La biométrie : Etude de la variation des caractères quantitatifs

Introduction :

La génétique mendélienne s'intéresse à l'étude des caractères qualitatifs (la couleur, le forme, groupes sanguins …) qui sont faciles à distinguer. Cependant, il existe d'autres caractères dits quantitatifs (la taille, le poids, nombre des œufs pondues…) non soumis aux lois de *Mendel* et qui sont des caractères mesurables (en kg, en m, en l, ...) donc qui peuvent prendre différentes valeurs

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- Notions de variation continue et variation discontinue

1. Variation discontinue et sa représentation graphique :

Document 1

On s'intéresse à l'étude du nombre des nouveau-nés après chaque grossesse chez une population de 100 souris, et on obtient les résultats suivants :

Variable x_i : nombre des nouveau-nés	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Effectif f_i : nombre des femelles	2	8	12	16	23	18	10	7	1

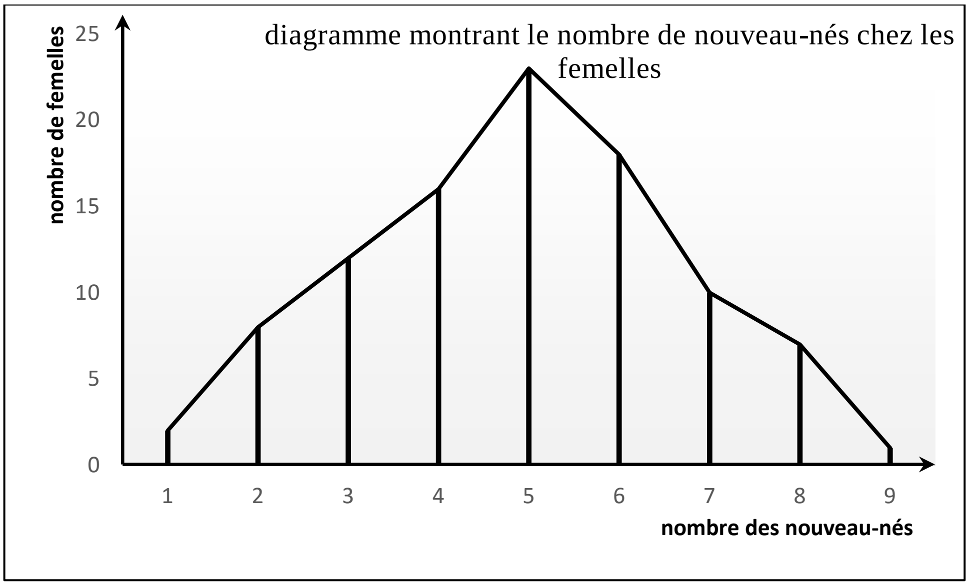
1. Déterminez le type de variation étudiée. Justifiez votre réponse

2. Représentez graphiquement la répartition des fréquences sous forme de diagramme en bâtons et polygone de fréquences.

1. Il s'agit d'une variation **discontinue** car il prend des valeurs limitées exprimées par des nombres entiers.

2.Dans le cas d'une variation discontinue, on représente les résultats sous forme de diagramme en bâtons.

On relie les points du sommet des traits verticaux par des segments de droites pour obtenir le polygone de fréquence.



2. Variation continue et sa représentation graphique :

Document 2

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

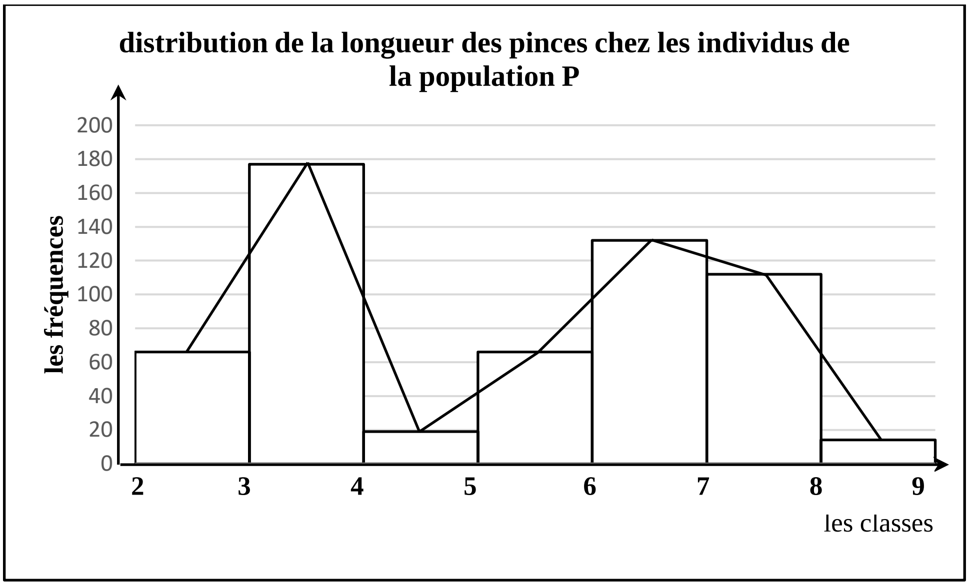
2.Dressez l'histogramme et le polygone de fréquence de la distribution de la longueur des pinces chez les individus de la population P**. (2 pts)**

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

Dans le cas variation continue, on peut construire une série de rectangles dont la base a pour valeur les intervalles des classes et comme hauteur la fréquence de la classe correspondante

1.Il s'agit d'une variation continue

Justification : la longueur des pinces peut prendre toutes les valeurs possibles dans un intervalle donné.



Les tableaux et les représentations graphiques sont difficiles à exploiter pour décrire une distribution de fréquence. Ainsi les généticiens utilisent des paramètres mathématiques qui leur permettent de décrire et comparer les distributions de fréquences.

II- Les paramètres caractéristiques d'une distribution de fréquence

1. Les paramètres de position

Ils correspondent aux valeurs centrales autour desquelles se répartissent les valeurs des séries étudiées. On en distingue deux types :

a. Le mode :

✍ 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 désigne donc la valeur dominante c-à-d la valeur la plus représentée d'une variable dans la population étudiée.

➪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**.

Ainsi dans l'exemple de nombre de nouveau nés chez les femelles, la plus grande fréquence est 23 et correspond à 5 nouveau nés (c-à-d que la plupart des femelles (23) possèdent 5 nouveau nés). Le polygone de fréquence est unimodale donc la population est **homogène**.

Dans l'exemple de la distribution de la longueur des pinces, le polygone de fréquence montre deux modes ; le premier mode à 3.5 mm et le deuxième mode à 6.5mm. La population est donc **hétérogène**.

b. La moyenne arithmétique $\overline{\mathrm{X}}$ (lire X barre)

Elle nous renseigne sur la valeur centrale du variable tenant compte des effectifs. Elle est calculée par la formule suivante :

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

Avec :

- *xi* :la valeur de la variable (dans une variation discontinue) ou le centre de la classe (dans une variation continue).
- f_i : la fréquence de la variable.
- *n* : le nombre total d'individus dans la population étudiée.

Ainsi, dans l'exemple de nombre de nouveau nés chez les femelles de souris, la moyenne est :

$$\overline{\mathbf{X}} = \frac{\cdots\cdots}{\cdots..} = \cdots.$$

Document 3

Calcul de $\overline{\mathbf{X}}$ 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	….	….	….	….	….	….	….	….	….	$\Sigma\, xi.fi = ...$

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

Et dans l'exemple de la distribution de la longueur des pinces, la moyenne est : $\overline{\mathbf{X}} = \frac{\cdots}{\cdots} = \cdots$

Document 4

Calcul de $\overline{\mathbf{X}}$ pour la distribution de la longueur des pinces

Les classes	[2-3[	[3-4[	[4-5[	[5-6[	[6-7[	[7-8[	[8-9]	
Centre des classes (mm) (*xi*)	2.5	3.5	4.5	5.5	6.5	7.5	8.5	
Les fréquences (*fi*)	66	177	19	66	132	112	14	$n = \Sigma fi = \ldots$
xi . fi	…	…	…	…	…	…	…	$\Sigma\, xi.fi = \ldots$

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

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	…	…	…	…	…	…	…

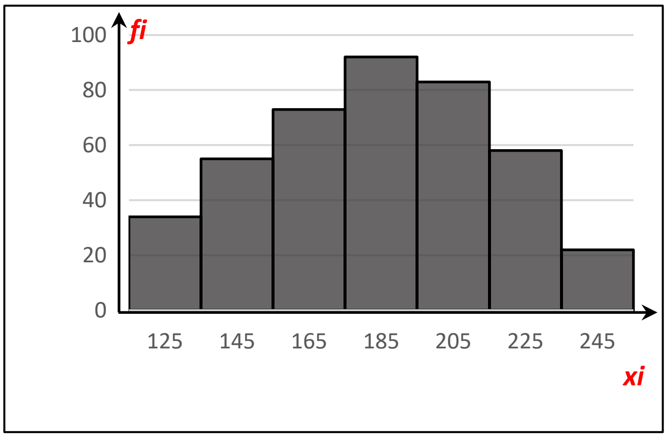
▶Echantillon du champ 2 :

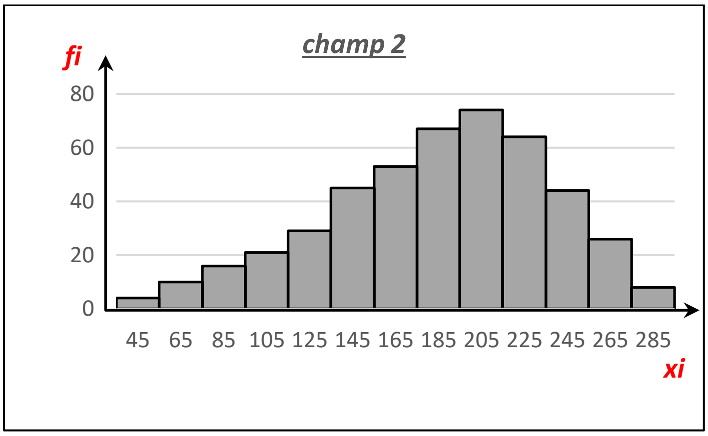
Masse des tubercules de pomme de terre en g	35-55	55-75	75-95	95-115	115-135	135-155	155-175	175-195	195-215	215-235	235-255	255-275	275-295
Centre des classes (*xi*)	…	…	…	…	…	…	…	…	…	…	…	…	…
Nombre de tubercules = fréquences (*fi*)	4	10	16	21	29	45	53	67	74	64	44	26	8
xi . fi	…	…	…	…	…	…	…	…	…	…	…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

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.

On déduit que la moyenne arithmétique est insuffisante pour décrire la distribution d'un caractère quantitative.

Conclusion :

Il se peut qu'une variable ait la même moyenne dans deux distributions, mais les valeurs se présentent avec des dispersions très différentes. Afin d'estimer l'écart de la variable par rapport à la moyenne, les généticiens utilisent d'autres paramètres appelés paramètres de dispersion.

2. Les paramètres de dispersion

a. La variance (V)

La variance est un paramètre permettant de mesurer le degré de dispersion d'une distribution.

La variance se calcule par la formule suivante :

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

b. L'écart type (σ)

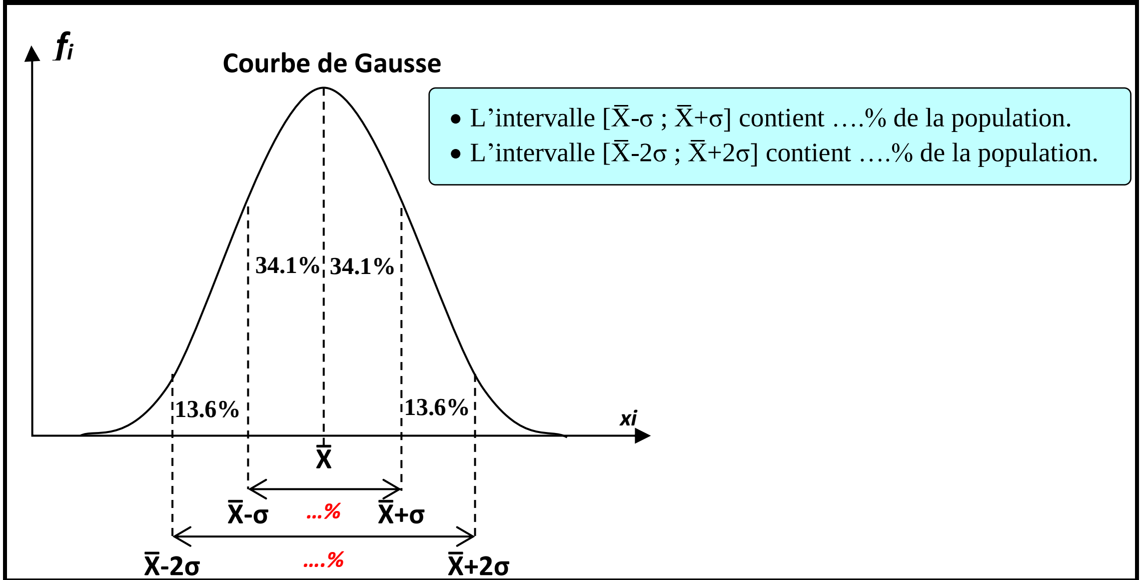
L'écart type est défini comme la racine carrée de la variance et se calcule par la formule suivante :

$$\sigma = \sqrt{V} \text{ soit } \sigma = \sqrt{\frac{\sum_1^i fi\,(xi - \bar{X})^2}{n}}$$

Plus l'écart type est grand, plus les valeurs sont dispersées, et plus l'homogénéité de la population diminue

L'écart type permet de définir ce qu'on appelle le **domaine de confiance**.

Document 5



→ Dans l'intervalle $[\bar{X}-\sigma\ ;\ \bar{X}+\sigma]$: On trouve les 2/3 c.à.d. 68% des individus de la population ;

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

▶**Application**

Document 6

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

xi	*fi*	$(x_i-\bar{X})$	$(x_i-\bar{X})^2$	$(x_i-\bar{X})^2.f_i$
1	2	…	…	…
2	8	…	…	…
3	12	…	…	…
4	16	…	…	…
5	23	…	…	…
6	18	…	…	…
7	10	…	…	…
8	7	…	…	…
9	1	…	…	…

$n = \Sigma fi = ...$	$\Sigma(x_i-\bar{X})^2.f_i = ...$
$\sigma = \sqrt{\Sigma(xi-\bar{X})^2.fi/n} = \ldots$	

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

xi	*fi*	$(x_i-\bar{X})$	$(x_i-\bar{X})^2$	$(x_i-\bar{X})^2.f_i$
2.5	66	…	…	…
3.5	177	…	…	…
4.5	19	…	…	…
5.5	66	…	…	…
6.5	132	…	…	…
7.5	112	…	…	…
8.5	14	…	…	…

$n = \Sigma fi = ...$	$\Sigma(x_i-\bar{X})^2.f_i = ...$
$\sigma = \sqrt{\Sigma(xi-\bar{X})^2.fi/n} = \ldots$	

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

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

Remarque :

Il existe un 3ème paramètre de dispersion appelé **Coefficient de variation**, sa formule est : $CV = \frac{\sigma}{\bar{X}} \times 100$

En effet : - Si **CV < 15% :** On dit que la dispersion des valeurs autour de la moyenne est petite, donc la population est homogène.

- Si **15 ≤ CV < 30% :** On dit que les valeurs sont moyennement dispersées, donc une homogénéité moyenne de la population.
- Si **30 ≤ CV ≤ 100% :** On dit que les valeurs sont trop dispersées autour de la moyenne, donc la population est hétérogène.

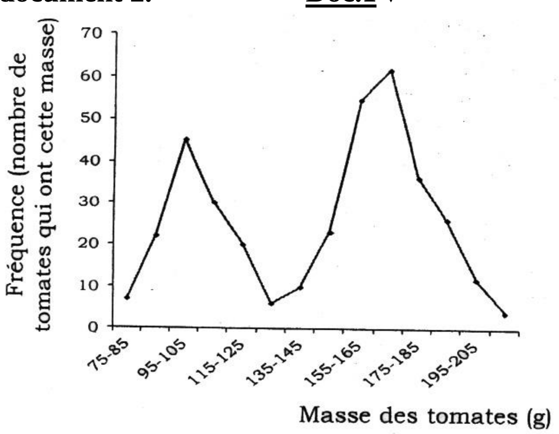
III- Sélection et notion de race pure.

1. Sélection à partir d'un groupe hétérogène.

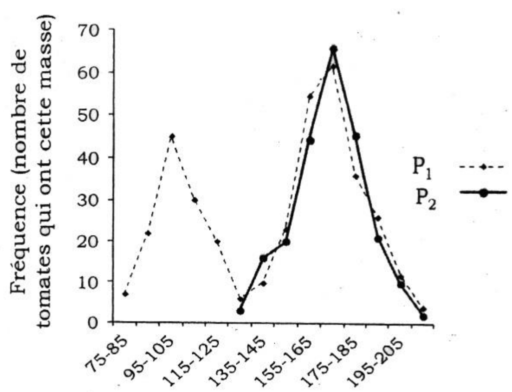
Document 7

La mesure de la masse des tomates d'une 1ère population P_1 a permis de tracer le polygone de référence représenté dans le **document** 1. Dans un deuxième temps, on a isolé les tomates de la classe [195 ;205] qu'on a cultivées dans des conditions adéquates ce qui a permis d'obtenir une 2ème population P_2 dont la distribution est représentée dans le polygone de fréquence du **document 2.**

Doc.1▼



Doc.2▼



Comparez la distribution des deux populations. Quel a été l'effet de la sélection ?

On constate que le polygone de fréquence de la population P1 est bimodale ce qui signifie que cette population est hétérogène. En revanche, le polygone de la population P2 est unimodale et donc cette population est homogène.

L'opération qui consiste à isoler une classe de la population et laisser ses individus se croiser entre eux (ou exclure de la reproduction certains individus) s'appelle **sélection artificielle** (exercée par l'homme). Elle a permis d'obtenir une population homogène.

2. Sélection à partir d'un groupe homogène.

Document 8

✍Dans une race de haricots, Wilhelm Johannsen a remarqué une différence au niveau de la taille des graines. Le tableau suivant représente les résultats d'étude biométrique de la masse des graines de haricot menée sur une population ***P*** de 1337 graines.

masse des graines en *cg*	**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**
Fréquence	2	14	32	89	182	293	267	209	130	66	26	17	9	1

1.a/ Dressez l'histogramme et le polygone de fréquence de la distribution de la masse des graines de la population P.

b/ Déterminez le mode et **calculez** la moyenne arithmétique $\overline{X}$

✍En 1913, Johannsen a isolé les graines d'haricot de la classe [21-25] (graines légères) et celles de la classe [86-90] (graines lourdes) puis il a cultivé séparément les graines de chaque classe. Après germination des graines et formation des plantes, il a laissé ces dernières s'autoféconder et a obtenu deux populations (P_1 et P_2) de graines dont les distributions sont résumées dans les tableaux suivants :

Distribution des effectifs de la population *P_1* (les graines légères)	masse des graines en *cg*	**21-25**	**26-30**	**31-35**	**35-40**	**41-45**	**46-50**	**51-55**	**56-60**	**61-65**
	Fréquence	2	7	18	23	20	16	10	5	2

Distribution des effectifs de la population *P_2* (les graines lourdes)	masse des graines en *cg*	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	**61-65**	**66-70**	**71-75**	**76-80**	**81-85**	**86-90**
	Fréquence	2	5	9	14	21	22	24	23	17	6	2

2.a/ Tracez, dans un même diagramme, le polygone des fréquences pour chaque population.

b/ Déduisez le mode et calculez la moyenne arithmétique $\overline{X}$ dans chaque cas.

3.Que déduisez-vous de la **comparaison** de ces paramètres avec ceux de la population d'origine (population P)

✍Dans une expérience similaire, Johannsen isole les grosses graines d'haricot des petites graines qu'il sème ensuite séparément. Après autofécondation, il a obtenu deux groupes de graines.

- ☑ Les graines légères donnent une distribution identique à celle de la population P1
- ☑ Les graines lourdes donnent une distribution identique à celle de la population P2

4.Expliquez ces résultats expérimentaux.

1.a/ voir papier millimètre

1.b/ M=48 et $\overline{X}$= 52.18

2.a/ voir papier millimètre

2.b/ M_1= 38 et $\overline{X}_1$= 41.09

M_2 = 68 et $\overline{X}_2$ = 64.18

3. On constate que le mode et la moyenne arithmétique de la population d'origine sont intermédiaires entre ceux deux populations.

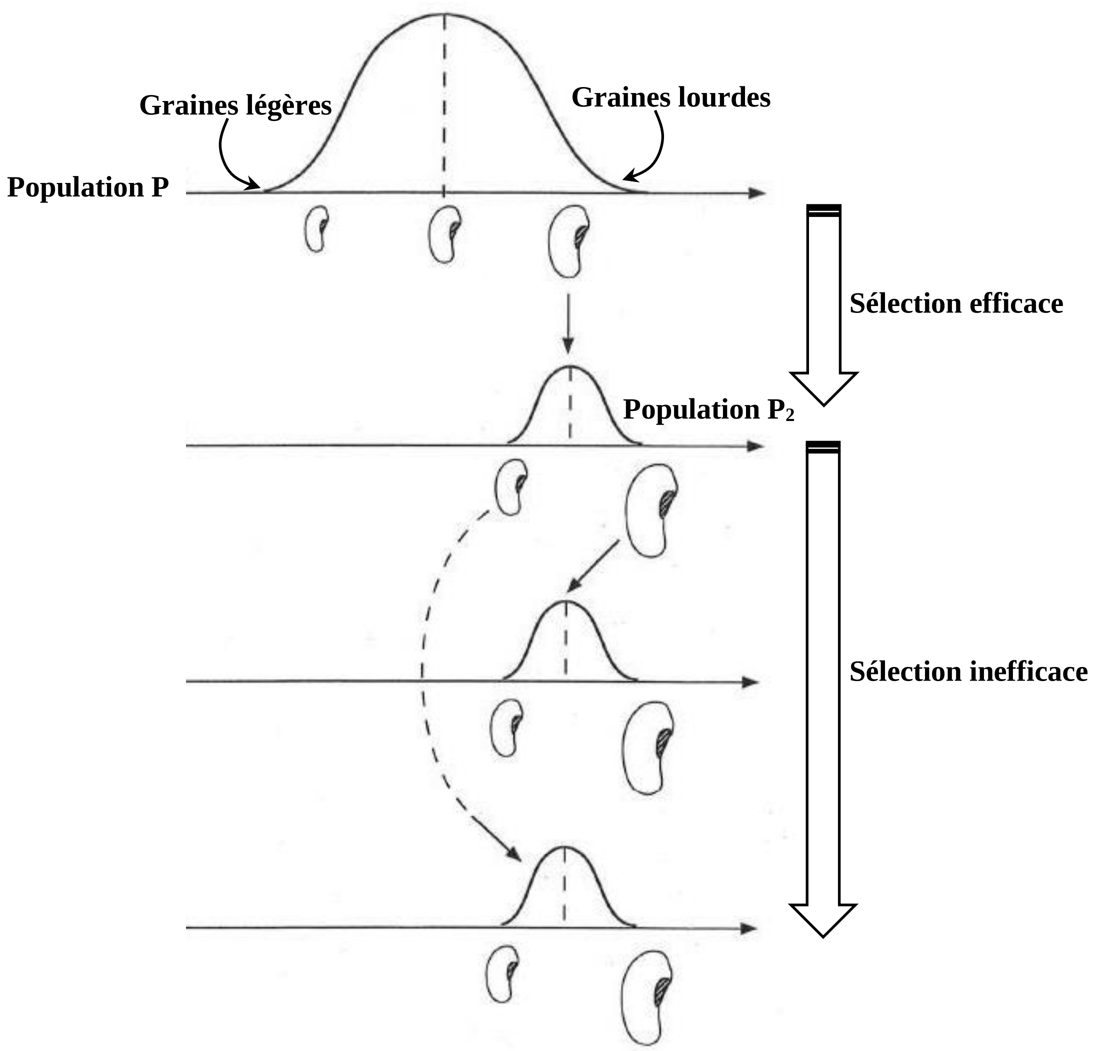
On déduit que la population d'origine était composée de deux groupes différents mais non détectables par l'étude statistique de la variation mais qu'on a pu mettre en évidence grâce à la sélection artificielle. Et donc cette sélection a été efficace.

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.



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.

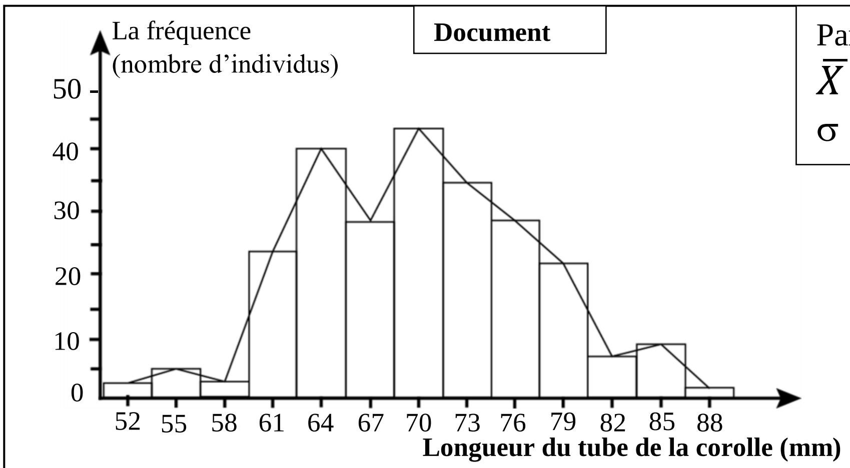
▪**La race pure :** ensemble d'individus (population) de même phénotype, la sélection au sein de cette population est inefficace, puisqu'on obtient chez la descendance après chaque croisement la même distribution des fréquences caractérisée par un mode constant, ce qui traduit son homogénéité.

Exercice d'application

Afin de favoriser l'amélioration de la longueur du tube de la corolle chez Longiflora (plante angiosperme), on a eu recours à la technique de la sélection artificielle, qui consiste à isoler les individus caractérisés par des tubes de corolle longs et les soumettre à des croisements aléatoires entre eux. Pour mettre en évidence l'efficacité de cette sélection, on propose l'étude des données expérimentales suivantes :

✍L'étude statistique de la distribution de la longueur du tube de la corolle dans la population mère (P_1) de Longiflora, a permis l'établissement de l'histogramme de fréquence, du polygone de fréquence et le calcul des paramètres statistiques $\overline{X}$ et σ. **(Voir document ci-dessous)**

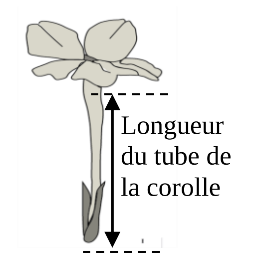
Document



Paramètres statistiques :

$\overline{X} = 70.05$

$\sigma = 7.16$



1- En **exploitant** le document ci-dessus :

a- **Déterminez** le type de variation étudiée.

b- **Décrivez** la distribution de fréquence dans cette population. **Qu'en déduisez- vous** ?

✍Les individus caractérisés par des tubes de corolle dont la longueur est égale ou supérieure à 79mm sont isolés et croisés entre eux au hasard. On a obtenu de ces croisements une population fille (P_2). Le tableau ci-dessous représente la distribution des fréquences de la longueur du tube de la corolle chez la population P_2.

La moyenne des classes (mm)	52	55	58	61	64	67	70	73	76	79	82	85	88	91
Nombre d'individus	0	0	0	1	1	1	15	20	28	41	18	3	3	2

2- **Calculer** la moyenne arithmétique et l'écart-type de cette distribution. **Utilisez** un tableau d'application pour calculer ces paramètres.

3- **Comparer** les paramètres statistiques $\overline{X}$ et σ des deux populations P_1 et P_2. **Qu'en déduisez-vous** à propos de l'efficacité de la sélection artificielle effectuée ?

Correction

1.a/ il s'agit d'une variation continue
Justification : la longueur du tube de la corolle peut prendre toutes les valeurs
1.b/ - Le polygone de fréquence montre deux modes ; le premier mode à 64mm et le deuxième mode à 70mm
- Un grand écart entre les valeurs de la longueur du tube de la corolle et la moyenne arithmétique
- Déduction : la population est donc hétérogène

2.

la moyenne des classes	fi	$fixi$	$xi - \bar{X}$	$(xi - \bar{X})^2$	$fi\,(xi - \bar{X})^2$
52	0	0	-25,04	626,88	0
55	0	0	-22,04	485,66	0
58	0	0	-19,04	362,43	0
61	1	61	-16,04	257,20	257,20442
64	1	64	-13,04	169,98	169,97886
67	1	67	-10,04	100,75	100,75329
70	15	1050	-7,04	49,53	742,91594
73	20	1460	-4,04	16,30	326,04330
76	28	2128	-1,04	1,08	30,14484
79	41	3239	1,96	3,85	157,89253
82	18	1476	4,96	24,63	443,25852
85	3	255	7,96	63,40	190,19973
88	3	264	10,96	120,17	360,52304
91	2	182	13,96	194,95	389,89756
la somme	133	10246			3168,81

$\bar{X}$ = 10246 /133 = 77.04mm

$\sigma = \sqrt{3168,81 / 133} = 4.88$

3.

- La moyenne arithmétique de la population fille P_2 est supérieure à celle de la population mère P_1
- L'écart- type de la population fille P_2 est inférieur à celui de la population mère P_1
- La sélection effectuée est efficace car chez la population P_2 la longueur du tube de la corolle s'est améliorée et la dispersion de la population a diminué

M.jaouani