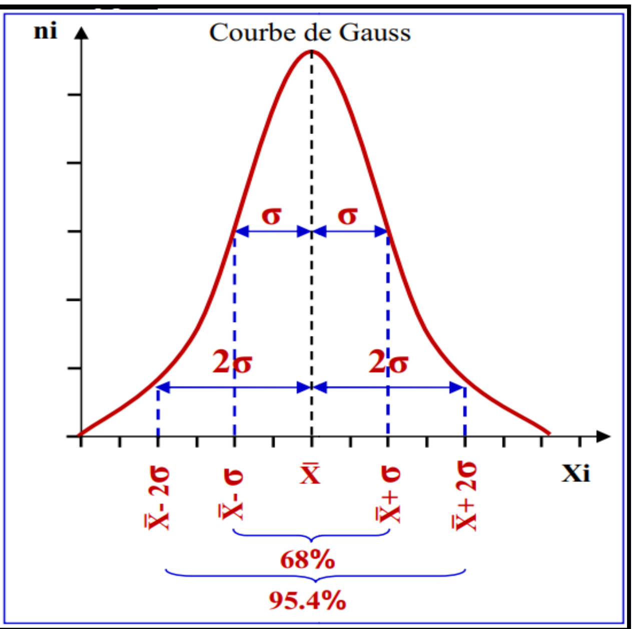
Remarque :

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 [$\bar{X}$-σ ; $\bar{X}$+σ] : On trouve 68% des individus de la population ;
- ⇨ Dans l'intervalle [$\bar{X}$-2σ ; $\bar{X}$+2σ] : 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 [$\bar{X}$-σ ; $\bar{X}$+σ] et 95,4% dans l'intervalle [$\bar{X}$-2σ ; $\bar{X}$+2σ].

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.



d - Le coefficient de variation K:

Le coefficient de variabilité **K** représente la relation entre **σ** et la moyenne arithmétique. 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 :

Coefficient de variation = $\frac{\sigma}{\overline{X}}100$

- ✓ $\overline{X}$: moyenne arithmétique.
- ✓ **σ:** L'écart type.

Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande.

- si $k \leq 15\%$, la population est homogène et la dispersion est faible ;
- si $15\% < k \leq 30\%$, 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.

*Application 1 :

Calcul de l'écart type dans la distribution de la longueur des pinces

xi	*fi*	*(xi-X̄)*	*(xi-X̄)²*	*(xi-X̄)².fi*
2.5	**66**	…	…	…
3.5	**177**	…	…	…
4.5	**19**	…	…	…
5.5	**66**	…	…	…
6.5	**132**	…	…	…
7.5	**112**	…	…	…
8.5	**14**	…	…	…

$n = \Sigma fi = ...$	$\Sigma(xi-\overline{X})^2.fi = ...$
$\sigma = \sqrt{\Sigma(xi-\overline{X})^2.fi/n} = ...$	

Le domaine de confiance $[\overline{X}-\sigma ; \overline{X}+\sigma]$ = […-… ;…-…] = [;] contient 68% des insectes ayant une longueur de pince comprise entre … et… . de même le domaine de confiance $[\overline{X}-2\sigma ; \overline{X}+2\sigma]$ = […-… ;…-…] = [;] contient 95% des insectes c.à.d 95% des insectes ont une longueur de pince comprise entre … et …

*Application 2 :

Calcul de l'écart type dans la distribution du nombre de nouveau nés chez les femelles de souris

xi	*fi*	*(xi-X̄)*	*(xi-X̄)²*	*(xi-X̄)².fi*
1	**2**	…	…	…
2	**8**	…	…	…
3	**12**	…	…	…
4	**16**	…	…	…
5	**23**	…	…	…
6	**18**	…	…	…
7	**10**	…	…	…
8	**7**	…	…	…
9	**1**	…	…	…

$n = \Sigma fi = ...$	$\Sigma(xi-\overline{X})^2.fi = ...$
$\sigma = \sqrt{\Sigma(xi-\overline{X})^2.fi/n} = ...$	

Le domaine de confiance $[\overline{X}-\sigma ; \overline{X}+\sigma]$ = […-… ;…-…] = [;] contient 68% des femelles ayant un nbre de nouveau nés compris entre … et… . de même le domaine de confiance $[\overline{X}-2\sigma ; \overline{X}+2\sigma]$ = […-… ;…-…] = [;] contient 95% des femelles c.à.d 95% des femelles ont un nombre de nouveau nés compris entre … et …

III- La sélection et la notion de race pure.

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**

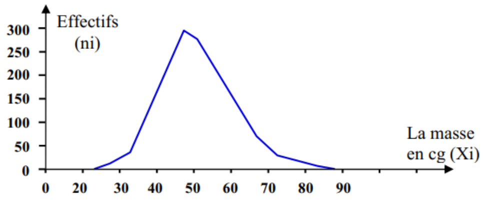
Classes	*21-25*	*26-30*	*31-35*	*36-40*	*41-45*	*46-50*	*51-55*	*56-60*	*61-65*	*66-70*	*71-75*	*76-80*	*81-85*	*86-90*
Fréquences	2	14	32	89	182	293	267	209	130	66	26	17	9	1

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** ?

Correction

R1: Il s'agit d'une variation continue car le variable qui est le poids des graines prend toutes les valeurs possibles.

R2: a. Réalisation du polygone de fréquence.
b. Le polygone de fréquence est unimodal, on peut supposer que la population **P** est homogène.



- ***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

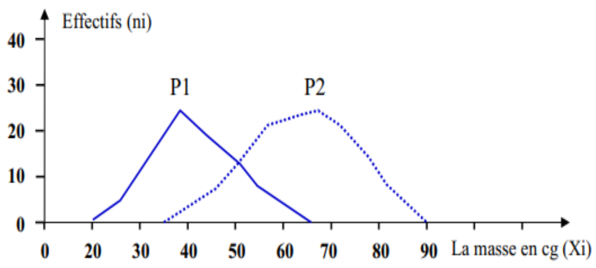
Classes	*36-40*	*41-45*	*46-50*	*51-55*	*56-60*	*61-65*	*66-70*	*71-75*	*76-80*	*81-85*	*86-90*
Fréquences	2	5	9	14	21	22	24	23	17	6	2

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** ?

Correction

R3: Représentation graphique de la variation du poids des graines chez les sous populations P_1 et P_2.

R4 : a. Pour **P**: 48cg. Pour P_1: 38cg. Pour P_2: 68cg.
b. La sélection artificielle effectuée sur la population **P** a été efficace, elle a permis d'isoler deux populations qui différent par la distribution et le mode. On déduit que la population **P** est hétérogène pour le caractère quantitatif étudié, elle comprend au moins deux lignées (deux races).



Remarque :

Dans ce cas la sélection a été efficace car on a obtenu deux sous populations différentes.
Pour déterminer si les deux populations P1 et P2 sont homogène, on refait la même expérience sur chaque population.

- Si on obtient un même mode que celui de la population sélectionnée, on dit que la sélection est inefficace. Et la population est constituée de race pure.
- Si on obtient deux modes différents de celui de la population sélectionnée, on dit que la sélection est efficace. Et la population est constituée de races différentes.

- ***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 ?

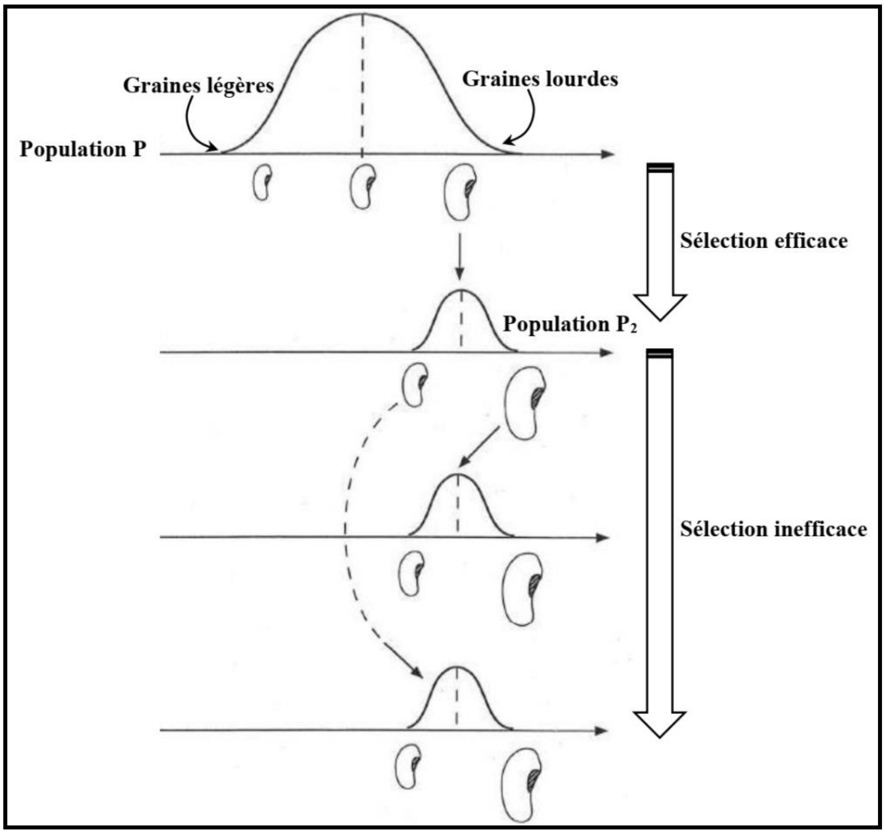
Correction

4. Cette deuxième sélection artificielle n'a entrainé aucune modification de la distribution des deux populations P1 et P2. Ceci peut être expliqué par le fait que les deux populations constituent deux races pures différentes :

- race des graines légères possédant le même génotype qui reste invariable d'une génération à une autre

- race des graines lourdes qui possèdent un génotype différent de celui des graines légères et qui reste stable également d'une génération à une autre.

Dans ce cas la sélection artificielle a été inefficace.



Remarque : La sélection artificielle réalisée par W. Johannsen est qualifiée de sélection conservatrice, elle a permis d'isoler et de conserver une souche bénéfique (souche à graines lourdes).

Bilan :

▪La sélection artificielle est un processus dont le but est d'isoler les individus possédant un phénotype recherché et le génotype qui en est responsable () ainsi la sélection artificielle permet l'amélioration de la production animale et végétale quand elle est effectuée dans une population composée de deux ou de plusieurs races alors qu'elle reste inefficace dans une race pure.

▪**La race pure :** ensemble d'individus (population) de même phénotype, la sélection au sein de cette population est inefficace, puisqu'on obtient chez la descendance après chaque croisement la même distribution des fréquences caractérisée par un mode constant, ce qui traduit son homogénéité.