

Apuntes de Genética Cuantitativa

J. Ignacio Lucas Lledó

13/12/2022

Introducción

El desarrollo de la genética cuantitativa en la primera mitad del siglo XX contribuyó al éxito de las campañas de mejora genética. Éstas, junto con la mecanización y el uso de agroquímicos, posibilitaron el mayor incremento de la productividad agropecuaria de la historia de la humanidad: la *revolución verde*. Actualmente, el aparato estadístico de la genética cuantitativa se combina con la abundancia de datos genómicos para continuar aumentando la productividad agrícola y ganadera. El desarrollo de métodos como la selección genómica y los estudios de asociación genómica hacen que los conceptos básicos de genética cuantitativa sean fundamentales para entender y participar en el progreso de la biomedicina, la biología evolutiva y la biotecnología.

El esquema a seguir para introducir estos conceptos es el siguiente:

1. Un gen. Con el caso más simple posible se pretende explicar en qué consiste la descomposición del valor genotípico, antes incluso de saber para qué sirve esta descomposición.
2. La población. Solamente un inciso para observar que la descomposición del valor genotípico sólo tiene sentido en referencia a una población concreta. El valor reproductivo de un individuo, así como el efecto de la dominancia no son propiedades inherentes del gen, ni del individuo en sí, sino del gen o del individuo *y de la población*.
3. Análisis de la varianza genética. Una primera motivación para haber descompuesto el valor genotípico es la partición de la variancia genética en variancia aditiva y variancia de dominancia.
4. Herencia poligénica. La regresión lineal nos permite modelar también los efectos simultáneos de diversos genes sobre un carácter cuantitativo. Pero el análisis se complica, porque en muchos casos hay epistasia: relaciones no lineales entre la acción de los genes y el carácter.
5. Fuentes de variación ambiental. El fenotipo depende de factores genéticos y ambientales, y también es posible incorporar el efecto del ambiente en una regresión lineal más general. El objetivo es comprender el significado de los términos en esta regresión.
6. Heredabilidad. Definimos la heredabilidad y poco más.

Un gen

Normalmente la genética cuantitativa es la genética de los caracteres **poligénicos**, tanto si el fenotipo se puede medir en una escala continua como si es binario (e.g., enfermo o sano). Pero imaginemos ahora que el carácter determinado por este único locus bialélico se mide en una escala continua. Si ningún otro gen ni el ambiente influyen en el carácter, habrá dos (con dominancia completa) o tres fenotipos posibles.

El valor del fenotipo determinado por el genotipo se representa con el símbolo G , y se le llama *valor genotípico*, pero se mide en unidades del **fenotipo**. La escala en que se mide el fenotipo es arbitraria. Por ejemplo, dos escalas útiles para los cálculos teóricos son las siguientes (reproducción de la figura 4.4 de Lynch and Walsh (1998), p. 62):

Genotipo	B_1B_1	B_1B_2	B_2B_2
Valor genotípico	$-a$	d	$+a$

Genotipo	B_1B_1	B_1B_2	B_2B_2
Valor genotípico	0	$(1+k)a$	$2a$

En presencia de dominancia, la relación entre el valor genotípico y el número de alelos de un cierto tipo no es lineal. Pero a Fisher (1918) se le ocurrió la buena idea de considerar una aproximación lineal a la relación entre el contenido génico (número de alelos B_2 , por ejemplo) y el valor genotípico, G . Esta *regresión* permite descomponer conceptualmente el valor genotípico en una parte atribuible al efecto *aditivo* o heredable de los alelos (el valor genotípico *estimado*, \hat{G}) y una desviación causada por la dominancia, δ_{ij} :

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

Una forma de expresar la regresión es esta:

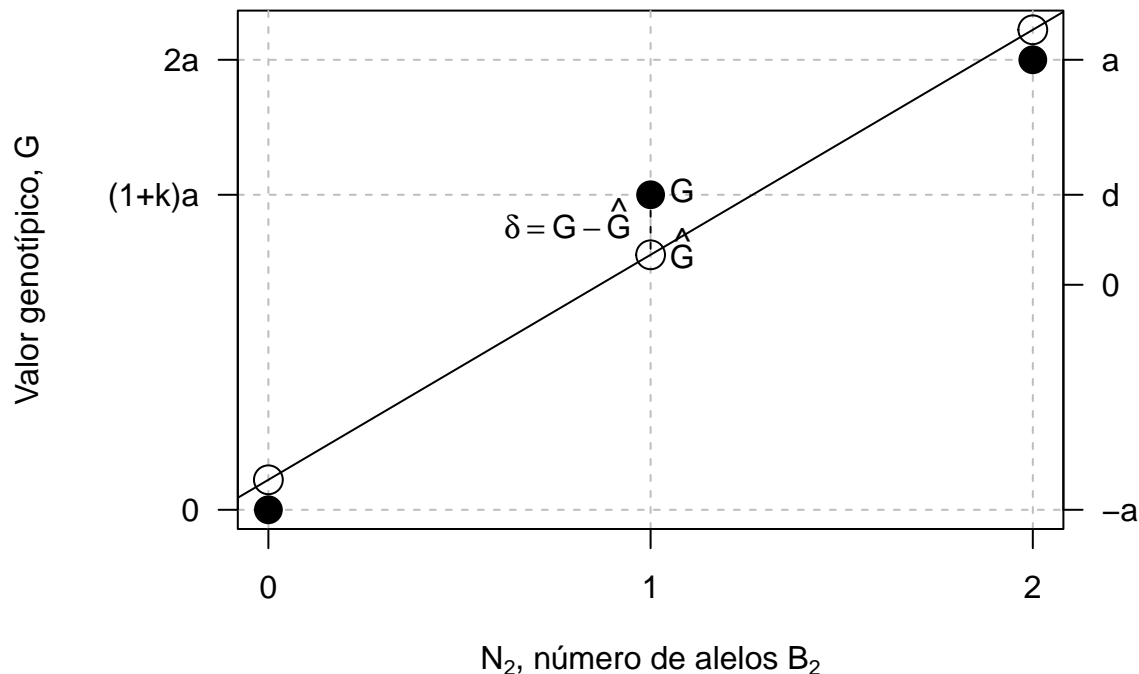
$$G_{ij} = \mu_G + \alpha_1 N_1 + \alpha_2 N_2 + \delta_{ij} \quad (2)$$

Donde μ_G sería el valor genotípico medio en la población, α_1 y α_2 son los coeficientes de regresión parciales, correspondientes a los efectos de los alelos de un tipo (B_1) y del otro (B_2). Y N_1 y N_2 son el número de alelos B_1 o B_2 presentes en el genotipo; estas son las variables predictivas. Los coeficientes α_1 y α_2 se conocen como los **efectos aditivos**. Si el genotipo es diploide, $N_1 = 2 - N_2$, porque sólo puede haber dos alelos. Aprovechando esta relación, podemos transformar la regresión múltiple en una regresión simple:

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

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

Así, la pendiente de la regresión es ahora $\alpha = \alpha_2 - \alpha_1$, y la intersección con el eje de ordenadas, $\mu_G + 2\alpha_1$, que equivale al valor genotípico predicho para el genotipo B_1B_1 .



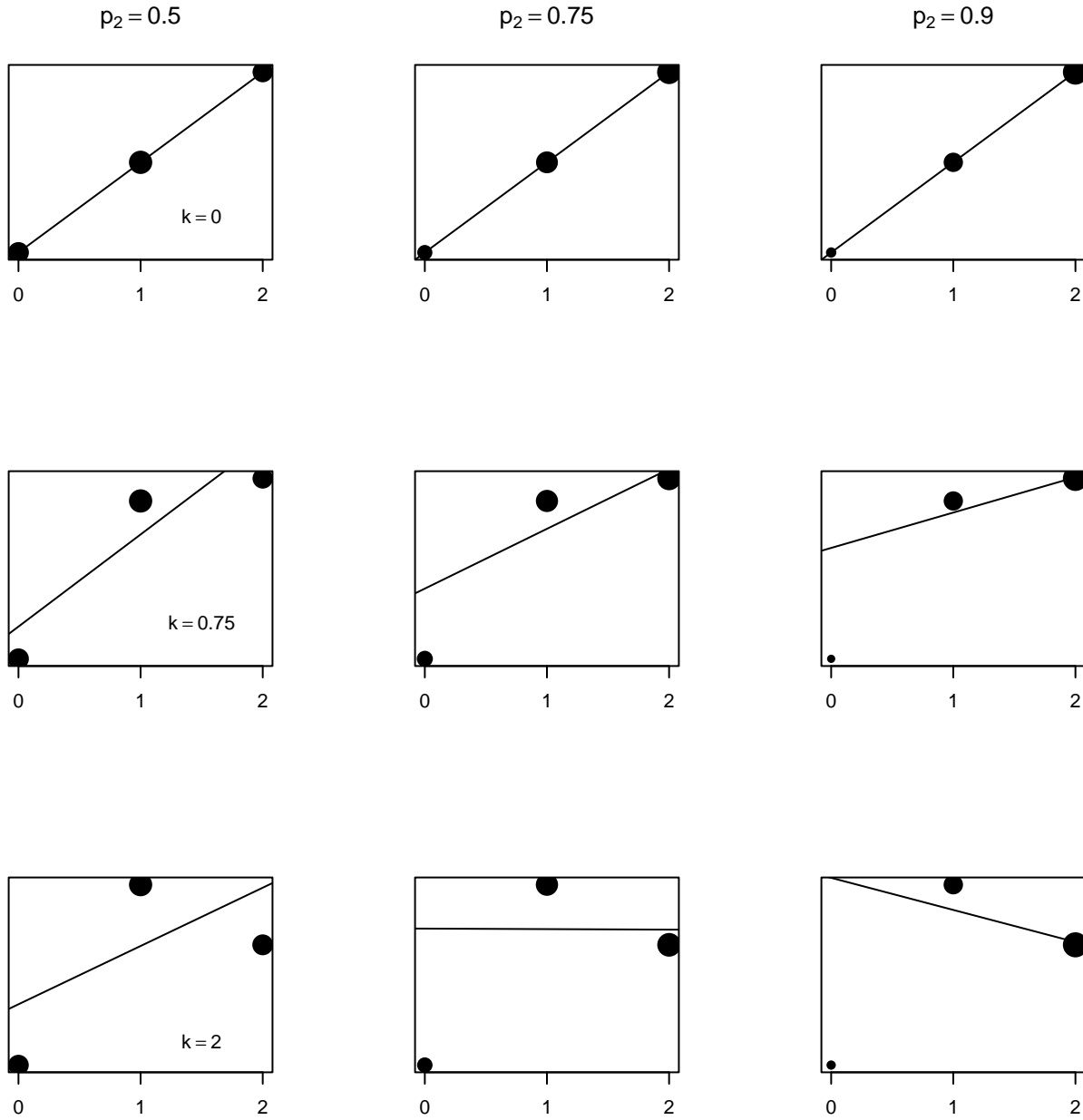
(Modificado de la figura 4.6 de Lynch y Walsh (1998), página 66.)

Cabría preguntarse el por qué de todo esto. Obviamente, si sabemos el fenotipo que corresponde a cada genotipo, ¡no necesitamos hacer una regresión! El motivo es que en una población con apareamiento aleatorio, cada progenitor transmite un único alelo por locus, el cual se combina al azar con cualquier otro. En estas condiciones, en la descendencia se manifiestan los efectos aditivos heredados de los progenitores. En definitiva, la regresión nos permite descomponer el efecto genotípico en términos que nos ayuden a comprender la herencia en el contexto de una población concreta. Esto es muy importante, por ejemplo, para la selección artificial, donde es necesario escoger los individuos que se reproducirán en cada generación y predecir qué mejora de la productividad podemos esperar.

La población

En la gráfica del apartado anterior, la regresión se ha hecho con sólo tres puntos, como si tuviéramos sólo un individuo de cada genotipo. Pero el valor genotípico medio y la pendiente de la regresión, α , dependen del número de individuos de cada tipo. A efectos de mostrar cómo funciona la regresión lineal o la descomposición del valor genotípico, era suficiente imaginar tres individuos. Pero este análisis sólo puede tener alguna utilidad práctica si se aplica a una población concreta. El hecho de que la composición genética de la población afecte el resultado del análisis nos indica que el objetivo **no** es describir una relación funcional general entre el genotipo y el fenotipo, aplicable a cualquier individuo o población. Tampoco en cualquier ambiente. Una relación funcional de ese tipo sería descrita en todo caso por la **norma de reacción**, pero no por la descomposición del valor genotípico. El objetivo es caracterizar el efecto de la diversidad genética de una población concreta en algún fenotipo.

La figura siguiente reproduce la figura 4.7 de Lynch and Walsh (1998), y demuestra que la descomposición del valor genotípico es un **análisis local**, en el sentido de que sólo es válido en una población concreta. El tamaño de los puntos en esta gráfica es proporcional al logaritmo del número de individuos de cada tipo.



La pendiente de la regresión, α , es *el efecto promedio de la sustitución alélica*. Es decir, representa el cambio promedio en el valor genotípico que resultaría cuando un alelo B_1 escogido al azar fuera sustituido por un alelo B_2 . Sin dominancia, $\alpha = a$ (a es la mitad de la diferencia del valor genotípico entre los dos homocigotos). Pero con dominancia α depende de la composición genética de la población. Con apareamiento aleatorio, se puede demostrar que podemos expresar α en términos de los parámetros a y k , con los que habíamos descrito la escala del valor genotípico, así:

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

Donde p_1 es la frecuencia de los alelos B_1 , y p_2 , la de los B_2 .

Antes habíamos definido α como $\alpha_2 - \alpha_1$, y hemos dicho que α_1 y α_2 son los *efectos aditivos*. Con reproducción aleatoria, se puede demostrar que:

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

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

Por último, hay que definir el **valor reproductivo** (*breeding value*) de un individuo como la suma de los valores aditivos de sus genes. Representamos el valor reproductivo con el símbolo A . En relación con un único locus, el valor reproductivo de un homocigoto B_1B_1 será $2\alpha_1$; el de un heterocigoto, $(\alpha_1 + \alpha_2)$; y el de un homocigoto B_2B_2 , $2\alpha_2$. Con reproducción aleatoria, una consecuencia importante de estas definiciones es que *el valor reproductivo de un individuo en una población con apareamiento aleatorio es el doble de la desviación esperada entre el fenotipo promedio de su descendencia y el valor fenotípico medio de la población.*

Análisis de la varianza genética

La abstracción matemática de descomponer el valor genotípico en una parte aditiva y un efecto de la dominancia, $G = \hat{G} + \delta$, nos permite descomponer la varianza genética total de la manera siguiente:

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

Que de acuerdo con las propiedades de la varianza equivale a:

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

La regresión por mínimos cuadrado tiene la propiedad de que la predicción **no** está correlacionada con el error residual. Es decir, en este caso, $\sigma(\hat{G}, \delta) = 0$. Por tanto, podemos decir:

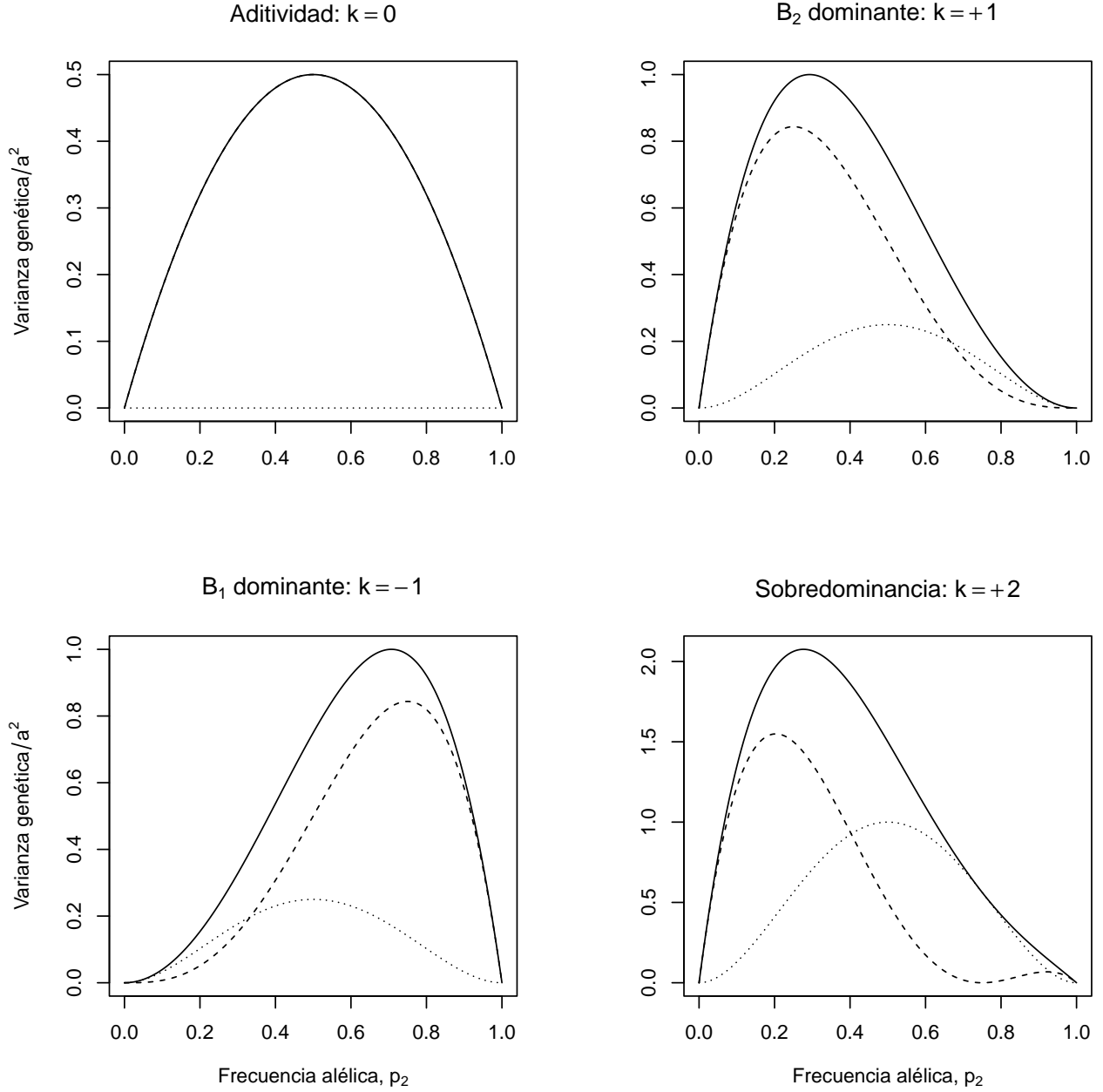
$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 \quad (9)$$

Donde σ_A^2 es la varianza genética explicada por la regresión sobre N_2 (sería lo mismo hacer la regresión sobre N_1). Se le llama **varianza genética aditiva**. Mientras que σ_D^2 es la varianza residual de la regresión o la varianza genética atribuible a la dominancia. Si la población está en equilibrio de Hardy-Weinberg:

$$\sigma_A^2 = 2p_1p_2\alpha^2 \quad (10)$$

$$\sigma_D^2 = (2p_1p_2ak)^2 \quad (11)$$

A continuación reproduzco la figura 4.8 de Lynch and Walsh (1998), que representa las varianzas genéticas total (línea continua), aditiva (línea discontinua) y de dominancia (punteada) en diferentes situaciones de dominancia, y con diferentes frecuencias alélicas.



La partición de la varianza permitió a Fisher (1918) explicar por qué en una regresión la altura de los progenitores no podía explicar más del 50% de la varianza en altura entre los descendientes, a pesar de existir evidencias de que la varianza no explicada por la regresión no podía ser toda de origen ambiental.

Herencia poligénica

Cuando hay dos genes influyendo un carácter, podemos utilizar la regresión lineal para separar los efectos *aditivos* de los dos genes del efecto atribuible a la interacción entre ellos. El efecto aditivo de cada gen se puede descomponer a su vez en los efectos aditivos de los alelos y el de la dominancia.

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

Aquí, G_{ijkl} es el valor genotípico del genotipo compuesto por los alelos i y j en un locus y los alelos k y l en otro locus. En principio, sabemos que modelar el efecto conjunto de los dos genes sobre un fenotipo con

una regresión *lineal* (como si sus efectos fueran aditivos o independientes entre ellos) es una aproximación imperfecta, y en la mayor parte de los casos esperamos que el valor predicho por la regresión será diferente del observado. En la ecuación anterior, el término ϵ_{ijkl} (la diferencia entre el valor observado y la predicción lineal) incluye los efectos posibles de todas las *interacciones* entre las variables explicativas. Para comprender qué y cuáles son esas interacciones hace falta definir la *media fenotípica condicional*. Por ejemplo, $G_{i...}$ sería el fenotipo medio condicional de los individuos que tienen al menos un alelo i en el primer locus, independientemente de qué otro alelo tenga en ese mismo locus o de qué genotipo tenga en el otro locus. Así, podemos definir todos los *efectos* incluidos en una regresión lineal. El efecto aditivo del alelo i del primer locus es la diferencia entre la media condicional de los individuos con un alelo i y la media poblacional:

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

Y de la misma manera, los efectos de dominancia son lo que queda por explicar después de haber tenido en cuenta los efectos aditivos de los alelos:

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

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

Y a continuación vendrían todas las interacciones posibles, que son de tres tipos: interacciones entre los efectos aditivos ($\alpha\alpha$), entre efectos aditivos y de dominancia ($\alpha\delta$), y entre efectos de dominancia ($\delta\delta$):

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

Así pues, el valor genotípico determinado por dos (o más) genes se puede descomponer en una serie de efectos aditivos, de dominancia, aditivos \times aditivos, aditivos \times dominancia, y de dominancia \times dominancia. La regresión atribuye primero la mayor parte de la varianza posible a los efectos aditivos, y va repartiendo la varianza que queda por explicar entre los efectos de orden mayor hasta llegar a $(\delta\delta)_{ijkl}$. Si hubiera más de dos genes influyendo en el carácter, se podría continuar descomponiendo el valor genotípico en cada vez más efectos de interacciones, lo cual se representa a continuación con los puntos suspensivos:

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

Como consecuencia de la repartición que se hace de la varianza entre los factores considerados en una regresión, tenemos la garantía de que los factores o efectos, definidos así a partir de una regresión, contribuyen a la

varianza fenotípica total de forma independiente unos de otros. Por tanto, con más de un gen influyendo en un carácter, también podemos distinguir diferentes fuentes de variación genética:

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

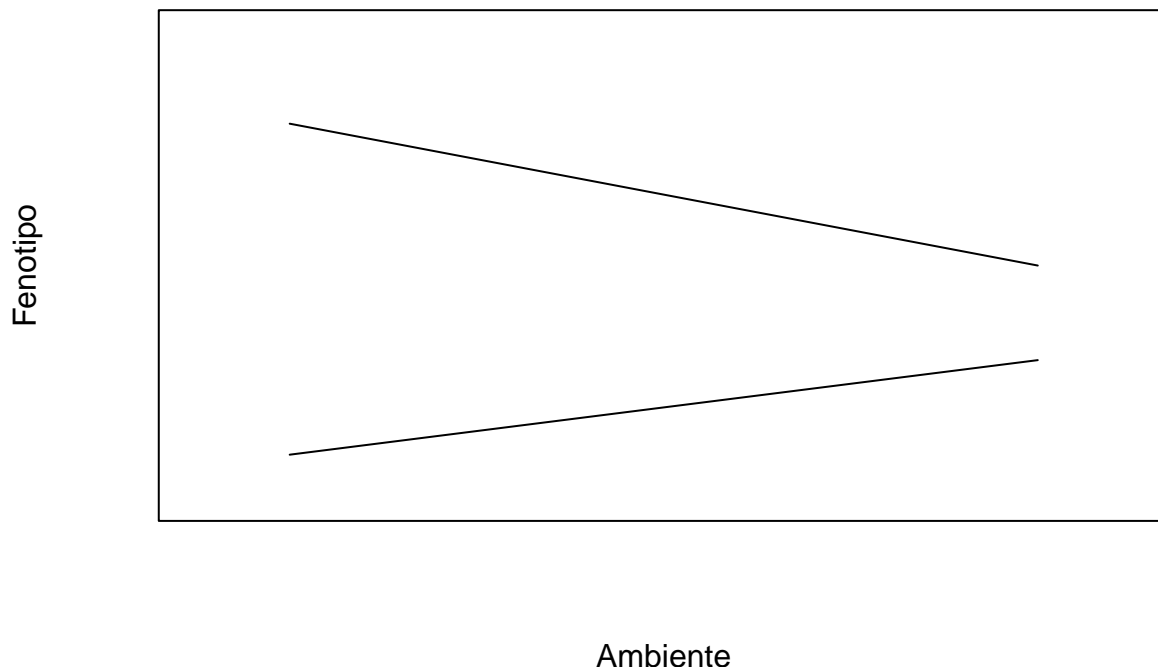
En genética cuantitativa se llama **epistasia** a todas estas interacciones entre los genes, ya sean interacciones entre los componentes aditivos o entre los aditivos y los de dominancia o entre los de dominancia. Hay que tener en cuenta que con tres genes puede haber interacciones entre los efectos aditivos de cualquier par de genes, así como la interacción triple entre los componentes aditivos de los tres genes. En general, cuanto mayor es el número de genes implicados en un carácter cuantitativo, mayor es el número de términos no aditivos de la regresión. En definitiva, es muy probable que las interacciones (la epistasia) constituyan una parte importante de la varianza genética de cualquier carácter cuantitativo, y en general es incorrecto asumir que los efectos epistáticos (no aditivos) son despreciables.

Fuentes de variación ambiental

El fenotipo de un individuo depende normalmente del ambiente en que se ha desarrollado. El efecto del ambiente se puede incorporar al modelo, y es especialmente útil diferenciar dos tipos de efectos: el del ambiente común a un conjunto de individuos, y los efectos del ambiente particular y único que experimenta sólo el individuo. Como ejemplos del ambiente común a un grupo de individuos, podemos pensar en el efecto de la parcela concreta donde crecen unas plantas (el *plot*), o el establo concreto donde se han criado las vacas, la densidad del gallinero, o también el efecto materno sobre una camada. El efecto ambiental común lo podemos representar con el símbolo E , y el efecto ambiental *especial* o individual, con la letra minúscula e .

Podríamos, por tanto, suponer que el fenotipo, que a partir de ahora llamaremos z , se puede descomponer en un efecto genético G (el valor genotípico, que ya sabemos descomponer en efectos aditivo y de dominancia) y un par de efectos ambientales, E y e . Pero resulta que no todos los genotipos responden igual a los cambios ambientales. Es decir, la diferencia que provoca un cambio ambiental sobre el fenotipo de individuos de un cierto genotipo puede ser diferente al efecto que el mismo cambio ambiental tenga sobre otros genotipos. En términos de la *norma de reacción*, esto implica que no sean paralelas.

Normas de reacción con $G \times E$



Por tanto, un modelo lineal general, que tenga en cuenta la interacción entre el ambiente y el genotipo debería de tener los siguientes componentes:

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

Esto se lee así: El fenotipo (z) del individuo k , en el ambiente j y con el genotipo i es la suma del efecto del genotipo i (G_i) más el efecto del ambiente j (E_j), más la interacción entre el genotipo i y el ambiente j (I_{ij}), más una desviación particular del individuo k en el ambiente j y con el genotipo i . Ya sabemos que el componente genético G_i es el valor genotípico, y que a su vez se puede descomponer en un valor reproductivo, A_i , y un efecto de la dominancia D .

Al aplicar la fórmula de la varianza de una suma a la ecuación anterior, encontramos que la varianza del fenotipo en la población, σ_P^2 , se descompone en los siguientes componentes:

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

De todas las covarianzas que podría haber entre componentes del fenotipo, sólo la covarianza entre el genotipo y el ambiente no tiene por qué ser cero. Las otras, por la manera en que se construye el modelo, tiene que ser cero. La covarianza entre el genotipo y el ambiente representa una asociación estadística entre genotipos y ambientes. Es decir, una distribución no aleatoria de los genotipos entre los diferentes ambientes. Este componente de la varianza fenotípica es muy difícil de distinguir de la varianza genética, y siempre puede contribuir a la varianza fenotípica, aunque no podamos medirlo. Un ejemplo de esta covarianza lo constituyen los efectos maternos, que son efectos ambientales pero que afectan a individuos emparentados. Esta covarianza hace que en la población total la varianza fenotípica sea mayor de la que sería si todos los genotipos estuvieran distribuidos al azar entre todos los ambientes disponibles.

Hay que diferenciar la covarianza $\sigma_{G,E}$ de la varianza atribuible a la interacción entre el genotipo y el ambiente, σ_I^2 . Este componente de la varianza fenotípica mide la variación de la no-aditividad entre los efectos ambiental y genético. σ_I^2 se puede entender como la diversidad de pendientes en las normas de reacción: algunas normas de reacción son más paralelas entre ellas que otras. Mientras que $\sigma_{G,E}$ depende de la frecuencia con que encontramos individuos de un cierto genotipo en un cierto ambiente.

La varianza ambiental más fina que se puede medir es la varianza entre los lados derecho e izquierdo de un animal con simetría bilateral. Esta varianza, a veces importante, forma parte de la varianza ambiental especial, e .

Heredabilidad

La heredabilidad (en sentido estricto) de un carácter cuantitativo en una población es la proporción de la varianza fenotípica observada que puede atribuirse a la varianza de los valores reproductivos (la parte aditiva de la varianza genética). La heredabilidad es lo que determina la semejanza entre individuos emparentados. De hecho, la heredabilidad de un carácter cuantitativo se puede medir estudiando la semejanza entre individuos emparentados.

Por motivos históricos, se suele utilizar el símbolo h^2 para representar la heredabilidad:

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

La heredabilidad se puede entender también como la pendiente de una regresión entre el fenotipo y el valor reproductivo. Indica hasta qué punto el fenotipo observado es un buen indicador del valor reproductivo. Hay que tener en cuenta que la heredabilidad no es una propiedad del carácter en sí, sino que también lo es del ambiente y de la composición genética particular de una población. Es decir, la heredabilidad nos ayuda a entender cómo afectan los factores genéticos a las *diferencias* entre individuos de una población concreta; pero no nos ayuda a explicar el fenotipo concreto de ningún individuo.

A partir de la heredabilidad, se puede estimar la *respuesta a la selección*, a través de lo que se conoce como la **ecuación del ganadero**:

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

Donde R es la respuesta a la selección, h^2 es la heredabilidad y S es el *diferencial de selección*. Más concretamente:

- **Diferencial de selección:** la diferencia entre el fenotipo medio de los progenitores seleccionados y el fenotipo medio de la población.
- **Respuesta a la selección:** la diferencia entre el fenotipo medio de los descendientes y el fenotipo medio de la población 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.