

NOTES POUR L'UE BIOS 345

**GÉNÉTIQUE
QUANTITATIVE**

DU

Dr NGALLE Hermine BILLE

Chargée de Cours

2020-2021

SOMMAIRE

| | |
|---|----|
| CHAPITRE 1. INTRODUCTION | 1 |
| CHAPITRE 2. MESURE DE LA VARIATION QUANTITATIVE | 6 |
| 2.1. Quelques exemples de caractères quantitatifs | 6 |
| 2.2. Types de caractères quantitatifs..... | 6 |
| 2.3. Outils statistiques..... | 8 |
| 2.3.1. <i>La moyenne</i> | 8 |
| 2.3.2. <i>La variance</i> | 9 |
| 2.3.3. <i>La distribution normale</i> | 10 |
| 2.3.4. <i>Études de cas</i> | 12 |
| CHAPITRE 3. MODÈLE GÉNÉTIQUE..... | 18 |
| 3.1. Variables non corrélées..... | 18 |
| 3.1.1. <i>Écarts génétiques et environnementaux</i> | 18 |
| 3.1.2. <i>Variances génétiques et environnementales</i> | 22 |
| 3.2. Variables corrélées..... | 24 |
| 3.2.1. <i>Définitions</i> | 24 |
| 3.2.2. <i>Exemples de corrélations naturelles</i> | 28 |
| 3.2.3. <i>Corrélations, écarts et variances génétiques</i> | 29 |
| CHAPITRE 4. HÉRITABILITÉ..... | 32 |
| 4.1. Héritabilité au sens large : l'inné et l'acquis | 32 |
| 4.1.1. <i>Définition</i> | 33 |
| 4.1.2. <i>Mesure de H^2 chez les plantes : cas du maïs</i> | 34 |

| | |
|---|----|
| 4.1.3. <i>Mesure de H^2 chez les humains : cas des jumeaux</i> | 36 |
| 4.2. Héritabilité au sens strict : prédiction des phénotypes | 38 |
| 4.2.1. <i>Action des gènes et transmission de la variation</i> | 39 |
| 4.2.2. <i>Action des gènes et transmission de la variation génétique</i> | 40 |
| 4.2.3. <i>Effets additifs et effets de dominance</i> | 42 |
| 4.2.4. <i>Modèle génétique pour l'additivité et la dominance</i> | 45 |
| 4.2.5. <i>Définition de l'héritabilité au sens strict</i> | 46 |
| 4.2.6. <i>Estimation de h^2 chez les êtres humains</i> | 47 |
| 4.2.7. <i>Prédiction des phénotypes des descendants</i> | 48 |
| 4.2.8. <i>Sélection sur des caractères complexes</i> | 49 |
| 4.3. Méthode pratique d'analyse génétique : mesure de l'héritabilité..... | 51 |

CHAPITRE 1. INTRODUCTION

La **génétique quantitative** peut être définie comme l'étude de la transmission des caractères dont l'**observation** passe par une **mesure**. Autrement dit, c'est la partie de la génétique qui étudie la **transmission des différences individuelles** à l'aide des modèles mathématiques et des statistiques.

Certains de ces caractères peuvent prendre un **nombre infini de valeurs**, donc présentent une **variation continue**. C'est le cas, entre autres, de la **taille** ou de la **pression artérielle** chez un groupe d'individus (femmes et/ou hommes), de la **masse des graines par plante** chez une variété de maïs et de la **production laitière** dans un troupeau de vaches. D'autres caractères quantitatifs s'expriment par un **nombre entier**, mais présentent **plusieurs valeurs possibles**, comme le **nombre de graines** par épi chez une variété de sorgho, la **fécondité** chez une espèce animale, la **précocité de la floraison** chez une variété de haricot et le **nombre d'écailles** par individu chez une espèce de poisson. Ces caractères à **variation discontinue** (mais **pas binaire**) sont en fait considérés comme présentant une variation continue. En bref, les caractères quantitatifs possèdent un **déterminisme complexe**, contrairement aux caractères qualitatifs (comme la graine lisse ou ridée des plants de pois de Gregor Mendel) dont la **transmission** est **simple**.

Au sein de n'importe quel groupe important d'êtres humains, il existe une **gamme importante de tailles** : certains sont petits, d'autres grands et d'autres encore sont dans la moyenne. Dans certaines familles, les parents et les enfants adultes sont tous assez grands, alors que dans d'autres, ces mêmes individus (parents et enfants adultes) sont plutôt petits. Tout ceci suggère que les **gènes interviennent** dans la détermination de la taille d'un humain. Cependant,

les individus ne se **catégorisent pas clairement** en « grand » ou « petit ». Ce caractère, en dépit du fait qu'il soit manifestement héritable, ne semble pas suivre les lois de Mendel. Dans ce type de caractères, il est courant d'observer ce que l'on qualifie de **transgression génétique** : un enfant, à l'âge adulte peut être plus grand que son parent le plus grand !

Ces caractères, présentant une variation continue et ne se comportant pas suivant un mode mendélien simple, sont donc qualifiés de **quantitatifs** (ou **caractères complexes**). En effet, ils sont sous le contrôle d'un « **complexe** » de **facteurs génétiques** et **environnementaux**. La taille qu'un individu atteindra à l'âge adulte par exemple, dépendra des **gènes** qui lui ont été transmis par ses parents. Celui-ci dépendra également de la **nutrition reçue** et de la **survenue** d'éventuelles **maladies** au cours de sa vie. En génétique quantitative, il s'agit donc concrètement de **séparer** les **contributions génétiques** des **apports environnementaux** dans un **phénotype individuel** donné.

Les travaux de Mendel sur la couleur de la fleur chez le haricot, ont mis en évidence la « **variation quantitative** ». En effet, en croisant des variétés à fleurs blanches et à fleurs rouges, il avait observé « *toute une série de couleurs* », allant du pourpre au blanc, en passant par le violet pâle. Pour expliquer ce phénomène, Mendel proposa de considérer que la couleur de la fleur du haricot est un **mélange de caractères élémentaires**, la transmission de chacun suivant les mêmes règles que celles mises en évidence chez le pois. Cette explication suggérait ainsi ce qui est décrit aujourd'hui comme l'**intervention de plusieurs loci**, pour rendre compte de la variation continue de certains caractères. On parle également de **génétique multifactorielle** ou d'**hérédité polygénique**.

Les travaux des sélectionneurs de plantes, comme l'Américain East (1903) ou le Suédois Nielson Ehle (1908) étayèrent cette **hypothèse**

multifactorielle. Certains biométriciens dont l'anglais Ronald Fisher (1918), proposèrent un modèle de synthèse, rendant à la fois compte des **lois de Mendel** et des **relations biométriques entre apparentés**.

L'hypothèse multifactorielle, finalement émise en 1920, est ainsi venue lever la controverse sur l'applicabilité des lois de Mendel aux caractères à variation continue. Selon cette hypothèse, les caractères continus sont contrôlés par des **facteurs environnementaux**, ainsi que par « **une combinaison de multiples loci mendéliens** », chacun de ceux-ci exerçant un petit effet sur le phénotype global.

Il s'agit en définitive de définir l'**architecture génétique** des caractères complexes. En d'autres termes, il faut décrire tous les facteurs génétiques et environnementaux qui influencent un caractère donné. Cette architecture comprend :

- le **nombre de gènes** impliqués dans l'expression du caractère considéré. Contrairement aux caractères qualitatifs, les caractères quantitatifs sont gouvernés par un grand nombre de gènes ;
- la **contribution relative** de chaque gène impliqué. Un gène peut avoir un effet soit important ou faible. Un autre peut avoir une action arithmétique ou géométrique. Pour ce dernier, au lieu d'ajouter ou soustraire une valeur constante au génotype résiduaire, le gène agit en multipliant ou en divisant ce qui existe déjà par un certain facteur donné. Toutefois, il existe pour certains caractères à variation continue, des gènes dits **majeurs**, dont l'action est facilement décelable au travers des **grandes différences** qu'ils induisent en regard de la variation totale observée.

L'architecture génétique est donc la **propriété** d'une **sous-population** bien spécifique et elle peut **varier** parmi les **populations** d'une **espèce donnée**.

Par exemple, l'architecture génétique d'un caractère tel que le rendement en huile de palme chez *Elaeis guineensis* diffère d'une palmeraie à une autre. Des allèles différents ségrégent dans des palmeraies (populations) bien distinctes et des palmeraies différentes existent dans des environnements variés (le rendement en huile d'un même hybride diffère suivant que celui-ci est cultivé en Indonésie ou au Cameroun ; au Cameroun même entre les plaines côtières et les hauts plateaux de l'Ouest). Il peut ainsi y avoir des architectures différentes pour ce caractère particulier qu'est le rendement en huile chez le palmier.

La génétique quantitative s'appuie ainsi sur la **génétique des populations** et les **statistiques**. L'analyse des **loci de caractères quantitatifs** (QTL : Quantitative Trait Loci) est un ajout plus récent à l'étude de la génétique quantitative. Elle consiste à déterminer le **nombre de régions** du génome impliquées, ainsi que leurs **positions** et leurs **effets** sur le phénotype. Cette approche se base sur l'analyse combinée de l'information moléculaire et d'un caractère quantitatif dans une descendance en ségrégation. Elle permet de tester statistiquement le lien entre variation génétique (comme celle de marqueurs moléculaires) et variation phénotypique. Si ce test est **significatif**, on met en évidence **un QTL**.

Partant d'une analyse de « marqueur simple » [par exemple, si un marqueur moléculaire est présent, l'individu produit l'huile acide ; si le marqueur est absent, l'individu produit l'huile non acide (analyse de variance à un facteur)], l'analyse QTL a évolué vers des méthodes de plus en plus complexes telles que l'« Interval Mapping », **IM** (utilisant l'information de la carte génétique pour positionner les QTL le long du génome) ou le « Multiple Interval Mapping », **MIM** (prenant en compte des cofacteurs, marqueurs significativement associés au caractère).

Il semble crucial de comprendre la transmission des caractères quantitatifs car ils présentent une importance capitale en agriculture et en médecine. Chez l'Homme par exemple, la transmission des caractères quantitatifs tels que le taux de cholestérol ou le risque de développer un cancer ou une autre maladie est presque inconnue ; alors que l'hérédité des maladies telles que la mucoviscidose ou la drépanocytose (caractères qualitatifs) est clairement établie. Dans le cas des plantes cultivées, la résistance ou la tolérance aux pathogènes et/ou aux stress (sécheresse, froid...), l'efficacité de l'apport d'engrais représentent quelques caractères dont la transmission reste à élucider. Enfin, en ce qui concerne l'élevage, il est important d'appréhender la transmission des caractères comme le nombre de petits dans une portée chez les truies, la fréquence et le nombre d'œufs pondus chez une poule ou la production quotidienne des vaches laitières.

Questions clés

1. Pour un caractère donné dans une population, quelle est la **fraction de la variation** due à des **facteurs génétiques** et celle qui est due à des **facteurs environnementaux** ?
2. Comment utiliser les renseignements que l'on possède sur les **phénotypes parentaux pour prédire les phénotypes de leurs descendants** ?
3. **Combien de gènes** contribuent à la **variation génétique** d'un caractère quantitatif ?
4. Quels sont les **gènes spécifiques** qui contribuent à la **variation des caractères quantitatifs** ?

CHAPITRE 2. MESURE DE LA VARIATION QUANTITATIVE

A la fin de ce chapitre, l'étudiant doit être capable d'**identifier** de manière claire un **caractère complexe**, de **préciser** s'il est **discontinu** ou **continu** et de **décrire sa distribution** dans une population donnée.

2.1. Quelques exemples de caractères quantitatifs

Les caractères quantitatifs appartiennent à diverses catégories. En effet, il existe des **caractères biométriques** (taille, poids ou croissance des individus ; pression artérielle, taux de cholestérol, glycémie, nombre de soies de l'abdomen de la drosophile), des **caractères agronomiques** (taille de portée chez les animaux, production laitière, teneur en huile chez le soja ou le palmier à huile, nombre de graines par épi de maïs, date de floraison chez le blé). On peut également citer les **maladies multifactorielles** comme le diabète, la prédisposition à l'obésité ou à développer un cancer, l'hypertension ou l'hypotension artérielle, l'hypercholestérolémie. On peut enfin mentionner les **caractères impliqués dans l'adaptation** dont la précocité de la floraison, la fertilité, la tolérance aux facteurs du milieu (stress hydrique, froid...).

2.2. Types de caractères quantitatifs

La **génétique qualitative** étudie la transmission des **caractères discontinus**, c'est-à-dire les caractères à **variation binaire**. Pour ce type de caractères, les individus d'une population donnée peuvent être triés en catégories distinctes. C'est le cas de la drépanocytose (on est sain ou drépanocytaire) ou de la couleur blanche ou violette des fleurs des plants de pois de Mendel. Les caractères qui peuvent être ainsi catégorisés ont en général une **transmission simple** et dans les descendance produites, on observe les rapports mendéliens F_2 standards, comme **3:1** ou **15:1** (un gène ou deux gènes impliqués).

Cependant, certains caractères discontinus ne présentent pas une transmission simple et peuvent de ce point de vue, être placés dans le cadre de la génétique quantitative. Dans le cas de nombreuses maladies humaines par exemple, les individus peuvent être classés « sains » ou « affectés ». Toutefois, il existe de multiples facteurs génétiques et environnementaux qui confèrent à chaque individu le risque de développer l'une ou l'autre de ces maladies. Les individus possédant un certain nombre de ces facteurs de risque dépasseront le seuil de déclenchement et deviendront malades. Cette forme de caractères discontinus a un nom : **caractères à effet seuil**. Exemples : le diabète de type 2, la maladie d'Alzheimer, l'asthme, le cancer du sein ou encore les maladies cardiovasculaires.

Par ailleurs, d'autres caractères discontinus peuvent prendre un ensemble de valeurs discrètes limitées : ce sont des **caractères méristiques** ou **caractères numériques discontinus**. En effet, un individu peut par exemple avoir 1, 2, 3, 4 enfants ou davantage, mais ne peut pas en avoir 3,65. C'est également le cas pour le nombre de graines dans un épi de maïs, le nombre de soies de l'abdomen chez la drosophile ou le nombre d'écailles chez un poisson.

Quant aux **caractères continus**, ils peuvent prendre une gamme continue de valeurs. Le nombre de valeurs possibles est donc potentiellement infini. Si l'on prend par exemple la taille chez l'Homme adulte et sain (sans maladie), elle varie de 140 à 230 cm. Toutefois, la mesure précise de cette variable donne lieu à des possibilités infinies. Une personne peut avoir exactement 155 cm, tandis que d'autres ont 155,7 ou 155,07 ou encore 155,0047 cm. Il en est de même pour la taille d'une plante, le poids des bébés à la naissance et la quantité de lait quotidiennement produite par une vache.

Remarque

Les caractères quantitatifs, continus ou discontinus présentent typiquement une transmission complexe impliquant plusieurs gènes et sous

l'influence significative des facteurs environnementaux. Ils peuvent donc tous être qualifiés de caractères complexes.

2.3. Outils statistiques

L'étude de la transmission des caractères complexes nécessite l'utilisation de certains outils **statistiques**. Plusieurs sont utilisés, comme la covariance, la régression, la corrélation ou l'analyse de variance (ANOVA). Les outils statistiques **élémentaires** couramment utilisés dans l'évaluation de la variation quantitative sont :

- la **moyenne** qui peut permettre de décrire des populations et de comparer des différences au sein de celles-ci,
- la **variance** au travers de laquelle la variation existant au sein d'un groupe peut être quantifiée,
- la **distribution normale** grâce à laquelle il est possible de prédire des phénotypes avec des pourcentages précis. Elle est au centre de la compréhension de la variation quantitative dans les populations.

Généralement, cette étude se fait à partir d'un groupe particulier d'individus ou **population**, qui partagent certaines caractéristiques telles que l'âge, le sexe, l'ethnie ou l'origine géographique, la famille ou l'espèce, etc. Seulement, l'étude à l'échelle de la population toute entière est une tâche herculéenne. En général, c'est un **sous-groupe** ou un **échantillon de la population totale** qui sert dans l'étude à réaliser. Celui-ci doit être constitué au hasard et être suffisamment représentatif de la population totale. A partir d'un tel échantillon, les déductions sur la population entière pourront être légitimes.

2.3.1. La moyenne

Le caractère mesuré est une **variable aléatoire**, notée X , qui peut prendre différentes valeurs. Lorsque l'on sélectionne au hasard un individu de la population, la valeur observée est gouvernée par un élément aléatoire. La

moyenne (m ou \bar{X} ou encore μ) est la somme de **toutes les mesures individuelles** effectuées, divisée par la **taille de l'échantillon**. C'est donc la somme des x_i résultats divisée par l'effectif total n .

$$m = \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad 1 \leq i \leq n$$

Si les différentes valeurs possibles de la variable sont obtenues plusieurs fois, on a :

$x_1, x_2 \dots x_k$: valeurs possibles

$r_1, r_2 \dots r_k$: nombre de répétitions

$f_1, f_2 \dots f_k$: fréquences

$$\text{Avec } n = r_1 + r_2 + \dots + r_k = \sum_{i=1}^k r_i$$

Et la moyenne

$$m = \frac{1}{n} \sum_{i=1}^k r_i x_i = \sum_{i=1}^k \frac{r_i}{n} x_i = \sum_{i=1}^k f_i x_i$$

Il faut enfin signaler que la moyenne d'une variable aléatoire représente l'**attente** ou la **valeur attendue** ou encore l'**espérance mathématique** pour cette variable. La valeur attendue est la moyenne de toutes les valeurs que l'on observerait si l'on mesurait X de nombreuses fois. Cette espérance est symbolisée par E ou $E(X)$ pour signifier « la valeur attendue de X ». On écrit : $E(X) = m$.

2.3.2. La variance

L'ampleur de la variation présente dans une population peut être visualisée au travers d'une représentation graphique. Elle peut par exemple être représentée sur un **histogramme de fréquences**. Si la variation est importante, les valeurs s'étaleront le long de l'axe des abscisses. Par contre, les valeurs se regrouperont fortement autour de la moyenne si la variation est faible.

La **variance** peut également servir à la quantification plus directe de la variation dans une population. Elle permet en effet de mesurer l'**écart** des individus de la population par rapport à la moyenne de celle-ci, offrant ainsi une mesure de la dispersion par rapport à la moyenne. Certains individus auront des valeurs au-dessus de la moyenne et de ce fait bénéficieront d'un écart positif ; d'autres auront des valeurs inférieures à la moyenne, et donc disposeront d'un écart négatif. Ainsi, pour l'ensemble de la population, l'écart par rapport à la moyenne est nul.

Lorsque la variance est importante, les valeurs individuelles s'écartent fortement de la moyenne. Par contre, si elle est faible, les valeurs individuelles se regroupent près de la moyenne. Mathématiquement, la variance (V) est la **moyenne des carrés des écarts** avec la moyenne m . Soient x_i les valeurs possibles de la variable, r_i leur nombre de répétitions et f_i leur fréquence, la

$$\text{variance est : } V = \frac{1}{n} \sum_{i=1}^k r_i (x_i - m)^2 = \sum_{i=1}^k f_i (x_i - m)^2$$

Concrètement, c'est l'**écart-type** qui sert à quantifier l'**ampleur** de la variation par rapport à la moyenne de la population. Il est égal à la racine carré de la variance.

$$\sigma = \sqrt{V} .$$

2.3.3. La distribution normale

La distribution **normale** ou **gaussienne** est une **distribution continue** des fréquences, **similaire** à un **histogramme de fréquences** et qui s'applique parfaitement aux caractères continus. Ceci parce que la répartition des fréquences de ce type de caractères suit une distribution normale. La courbe qui caractérise la distribution normale a une **forme en cloche** (Figure 1).

Cette forme est définie par une équation appelée **fonction de densité de probabilités** de la loi normale. La forme de la courbe est essentiellement déterminée par **deux facteurs**, la moyenne et l'écart-type.

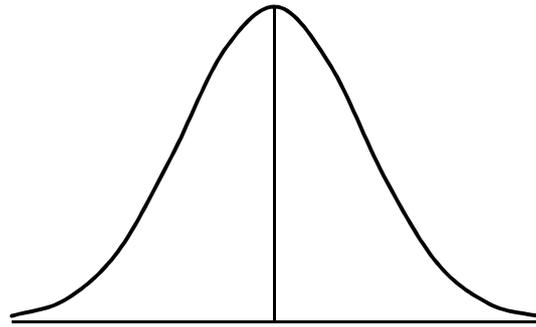


Figure 1. Morphologie de la courbe de la loi normale.

La distribution normale permet ainsi de prédire, d'une part la forme de la distribution du caractère étudié dans la population globale, et d'autre part la **probabilité** d'observer certaines valeurs en prélevant un échantillon dans ladite population, ceci à partir des données standards de la courbe de Gauss (Figure 2).

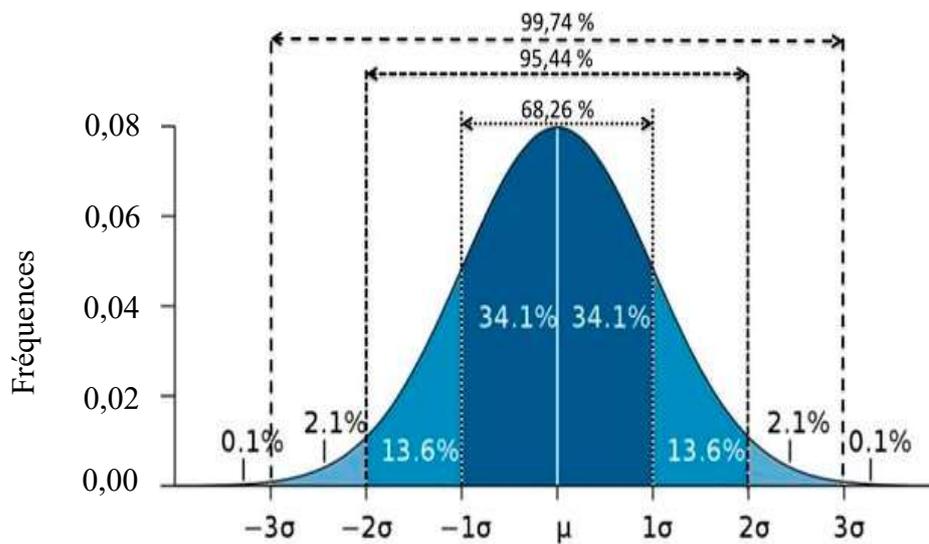


Figure 2. Normes de la loi normale : pourcentages prédits.

On montre également que la forme générale de la courbe **dépend beaucoup de l'écart-type (σ)**.

- Si σ est **grand**, la courbe est comme **aplatie**. Les **écarts importants** par rapport à la moyenne de la population ont des **probabilités non nulles**.
- Si par contre σ est **petit**, la courbe est **effilée**. Les **écarts importants** par rapport à la moyenne ont des **probabilités quasi nulles** (Figure 3).

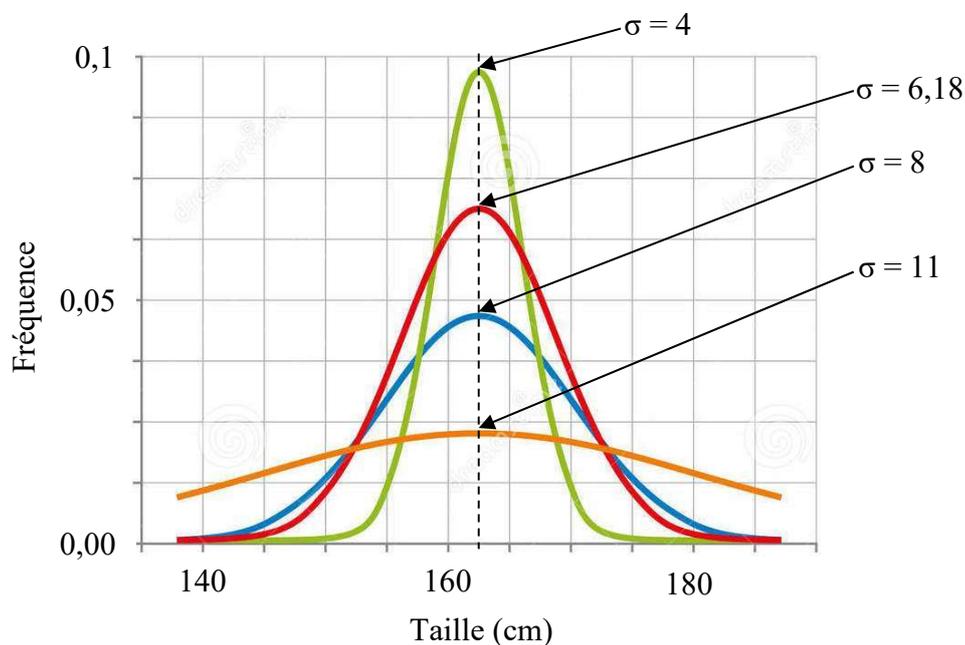


Figure 3. Courbes normales avec la même moyenne mais des écart-types différents.

2.3.4. Études de cas

1^{er} cas : utilisation de la moyenne et l'écart-type pour décrire la distribution d'un caractère dans une population

S'il faut par exemple réaliser l'étude de la taille pour la population des hommes adultes de Shanghai (Chine), ce serait fastidieux, puisque cette ville compte plus de 5 millions d'hommes adultes. Un échantillon aléatoire de 100 hommes adultes a été retenu, afin de faciliter l'évaluation. Puis, la taille de

chacun des individus constituant ledit échantillon a été mesurée et consignée au Tableau I. Les tailles relevées étaient comprises entre 156 et 184 cm. L'expérimentateur a choisi de retenir des chiffres entiers, arrondis à l'unité près.

Tableau I. Données simulées de la taille de 100 hommes à Shanghai.

| Taille (cm) | Nombre d'individus | Taille (cm) | Nombre d'individus | Taille (cm) | Nombre d'individus |
|-------------|--------------------|-------------|--------------------|-------------|--------------------|
| 156 | 1 | 166 | 1 | 175 | 6 |
| 157 | 2 | 167 | 6 | 176 | 3 |
| 158 | 1 | 168 | 9 | 177 | 4 |
| 159 | 2 | 169 | 7 | 178 | 2 |
| 160 | 1 | 170 | 9 | 179 | 2 |
| 161 | 1 | 171 | 5 | 180 | 2 |
| 162 | 2 | 172 | 5 | 181 | 2 |
| 164 | 7 | 173 | 6 | 184 | 2 |
| 165 | 7 | 174 | 5 | | |

Comme on peut le constater, certaines valeurs n'apparaissent qu'une seule fois. C'est le cas de 156 cm ou encore de 166 cm ; tandis que l'on recense 7 individus présentant soit une taille de 164 cm ou de 169 cm. Globalement, le nombre de répétitions maximum est de 9. Par contre, d'autres valeurs n'apparaissent pas du tout dans cet échantillon. On peut citer 163 et 182 cm.

Les calculs effectués à partir du Tableau I sont consignés au Tableau II. Ils permettent d'obtenir la **moyenne** de la taille des hommes de l'échantillon, ainsi que la **variance** et l'**écart-type**.

Commentaires

- Les hommes de l'échantillon ont en moyenne une taille de **170 cm**. Ce qui revient à dire que cette valeur représente la taille moyenne des hommes adultes de Shanghai.

- La variation par rapport à la moyenne, l'écart-type, vaut sensiblement **6** (exactement **5,92**) Ceci signifie que dans cette population, la taille des individus varierait de 164 à 176 cm.
- La taille moyenne des hommes adultes de Shanghai peut en définitive s'exprimer par la valeur **170 ± 6 cm**.

L'ampleur de la variation peut également s'exprimer de **manière visuelle**, au travers d'un **histogramme de fréquences**. En effet, l'on peut regrouper les tailles relevées dans l'échantillon en **classes**. Si l'on rassemble par exemple les tailles de ces hommes adultes en groupes de **4 cm**, on constituera ainsi **8 classes** de tailles, comme représentées à la Figure 4.

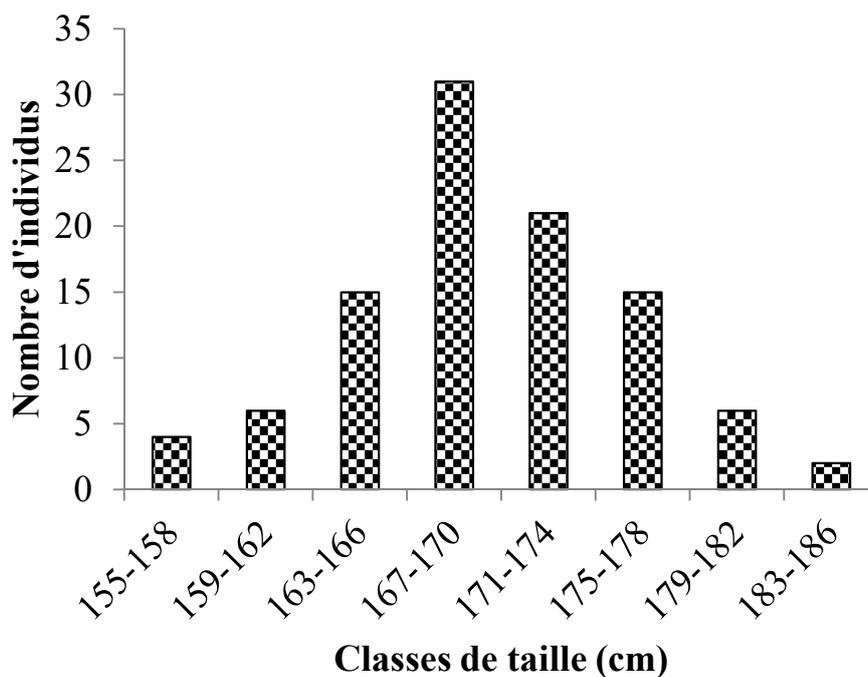


Figure 4. Histogramme de fréquences de la taille des hommes adultes de Shanghai.

Tableau II. Calculs pratiques de la moyenne et de l'écart-type.

| n_i | x_i | x_i^2 | $n_i x_i$ | $n_i x_i^2$ | |
|--------------|------------|---------|---------------|------------------|--|
| 1 | 156 | 24336 | 156 | 24336 | |
| 2 | 157 | 24649 | 314 | 49298 | |
| 1 | 158 | 24964 | 158 | 24964 | |
| 2 | 159 | 25281 | 318 | 50562 | |
| 1 | 160 | 25600 | 160 | 25600 | $n = \sum n_i = \mathbf{100}$ |
| 1 | 161 | 25921 | 161 | 25921 | |
| 2 | 162 | 26244 | 324 | 52488 | $T_1 = \sum n_i x_i = \mathbf{17\ 000}$ |
| 7 | 164 | 26896 | 1148 | 188272 | |
| 7 | 165 | 27225 | 1155 | 190575 | $m = 17\ 000/100 = \mathbf{170}$ |
| 1 | 166 | 27556 | 166 | 27556 | $T_2 = \sum n_i x_i^2 = \mathbf{2\ 893\ 506}$ |
| 6 | 167 | 27889 | 1002 | 167334 | |
| 9 | 168 | 28224 | 1512 | 254016 | $V = \frac{T_2}{n} - \left(\frac{T_1}{n}\right)^2$ |
| 7 | 169 | 28561 | 1183 | 199927 | $= \frac{2893506}{100} - \left(\frac{17000}{100}\right)^2$ |
| 9 | 170 | 28900 | 1530 | 260100 | |
| 5 | 171 | 29241 | 855 | 146205 | $V = \mathbf{35,06}$ |
| 5 | 172 | 29584 | 860 | 147920 | |
| 6 | 173 | 29929 | 1038 | 179574 | $\sigma = \sqrt{V} = \sqrt{35,06}$ |
| 5 | 174 | 30276 | 870 | 151380 | |
| 6 | 175 | 30625 | 1050 | 183750 | $\sigma = \mathbf{5,92} \approx \mathbf{6}$ |
| 3 | 176 | 30976 | 528 | 92928 | $m = \mathbf{170} \pm \mathbf{6\ cm}$ |
| 4 | 177 | 31329 | 708 | 125316 | |
| 2 | 178 | 31684 | 356 | 63368 | |
| 2 | 179 | 32041 | 358 | 64082 | |
| 2 | 180 | 32400 | 360 | 64800 | |
| 2 | 181 | 32761 | 362 | 65522 | |
| 2 | 184 | 33856 | 368 | 67712 | |
| Total | 100 | | 17 000 | 2 893 506 | |

2nd cas : utilisation de la loi normale pour décrire la distribution d'un caractère dans une population

Les données de la taille de 660 femmes vivant aux Etats-Unis sont recueillies et présentées à la Figure 5.

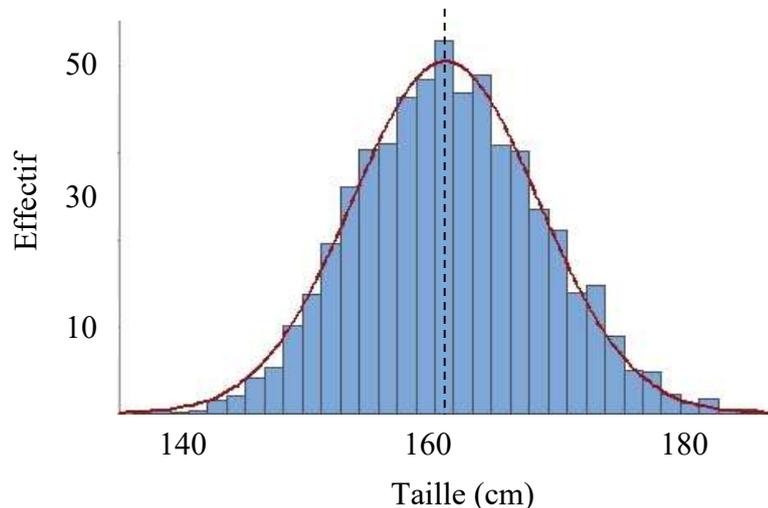


Figure 5. Histogramme de fréquences de la taille des femmes adultes aux Etats-Unis.

L'histogramme des fréquences montre la forme classique d'une **courbe en cloche** (en lissant la courbe issue de celui-ci), avec un **pic proche** de la **valeur moyenne** (164,40 cm). Les valeurs qui sont différentes de la moyenne sont distribuées de **manière symétrique** autour de celle-ci.

Il est possible de tracer une courbe normale de la distribution de la taille de ces femmes en utilisant **uniquement la moyenne** (164,40 cm) et **l'écart-type** (6,18 cm). Il serait ainsi possible de **prédire les pourcentages** et donc les effectifs des femmes ayant des tailles situées à une certaine **distance de la moyenne**, selon des normes de la loi normale, comme indiqué à la Figure 2.

Ainsi, selon les données, **449** femmes ont une taille comprise entre 158,22 et 170,58 cm ($\mu \pm 1 \sigma$), soit **71 %** ; et **633** femmes ont une taille comprise entre 152,04 et 176,76 cm ($\mu \pm 2 \sigma$), soit **96 %**. Ces valeurs sont très proches des prédictions de **68,26 %** et de **95,44 %** basées sur la courbe normale (Figure 2), soit exactement **450,516** et **629,904** femmes respectivement.

Si l'on connaît la moyenne et l'écart-type d'un caractère, il est également possible de **prédire la forme de sa distribution** dans la population, ainsi que la probabilité d'observer certaines valeurs à partir d'un échantillon de cette population. Ainsi, dans le cas présent, seuls **2,28 %** $[(100 - 95,44)/2]$ des femmes mesurent plus de 176,76 cm (plus de $\mu + 2\sigma$). Si l'écart-type valait par exemple 8 cm (au lieu de 6,18 cm), la courbe serait plus plate, mais il **y aurait toujours 2,28 % de femmes à plus de 2 σ de la moyenne**, c'est-à-dire ici à plus de 180,4 cm.

Commentaires

- La courbe de la loi normale permet de prédire le pourcentage des phénotypes situés à une certaine distance de la moyenne : 68,26 % des phénotypes observables par exemple se situent à $(\mu \pm 1\sigma)$.
- Quelle que soit la valeur de σ , le pourcentage de phénotypes situés à une certaine distance de la moyenne reste le même.

Conclusion : *un caractère complexe est un caractère qui ne présente pas une transmission simple, comme les caractères « mendéliens ». Il peut être continu ou discontinu, agronomique, biométrique, impliqué dans l'adaptation ou encore une maladie multifactorielle. La variation observée pour ces caractères complexes s'étudie à l'aide d'outils statistiques élémentaires que sont la moyenne, la variance et la distribution normale.*

CHAPITRE 3. MODÈLE GÉNÉTIQUE

Nous avons vu que les **caractères complexes** sont contrôlés par des **facteurs génétiques** et **environnementaux**. Le **challenge** de la **génétique quantitative** est de **séparer**, dans un **phénotype donné**, les contributions **génétiques** des apports **environnementaux**. Cette décomposition phénotypique prendra en compte le statut des deux variables impliquées. En effet, le fonctionnement du modèle génétique proposé dans ce chapitre dépend essentiellement de l'**absence** ou de l'**existence** de **corrélation** (relation, interaction) entre les deux variables impliquées ici, le **génotype** et l'**environnement**.

A l'issue de ce chapitre, l'étudiant doit donc être en mesure de **décomposer** un phénotype, quel qu'il soit, en ses **composantes génétique** et **environnementale**. Ces deux composantes phénotypiques peuvent être exprimées soit sous forme d'**écarts**, soit sous forme de **variances**.

3.1. Variables non corrélées

3.1.1. Écarts génétiques et environnementaux

Ces écarts sont évalués à l'aide d'un **modèle mathématique simple** qui utilise **deux outils statistiques élémentaires** :

- la **moyenne**, utile pour **décrire** une population donnée (la taille moyenne des ♂ de Shanghai, par exemple, est de **170 cm**) et **comparer** les différences au sein des populations (la taille moyenne des ♂ en zone urbaines chinoises, comme Shanghai, est de **170 cm** ; tandis qu'en zones rurales les ♂ ont plutôt une taille moyenne de **166 cm**) ;
- et la **variance** (ou l'**écart-type**) qui mesure l'**écart** par rapport à la **moyenne**. En d'autres termes, elle permet d'appréhender

l'**ampleur** de la **variation** qui existe dans les populations. Elle peut être élevée ou faible.

Ainsi, le phénotype d'un individu, pour un caractère donné (valeur du caractère), peut s'exprimer par la moyenne de la population et les écarts dus aux facteurs génétiques et environnementaux. La valeur de ce phénotype peut donc être notée par la formule $X = M + g + e$, dans laquelle :

X = valeur de l'individu

M = valeur moyenne de la population

g = écart dû aux effets génétiques

e = écart dû aux effets environnementaux

3.1.1.1. Étude d'un cas hypothétique

Yao Ming, un pivot professionnel Chinois de l'équipe de basketball des Houston Rockets aux Etats-Unis, mesure $X = 229$ cm. La taille moyenne de la population de Shanghai, sa **ville natale**, est $M = 170$ cm.

Comme pour chacun de nous, la taille de Yao Ming résulte des **gènes** qu'il a hérités de ses parents et de l'**environnement** dans lequel il a grandi. Imaginons une expérience afin de séparer les contributions génétique et environnementale dans la taille exceptionnelle de cet homme. La taille de Yao Ming est supérieure à celle de la population de **59 cm**.

Le surplus observé représente d'une part les **contributions** de la **génétique** (écart dû aux gènes : g) [Yao a des gènes qui lui sont **propres**, qu'il a reçus de ses parents] et d'autre part celles de l'**environnement** (écart dû à l'environnement : e) [Yao a été **élevé** dans un environnement bien précis]. Ce surplus peut être représenté par l'équation : $x = g + e$, x étant l'**écart phénotypique individuel** de Yao Ming ($x = 59$ cm).

Comment déterminer g et e pour Yao Ming ? Ceci serait facile si l'on disposait de **plusieurs clones** de Yao Ming, c'est-à-dire de plusieurs Yao Ming, tous des individus **génétiquement identiques** (même génotype). Supposons que

ces clones **existent**. Imaginons qu'ils soient répartis dans **plusieurs** foyers de Shanghai à leur naissance et que l'on mesure leur taille à l'âge de **21 ans**. On détermine que leur **taille moyenne** est de **212 cm**.

Certains des foyers seraient des environnements **positifs (+e)** pour le caractère étudié (la taille) et d'autres **négatifs (-e)**. La valeur attendue pour *e* dans les nombreux environnements dans lesquels les clones ont grandi est égale à **zéro**. Par conséquent, l'**écart génotypique** serait représenté par la différence entre la **moyenne des clones** et la **moyenne de la population**, soit :

$$g = 212 - 170 = 42 \text{ cm.}$$

Les **17 cm restants** de l'écart phénotypique correspondent à l'**écart** pour l'**environnement spécifique** dans lequel Yao Ming a été élevé, soit :

$$e = 59 - 42 = 17 \text{ cm.}$$

La **valeur phénotypique** de Yao Ming, pour ce qui est du caractère « **taille** », peut donc finalement s'écrire : **229 = 170 + 42 + 17**.

En conclusion, la **taille exceptionnelle** de Yao Ming est essentiellement due **aux gènes extraordinaires** dont il a hérité de ses parents ($g = 42 \text{ cm}$). Toutefois, l'environnement dans lequel cet homme a grandi a également stimulé sa taille ($e = 17 \text{ cm}$).

Le cas hypothétique traité bute sur l'existence de **plusieurs Yao Ming** (clones d'un génotype donné chez l'homme). Chez les végétaux et certaines espèces animales (surtout des animaux d'élevage tels que vaches, poules, canards, cochons, souris...), le clonage permet de disposer d'**individus génétiquement identiques**. Chez les animaux, il suffit de placer, dans un ovule énucléé, le noyau d'une des cellules de l'individu à cloner et ensuite de l'implanter chez une mère porteuse. Quant aux plantes, l'on peut procéder par bouturage, marcottage ou greffage, et même par production de lignées pures. Dans ces cas-là, il est ainsi possible d'estimer, pour un caractère donné, les contributions génétiques et environnementales, en élevant les clones dont on dispose dans différents environnements.

3.1.1.2. Étude d'un cas réel

Considérons par exemple 10 lignées de maïs, cultivées dans trois environnements différents. Si l'on compte les jours qui séparent le semis et la première libération de pollen, chez ces différentes lignées pures, les résultats possibles sont présentés au Tableau III.

Tableau III. Données simulées du nombre de jours précédents la libération de pollen pour dix lignées de maïs cultivées dans trois environnements. LP : lignées pures (A - J) ; E : environnements (1 - 3) ; M1 : moyenne de toutes les lignées dans un environnement donné ; M2 : moyenne d'une lignée donnée dans les 3 environnements.

| E \ LP | A | B | C | D | E | F | G | H | I | J | M₁ |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------------------|
| E1 | 62 | 64 | 66 | 66 | 68 | 68 | 70 | 70 | 72 | 74 | 68 |
| E2 | 64 | 66 | 68 | 68 | 70 | 70 | 72 | 72 | 74 | 76 | 70 |
| E3 | 66 | 68 | 70 | 70 | 72 | 72 | 74 | 74 | 76 | 78 | 72 |
| M₂ | 64 | 66 | 68 | 68 | 70 | 70 | 72 | 72 | 74 | 76 | 70 |

Commentaires

- La moyenne générale de toutes les lignées pour tous les environnements confondus est de **70 jours**.
- Les moyennes des lignées sont de **68, 70 et 72 jours** dans les milieux E1, E2 et E3 respectivement. Celles-ci présentent des écarts à la moyenne générale, respectivement **négatif, nul et positif**. Ainsi, pour les trois environnements confondus,
 - les lignées A à D ont, des écarts à la moyenne générale négatifs, leurs moyennes s'étendant de 64 à 68 jours ;
 - les lignées E et F, avec des valeurs égales à la moyenne générale, 70 jours, ont des écarts nuls ;
 - les lignées G à J ont des écarts à la moyenne générale positifs, leurs moyennes se situent entre 72 et 76 jours.

Considérons par exemple la lignée A. Son phénotype moyen lorsqu'elle est cultivée dans E1 est de **62 jours**. **Comment expliquer cette valeur à partir des écarts environnementaux et génétiques ?**

La moyenne générale (toutes les lignées dans tous les environnements) est de **70 jours** et la moyenne de toutes lignées dans E1 est de **68 jours**, soit **deux jours de moins**. L'écart environnemental (e pour E1) est donc $e = 68 - 70 = -2$.

Le phénotype moyen de la **lignée A**, pour tous les environnements confondus est de **64 jours**, soit **6 jours de moins** que la moyenne générale. L'individu de la lignée A présente un écart, à cause de ses gènes, de 6 jours par rapport à une plante prise au hasard dans la population. L'écart génotypique (g pour A) est donc $g = 64 - 70 = -6$.

Le phénotype moyen de la **lignée A** lorsqu'elle est plantée dans **l'environnement 1** peut donc se **décomposer**, en se référant à la formule mathématique $X = M + g + e$, de la manière ci-dessous :

$$62 = 70 + (-6) + (-2).$$

On pourrait ainsi faire des calculs identiques pour les 9 autres lignées pures. On obtiendrait *in fine*, pour chacune d'elles, les écarts génétiques et environnementaux.

Conclusion : pour un caractère donné, le phénotype d'un individu peut s'exprimer d'après son écart par rapport à la moyenne de la population. Cet écart phénotypique est composé de l'écart génétique et de l'écart environnemental.

3.1.2. Variances génétiques et environnementales

Le modèle simple défini plus haut, à savoir $x = g + e$ peut servir de base pour réfléchir en profondeur à la variance des caractères quantitatifs. A titre de rappel, la variance mesure l'ampleur de l'écart de certains individus par rapports à la moyenne de la population. Selon ce modèle, la variance d'un

caractère donné (**variance phénotypique**) peut être répartie en **variance génétique** et **variance environnementale**, c'est-à-dire $V_X = V_g + V_e$.

Reconsidérons le cas réel étudié plus haut (Tableau III). A partir des données de ce tableau, on peut calculer :

- la variance phénotypique (V_X), en utilisant les moyennes des **dix lignées** dans chacun des **trois environnements**, soit au total **trente valeurs** phénotypiques (toutes les valeurs sauf les moyennes, dernière ligne et dernière colonne). Le résultat est $V_X = 14,67$ jours² ;
- la variance génotypique ou génétique (V_g), qui correspond à la variance de la moyenne entre les **dix lignées** pures, soit un total de **dix valeurs** phénotypiques (dernière ligne du tableau). Le résultat est $V_g = 12,00$ jours² ;
- la variance environnementale (V_e), représentant la variance de la moyenne des dix lignées dans chacun des **trois environnements**, soit **trois valeurs** phénotypiques. Le résultat est $V_e = 2,67$ jours².

La **variance phénotypique** est donc la **somme** des variances des composantes **génétique** et **environnementale** : $14,67 = 12,00 + 2,67$.

Conclusion

- *S'il existe de nombreuses sources influençant la variance phénotypique, la variance totale est la somme des variances de chacune des composantes. Ainsi la variance phénotypique V_X est la somme de la variance due aux gènes V_g et de la variance due à l'environnement V_e .*
- *Si l'on considère des lignées pures la variance génétique est égale à zéro. Ceci est également vrai pour la F_1 issue du croisement entre ces deux lignées.*
- *En F_2 on retrouvera les variances environnementale et génétique. La variance génétique entre les deux lignées parentales peut être calculée en soustrayant la variance environnementale de la*

variance phénotypique. La variance des lignées parentales et la variance des F_1 permettent d'avoir une estimation de la variance due à l'environnement.

- *La variance génétique reflète la somme des différentes sources à l'origine de la variation due aux gènes, à savoir la **variance génétique additive**, la **variance génétique dominante** et la **variance due aux interactions entre gènes (ici l'épistasie)**. Ces différentes composantes de la variance génétique totale seront détaillées dans le cadre de l'héritabilité.*

3.2. Variables corrélées

3.2.1. Définitions

Jusqu'à présent, nous nous sommes intéressés à des questions au travers desquelles le comportement statistique d'une seule variable était étudié. Ces questions étaient du type « quelle est la taille moyenne des adultes de Shanghai ? » ou « combien de jours en moyenne précèdent la libération du pollen après le semis chez une lignée de maïs ? » ou encore « combien de femmes vivant aux États-Unis ont une taille supérieure à 177 cm ? ». Seul le comportement statistique d'une variable est étudié. Il existe cependant toute une gamme de problèmes statistiques où l'on s'intéresse à la relation entre plusieurs variables.

Exemples

- Les individus les plus grands sont-ils les plus lourds ?
- Le revenu d'une famille a-t-il une influence sur les résultats scolaires des enfants ?
- Y a-t-il une relation entre le tabagisme et le cancer du poumon ?
- Le rendement en céréales dépend-il de la quantité d'engrais utilisée ?
- La productivité d'une entreprise est-elle liée au salaire des employés ?

Dans ces questions, nous désirons savoir si le comportement d'une variable est influencé par la valeur d'une autre variable. Pour étudier les relations ou **corrélations** entre deux variables statistiques, on peut les porter sur un graphique. Intéressons-nous par exemple à la première question posée plus haut, à savoir « les individus **les plus grands** sont-ils les **plus lourds** ? » Autrement dit, la taille est-elle corrélée au poids chez l'être humain ? Le degré de corrélation entre deux variables chez trente individus peut se visualiser au travers d'un diagramme de dispersion, comme présenté ci-dessous (Figure 5).

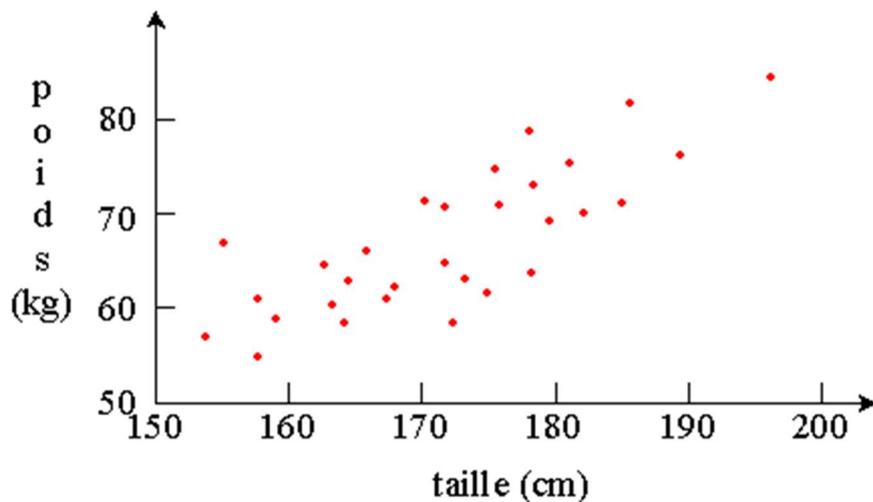


Figure 6. Relation entre le poids et la taille dans un échantillon de 30 individus.

On peut, par la pensée ou réellement, tracer une droite qui passe au mieux par les points représentés sur ce diagramme. En d'autres termes, cette droite doit passer au milieu du « nuage » de points. Si cette droite « monte », comme c'est le cas ici, on dira qu'il y a **corrélation positive** entre les deux variables. Ici, la **taille est donc corrélée au poids** : plus on est grand, plus on est lourd et vice versa (Figure 6A).

La droite peut « descendre » : la **corrélation** entre les deux variables est **négative**. Dans l'exemple traité, la taille serait négativement corrélée au poids. Dans ce cas, un individu de grande taille aurait plutôt tendance à avoir un petit poids et vice versa (Figure 6B).

Si la droite est plutôt « horizontale » ou s'il est impossible de la matérialiser, c'est qu'il y a absence de corrélation : la **corrélation** est donc **nulle**. Ici, la taille d'un individu serait aléatoire par rapport à son poids : s'il est grand, il peut aléatoirement être lourd ou de faible poids (Figure 6C).

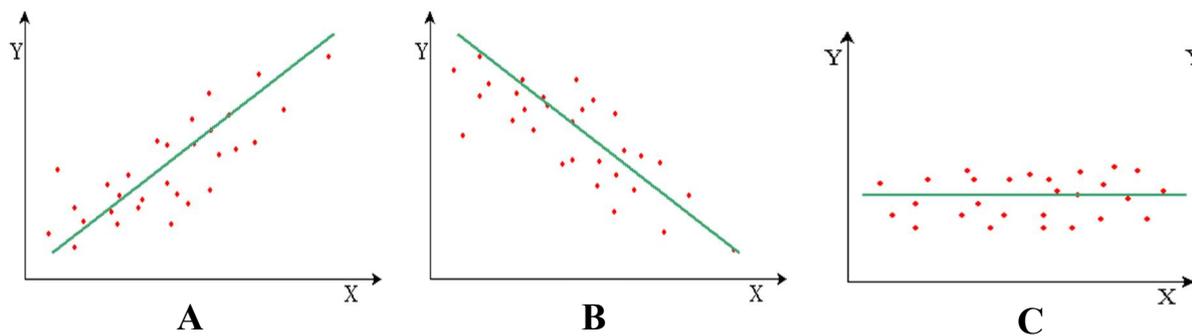


Figure 7. Diagrammes de dispersion des points pour des corrélations positive (A), négative (B) et nulle (C).

Statistiquement, l'association entre deux variables se mesure à l'aide du **coefficient de corrélation** (r) dont la formule et le calcul sont complexes. Il faut simplement retenir qu'il a une valeur comprise entre (-1) et (+1). Quand il est nul ($r = 0$), la corrélation est dite nulle (Figure 7).

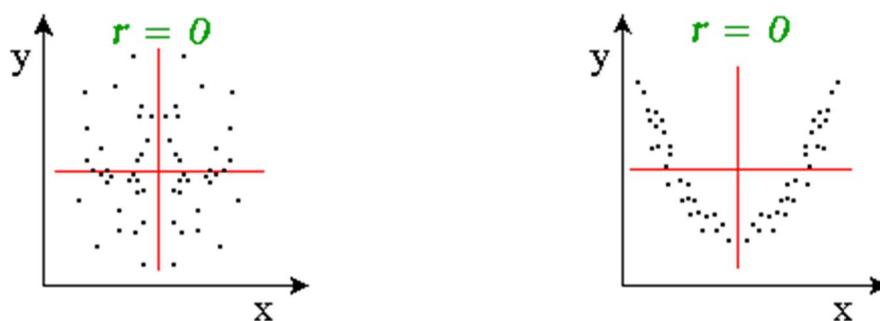


Figure 8. Diagrammes de dispersion des points pour une corrélation nulle.

Le coefficient de corrélation peut être négatif ($-1 < r < 0$) ou alors positif ($0 < r < +1$). On a respectivement une corrélation **négative** (Figure 8) ou

positive (Figure 9). Plus les points se dispersent, plus le coefficient de corrélation est proche de zéro et plus l'intensité de la corrélation diminue. Ainsi, la corrélation peut être **forte** (Figures 8A et 9A), **moyenne** (Figures 8B et 9B) ou **faible** (Figures 8C et 9C).

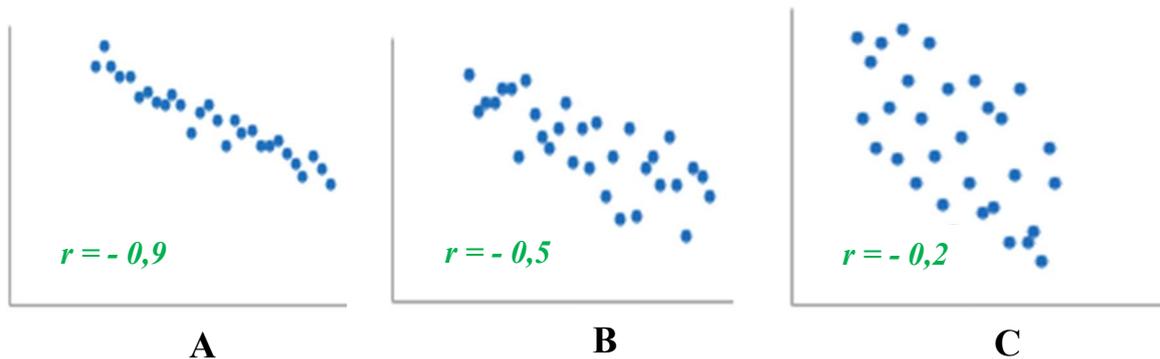


Figure 9. Diagrammes de dispersion des points pour une corrélation négative forte (A), moyenne (B) et faible (C).

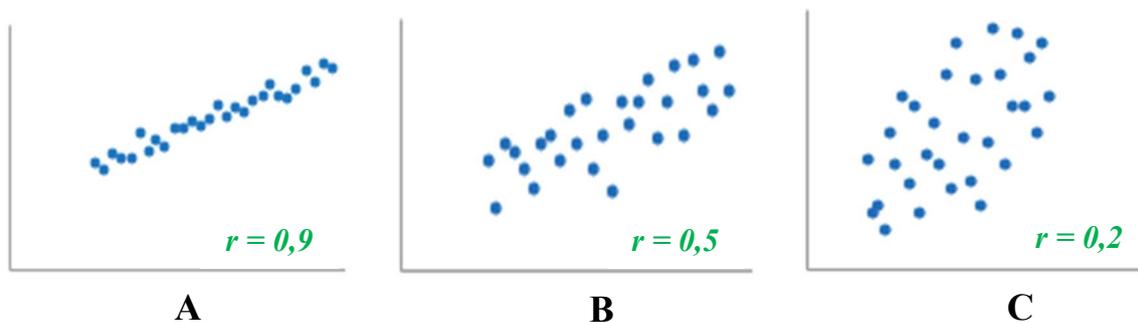


Figure 10. Diagrammes de dispersion des points pour une corrélation positive forte (A), moyenne (B) et faible (C).

Il faut enfin signaler que lorsque $r = 1$ (ou $r = -1$), la **corrélation** est **parfaite** (Figure 10).

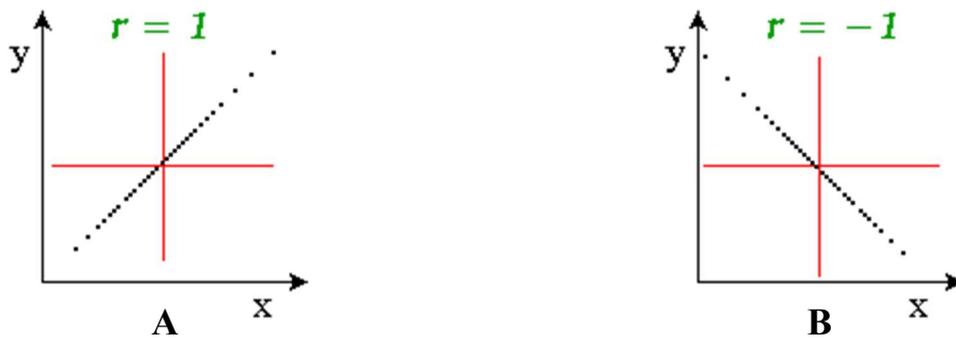


Figure 11. Diagrammes de dispersion des points pour une corrélation parfaite positive (A) et négative (B).

3.2.2. Exemples de corrélations naturelles

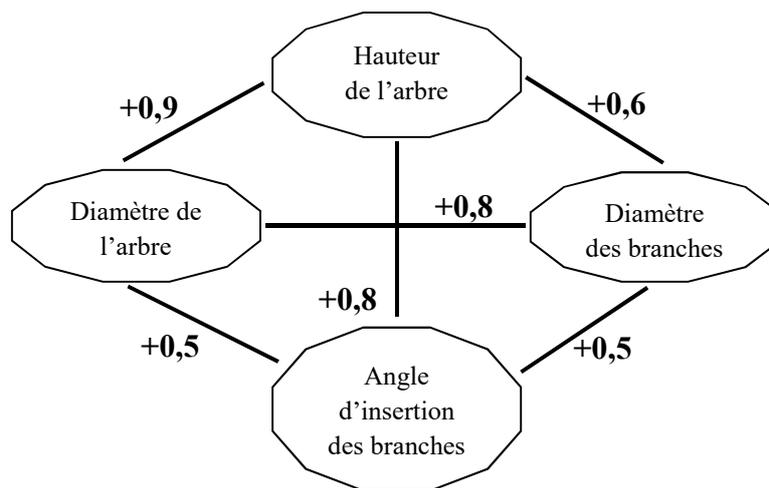


Figure 12. Corrélations génétiques chez le cèdre

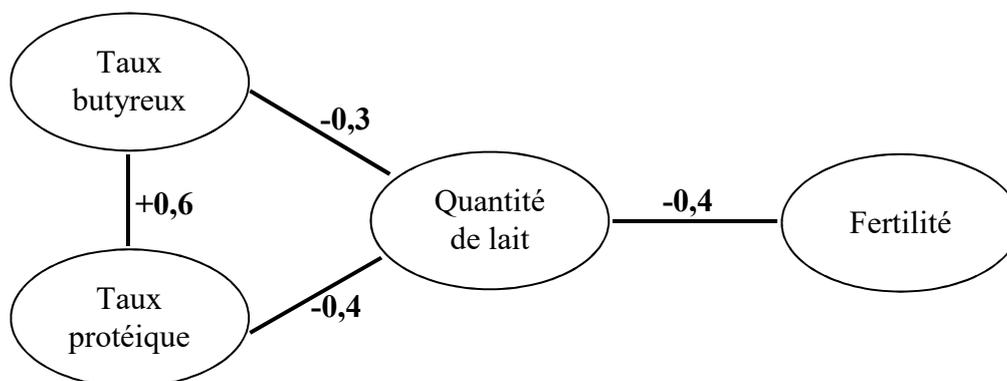


Figure 13. Corrélations génétiques chez la vache laitière.

3.2.3. *Corrélations, écarts et variances génétiques*

Regardons concrètement ce qui se produit au niveau des écarts et des variances si le **génotype** et l'**environnement** sont **corrélés**. En fait, les équations du modèle génétique devraient, pour prendre en compte la possible corrélation entre génotype et environnement (donc interactions génotype x environnement), s'écrire comme suit :

- pour les écarts, $x = g + e + (g \times e)$;
- pour les variances, $V_X = V_g + V_e + V_{(g \times e)}$.

Cette interaction génotype x environnement n'est pas intégrée dans le modèle, tel qu'il a été présenté ci-dessus. En fait, on fait l'**hypothèse implicite** que **celle-ci n'existe pas**.

L'expérience imaginée pour illustrer la corrélation génotype-environnement concerne la course des chevaux de Kentucky. Elle a lieu entre neuf chevaux pur-sang. On suppose que l'on connaît tous les écarts par rapport à la moyenne de la population, à savoir :

- les écarts génétiques (g) du temps nécessaire aux neuf chevaux pour parcourir la distance prévue ;
- les écarts environnementaux (e) de la contribution de leurs entraîneurs respectifs, au temps que chaque cheval met à parcourir la distance de la course. Il est à noter que l'entraînement est considéré comme la seule source de variation environnementale.

Par ailleurs, l'on attribue les meilleurs chevaux aux meilleurs entraîneurs et les plus mauvais chevaux aux entraîneurs les moins qualifiés. Ceci crée une **corrélation** (relation non aléatoire) entre le **génotype** (les chevaux) et l'**environnement** (les entraîneurs). La moyenne de la population de ce groupe de pur-sang pour cette course est de **123 secondes**. Le tableau IV résume les données simulées de cette course.

Tableau IV. Données simulées du temps (en secondes) que les chevaux mettent pour parcourir la distance prévue. X : valeur de l'individu ; g : écart génétique ; e : écart environnemental ; x : écart phénotypique individuel.

| Cheval | Entraîneur | X | g | e | x |
|--------------------|------------|-------------|-------------|-------------|-------------|
| <i>Secretariat</i> | Lucien | 119 | -2 | -2 | -4 |
| <i>Decidedly</i> | Horatio | 120 | -2 | -1 | -3 |
| <i>Barbaro</i> | Mike | 121 | -1 | -1 | -2 |
| <i>Unbridled</i> | Carl | 122 | -1 | 0 | -1 |
| <i>Ferdinand</i> | Charlie | 123 | 0 | 0 | 0 |
| <i>Cavalcade</i> | Bob | 124 | 1 | 0 | 1 |
| <i>Meridian</i> | Albert | 125 | 1 | 1 | 2 |
| <i>Whiskery</i> | Fred | 126 | 2 | 1 | 3 |
| <i>Gallant Fox</i> | Jim | 127 | 2 | 2 | 4 |
| Moyenne (s) | | 123 | | | |
| Variance (s^2) | | 6,67 | 2,22 | 1,33 | 6,67 |

On peut remarquer que V_X (6,67) n'est pas égale à la somme de V_g (2,22) et V_e (1,33). Comme le génotype et l'environnement sont corrélés, l'hypothèse de l'équation qui spécifie que $V_X = V_g + V_e$ a été enfreinte. **Cette équation ne fonctionne donc pas lorsque le génotype et l'environnement sont corrélés.**

Il en est de même pour les écart-types (racine carré de la variance) : les variances s'additionnent ce qui n'est pas le cas pour les écart-types. En effet, l'écart-type phénotypique (**3,83**) calculé à partir des données du tableau IV, n'est pas égal à la somme des écart-types, génétique (**3,46**) et environnemental (**1,63**). Les variances peuvent être décomposées en plusieurs sources. Les **écart-types ne peuvent pas se décomposer de cette manière**. Cette propriété de la variance sera largement exploitée dans la quantification de l'ampleur de l'**héritabilité** de la variation d'un caractère donné.

Lorsque le génotype et l'environnement sont corrélés (ne sont pas indépendants), l'équation $V_X = V_g + V_e$ ne s'applique plus. Le concept de **corrélacion**, c'est-à-dire l'**existence d'une relation entre deux variables** est

critique pour la génétique quantitative. En effet, celle-ci peut faire varier les valeurs phénotypiques des deux variables dans le **même sens** (corrélation **positive**) ou en **sens opposé** (corrélation **négative**).

Conclusion : *lorsque les génotypes sont corrélés aux environnements, la décomposition de la variance phénotypique en variances génétique et environnementale n'est plus possible.*

CHAPITRE 4. HÉRITABILITÉ

Nous savons à présent que les **phénotypes observés** pour un **caractère quantitatif** donné résultent des **apports génétiques** et **environnementaux**. Nous savons à présent décomposer les phénotypes en leurs composantes génétique et environnementale, ceci en **termes d'écarts et de variances**. La question qui sous-tendait cette décomposition était la suivante : quelle fraction de la variation phénotypique est due à des facteurs génétiques et laquelle est due à l'environnement ?

Dans le présent chapitre, il s'agira d'aller plus en profondeur en ce qui concerne les apports génétiques. La variation due aux gènes est-elle prédominante sur celle résultant de l'environnement ? Comment cette variation génétique est-elle transmise au fil des générations ? En d'autres termes, quelle est l'héritabilité d'un caractère complexe donné ? Toute la variation génétique est-elle transmissible ? Des outils mathématiques, dont l'**héritabilité au sens large** et l'**héritabilité au sens strict**, ont été développés pour apporter des réponses à ces questions.

A l'issue de ce chapitre, tout étudiant doit être capable d'estimer l'**héritabilité** d'un caractère, quel qu'il soit. Autrement dit, il doit être à même de « mesurer » l'**influence** de l'**inné** (des **gènes**) sur un phénotype quelconque.

4.1. Héritabilité au sens large : l'inné et l'acquis

Il s'agira ici de préciser l'apport des facteurs innés (**génétiques**) dans la variation observée. Le problème se pose ainsi en termes de **proportion (pourcentage)**. Pour certains sujets controversés comme le QI, les comportements antisociaux (alcoolisme, violence...) ou criminels, le débat « inné contre acquis » est très houleux. Les « pro-génétiques » pensent que les humains sont essentiellement prédéterminés par leurs gènes pour ces caractères,

tandis que pour les autres, c'est l'environnement qui explique la variation comportementale observée dans les populations humaines.

Les réponses à ces questions de l'inné contre l'acquis ont une importance pratique. Si la tension artérielle élevée par exemple est principalement due au style de vie, (environnement), alors des changements du régime alimentaire ou des habitudes d'activité physique seraient les plus appropriés pour adresser le problème. Si par contre ce caractère (tension artérielle élevée) est largement prédéterminé par nos gènes, un traitement médicamenteux serait plus adéquat.

Par ailleurs chez les sélectionneurs (des variétés animales ou végétales), l'appréciation des effets génétiques et environnementaux d'un caractère est essentielle. Si les effets génétiques sont importants, on peut envisager la création de nouveaux génotypes améliorés. Si les effets environnementaux sont les plus importants, il faut plutôt proposer des pratiques agronomiques améliorées.

4.1.1. Définition

L'héritabilité au sens large représente la partie de la variance phénotypique due aux **différences génétiques** entre les individus d'une population. En d'autres termes, elle correspond au **rapport** entre la **variance génétique** et la **variance totale**. Elle mesure donc **toute la variance génétique**.

Elle se calcule en utilisant la formule :
$$H^2 = \frac{V_g}{V_X} \quad (V_X = V_g + V_e)$$

L'héritabilité au sens large peut varier de **0** à **1** ($0 < H^2 < 1$).

Lorsque dans une population **toute la variation** observée a une origine **environnementale**, $H^2 = 0$: il n'y a donc pas de variation génétique dans cette population ($V_g = 0$ et $V_e = V_X = 1$).

Lorsque **toute la variation** observée dans une population a des origines **génétiques**, alors $H^2 = 1$: il n'y a que la variation génétique ($V_g = V_X = 1$ et $V_e = 0$).

L'héritabilité au sens large englobe **plusieurs moyens différents** par lesquels les gènes peuvent contribuer à la variation. Ces diverses contributions peuvent concerner les **apports individuels** des gènes et leur **fonctionnement global** (interactions alléliques et/ou géniques). Il existe donc plusieurs composantes de la variance génétique.

4.1.2. Mesure de H^2 chez les plantes : cas du maïs

Reconsidérons l'étude de cas réel du chapitre précédent (Tableau III). En imaginant un nombre de jours précédant la libération de pollen par des lignées pures de maïs, $V_X = 14,67$ jours², $V_g = 12,00$ jours² et $V_e = 2,67$ jours².

$$\text{Par conséquent, } H^2 = \frac{V_g}{V_X} = \frac{12}{14,67} = 0,82 \quad \text{Soit } \mathbf{82 \%}.$$

Cette valeur indique que les gènes contribuent à la plus grande partie de la variation et que l'environnement intervient pour une part beaucoup plus modeste (18 %). Au vu de ce résultat, le nombre de jours précédant l'apparition du pollen est un caractère **hautement héréditaire chez le maïs**.

Les dix lignées pures de l'expérience référée plus haut (Tableau III) peuvent être placées dans des milieux complètement différents des trois précédents. Les résultats ainsi obtenus sont présentés au Tableau V.

Les variances calculées sont indiquées ci-dessous.

$V_g = 12,00$ jours², comme précédemment : les génotypes étant les mêmes, la variance génétique ne change pas. Par contre, $V_e = 24$ jours², contre 2,67 jours² pour la première expérience. Les environnements de cette seconde expérience semblent donc **plus extrêmes**, d'où une variance environnementale **plus importante**. Dans ce cas, $V_X = 12 + 24 = 36$ jours².

$$\text{Par conséquent, } H^2 = \frac{V_g}{V_X} = \frac{12}{36} = 0,33 \quad \text{Soit } \mathbf{33 \%}.$$

Tableau V. Données simulées du nombre de jours précédents la libération de pollen pour dix lignées de maïs cultivées dans trois environnements. LP : lignées pures (A - J) ; E : environnements (4 - 6) ; M1 : moyenne de toutes les lignées dans un environnement donné ; M2 : moyenne d'une lignée donnée dans les 3 environnements.

| LP | A | B | C | D | E | F | G | H | I | J | M ₁ |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------------|
| E4 | 58 | 60 | 62 | 62 | 64 | 64 | 66 | 66 | 68 | 70 | 64 |
| E5 | 64 | 66 | 68 | 68 | 70 | 70 | 72 | 72 | 74 | 76 | 70 |
| E6 | 70 | 72 | 74 | 74 | 76 | 76 | 78 | 78 | 80 | 82 | 76 |
| M | 64 | 66 | 68 | 68 | 70 | 70 | 72 | 72 | 74 | 76 | 70 |

L'estimation de H^2 se trouve ici du côté réduit, plus proche de 0 que de 1. Cette valeur indique que l'environnement intervient pour une part **beaucoup plus importante** (67 %) que les gènes. On peut donc conclure que le nombre de jours précédant l'apparition du pollen **n'est pas** un caractère **hautement héréditaire chez le maïs** (il est même peu héritable).

Le **contraste** entre ces deux estimations pour le **même groupe** de lignées pures (82 % contre 33 %) souligne le fait que l'**héritabilité est bien la proportion** de la variance phénotypique **due** à la **génétique**. Plus l'environnement a une part importante dans la variation, moins la génétique est importante et plus la valeur de H^2 est faible ; et vice-versa.

H^2 est donc une cible **mouvante** : une **estimation** de H^2 s'applique à une **population précise** et à l'**environnement dans lequel** cette estimation **a été effectuée**. Les résultats d'une étude ne s'appliquent pas automatiquement à une autre étude. H^2 est donc la **propriété d'une population et d'un environnement particulier**.

4.1.3. Mesure de H^2 chez les humains : cas des jumeaux

L'estimation de l'héritabilité souligne la nécessité de disposer de lignées pures, c'est-à-dire d'individus génétiquement identiques. Chez les êtres humains, ce sont les **jumeaux monozygotes** qui sont assimilables aux lignées pures. Toutefois, la majorité des vrais jumeaux sont élevés dans le même foyer, donc dans le même environnement. Les études les utilisant comme modèle doivent impliquer des jumeaux monozygotes ayant grandi séparément. Ceci, pour ne pas enfreindre l'hypothèse du **modèle génétique** selon laquelle les **gènes et l'environnement ne sont pas corrélés**.

Dans ce cas précis, l'estimation de H^2 utilise la **covariance**, qui est **équivalente** à la **variance génétique**. La covariance fait partie des outils statistiques développés par les généticiens pour estimer H^2 à l'aide de la **corrélacion** entre **personnes apparentées** (vrais ou faux jumeaux, frères, sœurs).

$$H^2 = \frac{COV_{X', X''}}{V_X}$$

Les vrais jumeaux partagent 100 % de leurs gènes, alors que les jumeaux dizygotes, des frères ou des sœurs partagent 50 % de leur patrimoine génétique. La force de la corrélation entre ces apparentés peut être pondérée pour la proportion des gènes qu'ils partagent. Les résultats ainsi obtenus peuvent être utilisés pour estimer les contributions génétique et environnementale à la variation des caractères. Différentes études impliquant des personnes apparentées ont renseigné sur bon nombre de caractères humains.

Le Tableau VI indique quelques résultats d'études impliquant des jumeaux. Certains aspects de ce que nous sommes en tant que personnes (taille, cheveux, yeux...) sont fortement influencés par les gènes : ils sont le résultat des processus biochimiques et de développement, sous contrôle génétique. Et dans ce contexte, il n'est pas surprenant que la génétique contribue à la variance des caractères tels que les attitudes sociales ou les croyances religieuses.

Tableau VI. L'héritabilité au sens large pour certains caractères humains.

| Caractère | H^2 |
|---|-------------------------|
| <i>Attributs physiques</i> | |
| Nombre de boucles dans les empreintes digitales | 0,97 |
| Taille | 0,88 |
| Pression sanguine systolique | 0,64 |
| Tour de poitrine | 0,61 |
| Rythme cardiaque | 0,49 |
| Tour de taille | 0,25 |
| <i>Caractéristiques mentales</i> | |
| QI | 0,69 |
| Vitesse de traitement de l'information | 0,56 |
| Vitesse de traitement spatial | 0,36 |
| Vitesse d'acquisition de l'information | 0,20 |
| <i>Troubles psychiatriques</i> | |
| Autisme | 0,90 |
| Schizophrénie | 0,80 |
| Alcoolisme | 0,50 - 0,60 |
| Dépression grave | 0,37 |
| Anxiété | 0,30 |
| <i>Caractéristiques de la personnalité</i> | |
| Extraversion | 0,54 |
| Relations chaleureuses | 0,50 |
| Conscience | 0,49 |
| Névrose | 0,48 |
| Comportement antisocial chez les adultes | 0,41 |
| <i>Croyances et attitudes politiques</i> | |
| Opinion conservatrice à l'âge adulte | 0,45 - 0,65 |
| Avis sur la prière à l'école | 0,41 |
| Avis sur le pacifisme | 0,38 |
| Pratique d'une religion à l'âge adulte | 0,30 - 0,45 |

Les jumeaux utilisés dans de nombreuses études ont été séparés dès la naissance et placés dans des familles adoptives. Cependant, ces bébés sont adoptés par des foyers émotionnellement, socialement et économiquement stables. Dans ces cas, V_e est sous-estimée, tandis que V_g est surestimée. On aboutit donc à une **surestimation** de H^2 .

Il faut enfin signaler que **H^2 n'est pas utile** pour interpréter les différences dans la moyenne des caractères **entre des populations**. Lorsque H^2 est élevée, comme c'est le cas pour la taille (0,88), cette valeur ne renseigne pas sur l'origine de la différence de taille dans le groupe.

Aux alentours de 1800, les hollandais mesuraient en moyenne 168 cm, contre 184 cm actuellement. Les 16 cm de différence ne peuvent pas s'expliquer par la génétique : en effet, le pool de gènes de cette population ne peut pas avoir changé de manière appréciable en 200 ans. Cependant, les améliorations en matière de santé et de nutrition peuvent aisément expliquer cette différence.

Conclusion : *l'héritabilité au sens large (H^2) indique la proportion de la variance dans une population due aux facteurs génétiques et ce, en une seule génération. Elle exprime donc le degré auquel les différences phénotypiques entre individus d'une génération donnée sont déterminées par des différences dans leurs génotypes. Les estimations de H^2 (V_g/V_X) s'appliquent à la population et à l'environnement dans lequel elles ont été effectuées. H^2 n'est donc pas utile pour interpréter les différences dans la moyenne des caractères entre des populations.*

4.2. Héritabilité au sens strict : prédiction des phénotypes

La variation génétique observée à une génération donnée et mesurée à l'aide de H^2 peut ne pas exister à la génération suivante. Si elle existe, elle peut ne pas être au même niveau. Sa transmission à la génération suivante peut ne pas être prévisible. Mesurer la variation génétique d'un parent à l'aide de H^2 ne

permet donc pas de prévoir la variation transmise à ses descendants. L'**héritabilité au sens strict** (h^2) le permet. Toutefois, avant de présenter cette seconde forme d'héritabilité, il est nécessaire de rappeler les différents modes d'action des gènes qui sont au cœur de la compréhension de cette notion.

4.2.1. Action des gènes et transmission de la variation

Considérons par exemple un cultivateur qui livre régulièrement des fleurs (roses, becs de canard...) à des revendeurs. Le nombre de fleurs par plante représentant un élément important dans sa production, il cherchera donc naturellement à sélectionner et à reproduire les variétés ayant un grand nombre de fleurs. Supposons que le nombre de fleurs par plante soit contrôlé par un locus ayant deux allèles, B_1 et B_2 . Il existe donc trois génotypes à ce locus, B_1B_1 , B_1B_2 et B_2B_2 , qui peuvent, selon les interactions inter-alléliques, conduire à des phénotypes différents.

Supposons que les plantes de génotypes B_1B_1 , B_1B_2 et B_2B_2 aient respectivement une, deux et trois fleurs. La valeur de l'hétérozygote est à « mi-chemin » entre celles des homozygotes : l'**action** du gène est définie comme étant **additive** (Figure 14A).

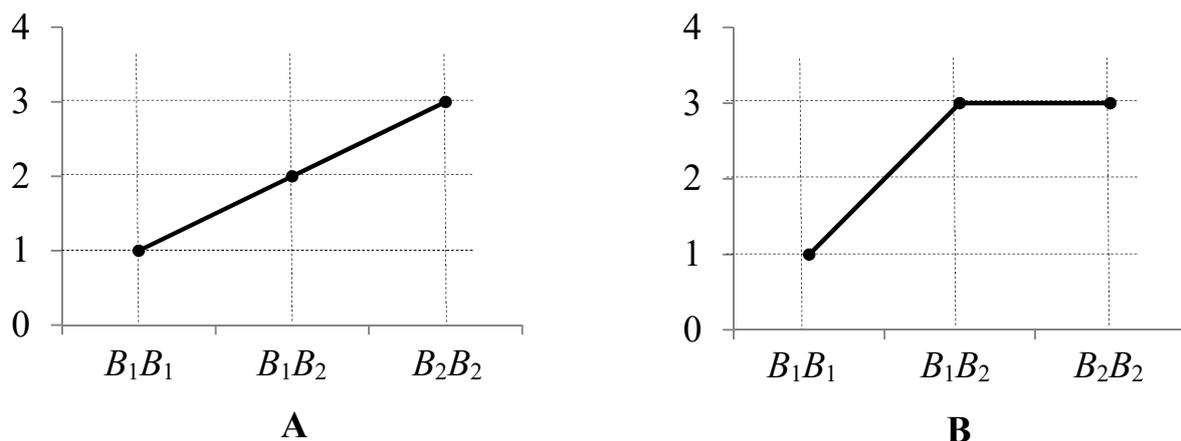


Figure 14. Représentation graphique des actions additive (A) et dominante (B) d'un gène donné.

Par contre si la valeur de l'hétérozygote vaut celle de l'homozygote B_2B_2 , c'est-à-dire trois fleurs, l'allèle B_2 est dominant par rapport à l'allèle B_1 . Dans un cas comme celui-ci, l'**action** du gène est qualifiée de **dominante** (Figure 14B). Si l'on prend plutôt en compte l'allèle B_1 qui est récessif par rapport à l'allèle B_2 , l'action du gène pourrait être définie comme **récessive**.

Il faut signaler que l'action des gènes n'est pas nécessairement purement additive ou purement dominante ; elle peut également être **partiellement dominante**. Si par exemple l'hétérozygote produit en moyenne 2,5 fleurs/plante, l'allèle B_1 présenterait une **dominance partielle**.

4.2.2. Action des gènes et transmission de la variation génétique

Considérons toujours l'exemple du cultivateur de fleurs, le nombre de fleurs étant contrôlé par les allèles B_1 et B_2 , de même fréquence (0,5). Supposons également que la population constituée soit en équilibre (0,25 ; 0,50 ; 0,25) et que le cultivateur veuille créer une population améliorée avec davantage de fleurs par plante.

Reconsidérons également les deux cas traités dans le mode d'action des gènes (gène déterminant le nombre de fleurs par plante) à savoir l'action additive (Figure 14A) et l'action de dominance (Figure 14B). Supposons enfin qu'il n'y ait pas d'effet environnemental : seul le génotype détermine le nombre de fleurs. Dans ce cas, $H^2 = 1$.

4.2.2.1. Cas de la dominance

Le nombre moyen de fleurs par plante dans la nouvelle population (avant les croisements) est de 2,5 (Tableau VII). Ce nombre moyen de fleurs par plante n'est pas à son maximum, puisqu'il est de 3 pour les meilleurs génotypes.

Tableau VII. Nombre moyen de fleurs par plante dans une population en équilibre génétique dans le cas de la dominance.

| Génotype | Fréquence | Valeur du Caractère | Contribution à la moyenne (fréquence x valeur) |
|----------|-----------|---------------------|---|
| B_1B_1 | 0,25 | 1 | $0,25 \times 1 = 0,25$ |
| B_1B_2 | 0,50 | 3 | $0,50 \times 3 = 1,50$ |
| B_2B_2 | 0,25 | 3 | $0,25 \times 3 = 0,75$ |
| | | | Moyenne = 2,50 |

Si le cultivateur veut obtenir le maximum de fleurs dans les descendance à créer, il devra uniquement sélectionner au sein de la population dont il dispose, les plantes à 3 fleurs. $2/3$ de ces plantes ont le génotype B_1B_2 et $1/3$ le génotype B_2B_2 . Dans ce cas, $4/9$ ($2/3 \times 2/3$) des croisements entre les plantes ainsi sélectionnées impliqueront des hétérozygotes. Ceux-ci engendreront $1/4$ de descendants de génotype B_1B_1 (donc à 1 fleur) et les autres $3/4$ de descendants de génotypes B_1B_2 ou B_2B_2 (donc à 3 fleurs).

La moyenne globale pour ces descendants est de **2,78** alors que celle des parents est de **3**. La sélection telle que pratiquée par le cultivateur fonctionne, mais pas de manière parfaite : **en cas de dominance, le phénotype n'est pas pleinement héritable.**

4.2.2.2. Cas de l'additivité

Le nombre moyen de fleurs par plante dans la nouvelle population (avant les croisements et toujours en équilibre génétique) est de 2 (Tableau VIII). Ce nombre moyen de fleurs par plante n'est pas à son maximum, puisqu'il est de 3 pour les meilleurs génotypes. Toutefois, cette moyenne correspond à la valeur de l'hétérozygote.

Tableau VIII. Nombre moyen de fleurs par plante dans une population en équilibre génétique dans le cas de l'additivité.

| Génotype | Fréquence | Valeur du Caractère | Contribution à la moyenne (fréquence x valeur) |
|----------|-----------|---------------------|---|
| A_1A_1 | 0,25 | 1 | $0,25 \times 1 = 0,25$ |
| A_1A_2 | 0,50 | 2 | $0,50 \times 2 = 1,00$ |
| A_2A_2 | 0,25 | 3 | $0,25 \times 3 = 0,75$ |
| | | | Moyenne = 2,00 |

Si le cultivateur veut obtenir le maximum de fleurs dans les descendance à créer, il devra uniquement sélectionner au sein de la population dont il dispose, les plantes à 3 fleurs, qui ont le génotype B_2B_2 . Celles-ci, croisées entre elles engendreront des plantes toutes à 3 fleurs. C'est l'objectif recherché.

La moyenne globale pour ces descendants est donc égale à celle des parents croisés, c'est-à-dire **3 fleurs/plante**. La sélection telle que pratiquée par le cultivateur fonctionne parfaitement : **en cas d'additivité, le phénotype est pleinement héritable.**

Conclusion : lorsqu'il y a dominance, on ne peut pas prédire les phénotypes des descendants à partir des phénotypes des parents. Une partie de la variation observée entre les parents est due aux interactions de dominance entre allèles. Puisque les parents transmettent à leurs descendants leurs gènes et pas leurs phénotypes, ces interactions de dominance ne sont pas transmises. La prédiction des phénotypes des descendants n'est possible qu'avec l'additivité.

4.2.3. Effets additifs et effets de dominance

Il faut être en mesure de quantifier la dominance et l'additivité. Considérons une fois de plus le locus B qui contrôle le nombre de fleurs sur une plante (mode d'action des gènes), avec le cas d'une action dominante et le cas d'une action additive.

L'**effet additif (A)** mesure le degré de changement dans le phénotype qui se produit lorsque l'allèle B_1 est remplacé par l'allèle B_2 . Il se calcule comme étant la moyenne de la différence entre les deux homozygotes (en général les parents), c'est-à-dire

$$A = \frac{X_{B_2B_2} - X_{B_1B_1}}{2}$$

Quant à l'**effet de dominance (D)**, il représente l'écart de l'hétérozygote par rapport au point central (ou encore à la moyenne) des deux homozygotes, c'est-à-dire

$$D = X_{B_1B_2} - \frac{X_{B_2B_2} + X_{B_1B_1}}{2}$$

Il est par ailleurs possible de mesurer le **degré de dominance**, ceci en utilisant le rapport entre l'effet de dominance et l'effet additif : **D/A**.

Si l'on calcule les effets additifs et les effets de dominance (A et D) pour les deux situations décrites à la figure 14, on aura :

- Pour la figure 14A : $A = \frac{3-1}{2} = 1$; $D = 2 - \frac{3+1}{2} = 2 - 2 = 0$. Dans ce cas,

le rapport **D/A = 0** : ceci indique une **additivité pure** et une **absence totale de la dominance**, c'est-à-dire **A = 1** et **D = 0**.

- Pour la figure 14B : $A = \frac{3-1}{2} = 1$; $D = 3 - \frac{3+1}{2} = 3 - 2 = 1$. Dans ce cas,

le rapport **D/A = 1** : ceci indique une **dominance complète**, c'est-à-dire **A = 1** et **D = 1**.

Un rapport **D/A = -1** indiquerait une récessivité complète. La **distinction** entre dominance et récessivité dépend donc de la façon dont les phénotypes sont codés et en ce sens, elle est **arbitraire**. Pour $0 < D/A < 1$, il s'agira de **dominance partielle** et pour $-1 < D/A < 0$, il s'agira de **récessivité partielle**. Illustrons ces différentes notions par des exemples.

Autre exemple

Chez les épinoches, un type de poissons, il existe des populations marines avec de longues épines pelviennes et des populations qui vivent près des

fonds des lacs d'eau douces à épines pelviennes fortement réduites. L'on a étudié la ségrégation dans une F₂, des deux allèles (*l* pour long et *s* pour 'short') de l'un des gènes (*Pitx1*) impliqués dans ce caractère, à savoir la taille des épines pelviennes. Les données ci-dessous représentent les valeurs moyennes en termes de proportion par rapport à la longueur totale du corps. Les calculs des effets additifs (A) et de dominance (D), ainsi que du rapport A/D sont présentés.

| <i>s/s</i> | <i>s/l</i> | <i>l/l</i> | |
|------------|------------|------------|--|
| 0,068 | 0,132 | 0,148 | $A = (0,148 - 0,068)/2 = 0,04$, soit 4 % de la longueur du corps $D = 0,132 - [(0,148 + 0,068)/2] = 0,024$ $D/A = 0,024/0,04 = 0,6 \leftrightarrow 0 < D/A < 1$ |

La valeur de 0,6 indique que l'allèle *l* est **partiellement dominant** sur l'allèle *s*.

L'exemple ci-dessus présente les calculs effectués à partir des valeurs concernant un seul des gènes impliqués dans le caractère étudié. Il est possible, avec des données appropriées, de calculer la moyenne des effets additifs et de dominance pour tous les gènes du génome qui affectent un caractère donné. Les valeurs ci-dessous sont celles du diamètre moyen de l'œil des poissons de grotte d'une part et des poissons apparentés de surface d'autre part. Les populations vivant dans des grottes ont des yeux fortement réduits (faible diamètre) comparés à ceux des populations de surface.

| Grotte | F ₁ | Surface | |
|--------|----------------|---------|---|
| 2,10 | 5,09 | 7,05 | $A = (7,05 - 2,10)/2 = 2,48$ $D = 5,09 - [(7,05 + 2,10)/2] = 0,52$ $D/A = 0,52/2,48 = 0,21 \leftrightarrow 0 < D/A < 1$ |

Au chiffre rond, l'action des gènes pour ce caractère est **proche** d'un état **purement additif**. Toutefois, comme **D/A** est quand même **différent de zéro**, le **génotype des populations de surface** pour le diamètre des yeux est donc **partiellement dominant**.

Conclusion : lorsque la valeur d'un caractère pour la classe hétérozygote est à mi-chemin entre celles des deux classes homozygotes, l'action

des gènes est qualifiée d'additive. Tout écart de l'hétérozygote par rapport à la valeur centrale entre les deux homozygotes indique un degré de dominance d'un des allèles impliqués. Les effets additifs (A) et de dominance (D), ainsi que leur rapport (D/A) fournissent une mesure pour quantifier le mode d'action des gènes.

4.2.4. Modèle génétique pour l'additivité et la dominance

Lorsque l'on veut prédire les phénotypes des descendants, il est important de **séparer** les contributions **additives** et de **dominance**. Pour cela, il faut modifier le modèle génétique présenté au Chapitre 3, à savoir $x = g + e$.

Reprenons le cas de la dominance étudié pour le livreur de fleurs. Les individus de génotypes B_1B_2 et B_2B_2 ont le même phénotype, 3 fleurs/plante. Si l'on laisse s'autopolliniser un individu B_2B_2 , toute la descendance sera B_2B_2 et la valeur moyenne du caractère de ces descendants sera de 3. Par contre, l'autopollinisation d'un individu B_1B_2 engendrera $\frac{1}{4} B_1B_1$, $\frac{1}{2} B_1B_2$ et $\frac{1}{4} B_2B_2$. La valeur moyenne du caractère de ces descendants sera de 2,75.

Même si les génotypes B_1B_2 et B_2B_2 ont le même phénotype, ils ne produisent pas des descendants équivalents, car l'origine de leur phénotype est différente. Le phénotype de B_1B_2 dépend de l'effet de dominance (D) alors que celui de B_2B_2 n'implique pas la dominance. Non seulement ils ont le même phénotype, mais ils ont également le même écart génotypique par rapport au génotype B_1B_1 :

$$\frac{X_{B_2B_2} - X_{B_1B_1}}{2} = \frac{X_{B_1B_2} - X_{B_1B_1}}{2} = \frac{3-1}{2} = 1.$$

Le modèle $x = g + e$ peut être étendu pour incorporer les contributions additive et de dominance : l'écart génotypique (g) est la somme de l'écart additif (a) et de l'écart de dominance (d). $g = a + d$. Dans ce cas le modèle simple peut s'écrire $x = a + d + e$. L'écart additif est transmis de manière prévisible du parent au descendant. L'écart de dominance n'est pas transmis car à chaque génération, de nouveaux génotypes, et donc de nouvelles interactions entre les

allèles sont créées. Les écarts génotypiques (g) sont calculés en soustrayant la moyenne de la population (2,5) de la valeur du caractère pour chaque génotype.

La variance génétique (V_g) peut également être répartie en variance additive (V_a) et variance de dominance (V_d) : $(V_g) = (V_a) + (V_d)$ et la variance phénotypique (V_X) peut globalement s'écrire $(V_X) = (V_a) + (V_d) + (V_e)$. (V_a) représente la variance des écarts additifs, autrement dit la partie de la variance génétique qui est transmise des parents à leur descendance et (V_d) est la variance des écarts de dominance.

NB : l'équation $(V_X) = (V_a) + (V_d) + (V_e)$ suppose que les composantes additive et de dominance ne sont pas corrélées.

Conclusion : la variance génétique (g) d'un individu par rapport à la moyenne de la population est composée de deux parties – son écart additif (a) et son écart de dominance (d). Seul l'écart additif d'un individu est transmis à ses descendants. De même, seule la variance additive, V_a , fraction de la variation génétique, est transmise du parent au descendant.

4.2.5. Définition de l'héritabilité au sens strict

L'héritabilité au sens strict, qui se symbolise par h^2 est définie comme le rapport de la variance additive à la variance phénotypique totale.

$$h^2 = \frac{V_a}{V_X} \quad \text{avec } V_X = V_a + V_d + V_e$$

Elle mesure le degré auquel la variation entre les individus d'une population est transmise de façon prévisible à leurs descendants.

L'héritabilité au sens strict peut varier de **0** à **1** ($0 < h^2 < 1$).

Lorsque $h^2 = 1$ ($V_a = V_X$), les phénotypes des descendants sont exactement égaux à la valeur correspondant aux deux parents. **Toute la variation** de la population est **additive** et **héritable** au sens strict.

Lorsque $h^2 = 0$ ($V_a = 0$), la valeur attendue pour le phénotype de n'importe quel descendant sera la moyenne de la population. **Toute la variation**

observée dans la population est due, soit à des facteurs de dominance, soit à des facteurs environnementaux. Celle-ci n'est donc **pas transmissible** à la descendance. Seule h^2 est **transmissible** et **accessible** à la **sélection** naturelle ou artificielle : elle permet de **prédire** si la sélection génétique sera efficace. On distingue ainsi :

- des caractères **peu héritable**s, pour lesquels [$h^2 < 0,2$] ;
- des caractères **moyennement héritable**s, avec [$0,2 < h^2 < 0,4$] ;
- des caractères **fortement héritable**s, pour lesquels [$h^2 > 0,4$].

En définitive, pour estimer h^2 , il faut impérativement mesurer la variance additive (V_a). Comment ? Il existe des méthodes développées par les généticiens pour mesurer cette V_a et ainsi estimer h^2 . Celles-ci sont spécifiques aux populations étudiées, humaines ou non.

4.2.6. Estimation de h^2 chez les êtres humains

Il a été possible d'évaluer H^2 chez les humains à partir de la covariance entre deux jumeaux monozygotes élevés séparément, dans des foyers adoptifs. Il est également possible d'estimer h^2 à l'aide de la covariance entre un parent (P) et son descendant (O, *offspring*), qui est égale à la moitié de la variance additive. $\text{COV}_{P,O} = 1/2 V_a$

Ceci s'explique par le fait que chaque descendant ne reçoit que la moitié des gènes de l'un de ses parents et l'autre moitié de l'autre parent. En combinant la formule ci-dessus avec celle de h^2 , on obtient :

$$h^2 = \frac{V_a}{V_X} = \frac{2\text{COV}_{P,O}}{V_X}$$

La mesure de V_a nécessite de contrôler les facteurs environnementaux dans les expériences. Mais les parents et les descendants ne sont pas élevés au même moment. Pour pallier à cet inconvénient, V_a est mesurée entre frères et sœurs consanguins ; dans ce cas, tous les individus de l'expérience peuvent être élevés au même moment et dans le même environnement. Ces demi-frères et

demi-sœurs partagent $\frac{1}{4}$ de leurs gènes, de sorte que V_a est égale à 4 fois la covariance entre demi-frère et demi-sœur.

L'héritabilité au sens strict de la taille chez les humains a des valeurs autour de 0,8, ce qui signifie que la variance est à 80 % additive, donc transmissible d'un parent à ses descendants. Toutefois, les résultats d'autres estimations de h^2 peuvent s'écarter de cette valeur, à cause, entre autres, de l'échantillonnage qui peut être petit et peu représentatif de la population dans laquelle la valeur de 0,8 a été obtenue ; des corrélations éventuelles entre les environnements des parents et ceux de leurs descendants.

4.2.7. Prédiction des phénotypes des descendants

L'héritabilité au sens strict est un concept important pour les agriculteurs (cultivateurs et éleveurs) : elle indique respectivement quels caractères peuvent être améliorés par sélection artificielle. h^2 est également essentielle pour les biologistes de l'évolution : elle permet de comprendre comment les populations changeront en réponse à la sélection naturelle imposée dans un environnement changeant.

Pour améliorer efficacement la réserve des caractères ayant une importance agronomique, l'agriculteur doit être capable de **prédire le phénotype** d'un **descendant** à partir de celui de ses parents. Pour cela, connaître les **écarts additifs des parents** et h^2 desdits caractères est essentiel. Comme il a déjà été indiqué plus haut :

- l'écart phénotypique individuel s'écrit $x = a + d + e$;
- les écarts de dominance (d) et environnemental (e) **ne sont pas transmis** à la descendance ;
- la **partie transmissible** aux descendants est l'**écart additif, a** .

L'on peut croiser des parents dont les écarts phénotypiques respectifs sont :

$$x' = a' + d' + e' \quad \text{et} \quad x'' = a'' + d'' + e''$$

L'écart phénotypique de la descendance (*offspring*) peut être exprimé comme la **moyenne de l'écart additif** des deux parents, puisque l'additivité est la seule partie transmissible : $x_{off} = (a' + a'')/2 = \bar{a}_{par}$.

Cependant, il n'est pas possible d'observer directement les écarts additifs des parents ; ils peuvent uniquement **être estimés** (\hat{a}), sachant que l'écart additif est la seule partie héritable du phénotype d'un individu : $\hat{a} = xh^2$.

De ce fait, la moyenne des écarts additifs des parents peut être estimée, comme étant le produit entre h^2 et la **moyenne** de leurs écarts phénotypiques :

$$\hat{x}_{off} = h^2[(x' + x'')/2] \quad \text{ou alors} \quad \hat{x}_{off} = h^2\bar{x}_{par}$$

4.2.8. Sélection sur des caractères complexes

La sélection artificielle est un processus dans lequel seuls les individus possédant certaines caractéristiques contribuent au pool génétique qui forme la génération suivante. Lorsqu'elle est appliquée par l'homme, elle est qualifiée de **sélection artificielle**, par opposition à la sélection naturelle.

L'on peut par exemple vouloir améliorer le contenu en provitamine A (précurseur dans la biosynthèse de la vitamine A) d'une population de maïs.

L'évaluation de la moyenne courante de la production de provitamine A de cette population indique une valeur $\mu = 1,25 \pm 0,24 \mu\text{g/g}$ de grains de maïs. Les plantes qui produisent **au minimum 1,5 μg** de provitamine A sont sélectionnées dans l'optique d'améliorer ce caractère, constituant ainsi un groupe de meilleures plantes productrices de la population. La moyenne pour le groupe sélectionné ($\mu' = 1,63 \mu\text{g/g}$) est supérieure à la moyenne de la population générale. Par ailleurs, la production de la descendance ($\mu'' = 1,44 \mu\text{g/g}$) qu'engendrent ces plantes « supérieures » dépasse celle de la population globale (Figure 15).

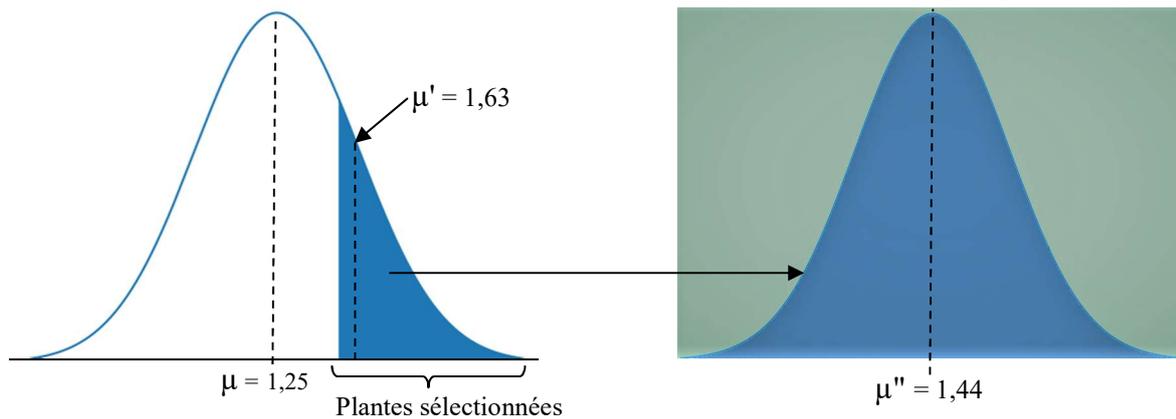


Figure 15. Distribution des valeurs pour la provitamine A dans les grains de maïs dans une population de départ (A) et une population de descendants (B) après une génération de sélection

Il faut enfin signaler que h^2 peut être évaluée à partir des données de l'expérimentation indiquées ci-dessus (production de la provitamine A). Dans ce cadre, la formule $\bar{x}_{off} = h^2 \bar{x}_{par}$ peut être réécrite comme suit $h^2 = \bar{x}_{off} / \bar{x}_{par}$.

Rappelons que \bar{x}_{par} est l'écart moyen des parents (ici les plantes sélectionnées) par rapport à la moyenne de la population. Il est également désigné **différentiel de sélection (S)** et vaut : $\bar{x}_{par} = 1,63 - 1,25 = \mathbf{0,38} = \mathbf{S}$.

De même, \bar{x}_{off} est l'écart moyen des descendants par rapport à la moyenne de la population. Il est encore appelé **réponse à la sélection (R)** et dans notre exemple, il vaut : $\bar{x}_{off} = 1,44 - 1,25 = \mathbf{0,19} = \mathbf{R}$. Avec ces deux valeurs, h^2 pour la production de la provitamine A dans cette population de maïs peut être calculée : $h^2 = \bar{x}_{off} / \bar{x}_{par} = R/S = 0,19/0,38 = \mathbf{0,5}$.

Conclusion : l'héritabilité au sens strict (h^2) indique la proportion de la variance phénotypique que l'on peut attribuer à des effets additifs. Cette forme d'héritabilité mesure le degré auquel la variation entre les individus est transmise de façon prévisible à leurs descendants. Les estimations de h^2 (V_d/V_X ou $2COV_{P,O}/V_X$ ou encore R/S) s'appliquent à la population et à l'environnement

dans lequel elles ont été effectuées. h^2 tient une place essentiel dans la sélection des caractères, aussi bien chez les plantes que les animaux.

4.3. Méthode pratique d'analyse génétique : mesure de l'héritabilité

L'analyse génétique des générations parentales (P), F_1 , F_2 et RC_1 encore appelée GMA (*Generations Means Analysis*) permet, entre autres, d'étudier les valeurs de l'héritabilité, le **nombre minimum de facteurs** (gènes) et le **mode d'action des gènes** impliqués dans un **caractère quantitatif**.

C'est une méthode d'analyse qui demande peu d'opérations d'hybridations contrôlées et fournit plus de renseignements quant au mode d'action des gènes (en détaillant les effets des diverses interactions intergéniques); mais elle exige l'analyse d'un **nombre élevé d'échantillons** pour chaque génération. Cette méthode d'analyse consiste à générer plusieurs types de générations, au travers de croisements divers, comme l'indique la Figure 16.

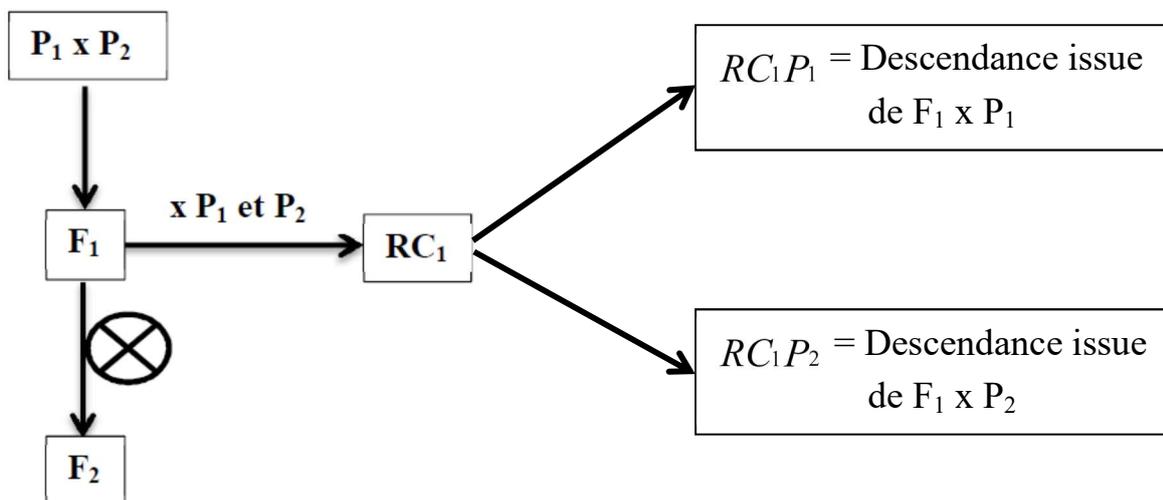


Figure 16. Obtention d'hybrides F_1 et des descendants F_2 et RC_1 .

Une fois que l'on dispose de ces cinq types de générations, l'évaluation se passe dans des conditions homogènes. Comme dans notre cas il s'agit de végétaux, les plantes de chacune de ces cinq générations sont plantées dans le même environnement, avec plusieurs répétitions. Les valeurs de diverses

variances et les comparaisons des générations permettent d'estimer les héritabilités.

Chez les plantes autogames comme le niébé [*Vigna unguiculata* (L.)] et dans les conditions de culture homogène, les **variations observées** entre les individus **parentaux homozygotes** ou entre les plantes **F₁ homogènes** (tous les individus de cette génération sont hybrides, les parents étant homozygotes) sont d'origine **environnementale** [**variance environnementale** (V_e)]. Par contre, entre les individus de la génération F₂, les différences notées sont d'origine **environnementale et génétique** [**variance phénotypique** (V_X)].

Rappel : $V_X = V_g + V_e$; dans ce cas, $V_g = V_X - V_e$.

En utilisant la GMA, il est ainsi possible de mesurer le degré relatif auquel un caractère donné est transmis, du parent à la progéniture, en évaluant les héritabilités (H^2 et h^2).

Diverses formules sont utilisées dans l'évaluation génétique par la GMA

$$V_X = V_g + V_e$$

$$V_g = V_X - V_e$$

$$H^2 = \frac{V_g}{V_X}$$

$$h^2 = \frac{V_a}{V_X}$$

$$GS = k \times \sqrt{V_X} \times h^2$$

$$V_X = V_{F_2}$$

$$V_e = \frac{V_{F_1} + V_{P_1} + V_{P_2}}{3}$$

$$V_a = 2V_{F_2} - (V_{RC_1P_1} + V_{RC_1P_2})$$

$$[a] = RC_1P_1 - RC_1P_2$$

$$[d] = [-0,5P_1 - 0,5P_2] + F_1 - 4F_2 + [2RC_1P_1 + 2RC_1P_2]$$

V_X = Variance phénotypique

V_e = Variance environnementale

V_g = Variance génétique totale

V_a = Variance génétique due aux effets additifs

V_{P_1} = Variance du parent P₁

V_{P_2} = Variance du parent P₂

V_{F_1} = Variance de la F₁

V_{F_2} = Variance de la F₂

$V_{RC_1P_1}$ = Variance de la RC₁P₁

$V_{RC_1P_2}$ = Variance de la RC₁P₂

[a] = additivité

[d] = dominance

H^2 = héritabilité au sens large

h^2 = héritabilité au sens strict

GS = gain de sélection

k = différentiel de sélection pour une intensité de sélection de 10 % ; ici, k = 1,75

Conclusion : l'héritabilité est une donnée statistique évaluant la part de la génétique dans la variation de l'expression phénotypique d'un caractère mesurable au sein d'une population donnée. Elle correspond ainsi à la proportion de la variance phénotypique qui est due à la variance génétique. Il existe deux expressions bien distinctes pour cette grandeur : l'héritabilité au sens large (H^2) et l'héritabilité au sens strict (h^2). L'héritabilité (H^2 ou h^2) n'est pas la propriété d'un caractère, mais la propriété d'une population dans un milieu donné. Par conséquent, un caractère peut avoir une héritabilité nulle pour une population et un environnement donnés. En outre, lorsque l'héritabilité est forte pour un caractère donné, cela ne signifie pas que celui-ci est fortement déterminé par les gènes, mais que les différences observées entre les individus pour cette population et dans ce milieu sont associées aux différences génétiques qui existent entre les individus de cette population. L'héritabilité ne représente donc pas une mesure absolue pour un caractère chez une espèce.

Document de base utilisé :

Griffiths A. J. F., Wessler S. R., Carroll S. B., Doebley J. 2013. *Introduction à l'analyse génétique*. 6^e édition. De Boeck. Bruxelles. 830 p.