

# Apunts de Genètica Quantitativa

02/10/2023

## Introducció

El desenvolupament de la genètica quantitativa en la primera meitat del segle XX va contribuir a l'èxit de les campanyes de millora genètica, les quals, junt amb la mecanització i l'ús d'agroquímics, van possibilitar el major increment de la productivitat agropecuària de la història de la humanitat: la *revolució verda*. Actualment, l'aparell estadístic de la genètica quantitativa es combina amb l'abundància de dades genòmiques per continuar augmentant la productivitat agropecuària. El desenvolupament de mètodes com la selecció genòmica i els estudis d'associació genòmica fan que els conceptes bàsics de genètica quantitativa siguin fonamentals per entendre i participar en el progrés de la biomedicina, la biologia evolutiva i la biotecnologia.

L'esquema a seguir per introduir aquests conceptes és el següent:

1. Un gen. Amb el cas més simple possible es pretén explicar en què consisteix la descomposició del valor genotípic, abans fins i tot de saber per a què serveix esta descomposició.
2. La població. Només un incís per observar que la descomposició del valor genotípic només té sentit en referència a una població concreta. El valor reproductiu d'un individu, així com l'efecte de la dominància no són propietats inherents del gen, ni de l'individu en sí, sinó del gen o l'individu *i de la població*.
3. Anàlisi de la variància genètica. Una primera motivació per haver descomposat el valor genotípic és la partició de la variància genètica en variància additiva i variància de dominància.
4. Herència poligènica. La regressió lineal ens permet modelar també els efectes simultanis de diversos gens sobre un caràcter quantitatiu. Però l'anàlisi es complica, perquè en molts casos hi ha epístasi: relacions no lineals entre l'acció dels gens i el caràcter.
5. Fonts de variació ambiental. El fenotip depèn de factors genètics i ambientals, i també és possible incorporar l'efecte de l'ambient en una regressió lineal més general. L'objectiu és comprendre el significat dels termes en esta regressió.
6. Heretabilitat. Definim l'heretabilitat i poc més.

## Un gen

Normalment la genètica quantitativa és la genètica dels caràcters **poligènics**, tant si el fenotip es pot mesurar en una escala contínua com si és binari (e.g., malalt o sa). Però imaginem ara que el caràcter determinat per aquest únic locus bial·lèlic es mesura en una escala contínua. Si cap altre gen ni l'ambient influeixen en el caràcter, hi hauria dos (amb dominància completa) o tres fenotips possibles.

El valor del fenotip determinat pel genotip es representa amb el símbol  $G$ , i se l'anomena *valor genotípic*, però es mesura en unitats de **fenotip**. L'escala en què es mesura el fenotip és arbitrària. Per exemple, dues escales útils per als càlculs teòrics són les següents (reproducció de la figura 4.4 de Lynch and Walsh (1998), p. 62):

Genotip	$B_1B_1$	$B_1B_2$	$B_2B_2$
Valor genotípic	$-a$	$d$	$+a$

Genotip	$B_1B_1$	$B_1B_2$	$B_2B_2$
Valor genotípic	$0$	$(1+k)a$	$2a$

En presència de dominància, la relació entre el valor genotípic i el nombre d'al·lels d'un cert tipus no és lineal. Però a Fisher (1918) se li va ocórrer la bona idea de considerar una aproximació lineal a la relació entre el contingut gènic (nombre d'al·lels  $B_2$ , per exemple) i el valor genotípic,  $G$ . Aquesta *regressió* permet descomposar conceptualment el valor genotípic en una part atribuïble a l'efecte *additiu* o heretable dels al·lels (el valor genotípic *estimat*,  $\hat{G}$ ) i una desviació causada per la dominància,  $\delta_{ij}$ :

$$G_{ij} = \hat{G}_{ij} + \delta_{ij} \quad (1)$$

Una forma d'expressar la regressió és esta:

$$G_{ij} = \mu_G + \alpha_1 N_1 + \alpha_2 N_2 + \delta_{ij} \quad (2)$$

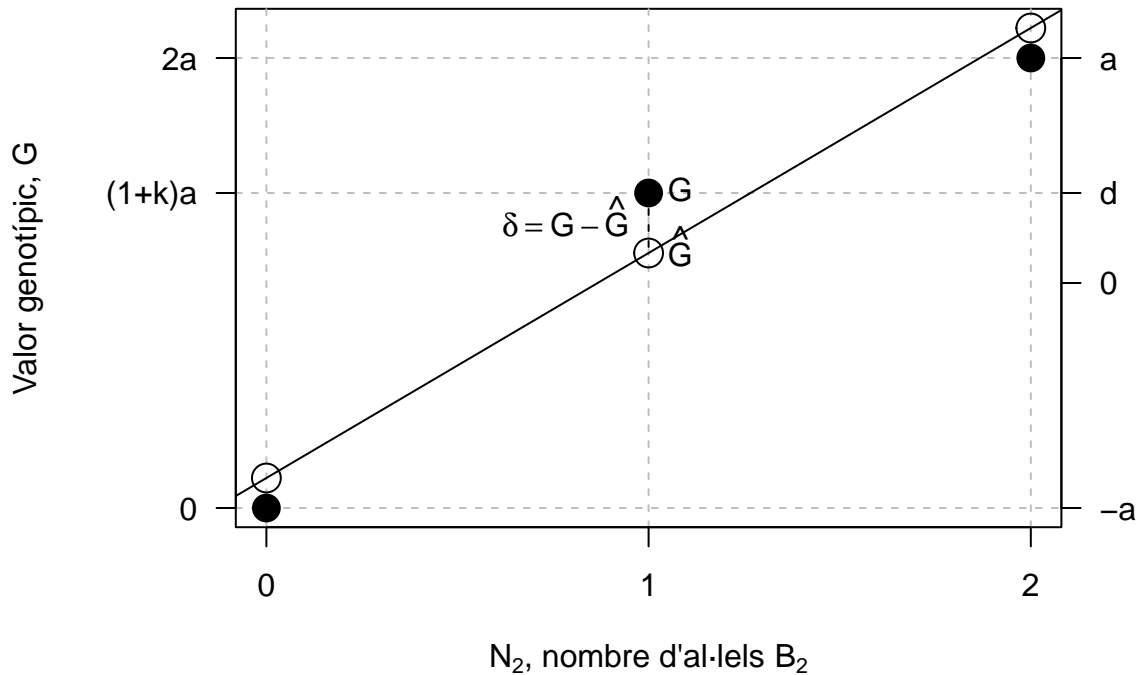
On  $\mu_G$  seria el valor genotípic mig en la població,  $\alpha_1$  i  $\alpha_2$  són els coeficients de regressió parcials, corresponents als efectes dels al·lels d'un tipus ( $B_1$ ) i de l'altre ( $B_2$ ). I  $N_1$  i  $N_2$  són el nombre d'al·lels  $B_1$  o  $B_2$  presents al genotip; aquestes són les variables predictores. Els coeficients  $\alpha_1$  i  $\alpha_2$  es coneixen com els **efectes additius**. Si el genotip és diploide,  $N_1 = 2 - N_2$ , perquè només poden haver dos al·lels. Aprofitant aquesta relació, podem transformar la regressió múltiple en una regressió simple:

$$G_{ij} = \mu_G + \alpha_1(2 - N_2) + \alpha_2 N_2 + \delta_{ij}$$

$$G_{ij} = \mu_G + 2\alpha_1 + (\alpha_2 - \alpha_1)N_2 + \delta_{ij}$$

$$G_{ij} = (\mu_G + 2\alpha_1) + \alpha N_2 + \delta_{ij} \quad (3)$$

Així, el pendent de la regressió és ara  $\alpha = \alpha_2 - \alpha_1$ , i la intersecció amb l'eix d'ordenades,  $\mu_G + 2\alpha_1$ , que equival al valor genotípic predit per al genotip  $B_1B_1$ .



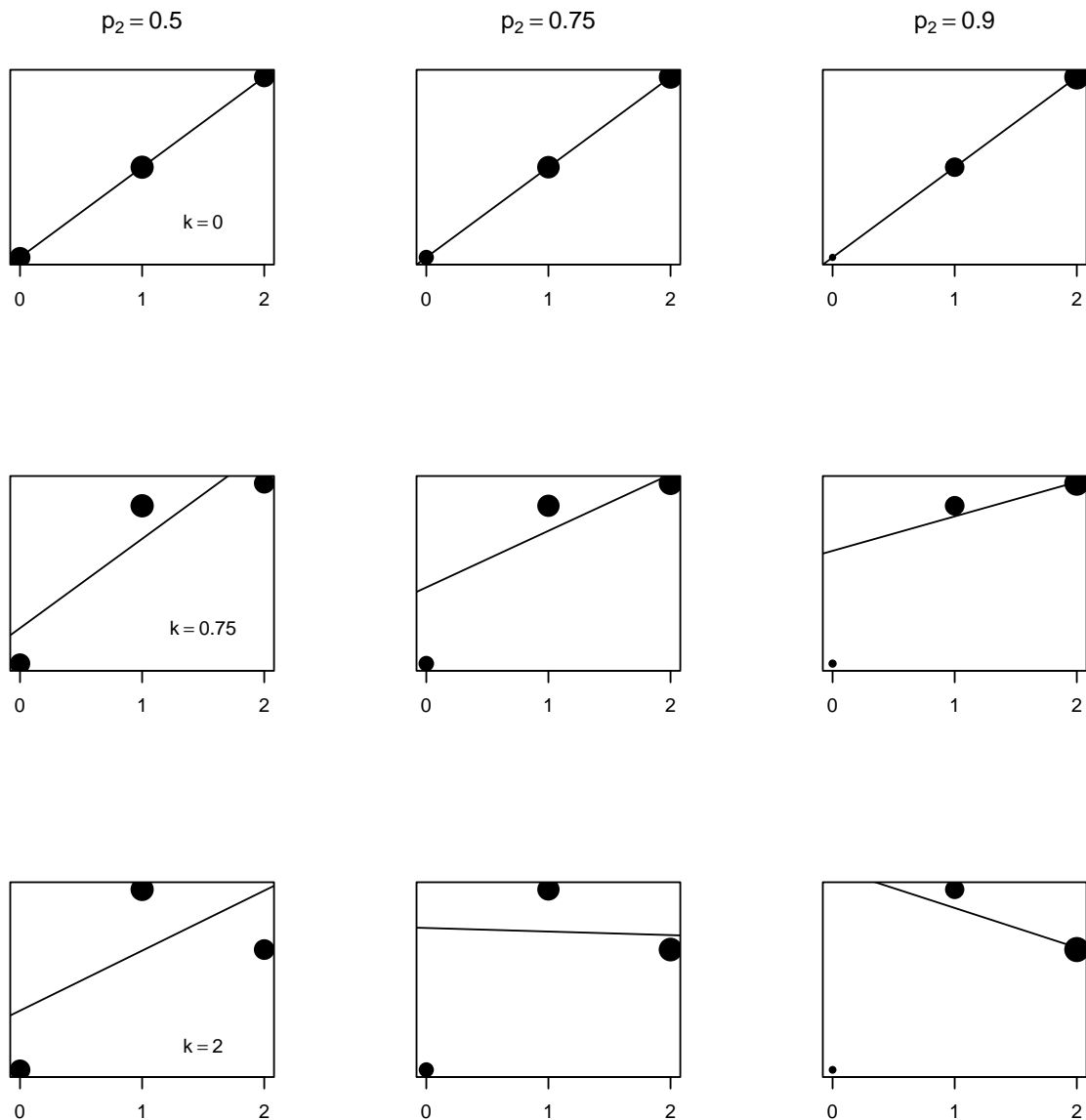
(Modificat de la figura 4.6 de Lynch i Walsh (1998), pàgina 66.)

Cabria preguntar-se el per què de tot açò. Òbviament, si sabem el fenotip que correspon a cada genotip, no necessitem fer una regressió! El motiu és que en una població amb aparellament aleatori, cada progenitor transmet un únic al·lel per locus, el qual es combina a l'atzar amb qualsevol altre. En estes condicions, en la descendència es manifesten els efectes additius heretats dels progenitors. En definitiva, la regressió ens permet descomposar l'efecte genotípic en termes que ens ajuden a comprendre l'herència en el context d'una població concreta. Açò és molt important, per exemple, per a la selecció artificial on cal triar els individus que es reproduiran en cada generació i predir quina millora de la productivitat podem esperar.

## La població

A la gràfica de l'apartat anterior, la regressió s'ha fet amb només tres punts, com si tinguérem només un individu de cada genotip. Però el valor genotípic mig i el pendent de la regressió,  $\alpha$ , depenen del nombre d'individus de cada tipus. A efectes de mostrar com funciona la regressió lineal o la descomposició del valor genotípic, era suficient imaginar tres individus. Però aquesta anàlisi només pot tenir alguna utilitat pràctica si s'aplica a una població concreta. El fet que la composició genètica de la població afecte el resultat de l'anàlisi ens indica que l'objectiu **no** és descriure una relació funcional general entre el genotip i el fenotip, aplicable a qualsevol individu o població, ni en qualsevol ambient. Una relació funcional d'aquest tipus seria descrita en tot cas per la **norma de reacció**, però no per la descomposició del valor genotípic. L'objectiu és caracteritzar l'efecte de la diversitat genètica d'una població concreta en algun fenotip.

La figura següent reproduïx la figura 4.7 de Lynch and Walsh (1998), i demostra que la descomposició del valor genotípic és una **anàlisi local**, en el sentit de què només és vàlida en una població concreta. La mida dels punts en aquesta gràfica és proporcional al logaritme del nombre d'individus de cada tipus.



El pendent de la regressió,  $\alpha$ , és *l'efecte promig de la substitució al·lèlica*. És a dir, representa el canvi promig en el valor genotípic que resultaria quan un al·lel  $B_1$  triat a l'atzar fóra substituït per un al·lel  $B_2$ . Sense dominància,  $\alpha = a$  ( $a$  és la meitat de la diferència de valor genotípic entre els dos homozigots). Però amb dominància  $\alpha$  depèn de la composició genètica de la població. Amb aparellament aleatori, es pot demostrar que podem expressar  $\alpha$  en termes dels paràmetres  $a$  i  $k$ , amb què havíem definit l'escala del valor genotípic, així:

$$\alpha = a(1 + k(p_1 - p_2)) \quad (4)$$

On  $p_1$  és la freqüència dels al·lells  $B_1$ , i  $p_2$ , la dels  $B_2$ .

Abans havíem definit  $\alpha$  com a  $\alpha_2 - \alpha_1$ , i hem dit que  $\alpha_1$  i  $\alpha_2$  són els *efectes additius*. Amb reproducció aleatòria, es pot demostrar que

$$\alpha_2 = p_1 a(1 + k(p_1 - p_2)) = p_1 \alpha \quad (5)$$

$$\alpha_1 = -p_2 a(1 + k(p_1 - p_2)) = -p_2 \alpha \quad (6)$$

Per últim cal definir el **valor reproductiu** (*breeding value*) d'un individu com a la suma dels valors additius dels seus gens. Representem el valor reproductiu amb el símbol  $A$ . En relació amb un únic locus, el valor reproductiu d'un homozigot  $B_1 B_1$  serà  $2\alpha_1$ ; el d'un heterozigot,  $(\alpha_1 + \alpha_2)$ ; i el d'un homozigot  $B_2 B_2$ ,  $2\alpha_2$ . Amb reproducció aleatòria, una conseqüència important d'aquestes definicions és que *el valor reproductiu d'un individu en una població amb aparellament aleatori és el doble de la desviació esperada entre el fenotip promig de la seua descendència i el valor fenotípic mitjà de la població*.

## Anàlisi de la variància genètica

L'abstracció matemàtica de descomposar el valor genotípic en una part additiva i un efecte de la dominància,  $G = \hat{G} + \delta$ , ens permet descomposar la variància genètica total de la manera següent:

$$\sigma_G^2 = \sigma^2(\hat{G} + \delta) \quad (7)$$

Que d'acord amb les propietats de la variància equival a:

$$\sigma_G^2 = \sigma^2(\hat{G}) + 2\sigma(\hat{G}, \delta) + \sigma^2(\delta) \quad (8)$$

La regressió per mínims quadrats té la propietat de què la predicció **no** està correlacionada amb l'error residual. És a dir, en este cas,  $\sigma(\hat{G}, \delta) = 0$ . Per tant, podem dir:

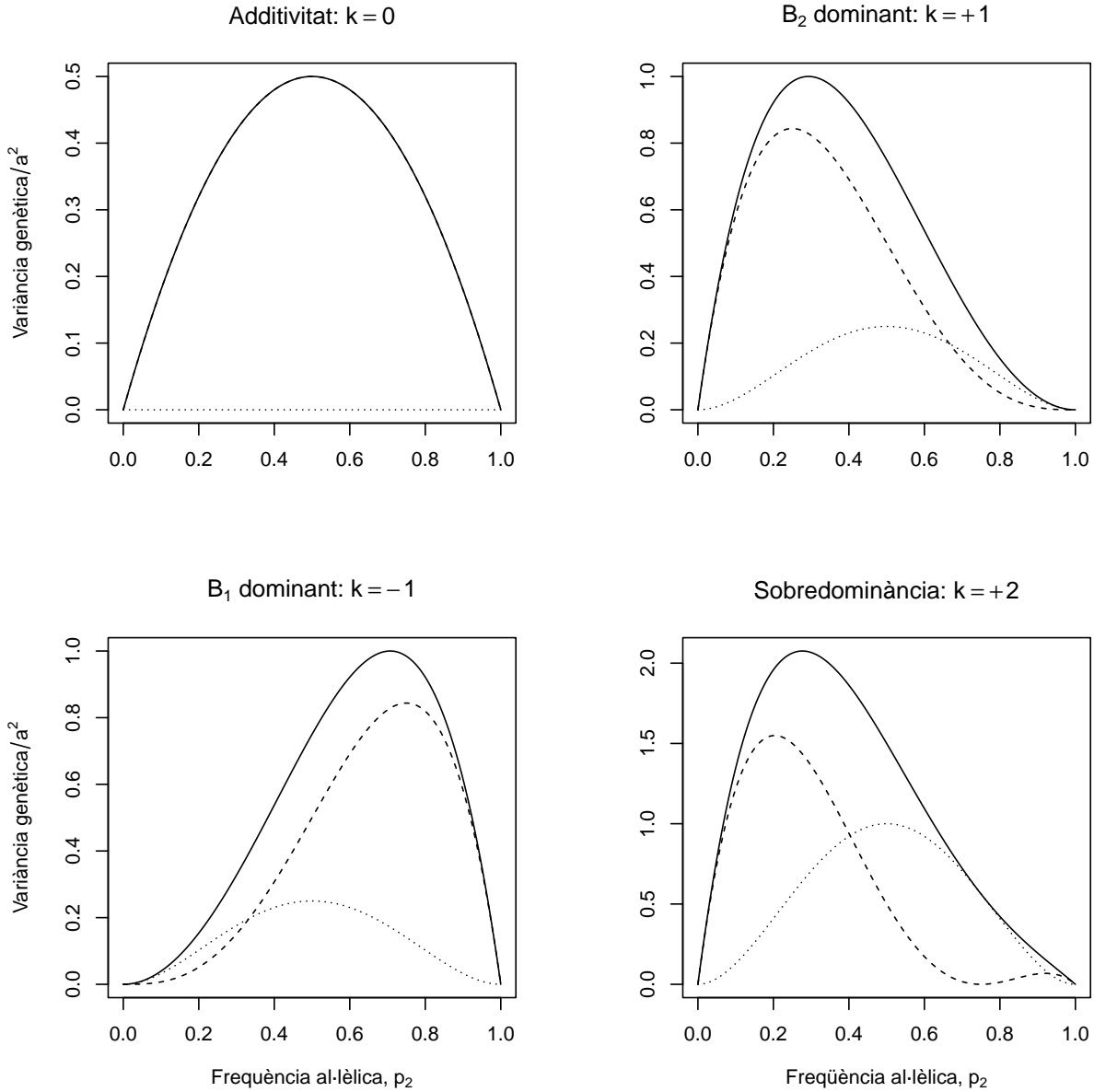
$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 \quad (9)$$

On  $\sigma_A^2$  és la variància genètica explicada per la regressió sobre  $N_2$  (seria el mateix fer la regressió sobre  $N_1$ ). Se l'anomena **variància genètica additiva**. Mentre que  $\sigma_D^2$  és la variància residual de la regressió o la variància genètica atribuïble a la dominància. Si la població està en equilibri de Hardy-Weinberg:

$$\sigma_A^2 = 2p_1 p_2 \alpha^2 \quad (10)$$

$$\sigma_D^2 = (2p_1 p_2 a k)^2 \quad (11)$$

A continuació reproduïsc la figura 4.8 de Lynch and Walsh (1998), que representa les variàncies genètiques total (línia contínua), additiva (línia discontinua) i de dominància (puntejada) en diferents situacions de dominància, i amb diferents freqüències al·lèl · liques.



La partició de la variància va permetre a Fisher (1918) explicar per què en una regressió, l'alçada dels progenitors no podia explicar més del 50% de la variància d'alçada entre els descendents, a pesar d'haver evidències de què la variància no explicada per la regressió no podia ser tota d'origen ambiental.

## Herència poligènica

Quan hi ha dos gens influint en un caràcter, podem utilitzar la regressió lineal per separar els efectes *additius* dels dos gens de l'efecte atribuïble a la interacció entre ells. L'efecte additiu de cada gen es podia descomposar alhora en els efectes additius dels al·lèls i el de la dominància.

$$G_{ijkl} = \mu_G + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + \epsilon_{ijkl} \quad (12)$$

Ací,  $G_{ijkl}$  és el valor genotípic del genotip compost pels al·lèls  $i$  i  $j$  en un locus i els al·lèls  $k$  i  $l$  en un altre locus. D'entrada sabem que modelar l'efecte conjunt de dos gens sobre un fenotip amb una regressió

*lineal* (com si els seus efectes foren additius) és una aproximació imperfecta, i en la major part dels casos esperem que el valor predit per la regressió serà diferent de l'observat. En l'equació anterior, el terme  $\epsilon_{ijkl}$  (la diferència entre el valor observat i la predicció lineal) inclou els efectes possibles de totes les *interaccions* entre les variables explicatives. Per comprendre què i quines són eixes interaccions cal definir la *mitjana fenotípica condicional*. Per exemple,  $G_{i...}$  seria la mitjana fenotípica condicional dels individus que porten almenys un al · lel  $i$  al primer locus, independentment de quin altre al · lel tinguen en eixe locus o de quin genotip tenen en l'altre locus. Així, podem definir tots els *efectes* inclosos en una regressió lineal. L'efecte additiu de l'al · lel  $i$  del primer locus és la diferència entre la mitjana condicional dels individus amb un al · lel  $i$  i la mitjana poblacional:

$$\alpha_i = G_{i...} - \mu_G \quad (13)$$

I de la mateixa manera, els efectes de dominància són el que queda per explicar després d'haver tingut en compte els efectes additius dels al · lels:

$$\delta_{ij} = G_{ij..} - \mu_G - \alpha_i - \alpha_j \quad (14)$$

$$\delta_{kl} = G_{..kl} - \mu_G - \alpha_k - \alpha_l \quad (15)$$

I a continuació vindrien totes les interaccions possibles, que són de tres tipus: interaccions entre els efectes additius ( $\alpha\alpha$ ), entre efectes additius i de dominància ( $\alpha\delta$ ), i entre efectes de dominància ( $\delta\delta$ ):

$$\begin{aligned} (\alpha\alpha)_{ik} &= G_{i..k} - \mu_G - \alpha_i - \alpha_k \\ (\alpha\alpha)_{jk} &= G_{.jk.} - \mu_G - \alpha_j - \alpha_k \\ (\alpha\alpha)_{il} &= G_{i..l} - \mu_G - \alpha_i - \alpha_l \\ (\alpha\alpha)_{jl} &= G_{.jl.} - \mu_G - \alpha_j - \alpha_l \\ (\alpha\delta)_{ikl} &= G_{i..kl} - \mu_G - \alpha_i - \alpha_k - \alpha_l - \delta_{kl} - (\alpha\alpha)_{ik} - (\alpha\alpha)_{il} \\ (\alpha\delta)_{jkl} &= G_{.jkl.} - \mu_G - \alpha_j - \alpha_k - \alpha_l - \delta_{kl} - (\alpha\alpha)_{jk} - (\alpha\alpha)_{jl} \\ (\alpha\delta)_{ijk} &= G_{ijk.} - \mu_G - \alpha_i - \alpha_j - \alpha_k - \delta_{ij} - (\alpha\alpha)_{ik} - (\alpha\alpha)_{jk} \\ (\alpha\delta)_{ijl} &= G_{ij.l} - \mu_G - \alpha_i - \alpha_j - \alpha_l - \delta_{ij} - (\alpha\alpha)_{il} - (\alpha\alpha)_{jl} \\ (\delta\delta)_{ijkl} &= G_{ijkl} - \mu_G - \alpha_i - \alpha_j - \alpha_k - \alpha_l - \delta_{ij} - \delta_{kl} \\ &\quad - (\alpha\alpha)_{ik} - (\alpha\alpha)_{il} - (\alpha\alpha)_{jk} - (\alpha\alpha)_{jl} \\ &\quad - (\alpha\delta)_{ikl} - (\alpha\delta)_{jkl} - (\alpha\delta)_{ijk} - (\alpha\delta)_{ijl} \end{aligned}$$

Així doncs, el valor genotípic determinat per dos (o més) gens es pot descomposar en una sèrie d'efectes additius, de dominància, additius  $\times$  additius, additius  $\times$  dominància, i de dominància  $\times$  dominància. La regressió atribueix primer la major part de la variància possible als efectes additius, i va repartint la variància que queda per explicar entre els efectes d'ordre major fins arribar a  $(\delta\delta)_{ijkl}$ . Si hi haguera més de dos gens influent el caràcter, es podria continuar descomposant el valor genotípic en cada vegada més efectes d'interaccions, la qual cosa es representa a continuació amb els punts suspensius:

$$\begin{aligned} G_{ijkl...} &= \mu_G + (\alpha_i + \alpha_j + \alpha_k + \alpha_l) + (\delta_{ij} + \delta_{kl}) \\ &\quad + ((\alpha\alpha)_{ik} + (\alpha\alpha)_{il} + (\alpha\alpha)_{jk} + (\alpha\alpha)_{jl}) \\ &\quad + ((\alpha\delta)_{ikl} + (\alpha\delta)_{jkl} + (\alpha\delta)_{ijk} + (\alpha\delta)_{ijl}) + (\delta\delta)_{ijkl} + \dots \end{aligned}$$

Com a conseqüència del repartiment que es fa de la variància entre els factors considerats en una regressió, tenim la garantia de què els factors o efectes, definits així a partir d'una regressió, contribueixen a la variància

fenotípica total de forma independent uns d'altres. Per tant, amb més d'un gen influent en un caràcter, també podem distingir diferents fonts de variació genètica:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 + \dots \quad (16)$$

En genètica quantitativa s'anomena **epístasi** a totes aquestes interaccions entre els gens, ja siguin interaccions entre els components additius o entre els additius i els de dominància o entre els de dominància. Cal tenir en compte que amb tres gens pot haver interaccions entre els efectes additius de qualsevol parell de gens, així com la interacció triple entre els components additius dels tres gens. En general, com major és el nombre de gens implicats en un caràcter quantitatiu, major és el nombre de termes no additius de la regressió. En definitiva, és molt probable que les interaccions (l'epístasi) constituïsquen una part important de la variància genètica de qualsevol caràcter quantitatiu, i en general és incorrecte assumir que els efectes epistàtics (no additius) són menypreables.

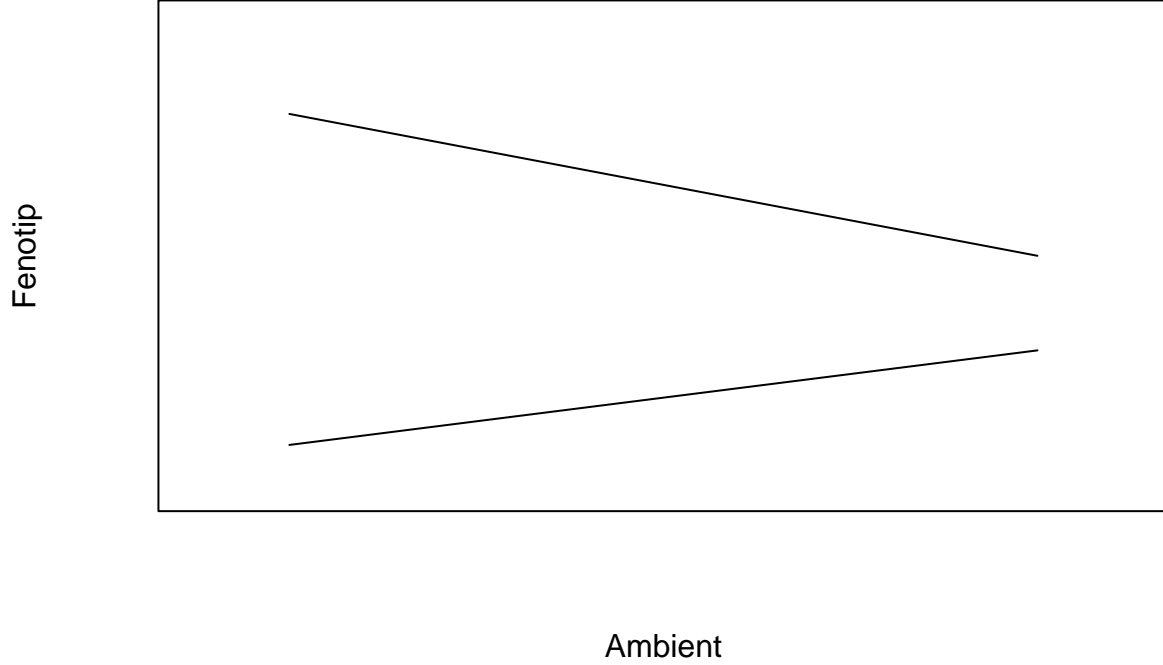
## Fonts de variació ambiental

El fenotip d'un individu depèn normalment de l'ambient en què s'ha desenvolupat. L'efecte de l'ambient es pot incorporar al model, i és especialment útil diferenciar dos tipus d'efectes: el de l'ambient comú a un conjunt d'individus, i els efectes de l'ambient particular i únic que experimenta només l'individu. Com a exemples de l'ambient comú a un grup d'individus, podem pensar en l'efecte de la parcel·la concreta en què creixen unes plantes (el *plot*, en anglès), o l'estable concret on s'han criat les vaques, la densitat del galliner, o també l'efecte matern sobre els seus descendents. L'efecte ambiental comú o general el podem representar amb el símbol  $E$ , i l'efecte ambiental *especial* (individual), amb la lletra minúscula  $e$ .

Podríem, per tant, suposar que el fenotip, que a partir d'ara anomenarem  $z$ , es pot descomposar en un efecte genètic,  $G$  (el valor genotípic, que ja sabem descomposar en efectes additiu i de dominància) i un parell d'efectes ambientals,  $E$  i  $e$ . Però resulta que no tots els genotips responen igual als canvis ambientals. És a dir, la diferència que provoca un canvi ambiental sobre el fenotip d'individus d'un cert genotip pot ser diferent a l'efecte que el mateix canvi ambiental tinga sobre altres genotips. En termes de la *norma de reacció*, això implica que no siguin paral·leles.



## Normes de reacció amb $G \times E$



Per tant, un model lineal general, que tinga en compte la interacció entre l'ambient i el genotip hauria de tenir els components següents:

$$z_{ijk} = G_i + E_j + I_{ij} + e_{ijk} \quad (17)$$

Açò es llig així: El fenotip ( $z$ ) de l'individu  $k$ , en l'ambient  $j$  i amb genotip  $i$  és la suma de l'efecte del genotip  $i$  ( $G_i$ ) més l'efecte de l'ambient  $j$  ( $E_j$ ), més la interacció entre el genotip  $i$  i l'ambient  $j$  ( $I_{ij}$ ), més una desviació particular de l'individu  $k$  en l'ambient  $j$  i amb el genotip  $i$ . Ja sabem que el component genètic,  $G_i$  és el valor genotípic, i que alhora es pot descomposar en un valor reproductiu,  $A_i$ , i un efecte de la dominància  $D$ .

En aplicar la fórmula de la variància d'una suma a l'equació anterior, trobarem que la variància del fenotip,  $\sigma_P^2$ , es descomposa en els components següents:

$$\sigma_P^2 = \sigma_G^2 + \sigma_I^2 + 2\sigma_{G,E} + \sigma_E^2 + \sigma_e^2 \quad (18)$$

De totes les covariàncies que podria haver entre components del fenotip, només la covariància entre el genotip i l'ambient no té per què ser zero. Les altres, per la manera en què es construeix el model, han de ser zero. La covariància entre el genotip i l'ambient representa una associació estadística entre genotips i ambients. És a dir, una distribució no aleatòria dels genotips entre els diferents ambients. Aquest component de la variància fenotípica és molt difícil de distingir de la variància genètica, i sempre pot estar present a pesar de què no podem mesurar-lo. Un exemple d'aquesta covariància són els efectes materns, que són ambientals, però que afecten a individus emparentats. Aquesta covariància fa que en la població total la variància fenotípica siga major de la que seria si tots els genotips estigueren distribuïts a l'atzar entre tots els ambients disponibles.

Cal diferenciar la covariància  $\sigma_{G,E}$  de la variància de la interacció entre genotip i ambient,  $\sigma_I^2$ . Aquest component de la variància fenotípica mesura la variació de la no-additivitat entre els efectes ambiental i

genètic.  $\sigma_I^2$  es pot entendre com la diversitat de pendents en les normes de reacció, amb algunes normes de reacció més paral·leles entre elles i d'altres més divergents. Mentre que  $\sigma_{G,E}$  depèn de la freqüència amb què trobem individus d'un cert genotip en un cert ambient.

La variància ambiental més fina que pot mesurar-se és la variància entre els costats dret i esquerre d'un animal amb simetria bilateral. Aquesta variància, a vegades important, forma part de la variància ambiental especial,  $e$ .

## Heretabilitat

L'heretabilitat (en sentit estricte) d'un caràcter quantitatiu en una població és la proporció de la variància fenotípica observada que es pot atribuir a la variància de valors reproductius (la part additiva de la variància genètica). L'heretabilitat és el que determina la semblança entre individus emparentats. De fet, l'heretabilitat d'un caràcter quantitatiu es pot mesurar estudiant la semblança d'individus emparentats.

Per motius històrics, se sol utilitzar el símbol  $h^2$  per representar l'heretabilitat:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} \quad (19)$$

L'heretabilitat es pot entendre també com el pendent d'una regressió entre el fenotip i el valor reproductiu. Indica fins a quin punt el fenotip observat és un bon indicador del valor reproductiu. Cal tenir en compte que l'heretabilitat no és una propietat del caràcter en sí, sinó que també ho és de l'ambient i la composició genètica particular d'una població. És a dir, l'heretabilitat ens ajuda a entendre com afecten els factors genètics a les *diferències* entre individus d'una població concreta; però no per explicar el fenotip concret de cap individu.

A partir de l'heretabilitat, es pot estimar la *resposta a la selecció*, a través del que es coneix com l'equació del ramader:

$$R = h^2 S \quad (20)$$

On  $R$  és la resposta a la selecció,  $h^2$  és l'heretabilitat i  $S$  és el *diferencial de selecció*. Més concretament:

- **Diferencial de selecció:** la diferència entre el fenotip mitjà dels progenitors seleccionats i el fenotip mitjà de la població.
- **Resposta a la selecció:** la diferència entre el fenotip mitjà dels descendents i el fenotip mitjà de la població parental.

## References

- Fisher, R. A. 1918. "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." *Trans. Royal Soc. Edinburgh* 52: 399–433.
- Lynch, M., and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA, U.S.A.: Sinauer Associates, Inc.