

FARMACOLOGIA

FÍSICA

GENÈTICA

DRET

GEOLÒGIA

INFORMÀTICA

FILOLOGIA

VETERINÀRIA

PEDIATRIA

PERIODISME

BIOLOGIA

SOCIOLÒGIA

BOTÀNICA

HISTÒRIA

ECONOMIA

MATEMÀTIQUES

MEDICINA

ART

FILUSOFIA

GEOGRAFIA

PEBAGOGIA

QUÍMICA

ELECTRÒNICA

MÚSICA

BIOTECNOLOGIA

ANTROPOLOGIA

ZOOLOGIA

PSICOLOGIA

PSIQUIATRIA



Universitat Autònoma de Barcelona
Servei de Publicacions

Facultat de Veterinària

Jesús Piedrafita Arilla

Notas sobre teoría de mejora genética

M

c o l · l e c c i ó
A T E R I A L S

49

Jesús Piedrafita Arilla

Notas sobre teoría de mejora genética

Facultat de Veterinària

Universitat Autònoma de Barcelona
Servei de Publicacions
Bellaterra, 2024

Primera edició (paper): octubre de 1998
Primera edició (digital): novembre de 2024

Edició:
Servei de Publicacions
Universitat Autònoma de Barcelona
Edifici A. 08193 Bellaterra (Cerdanyola del Vallès). Spain
Tel. 93 581 10 22
sp@uab.cat
<https://publicacions.uab.cat>

ISBN (digital): 978-84-10202-38-2



Aquest llibre està publicat amb una llicència Creative Commons CC-BY-NC-ND.
El titular de l'obra autoritza a utilitzar els continguts sempre que es reconegui l'autoria.
No es permet fer un ús comercial, ni la generació d'obres derivades.

Dedicado a Esperanza, Nicolás y Álvaro

Índice

PRESENTACIÓN.....	7
VARIACIÓN CONTINUA.....	9
VALORES Y MEDIAS.....	13
ESTUDIO DE LAS VARIANZAS. I. DESCOMPOSICIÓN DE LA VARIANZA FENOTÍPICA	19
ESTUDIO DE LAS VARIANZAS. II. DESCOMPOSICIÓN DE LA VARIANZA AMBIENTAL	25
PARECIDO ENTRE PARIENTES.....	31
HEREDABILIDAD	37
CARACTERES CORRELACIONADOS	43
EVALUACIÓN GENÉTICA DE REPRODUCTORES. I. UN SOLO CARÁCTER.....	49
Índices de selección	50
Mejor predicción lineal insesgada (BLUP)	62
EVALUACIÓN GENÉTICA DE REPRODUCTORES. II. CARACTERES MÚLTIPLES	77
SELECCIÓN ARTIFICIAL. I. PREDICCIÓN DE LA RESPUESTA.....	83
SELECCIÓN ARTIFICIAL. II. MEDIDA DE LA RESPUESTA.....	93
SELECCIÓN ARTIFICIAL. III. CARACTERES MÚLTIPLES.....	103
DIFUSIÓN DEL PROGRESO GENÉTICO	107
SISTEMAS DE APAREAMIENTO. I. REPRODUCCIÓN CONSANGUÍNEA	111
SISTEMAS DE APAREAMIENTO. II. CRUZAMIENTO	119
CARACTERES DE UMBRAL.....	129
REFERENCIAS BIBLIOGRÁFICAS.....	135
ANEXOS	
Propiedades de la esperanza	139
Análisis de la varianza	140
Equivalencia de la correlación entre clases e intraclase	144
Nociones de álgebra matricial.....	146

Presentación

La asignatura “Mejora genética” surgió como un curso cuatrimestral, de tres horas semanales de clases teóricas, como consecuencia de la adaptación a la normativa comunitaria del actual plan de estudios de la Facultad de Veterinaria de la Universidad Autónoma de Barcelona. Dicha asignatura incorpora los contenidos de la materia “Cría y salud animal”.

La teoría de mejora genética abarca conceptos tanto de genética mendeliana, como de poblaciones, molecular y cuantitativa, en su aplicación a la mejora del ganado. Dado que las otras ramas de la genética han sido cubiertas en un curso anterior de la licenciatura, estas notas sólo recogen conceptos fundamentales de la genética cuantitativa en su aplicación a la mejora, haciendo hincapié en los aspectos de evaluación, selección y métodos de apareamiento de los reproductores, pues son las ideas relevantes que debe conocer un veterinario que se desenvuelve en el contexto de la producción y sanidad del ganado.

Las “Notas sobre teoría de mejora genética” que presentamos son producto de la evolución de las materias explicadas a lo largo de los últimos años en una asignatura anual de genética. No pretenden tener un ámbito de aplicación amplio, sino servir de base a las clases teóricas para los alumnos que cursan los estudios de licenciatura en nuestra facultad. Se ha tratado de soslayar la exhaustividad; su finalidad principal es evitar el trabajo de amanuense que debe realizar el alumno, ofreciéndole las definiciones expresadas con concreción y precisión y, asimismo, las deducciones algebraicas con una nomenclatura uniformada y, en lo posible, libres de errores. Las notas carecen, en general, de ejemplos, reservados fundamentalmente para las clases de problemas, que son un complemento imprescindible para un entendimiento satisfactorio de la materia aquí expuesta.

El contenido esencialmente cuantitativo de estas notas requiere para su entendimiento el dominio de ideas básicas de estadística: medias, varianzas, correlación, regresión y análisis de la varianza. No obstante, a modo de anexos, se han incluido al final de las notas algunos aspectos del análisis de la varianza y del álgebra matricial no cubiertos en cursos anteriores.

Las notas son deudoras de excelentes libros y notas publicadas por diferentes autores: Falconer (1983, 1986), Johansson y Rendel (1971), Schaeffer y Mao (1987), Turner y Young (1969), Cunningham (1969) y Pirchner (1983). A todos ellos, así como a distintos autores citados a lo largo del texto, quiero mostrarles mi agradecimiento.

Quiero también señalar la aportación de Rosa Pujol y Raquel Quintanilla, pues leyeron con bastante detenimiento los primeros manuscritos, realizando múltiples sugerencias que sin duda ayudaron a comprender mejor el texto. No obstante, debo agradecer también la paciencia de los alumnos del curso 1992-1993, los cuales recibieron una primera entrega, con más errores de los debidos, que he tratado de subsanar en esta edición. De cualquier manera, soy consciente de que todavía pueden quedar errores, tanto de expresión, como de concepto o tipográficos, que agradecería me comunicara cualquier lector que los detectase.

Bellaterra, enero de 1994

Variación continua Son bien conocidos algunos caracteres morfológicos, como el color de la capa de los cerdos o de las vacas Holstein, reproductivos, tal como la hiperprolificidad de los merinos australianos, o patológicos, como la respuesta al anestésico halotano en los cerdos, que permiten asignar los animales a una de entre dos o varias clases bien diferenciadas. Se sabe, además, que muchos de estos **caracteres cualitativos** responden a modelos sencillos de herencia mendeliana.

A diferencia de dichos caracteres, la mayor parte de los caracteres productivos, tales como la producción de leche por lactación o el crecimiento diario, presentan unas diferencias de grado difícilmente asignables a clases discretas. A la variación que no presenta discontinuidades naturales se la conoce como **variación continua**, y a los caracteres que la exhiben se les denomina **caracteres cuantitativos** o **métricos**. La rama de la genética que se ocupa de los caracteres cuantitativos se conoce con el nombre de *genética cuantitativa*. Los principios genéticos en los que se sustenta la genética cuantitativa son básicamente los de la *genética de poblaciones*, explicada en un curso anterior. No obstante, como la segregación de los genes implicados no puede seguirse de forma individual, nos veremos obligados a recurrir a nuevos conceptos y técnicas, éstas fundamentalmente estadísticas.

La importancia de la genética cuantitativa se debe al hecho señalado de que la mayor parte de los caracteres productivos del ganado son caracteres métricos; pero, además, muchos de los caracteres implicados en la microevolución lo son también. La genética cuantitativa será, por lo tanto, de aplicación a la mejora genética del ganado y asimismo ayudará a entender aspectos de la teoría evolutiva.

Naturaleza hereditaria de la variación continua La naturaleza hereditaria de la variación continua es un asunto que ocupó las polémicas de los genéticos en los albores del siglo XX. Mientras que los “mendelianos”, encabezados por Bateson y De Vries, proponían que todas las diferencias hereditarias evolutivamente importantes eran cualitativas y discontinuas, negando la naturaleza genética de las pequeñas variaciones, que atribuían al ambiente, los “biómetras”, con Galton a la cabeza, afirmaban que la variación hereditaria es básicamente cuantitativa y continua, llegando a negar la existencia de los genes como entidades separables.

En aquellos momentos resultaba difícil conciliar los dos puntos de vista, pues no existían pruebas de que los genes, presumiblemente particulados y discontinuos, pudieran explicar la herencia de fenotipos intermedios y continuos. Pero, además, ¿cómo podía explicarse la herencia de los caracteres discontinuos por medio de los mecanismos necesarios de fusión que presumiblemente producían los caracteres continuos? La respuesta a esta aparente contradicción vino de la mano de dos experimentos realizados por Johannsen (1903) y Nilsson-Ehle (1909).

Johannsen estudió el peso de las semillas comerciales de alubias (*Phaseolus vulgaris*). Mediante autopolinización logró una serie de líneas altamente endogámicas que

presentaban un peso medio uniforme, a pesar de que existía dentro de cada línea un cierto grado de variabilidad, menor que en la población total, que explicaba por la acción del ambiente. De acuerdo con el experimento de Johanssen, podíamos concebir una población de individuos que varían respecto a un carácter cuantitativo como consistente en un cierto número de grupos genéticamente distintos. Dentro de cada grupo el carácter cuantitativo presentaría un margen de variación de medidas debido a las diferencias ambientales existentes entre los individuos. No obstante, las diferencias entre las líneas puras no podían atribuirse a genes determinados, por lo que la genética de la variación continua aún quedaba sin resolver.

Yule (1906) sugirió, como previamente había apuntado Mendel, que la variación cuantitativa podía ser producida por multitud de genes, cada uno de ellos con un efecto pequeño sobre el carácter medido. En este sentido, Nilsson-Ehle aportaría la primera prueba experimental, siendo el color del trigo el carácter analizado. Cruzando variedades de trigo rojo con trigo blanco, tras autopolinizar la F_1 , encontró una gradación de tonalidades que iban del rojo al blanco, siendo la frecuencia de esta última de aproximadamente $1/64$. Las clases intermedias en color eran claramente más frecuentes que las extremas (rojas o blancas). Estos resultados podían explicarse por la segregación conjunta de tres *loci* dialélicos (gen rojo y gen blanco), que actuarían aditivamente, en el sentido de que cada gen rojo sumaría un pequeño grado de color a la semilla de la planta.

La frecuencia génica de los dos tipos de alelos (rojos y blancos) en la F_1 era $p = q = 0,5$. Los gametos formados se distribuirían según los coeficientes de la binomial $(1/2 + 1/2)^N$, según el número de genes rojos o blancos que portaran, siendo $N (= 3)$, el número de *loci*. Los genotipos, consecuentemente, se distribuirían siguiendo el desarrollo binomial $(1/2 + 1/2)^{2N}$. Aparecerían siete genotipos que contendrían 6, 5, 4, 3, 2, 1 y 0 alelos blancos (rojos) en las siguientes proporciones: 1 : 6 : 15 : 20 : 15 : 6 : 1, que, representadas, darían una aproximación a una ley normal.

Dado el acuerdo entre la distribución observada y la esperada según este modelo, el experimento sirvió para que se aceptara que la variación continua podía explicarse a partir de la acción de genes que se transmitían independientemente. Se acuñó el concepto de **factor múltiple** o **factor polímero** –rebautizado posteriormente como **poligén** por Mather (1943)– para designar genes que presentan un efecto pequeño sobre un carácter determinado y que pueden suplementarse entre sí para producir cambios cuantitativos observables. Añadida a la acción de los genes estaría la acción de múltiples efectos ambientales. Generalizando el modelo a un mayor número de *loci*, como se muestra en la figura 1, las clases se van reduciendo y su distribución se acerca cada vez más a una normal, que es la distribución que presentan la mayoría de los caracteres productivos del ganado.

Las distribuciones de la figura 1 se han construido asumiendo que la frecuencia de todos los alelos de todos los *loci* que determinan el carácter es la misma, $p = q = 0,5$. Este supuesto raramente es esperable en situaciones reales. La distribución normal observada en los caracteres productivos del ganado puede explicarse, no obstante, recurriendo al *teorema central del límite*, que establece que cuando los valores observados de una variable (carácter métrico) son el resultado de un conjunto muy grande de causas independientes, siendo cada efecto individual de poca importancia respecto al conjunto –tal es el caso de los efectos de los genes y del ambiente–, es esperable que los valores sigan una distribución normal.

En resumen, lo expuesto hasta ahora nos permite suponer el valor observado de un carácter métrico como el resultado de la acción conjunta de un gran número de *loci* que actúan aditivamente, y de influencias ambientales. Éste es el modelo básico que asumiremos en las páginas que siguen. Se trata de un **modelo infinitesimal** cuyo desarrollo debemos a Fisher (1918), Haldane (resumido en 1932) y Wright (1921).

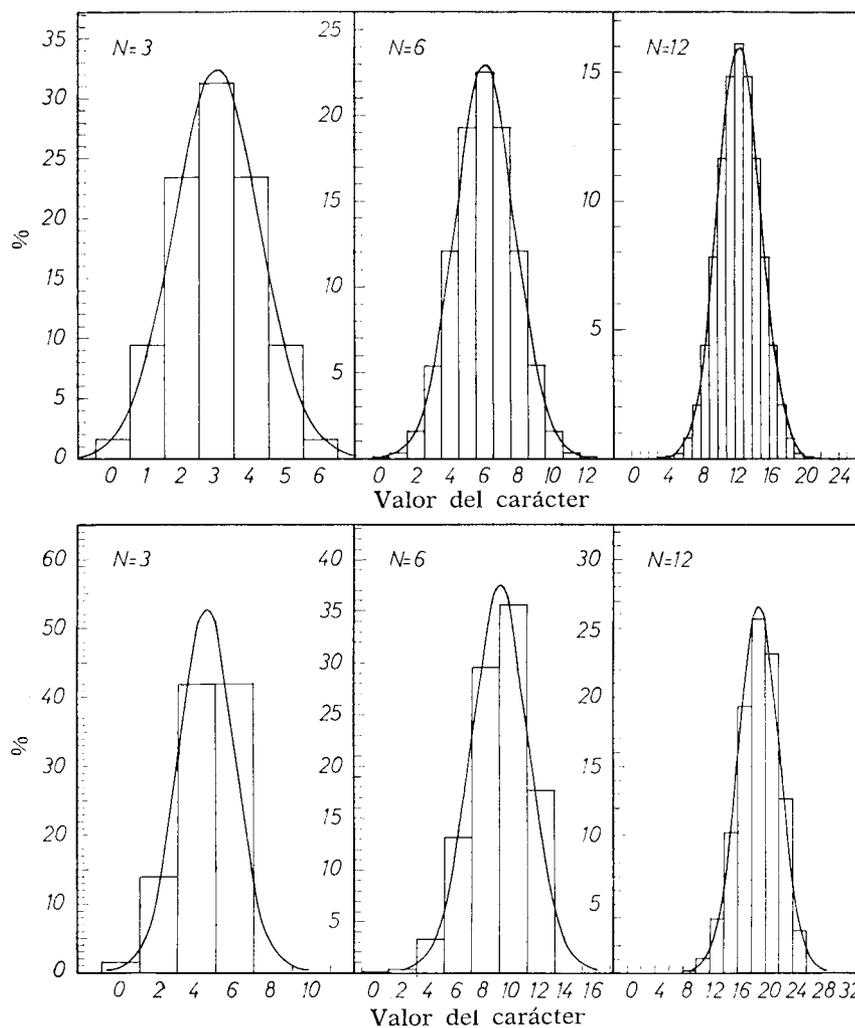


Figura 1. Distribución de diferentes fenotipos según el número de loci (N) que determinan el carácter y según no exista dominancia -parte superior- o esté presente -parte inferior. (Tomado de Johansson y Rendel, 1971.)

Valores y medias

El **valor fenotípico** de un individuo, P –usando la inicial inglesa de “phenotype”–, es el valor observado cuando se mide un carácter. Habitualmente, por razones de simplificación, se asume que el valor fenotípico es igual al valor **genotípico**, G , más la **desviación ambiental**, E –inicial de “environment”. De forma simbólica, podemos expresarlo como:

$$P = G + E$$

Los valores genotípico y ambiental son los componentes del valor fenotípico asociados con el genotipo –conjunto de genes que posee un individuo– y el ambiente, respectivamente. Dentro del ambiente se incluye todo lo que no sean causas genéticas, como son las condiciones climáticas, la iluminación, la temperatura, el manejo, el estado sanitario de la explotación, la alimentación, así como efectos del medio interno en el que se expresan los genes, p.e. la edad de la vaca, los problemas del desarrollo y, finalmente, los errores de medida.

El valor esperado de la desviación ambiental media se asume igual a cero, de manera que el valor fenotípico medio es igual al valor genotípico medio. La expresión **media de la población** se refiere por igual al valor fenotípico o al valor genotípico.

En principio el valor genotípico es medible, pero en la práctica no lo es, excepto cuando estamos tratando con un *locus* sencillo, donde los genotipos son distinguibles fenotípicamente. Utilizaremos este caso para deducir la relación entre la media de la población y los valores genotípicos y las frecuencias génicas de este *locus*.

Supongamos un *locus* dialélico, A , con los alelos A_1 y A_2 , que dan lugar a tres genotipos, A_1A_1 , A_1A_2 , y A_2A_2 . Podemos asociar a cada uno de los genotipos un valor de acuerdo con el siguiente esquema:

Genotipo	A_2A_2	A_1A_2	A_1A_1
	—+—	+ +—	+—
Valor genotípico	$-a$	$0 \quad d$	$+a$

El origen, o punto de valor cero, en esta escala se encuentra en el punto medio de los valores genotípicos de los dos homocigotos. El valor d del heterocigoto depende del grado de dominancia: si no hay dominancia, $d = 0$; si A_1 es dominante sobre A_2 , d es positivo, y si A_2 es dominante sobre A_1 , d es negativo. Si hay dominancia completa, $d = +a$ (ó $d = -a$), y si hay sobredominancia, d es mayor que $+a$ o menor que $-a$. El grado de dominancia se expresa como d/a . El intervalo total para el *locus* si no hay sobredominancia vale $2a$ (de $-a$ a $+a$).

Media de la población Supongamos una población grande sometida a apareamiento aleatorio, es decir, que las frecuencias son las esperadas en el equilibrio Hardy-Weinberg (H-W). Sean las frecuencias de A_1 y A_2 p y q , respectivamente.

La media de la población, M , puede deducirse como sigue:

Genotipo	Frecuencia	Valores	Frecuencia \times Valores
A_1A_1	p^2	$+a$	p^2a
A_1A_2	$2pq$	d	$2pqd$
A_2A_2	q^2	$-a$	$-q^2a$
			$M = (p^2 - q^2)a + 2pqd$
			$= a(p-q) + 2dpq$

donde $a(p-q)$ es la aportación de los homocigotos a la media y $2dpq$ es la aportación de los heterocigotos.

Generalizando al caso de que el carácter esté determinado por varios o un gran número de *loci*, la media de la población resultante de la acción conjunta de todos los *loci* sería $M = \Sigma a(p-q) + 2\Sigma dpq$, con una amplitud de variación total de $2\Sigma a$.

Efecto medio de un gen Como consecuencia de la meiosis, los individuos no transmiten los genotipos, sino los genes. El **efecto medio de un gen** es la desviación a la media de la población de los individuos que lo han recibido de un progenitor apareado con otro progenitor tomado al azar de la población. Dicho de otra manera, si gametos portadores del gen A_1 se unen al azar con gametos de la población, la desviación media de los genotipos resultantes con respecto a la media de la población es el efecto medio del gen. Si pudiésemos cambiar genes A_1 por genes A_2 tomados al azar en la población, podríamos cuantificar el cambio resultante, que sería el **efecto medio de sustitución del gen**.

Veamos ahora como se relaciona el efecto medio con los valores genotípicos y las frecuencias génicas:

Tipo de gameto	Valores y frecuencias de genotipos producidos			Valor medio de genotipos prod.	Media de la población a deducir	Efecto medio del gen
	A_1A_1	A_1A_2	A_2A_2			
	$+a$	d	$-a$			
A_1	p	q		$pa + qd$	$-[a(p-q)+2dpq]$	$q[a + d(q-p)]$
A_2		p	q	$-qa + pd$	$-[a(p-q)+2dpq]$	$-p[a+d(q-p)]$

Esta tabla expresa el hecho de que si los gametos A_1 se unen al azar con otros gametos de la población, una proporción p de las veces lo harán con gametos A_1 , mientras que una proporción q de las veces lo harán con gametos A_2 , originando genotipos A_1A_1 y A_1A_2 , respectivamente. La media de esos genotipos será $pa + qd$. La diferencia entre este valor medio y la media de la población será el efecto medio del gen A_1 , α_1 . Es decir,

$$\begin{aligned} \alpha_1 &= pa + qd - [a(p-q)+2dpq] \\ &= q[a + d(q-p)] \end{aligned}$$

De igual forma, el efecto medio del gen A_2 , α_2 , es:

$$\alpha_2 = -p[a + d(q-p)]$$

Consideremos ahora el efecto medio de la sustitución de un gen, α , suponiendo que el alelo A_2 es sustituido por el alelo A_1 .

	Cambio	Frecuencia	Cambio \times frecuencia
$A_1A_2 \rightarrow A_1A_1$	$(a-d)$	p	$p(a-d)$
$A_2A_2 \rightarrow A_1A_2$	$(d+a)$	q	$q(d+a)$
			$\alpha = p(a-d) + q(d+a)$ $= pa - pd + qd + qa$ $= a + d(q-p)$

Es fácil comprobar que $\alpha_1 = q\alpha$
 $\alpha_2 = -p\alpha$
 y por lo tanto, $\alpha = \alpha_1 - \alpha_2$

Es importante resaltar que el efecto medio de un gen, o de sustitución de un gen, depende de la frecuencia génica y de los valores genotípicos, y que es (son), por lo tanto, una propiedad de la población y del gen.

Valor mejorante

Hemos visto que los progenitores transmiten a su descendencia los genes, no los genotipos. Por ello, son los efectos medios de los genes de los progenitores los que van a determinar el valor genotípico medio de la descendencia.

Definimos como **valor mejorante** de un individuo el doble de la desviación media de su descendencia en relación con la media de la población, cuando el individuo se ha apareado con una muestra de individuos tomados al azar en dicha población. Se supone que el valor genotípico medio de los individuos tomados al azar, expresado como desviación a la media, es cero. Por ello, y dado que el progenitor solamente aporta la mitad de sus genes a su descendencia, debemos doblar la desviación media de ésta para obtener el valor mejorante.

Podemos definir teóricamente el valor mejorante de un individuo como la suma de los efectos medios de todos los *loci* que afectan al carácter. Por ejemplo, para el caso de un carácter determinado por un *locus* dialélico:

<u>Genotipo</u>	<u>Valor mejorante</u>
A_1A_1	$2\alpha_1 = 2q\alpha$
A_1A_2	$\alpha_1 + \alpha_2 = (q-p)\alpha$
A_2A_2	$2\alpha_2 = -2p\alpha$

Para generalizarlo al caso de múltiples *loci*, el valor mejorante de un individuo, definido en términos de efectos medios, es igual a la suma de los efectos medios de los genes de que es portador, considerando dos genes por cada *locus* y todos los *loci*. O lo que es igual, la suma de los valores mejorantes de cada uno de los *loci* que determinan el carácter.

El valor mejorante es, pues, una propiedad del individuo y de la población de la que se obtienen sus parejas. La consecuencia más importante de esto es que debemos referir el valor mejorante de un individuo a la población en la cual se realizó la medición.

En una población en equilibrio H-W el valor mejorante medio debe ser cero. Vamos a demostrarlo para un *locus*, siendo obvia la generalización al conjunto de los *loci*.

Genotipos	A_1A_1	A_1A_2	A_2A_2
Frecuencias	p^2	$2pq$	q^2
Valores mejorantes	$2q\alpha$	$(q-p)\alpha$	$-2p\alpha$
Suma valor mejorante \times frecuencia	$2p^2q\alpha + 2pq\alpha(q-p) + (-2)pq^2\alpha$ $= 2pq\alpha(p+q-p-q) = 0$		

El valor mejorante está asociado con el efecto aditivo de los genes, razón por la cual se le designa habitualmente como A .

Desviación dominante Normalmente no coinciden el valor mejorante, A , y el valor genotípico, G . La diferencia entre estos valores se conoce como **desviación dominante**, D , que surge debido a que los individuos tienen los genes por pares. Desde el punto de vista estadístico, la desviación de dominancia es una interacción entre alelos o *intra*locus. Cuando se considera un solo *locus*,

$$G = A + D$$

En ausencia de dominancia, el valor mejorante y el valor genotípico coinciden. Es fácil comprobar que las desviaciones dominantes dependen también de las frecuencias génicas. Recordemos que $M = a(p-q) + 2dpq$, $\alpha = a + d(q-p)$ y $a = \alpha - d(q-p)$. Supongamos el valor genotípico del genotipo A_1A_1 expresado como desviación a la media de la población en los siguientes términos

$$a - [a(p-q) + 2dpq] = a - ap + aq - 2dpq = a(1-p+q) - 2dpq = 2aq - 2dpq = 2q(a-dp)$$

y sustituyendo a por su valor (en términos de α) se obtiene,

$$2q(a - dp) = 2q(\alpha - dq + dp - dp) = 2q(\alpha - qd)$$

Si a este valor genotípico le sustraemos el valor mejorante, tenemos:

$$2q(\alpha - qd) - 2q\alpha = -2q^2d$$

Operando de forma similar en cada uno de los tres genotipos, puede construirse la siguiente tabla:

Genotipos	A_1A_1	A_1A_2	A_2A_2
Frecuencias	p^2	$2pq$	q^2
Valores genotípicos	$+a$	d	$-a$
Desviaciones a la media de la población			
Valor genotípico	$2q(a-pd)$ $2q(\alpha-qd)$	$a(q-p)+d(1-2pq)$ $(q-p)\alpha+2pqd$	$-2p(a+qd)$ $-2p(\alpha+pd)$
Valor mejorante	$2q\alpha$	$(q-p)\alpha$	$-2p\alpha$
Desviación dominante	$-2q^2d$	$+2pqd$	$-2p^2d$

Puede apreciarse que los tres valores de la desviación dominante son función de d . Si no hay dominancia, d es cero y todas las desviaciones dominantes son también cero. Pero las desviaciones dominantes dependen además de las frecuencias génicas; por lo tanto, son parcialmente propiedad de la población y no simples medidas del grado de dominancia.

Dentro de cada *locus*, la desviación dominante media es cero:

$$-2q^2d(p^2) + 2pqd(2pq) - 2p^2d(q^2) = 0$$

Hemos visto antes que el valor mejorante promedio era 0. Al serlo también la desviación dominante media, debemos concluir que dentro de un *locus* el valor genotípico medio es también 0.

La relación entre valores genotípicos, valores mejorantes y desviaciones de dominancia puede expresarse mediante el gráfico representado en la figura 2. El valor genotípico (círculos negros) se relaciona con el número de genes. Para ello se ajusta una recta de regresión (por mínimos cuadrados) de dichos valores genotípicos sobre el número de genes (A_1 o A_2). Sobre la recta de regresión quedan los valores mejorantes (círculos claros), siendo la pendiente de la recta de regresión el efecto de sustitución del gen, α . Podría entenderse la situación en el sentido de ver cómo cambia el valor genotípico a medida que aumenta el número de genes A_1 ; es decir, cero para el genotipo A_2A_2 , uno para el genotipo A_1A_2 , y dos para el genotipo A_1A_1 . Las diferencias entre los valores genotípicos y los valores mejorantes son las desviaciones de dominancia. El efecto medio de la sustitución de un gen fue definido por Fisher (1918) precisamente como la regresión lineal de los valores genotípicos sobre el número de genes.

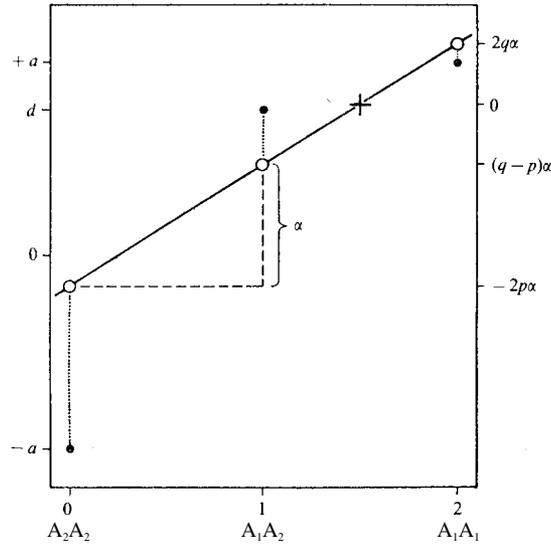


Figura 2. Representación gráfica de los valores genotípicos, valores mejorantes y efecto medio de sustitución en un locus dialélico. (Tomado de Falconer, 1983.)

Desviaciones de interacción

Cuando se considera el genotipo integrado por más de un locus, el valor genotípico puede contener una desviación adicional debida a una combinación no aditiva. Supongamos el caso sencillo de un carácter determinado por dos loci A y B, donde G_A y G_B son los valores genotípicos atribuibles a cada uno de los loci A y B. Entonces, el valor genotípico conjunto, G , atribuible a los dos loci será

$$G = G_A + G_B + I_{AB}$$

donde I_{AB} sería la desviación a la combinación aditiva de estos dos valores genotípicos. Cuando tratamos loci individualmente, supusimos que esta desviación era cero para todos los loci. Si este valor es distinto de cero, los loci se dice que interaccionan o que muestran epistasis. La desviación I se denomina **desviación de interacción** o **desviación epistática**.

La interacción puede darse entre dos, tres o más loci, incluso entre valores genotípicos y desviaciones de dominancia. No obstante, no nos preocuparemos de distinguir las, sino que asumiremos un modelo sencillo que engloba los distintos tipos de interacción. Dicho modelo lo simbolizaremos como

$$G = A + D + I$$

donde A es la suma de los valores mejorantes (valores aditivos) y D es la suma de las desviaciones dominantes. Si la desviación de interacción media es cero, diremos que los genes implicados actúan aditivamente entre loci. Debemos resaltar que el concepto de acción aditiva tiene dos significados diferentes. En el contexto de un locus significa ausencia de dominancia, mientras que cuando se refiere a genes de diferentes loci significa ausencia de epistasis.

Como resumen, podemos expresar el valor fenotípico de un individuo como sigue:

$$\begin{aligned} P &= G + E \\ &= A + D + I + E \end{aligned}$$

Estudio de las varianzas. I. Descomposición de la varianza fenotípica

La idea básica del estudio de la variación, medida a través de la varianza, es su descomposición en componentes atribuibles a diferentes causas. Veremos como la proporción relativa de las diferentes causas de variación respecto de la variación total determina las propiedades genéticas de la población y el grado de parecido entre parientes. En este apartado consideraremos la naturaleza de estos componentes, así como también la dependencia de las frecuencias génicas de los componentes genéticos.

La cantidad de variación se mide mediante la varianza: $V(x) = E(x^2) - E(x)^2$. Cuando expresamos los valores como desviaciones a la media de la población, la varianza es simplemente la media de los valores al cuadrado, $E(x^2)$.

Los componentes en los que se descompone la varianza son los mismos que los componentes del valor fenotípico descritos anteriormente. Podemos resumirlos de la siguiente manera:

<u>Componente de varianza</u>	<u>Símbolo</u>	<u>Componente del valor</u>
Fenotípica	V_P	P Valor fenotípico
Genotípica	V_G	G Valor genotípico
Aditiva	V_A	A Valor mejorante
De dominancia	V_D	D Desviación dominante
De interacción	V_I	I Desviación de interacción
Ambiental	V_E	E Desviación ambiental

La **varianza fenotípica** o total, asumiendo covarianzas nulas entre componentes, es la suma de componentes (véase el anexo 1):

$$\begin{aligned} V_P &= V_G + V_E \\ &= V_A + V_D + V_I + V_E \end{aligned}$$

Grado de determinación genética

La descomposición de la varianza nos permite cuantificar la importancia relativa de cada uno de los componentes por separado, es decir, saber en qué medida influyen los factores hereditarios o ambientales en la determinación de la variación total. La importancia relativa de una fuente de variación es el cociente entre dicha varianza y la varianza total.

La importancia relativa de la herencia en la determinación de los valores fenotípicos se denomina **heredabilidad** del carácter. No obstante, podemos distinguir dos significados distintos de heredabilidad según se haga referencia a los valores genotípicos o a los valores mejorantes. Definimos como **grado de determinación genética**, o heredabilidad en sentido amplio, H^2 , el cociente varianza genotípica / varianza fenotípica, V_G/V_P , y expresa la medida en que los fenotipos están determinados por los genotipos.

Componentes genéticos de la varianza

La heredabilidad en sentido estricto o simplemente **heredabilidad**, h^2 , se define como el cociente varianza aditiva / varianza fenotípica, V_A/V_P , que expresa el grado en que los fenotipos vienen determinados por los genes transmitidos por sus padres. En general se habla de heredabilidad referida a este último concepto.

Hemos visto cómo la varianza genotípica podía descomponerse en varianza aditiva, de dominancia y de interacción. De entre estos componentes, el componente de varianza aditiva es el más importante puesto que, como veremos, es la principal causa de parecido entre parientes, determinando las propiedades genéticas de la población y la respuesta a la selección. En la práctica, por lo tanto, lo importante es la proporción de la varianza aditiva frente al resto, es decir V_A/V_P , la heredabilidad.

Vamos a ver ahora como influyen las frecuencias génicas en los componentes genéticos de la varianza. Analizaremos el caso de nuevo con un *locus* dialélico. Hemos visto anteriormente que los valores mejorantes y las desviaciones dominantes pueden expresarse como desviaciones a la media de la población, desviaciones que tienen, obviamente, media cero. De esta manera, el cálculo de la varianza es muy sencillo, pues es el promedio de las desviaciones al cuadrado.

Varianza aditiva

La varianza de los valores mejorantes, o varianza aditiva, V_A , es:

$$\begin{aligned} V_A &= (2q\alpha)^2 \cdot p^2 + (q-p)^2\alpha^2 \cdot 2pq + (-2p\alpha)^2 \cdot q^2 \\ &= 2pq\alpha^2 (2pq + (q-p)^2 + 2pq) \\ &= 2pq\alpha^2 (p^2 + 2pq + q^2) \\ &= 2pq\alpha^2 \\ &= 2pq [a + d(q-p)]^2 \end{aligned}$$

Varianza de dominancia

Las desviaciones de dominancia dan lugar a la varianza de dominancia:

$$\begin{aligned} V_D &= (-2q^2d)^2 \cdot p^2 + (2pqd)^2 \cdot 2pq + (-2p^2d)^2 \cdot q^2 \\ &= d^2 (4q^4p^2 + 8p^3q^3 + 4p^4q^2) \\ &= 4p^2q^2d^2 (q^2 + 2pq + p^2) \\ &= (2pqd)^2 \end{aligned}$$

Las varianzas aditiva y dominante dependen de las frecuencias génicas y del efecto medio de sustitución y el valor d de la población; por lo tanto, son una propiedad de la población y del carácter.

En el caso de que el *locus* sea multialélico, las expresiones generales de la varianza aditiva y dominante son como sigue:

$$\begin{aligned} V_A &= 2\sum_i p_i\alpha_i^2 \\ V_D &= \sum_{ij} p_i p_j \delta_{ij}^2 \end{aligned}$$

donde δ_{ij} son las correspondientes desviaciones dominantes expresadas como diferencia con la media de la población.

La **varianza genética total** que se origina en un *locus*, dado que $G = A + D$, puede calcularse como:

$$V_G = V_A + V_D + 2 \text{cov}(A,D)$$

La teoría estadística nos dice que $\text{cov}(A,D) = E(AD) - E(A)E(D)$, y si expresamos los valores mejorantes y las desviaciones dominantes como desviaciones a la media de la población, su esperanza es cero. Vamos a demostrar que los valores mejorantes y las desviaciones de dominancia son variables estadísticamente independientes, es decir, que no están correlacionadas o, lo que es igual, que su covarianza es cero.

$$\begin{aligned} E(AD) &= 2q\alpha \cdot (-2q^2d) \cdot p^2 + (q-p)\alpha \cdot 2pqd \cdot 2pq + (-2p\alpha) \cdot (-2p^2d) \cdot q^2 \\ &= 4p^2q^2\alpha d(-q+q-p+p) = 0 \end{aligned}$$

La varianza genotípica total queda como sigue:

$$\begin{aligned} V_G &= V_A + V_D \\ &= 2pq [a + d(q-p)]^2 + (2pqd)^2 \end{aligned}$$

A veces se interpreta el concepto de varianza aditiva de una forma errónea. La existencia de varianza aditiva no conlleva que los genes actúen exclusivamente de una forma aditiva, es decir, que no muestren dominancia ni epistasia. La varianza aditiva puede surgir de genes que muestren cualquier grado de dominancia o epistasia; sólo si podemos comprobar que toda la varianza genotípica es aditiva podremos concluir que los genes no muestran dominancia ni epistasia.

La manera en que influyen las frecuencias génicas y el grado de dominancia en la magnitud de los componentes genéticos de la varianza puede expresarse gráficamente, según muestra la figura 3.

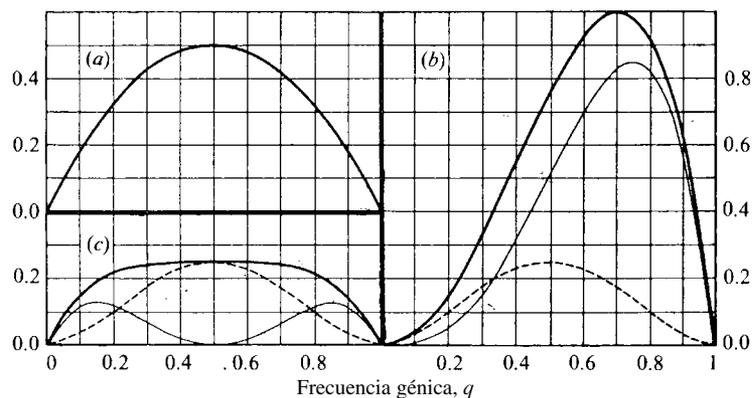


Figura 3. Magnitud de los componentes genéticos de la varianza originados en un *locus* dialélico según las frecuencias génicas. (Tomado de Falconer, 1983.)

Se estudian tres situaciones:

- No hay dominancia: la varianza genotípica –línea gruesa– es toda aditiva. Es máxima a una frecuencia intermedia, $p = q = 0,5$.
- Hay dominancia completa. La varianza de dominancia –línea discontinua– es máxima en frecuencias intermedias, $p = q = 0,5$; la varianza aditiva –línea fina– alcan-

za su máximo cuando la frecuencia del alelo recesivo es $q = 0,75$, y la varianza total es máxima cuando $q^2 = 0,5$, es decir, $q = 0,71$.

- c) Hay sobredominancia pura (los valores genotípicos de A_1A_1 y A_2A_2 son iguales), $a = 0$. La varianza de dominancia es la misma que en el segundo caso; la varianza aditiva es cero para $p = q = 0,5$ y tiene dos máximos, para $q = 0,15$ y $q = 0,85$. La varianza genotípica permanece prácticamente constante en un amplio margen de frecuencias génicas, aunque su composición varía profundamente.

Como conclusión general del estudio de esta figura, se deduce que los *loci* que presentan frecuencias génicas intermedias aportan más varianza que aquéllos cuyos alelos presentan frecuencias génicas extremas.

Para calcular los componentes de varianza tenemos que combinar los efectos de todos los *loci* que determinan la varianza. En una población en equilibrio H-W, las varianzas aditiva y dominante totales son la suma de las varianzas aditiva y de dominancia de cada uno de los *loci* implicados. Si existen desviaciones de interacción, aparece un nuevo componente que vamos a describir.

Varianza de la interacción

Cuando los *loci* muestran interacción, aparece un nuevo componente de varianza que denominamos varianza de interacción, V_I . El análisis de este componente requiere su subdivisión en nuevos componentes.

Según el número de *loci* implicados, hablaremos de interacción entre dos factores cuando son dos los *loci* implicados; de tres factores, cuando interaccionan tres *loci*, y así sucesivamente. No obstante, nos centraremos en la interacción de dos factores, puesto que las que implican un elevado número de *loci* tienen una aportación insignificante a la varianza de interacción.

Según sea la naturaleza de efectos que interaccionan, podemos distinguir:

1. Interacción entre dos valores mejorantes, que da lugar a un componente de varianza aditiva \times aditiva, V_{AA} .
2. Interacción entre el valor mejorante de un *locus* y la desviación dominante del otro, dando lugar a una varianza aditiva \times dominante, V_{AD} .
3. Interacción entre dos desviaciones dominantes, que da lugar a un componente de varianza dominante \times dominante, V_{DD} .

De esta manera, la varianza de interacción es la suma de los componentes:

$$V_I = V_{AA} + V_{AD} + V_{DD} + etc.$$

donde el término *etc.* se refiere a componentes similares que surgen de interacciones semejantes entre más de dos *loci*.

Aunque no cabe duda de que la interacción entre *loci* que controlan un carácter cuantitativo puede ser un fenómeno frecuente, se conoce poco acerca de la importancia relativa de la varianza de interacción como fuente de variación.

Correlación e interacción entre genotipo y ambiente

En relación con la partición de la varianza realizada pueden surgir dos complicaciones que pueden despreciarse normalmente sin que ello afecte a las conclusiones extraídas de la descomposición de la varianza, pero es importante conocer su existencia.

Correlación

Hasta ahora hemos asumido que los valores genotípicos y las desviaciones ambientales son independientes entre sí. En circunstancias reales podría, no obstante, surgir una correlación entre genotipo y ambiente si a los mejores genotipos se les diera los mejores ambientes. La producción de leche del ganado vacuno lechero constituye un ejemplo, ya que es práctica normal alimentar las vacas según su nivel de producción. Esta correlación entre valor fenotípico y desviación ambiental da lugar a una correlación entre valor genotípico y desviación ambiental.

Cuando existe una correlación de este tipo, la varianza fenotípica se ve aumentada por el doble de un componente de covarianza entre genotipo y ambiente, cov_{GE} :

$$V_P = V_G + V_E + 2cov_{GE}$$

Interacción

Hemos supuesto hasta ahora que una diferencia específica de ambiente tiene el mismo efecto sobre diferentes genotipos; o, en otras palabras, que podemos asociar una cierta desviación ambiental con una diferencia específica de ambiente, con independencia del genotipo sobre el que actúa. Cuando esto no sucede, existe una interacción estadística entre genotipo y ambiente.

Esta interacción puede presentarse de dos formas:

1. una diferencia específica de ambiente podría tener un mayor efecto sobre unos genotipos que sobre otros;
2. dicha diferencia produciría un cambio en el orden de mérito de una serie de genotipos que se evalúan en distintos ambientes.

La varianza fenotípica incluye en estos casos un componente de interacción, V_{GE} :

$$V_P = V_G + V_E + V_{GE}$$

Estudio de las varianzas. II. Descomposición de la varianza ambiental.

La varianza ambiental incluye la variación debida a una gran cantidad de causas de naturaleza no genética. En general, la varianza ambiental es una fuente de error que reduce la precisión de los estudios genéticos. Por ello, el mejorador se ve obligado a tratar de minimizarla a través de un manejo adecuado o aplicando técnicas estadísticas apropiadas.

Entre las causas de variación ambiental tenemos los factores nutritivos, de manejo y climáticos. Los efectos maternos, aunque de origen genético en la madre, son ambientales en su efecto sobre la descendencia, y se ha comprobado su incidencia tanto prenatal como postnatal. Los errores de medida son otra fuente habitual de varianza ambiental, generalmente despreciable. Se ha observado, no obstante, que aparatos de alta tecnología como los ecógrafos pueden dar valores muy dispares cuando la medición, por ejemplo del área del lomo en un cerdo, la realizan distintos operarios. Cuando la valoración de un carácter consiste en la asignación subjetiva de una puntuación en una escala, caso de la valoración morfológica de las vacas, estos errores pueden ser superiores. No obstante, en la práctica, con evaluadores avezados las variaciones suelen ser muy pequeñas, sobre todo cuando la escala es suficientemente amplia.

Frente a las causas de variación que acabamos de describir, existe también una variación intangible. Son factores ambientales que somos incapaces de conocer y no pueden eliminarse mediante el diseño experimental; podemos citar como ejemplo las variaciones del desarrollo que afectan a caracteres métricos relacionados con la estructura anatómica.

Repetibilidad

Se trata de un parámetro que puede estimarse cuando es posible hacer más de una medida de un carácter en un mismo individuo. La repetibilidad no se deriva de la descomposición de la varianza genética, sino de la de la varianza ambiental, aunque tiene algunas implicaciones prácticas para el análisis genético y la evaluación de los reproductores. Estadísticamente se define la **repetibilidad** como la correlación entre medidas repetidas de un mismo individuo.

Estimación de la repetibilidad

Una forma de medir la repetibilidad es la correlación entre clases (o de Pearson):

$$r = \frac{\text{COV}(y_1, y_2)}{\sigma_{y_1} \sigma_{y_2}}$$

donde y_1 e y_2 serían dos mediciones realizadas en el mismo individuo.

Otra manera equivalente de estimar la repetibilidad es a través de la correlación intra-clase, que designamos como r_I (véase el anexo 3). Su cálculo se deriva del análisis de la varianza, y se aplica cuando el número de observaciones por individuo es mayor que dos. Lo definiremos como el cociente de la varianza entre *–between–* grupos, σ_B^2 , a la varianza fenotípica o total, σ_P^2 (entre grupos, más dentro de *–within–* grupos, σ_W^2). Es decir,

$$r_I = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_P^2}$$

Supongamos un diseño de análisis de varianza jerárquico a dos niveles en el que hemos medido en kg la producción de las tres primeras lactaciones:

V_1	V_2	...	V_i	...	V_s
y_{11}	y_{21}		y_{i1}		y_{s1}
y_{12}	y_{22}		y_{i2}		y_{s2}
y_{13}	y_{23}		y_{i3}		y_{s3}

Asumamos que el modelo matemático que describe los datos es el siguiente modelo lineal:

$$y_{ij} = \mu + V_i + e_{ij}$$

donde

μ efecto de la media general

V_i efecto aleatorio de la vaca i -ésima ($i = 1, 2, 3, \dots, s$)

e_{ij} efecto del error o residuo no explicado por el modelo ($j = 1, 2, 3, \dots, n$).

Siguiendo el anexo 2, podemos construir una tabla de análisis de varianza para un diseño equilibrado:

Fuente de variación	Grados de libertad	Suma de cuadrados (SC)	Cuadrado medio (CM)	Esperanza de los cuadrados medios (ECM)
Entre (B)	$s-1$	$SC_B = \sum_i (y_i^2/n) - (y_{..}^2/ns)$	$SC_B / (s-1)$	$\sigma_W^2 + n\sigma_B^2$
Dentro de (W)	$s(n-1)$	$SC_W = \sum_{ij} y_{ij}^2 - \sum_i (y_i^2/n)$	$SC_W / (s(n-1))$	σ_W^2
Total	$sn-1$	$SC_P = \sum_{ij} y_{ij}^2 - (y_{..}^2/ns)$	$SC_P / (sn-1)$	

Los componentes σ_B^2 y σ_W^2 se denominan componentes observables, al estimarse a partir de los datos.

De una forma general, distinguiremos dos tipos de repetición de las medidas:

1. Medidas repetidas en el tiempo, repetición temporal, como por ejemplo la producción lechera de una vaca en lactaciones sucesivas o el número de lechones nacidos

vivos en sucesivos partos, entendido este último carácter como propio de la hembra. La varianza del carácter se puede descomponer en componente de varianza dentro de individuos que mide las diferencias entre los valores observados (producciones) en el mismo individuo, y componente de varianza entre individuos, que mide las diferencias permanentes entre individuos. El componente dentro de individuos es enteramente ambiental: riqueza del pienso dependiendo de la época, recursos forrajeros, patología temporal, etc. El componente entre individuos es en parte ambiental y en parte genético, debiéndose la parte ambiental a circunstancias que afectan de modo permanente. De esta manera puede separarse la varianza debida a circunstancias ambientales del resto.

2. Medidas repetidas en el espacio, repetición espacial, como por ejemplo el peso de los dos jamones de un cerdo o la terneza del músculo en diferentes muestras tomadas en la pechuga de un pavo. En este caso la varianza dentro de individuos es enteramente ambiental, representando variaciones del desarrollo que surgen de circunstancias localizadas; mientras que la variación entre individuos, en cambio, será debida a influencias no localizadas o permanentes.

A fin de discutir conjuntamente las repeticiones espacial y temporal, se usará en lo sucesivo los términos varianza ambiental especial, V_{Es} , para referirnos a la varianza dentro de individuos, que surge de circunstancias temporales o localizadas, y varianza ambiental general o permanente, V_{Eg} , que hace referencia al componente entre individuos que surge de circunstancias permanentes o no localizadas. V_{Es} y V_{Eg} son los componentes causales de la varianza en este caso.

Conceptualmente, la repetibilidad se define como el cociente entre la varianza genotípica más la varianza ambiental permanente —es decir, dos componentes asociados a causas que tienen un efecto permanente en el individuo— y la varianza fenotípica. Simbólicamente:

$$r = \frac{(V_G + V_{Eg})}{V_P}$$

De esta ecuación es de resaltar el que la repetibilidad marque el límite superior tanto del grado de determinación genética, como de la heredabilidad. El componente de varianza ambiental no permanente puede deducirse fácilmente:

$$V_{Es} / V_P = 1 - r$$

Hemos visto que la varianza fenotípica es igual a la suma de la varianza genotípica más la varianza ambiental general (permanente) y más la varianza ambiental especial (diferencias temporales y localizadas). Simbólicamente,

$$V_P = V_G + V_{Eg} + V_{Es}$$

Si comparamos esta expresión con la de la varianza fenotípica ($\sigma_p^2 = \sigma_b^2 + \sigma_w^2$), el componente de varianza entre grupos contendrá los componentes de varianza genotípica y varianza ambiental general, quedando igualado el componente dentro de grupos al componente de varianza ambiental especial. Dadas estas equivalencias, tal y como habíamos avanzado, la repetibilidad se estimará como el cociente de la varianza entre (σ_B^2) y la varianza fenotípica (σ_p^2):

$$r = \frac{V_G + V_{Eg}}{V_P} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_P^2}$$

En la idea de repetibilidad hay implícitos dos supuestos:

1. Igualdad de la varianza de diferentes medidas y que éstas tengan sus componentes en las mismas proporciones. Esto se entenderá mejor si analizamos lo descrito en el anexo 3.
2. Que las diferentes medidas reflejen el mismo carácter genético.

Valores estimados de repetibilidad

La repetibilidad varía mucho según la naturaleza del carácter y también, por supuesto, de acuerdo con las propiedades genéticas de la población. En la tabla 1 se presentan valores promedio de repetibilidad estimados en distintos caracteres y especies ganaderas.

Tabla 1. Estimaciones de repetibilidad referentes a distintos caracteres productivos del ganado. [Adaptado de Pirchner (1983) y Falconer (1983)]

Ganado vacuno

Producción de leche	0,35 – 0,60
Producción de grasa	0,35 – 0,45
Contenido graso	0,50 – 0,70
Duración del secado	0,15 – 0,25
Intervalo entre partos	0,01 – 0,10
Inseminaciones por concepción	0,01 – 0,05
Duración de la gestación	0,15 – 0,20

Ganado ovino

Peso del vellón limpio	0,50 – 0,70
Peso corporal	0,50 – 0,70
Diámetro de la fibra de lana	0,50 – 0,60

Ganado porcino

Número de nacidos vivos	0,12 – 0,18
Peso de la camada al destete	0,04 – 0,09

Ratones

Peso a las seis semanas	0,95
Tamaño de la camada	0,45

Aplicaciones de la repetibilidad

1. La repetibilidad marca el límite superior de la heredabilidad, como hemos comentado anteriormente. Normalmente, los valores estimados de repetibilidad son superiores a los de heredabilidad para un mismo carácter. La repetibilidad, no obstante, es más fácil de estimar que la heredabilidad, pues puede calcularse a partir de datos repetidos tomados en un mismo individuo. La heredabilidad, en cambio, necesita del conocimiento de la estructura genealógica de los individuos medidos. La repetibilidad nos orientará, en aquellos caracteres en los que pueda estimarse, del grado en que un carácter es hereditario. Si el valor estimado de la repetibilidad de un carácter en una población es bajo, sólo podremos esperar un valor de heredabilidad bajo.
2. Aumentar la precisión de la predicción de los valores genéticos de los individuos. Aunque este tema se abordará de nuevo en el capítulo de índices de selección, vamos a hacer algunas consideraciones. El único componente de varianza que se ve reducido por mediciones repetidas es el debido al ambiente especial, V_{Es} , siendo la magnitud de la reducción dependiente del número de medidas realizadas. Esta reducción representa la ganancia en precisión. Supongamos que medimos un individuo n veces y tomamos la media de esas n medidas como valor fenotípico del individuo, es decir, $P_{(n)}$. La varianza fenotípica estará compuesta por:

$$V_{P_{(n)}} = V_G + V_{Eg} + (1/n) V_{Es}$$

Para expresar esta ecuación en términos de repetibilidad basta multiplicar y dividir ordenadamente por V_P :

$$\begin{aligned} V_{P_{(n)}} &= [V_G + V_{Eg} + (1/n) (V_P - V_G - V_{Eg})] V_P / V_P \\ &= \left(r + \frac{1}{n} (1-r) \right) V_P \end{aligned}$$

Reordenando la ecuación anterior, podemos expresar la varianza fenotípica de n medidas en función de la varianza fenotípica total:

$$\frac{V_{P_{(n)}}}{V_P} = \frac{1 + r(n-1)}{n}$$

Cuando la repetibilidad es alta, y por lo tanto hay poca varianza ambiental especial, las mediciones múltiples dan poca ganancia en precisión. Cuando la repetibilidad es baja, las mediciones múltiples pueden conducir a una ganancia en precisión que merezca la pena. Por otra parte, según aumenta el número de medidas, el incremento en precisión cae rápidamente, por lo que a menudo no es conveniente hacer más de dos o tres mediciones.

3. Predecir las producciones futuras. Este apartado lo veremos más adelante en relación con los procedimientos para evaluar reproductores.

Parecido entre parientes

El grado de parecido proporciona el medio para estimar la magnitud de la varianza genética aditiva, cuya proporción relativa a la varianza fenotípica, la heredabilidad, determina principalmente el método de selección a utilizar en la mejora. El grado de parecido puede calcularse a partir de medidas simples realizadas en los individuos de la población.

Hemos visto como la varianza fenotípica puede descomponerse en componentes atribuibles a diferentes causas. A dichos componentes los hemos denominado componentes causales de la varianza, denotándolos mediante el símbolo V . A partir de la agrupación de individuos en familias, obtendremos unos componentes, denominados componentes observables de la varianza fenotípica, que denotamos con el símbolo σ^2 . Estos últimos componentes podrán ser estimados directamente a partir de los valores fenotípicos mediante técnicas estadísticas adecuadas, como hemos visto en el capítulo anterior.

Consideremos, por ejemplo, el agrupamiento en familias de hermanos carnales:

1	2	...	s
y_{11}	y_{21}		y_{s1}
y_{12}	y_{22}		y_{s2}
y_{13}	y_{23}		y_{s3}
y_{14}	y_{24}		y_{s4}

La varianza entre grupos, σ_B^2 , es la varianza de las medias verdaderas de estos grupos de hermanos. La varianza dentro de grupos, σ_W^2 , es la varianza de las fluctuaciones de la observación de cada individuo alrededor de la media verdadera de su grupo. La media verdadera de un grupo es la que se estimaría sin error a partir de un número muy elevado de individuos.

El parecido entre hermanos, t , puede entenderse como similitud de los individuos de un mismo grupo, o bien como diferencias entre individuos de diferentes grupos. Para una determinada varianza total, si los datos están estructurados como en el ejemplo anterior, cuanto más se parezcan los individuos de un mismo grupo, mayor será la diferencia entre grupos. Podremos medir el grado de parecido como la proporción de la varianza entre grupos respecto a la varianza total, que tiene, como hemos visto, estructura de correlación intraclase.

$$t = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

El componente entre grupos expresa la cantidad de variación común a los miembros de un grupo; puede entenderse igualmente como la covarianza esperada entre los miembros de un mismo grupo (véase el anexo 3).

Cuando se trata de averiguar el parecido fenotípico entre progenitores y descendientes directos, por ejemplo padres e hijos, es más conveniente recurrir a la regresión, pues sería difícil sustentar los supuestos de igualdad de varianza de padres e hijos, que a menudo no sucede. En este caso la estructura de los datos es:

1	2	...	s	Familia
P_1	P_2		P_s	Progenitor (P)
D_1	D_2		D_s	Descendencia (D)

P_i y D_i son los valores fenotípicos observados en el progenitor y el descendiente, respectivamente. El grado de parecido, definido como la regresión de los hijos sobre los padres, viene expresado como

$$b_{DP} = \frac{\text{cov}(D,P)}{\sigma_P^2}$$

donde σ_P^2 es la varianza de los progenitores. Nótese que tanto t como b_{DP} tienen una estructura semejante: el numerador es una covarianza y el denominador es una varianza fenotípica.

La covarianza entre individuos emparentados es una propiedad de la población que tenemos que deducir para investigar la causa del parecido entre parientes. Veremos que esta covarianza es una parte de la varianza fenotípica total y está integrada por los componentes de la varianza descritos, en proporciones que difieren según el tipo de parentesco. Conociendo la aportación de los componentes causales a la covarianza observada, utilizaremos ésta para estimar aquéllos. Los parentescos que se utilizan más comúnmente son los de hijos con padres, medios hermanos –es decir, hermanos de padre o de madre– y hermanos carnales.

La covarianza entre individuos emparentados se debe tanto a causas genéticas como ambientales. De hecho, la covarianza fenotípica es la suma de las covarianzas genética y ambiental. Analizaremos primero la covarianza genética, con las correlaciones y regresiones a que dan lugar, y comentaremos después las causas ambientales de parecido.

Covarianza genética

Vamos a deducir la covarianza entre los valores genotípicos de los individuos emparentados. Supondremos equilibrio H-W y por el momento ignoraremos los efectos de una posible interacción epistática. Consideraremos algunos de los parentescos más comunes.

Descendencia y un progenitor

Se trata de obtener la covarianza entre el valor genotípico del progenitor y el valor genotípico medio de su descendencia:

$$\begin{aligned} \text{cov}(P,D) &= \text{cov}(G, \frac{1}{2}A) = \text{cov}(A+D, \frac{1}{2}A) = \text{cov}(A, \frac{1}{2}A) + \text{cov}(D, \frac{1}{2}A) \\ &= \frac{1}{2}\text{cov}(A,A) + \text{cov}(A,D) = \frac{1}{2}\text{cov}(A,A) \\ &= \frac{1}{2}V_A \end{aligned}$$

La covarianza entre descendencia y progenitor es, por lo tanto, la mitad de la varianza aditiva de los padres. El grado de parecido, en ausencia de parecido ambiental, se estimará mediante una regresión, es decir:

$$\begin{aligned} b_{DP} &= \text{cov}(P,D) / \sigma_P^2 \\ &= \frac{1}{2}V_A / V_P \\ &= \frac{1}{2}h^2 \end{aligned}$$

Por lo tanto,

$$h^2 = 2b_{DP}$$

Vemos como el grado de parecido relaciona los componentes causales con los componentes observables, lo cual permite calcular fácilmente la heredabilidad.

Si sólo medimos un descendiente por progenitor, el valor de la covarianza genética es el mismo, ya que en términos de esperanzas, el valor aditivo esperado de un descendiente de dicho progenitor es el mismo que el esperado para la media de n descendientes.

Descendencia y progenitor medio

En este caso interesa la covarianza de la media de la descendencia con la media de ambos padres (progenitor medio). Simbólicamente,

$$\begin{aligned} \text{cov}(D, (P_1+P_2)/2) &= \text{cov}(D, \frac{1}{2}P_1) + \text{cov}(D, \frac{1}{2}P_2) \\ &= \frac{1}{4}V_A + \frac{1}{4}V_A \\ &= \frac{1}{2}V_A \end{aligned}$$

Asumimos para ello que la varianza aditiva de los dos sexos es la misma. Vemos, por otra parte, cómo la covarianza entre descendencia y progenitor medio es la misma que la covarianza entre descendencia y progenitor.

En este caso el grado de parecido se establece mediante la regresión de la descendencia sobre la media de los progenitores, P' , asumiendo igualdad de varianza fenotípica en los dos sexos:

$$\begin{aligned} b_{DP'} &= \text{cov}(D, P') / \sigma_{P'}^2 \\ &= \frac{1}{2}V_A / \frac{1}{2}V_P \\ &= h^2 \end{aligned}$$

Como se ve, $b_{DP'} = h^2$, estima directamente la heredabilidad.

Medios hermanos

En este contexto, los medios hermanos son individuos procedentes de un progenitor común, generalmente un macho, que se ha apareado con una muestra aleatoria de hembras de las que ha tenido un descendiente. A modo de ejemplo, en el ganado vacuno, dada la rareza de los partos gemelares y la generalización de la inseminación artificial, la población está estructurada prácticamente como familias (numerosas) de medios hermanos.

El valor medio de un grupo de medios hermanos es el valor medio de la descendencia de un progenitor apareado con una muestra aleatoria de individuos del otro sexo. Este valor medio es, por definición, la mitad del valor mejorante del progenitor. La covarianza entre medios hermanos es igual a la varianza de sus medias verdaderas, es decir:

$$\text{cov}(\text{MH}) = \sigma^2(1/2A) = 1/4V_A$$

El grado de parecido entre medios hermanos se expresa como la correlación intra-clase (el cociente de la varianza entre grupos, es decir, la covarianza, a la varianza fenotípica):

$$t_{\text{MH}} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{1/4V_A}{V_P}$$

y por lo tanto, $h^2 = 4t_{\text{MH}}$.

Hermanos carnales

La covarianza de hermanos carnales, o simplemente hermanos, es más compleja pues existe una aportación de la varianza de dominancia. Consideremos en primer lugar la covarianza debida al componente aditivo. Los hermanos tienen ambos padres en común y por tanto, el valor genotípico medio de un grupo de hermanos es igual al valor mejorante medio de sus padres. Sean A y A' los valores mejorantes de los progenitores; entonces

$$\begin{aligned} V(1/2(A+A')) &= 1/4(V_A + V_{A'}) \\ &= 1/4 \cdot 2V_A \\ &= 1/2 V_A \end{aligned}$$

para lo cual se ha asumido de nuevo que la varianza aditiva es la misma en los dos sexos.

Consideremos ahora la aportación de la varianza de dominancia. Supongamos que para un cierto *locus*, B , los genotipos de los padres son B_1B_2 y B_3B_4 . En la descendencia habrá, por lo tanto, cuatro genotipos posibles, B_1B_3 , B_1B_4 , B_2B_3 y B_2B_4 , cada uno de ellos con una probabilidad de $1/4$. Supongamos un descendiente de uno cualquiera de esos genotipos: la cuarta parte de los pares de hermanos tendrá el mismo genotipo y consecuentemente, la misma desviación de dominancia, D .

Dado que $\text{cov}(D,D) = \sigma^2(D) = V_D$ y que $\text{cov}(D,D') = 0$, la covarianza genética de los hermanos es, por lo tanto,

$$\text{cov}(\text{HC}) = 1/2V_A + 1/4V_D$$

El grado de parecido entre hermanos carnales, medido como una correlación intra-clase, es como sigue:

$$t_{\text{HC}} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{1/2V_A + 1/4V_D}{V_P}$$

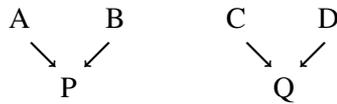
Caso general

Hemos visto que la covarianza entre individuos emparentados se expresa en general como fracciones sencillas de las varianzas aditivas y de dominancia. Estas fracciones están relacionadas con las coascendencias de los parientes y sus padres. Recordemos que la coascendencia entre dos individuos, f_{PQ} , es la probabilidad de que dos gametos tomados al azar, uno de cada uno, lleven alelos que sean idénticos por descendencia.

En general, si designamos como r a la fracción de la varianza genética aditiva y u a la de la varianza de dominancia, la expresión general de la covarianza entre dos parientes cualesquiera es

$$\text{cov}(P,Q) = rV_A + uV_D$$

Por otra parte, conocida la genealogía, por ejemplo la siguiente,



Crow y Kimura (1970) demuestran que

$$r = 2f_{PQ}$$

$$u = f_{AC}f_{BD} + f_{AD}f_{BC}$$

El valor r es conocido como coeficiente de parentesco y es igual al doble de la coascendencia entre dos individuos. Por su parte, el coeficiente u de la varianza de dominancia expresa la posibilidad de que los parientes considerados tengan el mismo genotipo, a través de la identidad por descendencia. Es cero a menos que los individuos tengan vías de coascendencia a través de sus dos progenitores.

Siguiendo a Falconer (1993), hemos mantenido el símbolo r para denotar el coeficiente de parentesco, puesto que se demuestra que es la correlación teórica o correlación entre valores mejorantes de los parientes. Es la correlación que se encontraría si toda la varianza fenotípica fuera genética aditiva. Se advierte al alumno que no debe confundirlo con el coeficiente de correlación de Pearson o la correlación intraclase. Cuando se conoce suficientemente la materia, es fácil deducir el significado de r según el contexto en el que se encuentre.

Interacción epistática

La interacción epistática aporta fracciones pequeñas a la covarianza genética entre individuos emparentados. La covarianza generalizada es según Kempthorne (1955):

$$\text{cov} = rV_A + uV_D + r^2V_{AA} + ruV_{AD} + u^2V_{DD} + r^3V_{AAA} + r^2uV_{AAD} + ru^2V_{ADD} + u^3V_{DDD} + \dots$$

Se desprende de esta expresión que los componentes de varianza de la interacción entre valores mejorantes están presentes siempre, por lo que cualquier estimación de la varianza aditiva contendrá siempre este término de interacción. Las interacciones que implican el componente de dominancia sólo existen si hay aportación inicial de la varianza de dominancia. En la práctica, se desconoce el valor real de la varianza de interac-

ción en caracteres del ganado, aunque en estudios realizados en siete caracteres del maíz mostró un valor relativo despreciable.

Covarianza ambiental

Existen causas ambientales que hacen que los parientes tiendan a parecerse. Esto sucede cuando se crían juntos los animales nacidos en un mismo parto, como es el caso de las camadas de cerdos o conejos. A la covarianza genética originada se le añadiría un componente de varianza ambiental, V_{Ec} , que denominamos **ambiente común**.

La varianza ambiental anteriormente definida puede descomponerse en dos partes: el ambiente común y un componente dentro de grupos, V_{Ed} , que no contribuye a la covarianza fenotípica entre los individuos. La varianza ambiental total es entonces:

$$V_E = V_{Ec} + V_{Ed}$$

El componente de varianza ambiental común va a depender de cómo agrupemos los individuos. Podemos hacerlo en familias de hermanos o de medios hermanos, aunque también en parejas progenitor-descendiente, por ejemplo. Cuando vayamos a estimar los componentes de varianza a partir de estadísticos tales como la regresión o la correlación intraclase, deberemos enjuiciar previamente si ha existido algún componente de covarianza ambiental que haya podido sesgar la estimación del componente genético.

Las causas de covarianza ambiental son variadas; podemos citar la nutrición, factores microclimáticos y el sistema de explotación. Un caso particular son los efectos maternos que pueden influir a lo largo de toda la vida del individuo. Así, el peso al destete de los terneros, influido claramente por la aportación lechera de la madre, habitualmente influye en el peso adulto.

Parecido fenotípico

La covarianza de los valores fenotípicos es la suma de las covarianzas originadas por los valores genéticos y los efectos ambientales. A efectos de estimar los componentes de varianza genética, y, como consecuencia de ello, la heredabilidad, asumiremos que el ambiente común no influye salvo en el caso de los hermanos carnales.

Heredabilidad

El coeficiente de heredabilidad, o más simplemente la heredabilidad, se ha definido previamente como el cociente de la varianza aditiva a la varianza fenotípica de un carácter:

$$h^2 = V_A / V_P$$

La importancia de la heredabilidad se debe a que constituye uno de los factores de la respuesta a la selección. Esto hace que la heredabilidad sea uno de los primeros parámetros a estimar cuando se pretenda instaurar un programa de selección. Además, la heredabilidad tiene valor predictivo, ya que permite predecir el valor mejorante de un individuo a partir de su valor fenotípico. Esto se debe a que bajo un modelo infinitesimal la heredabilidad es la regresión del valor mejorante sobre valor fenotípico:

$$h^2 = b_{AP} = \text{cov}(A,P) / \sigma_P^2$$

puesto que $\text{cov}(A,(A+D+I+E)) / \sigma_P^2 = V_A / V_P$

Por otra parte, la correlación entre el valor genético (o mejorante) y el valor fenotípico de un individuo es igual a la raíz cuadrada de la heredabilidad:

$$r_{AP} = b_{AP} (\sigma_P / \sigma_A) = h$$

La utilización de la heredabilidad para predecir los valores mejorantes a partir de los valores fenotípicos de los individuos entra de lleno en la teoría de índices de selección que desarrollaremos con amplitud más adelante.

La heredabilidad es una propiedad no solo del carácter, sino también de la población, pues depende de las frecuencias génicas así como de la magnitud de la varianza ambiental, de modo que en poblaciones de ambiente heterogéneo, la heredabilidad será menor, al ser mayor el componente de varianza ambiental. La magnitud del componente de varianza ambiental depende, no obstante, del modelo utilizado para la estimación de los componentes de la varianza fenotípica.

En estas notas solamente describiremos modelos sencillos relacionados con el análisis de la varianza para estimar los componentes observables. En las poblaciones animales, en condiciones reales, pueden incluirse en el modelo diversos factores no genéticos que tienden a disminuir la magnitud del componente de varianza ambiental. Si se define la heredabilidad como el cociente $\sigma_A^2 / (\sigma_A^2 + \sigma_E^2)$, el valor de heredabilidad estimado será dependiente también del modelo. Las estimaciones de heredabilidad de un carácter deberían utilizarse estrictamente sólo en la propia población y utilizar los componentes de varianza para evaluar los reproductores con el mismo modelo asumido para estimar los componentes.

La heredabilidad asimismo es una propiedad del carácter derivada de su arquitectura genética. Existen caracteres en los cuales la acción de los genes es principalmente aditiva. En otros, en cambio, la acción aditiva de los genes es poco importante, en relación con las desviaciones dominantes. Esto afectará a la magnitud de la varianza aditiva y, por lo tanto, a la heredabilidad.

Tabla 2. Valores estimados de heredabilidad en diferentes caracteres y especies de ganado. (Tomado de Pirchner, 1983.)

<u>Ganado vacuno</u>	
Producción de leche	0,20-0,40
Grasa %	0,30-0,80
Proteína %	0,40-0,70
Sólidos no grasos %	0,40-0,70
Persistencia	0,15-0,30
Velocidad de ordeño	0,50-0,80
Intervalo entre partos	0,00-0,05
Puntuación morfológica	0,20-0,30
Ganancia diaria	0,10-0,30
Piezas nobles %	0,20-0,50
<u>Ganado ovino</u>	
Peso del vellón	0,30-0,60
Longitud de la fibra	0,30-0,60
Finura de la fibra	0,20-0,50
Peso corporal	0,20-0,40
Tamaño de la camada	0,10-0,30
<u>Ganado porcino</u>	
Ganancia de peso: • alimentación individual	0,10-0,50
• en grupo	0,10-0,25
Eficiencia alimentaria: • alimentación individual	0,15-0,50
• en grupo	0,20-0,30
Longitud de la canal	0,30-0,70
Espesor de la grasa dorsal	0,30-0,70
Area del <i>longissimus dorsi</i>	0,20-0,60
Color de la carne	0,30-0,40
Tamaño de la camada	0,10-0,15
<u>Aves</u>	
Huevos / gallina alojada	0,05-0,15
Edad al primer huevo	0,20-0,50
Peso corporal	0,30-0,70
Peso del huevo	0,40-0,70
Altura del albumen	0,30-0,50
Fertilidad	0,05-0,15
Incubabilidad	0,05-0,20

Además de lo comentado, las estimaciones de heredabilidad están sometidas generalmente a grandes errores de muestreo. A pesar de ello, examinando la tabla 2, pueden establecerse tres grupos de caracteres según su heredabilidad:

- caracteres anatómicos o morfológicos, que muestran h^2 elevada, mayor que 0,45;
- caracteres de producción, con valores de h^2 medios, 0,25-0,45;
- caracteres relacionados con la eficacia reproductiva, *fitness*, de valores de h^2 bajos.

Estimación de la heredabilidad

La heredabilidad fundamentalmente se estima a partir del grado de parecido entre parientes.

$$h^2 = b_{DP} / r \quad \text{o también}$$

$$h^2 = t / r$$

donde r es el coeficiente de la varianza aditiva en la covarianza genética entre parientes, es decir, el coeficiente de parentesco entre los individuos considerados.

La elección de la clase de parientes a utilizar para la estimación de la heredabilidad depende de circunstancias externas, como son los datos disponibles. La siguiente tabla es un resumen de los tipos de parientes más comunes:

Parientes	Covarianza	Regresión (b) o Correlación (t)
Descendencia y un progenitor	$\frac{1}{2}V_A$	$b = \frac{1}{2}h^2$
Descendencia y progenitor medio	$\frac{1}{2}V_A$	$b = h^2$
Medios hermanos	$\frac{1}{4}V_A$	$t = \frac{1}{4}h^2$
Hermanos carnales	$\frac{1}{2}V_A + \frac{1}{4}V_D + V_{Ec}$	$t \geq \frac{1}{2}h^2$

A efectos de elegir el método de estimación de la heredabilidad debemos tener en cuenta dos parámetros: la precisión y el sesgo. En general, cuanto más estrecho es el parentesco, más precisa es la estimación.

Sesgo

El sesgo es habitualmente más importante que la precisión en la estimación de la heredabilidad. Son fuentes de sesgo la covarianza ambiental y el componente de dominancia en el caso de hermanos carnales. Por ello, en relación con el sesgo, los métodos más fiables son la regresión de descendiente sobre su progenitor y la correlación entre medios hermanos.

Precisión

La precisión de una estimación de heredabilidad viene dada por su error típico, que se obtiene fácilmente a partir del error típico de la regresión o correlación desde la que se ha estimado. Las estimaciones de heredabilidad suelen estar sometidas a considerables errores de muestreo (baja precisión) a menos que se estimen a partir de un elevado número de observaciones.

El número de observaciones necesarias para alcanzar un cierto grado de precisión va a depender del método (regresión o correlación intraclase de hermanos) y del diseño (número de animales medidos por familia). El método viene impuesto más por consideraciones prácticas y de sesgo que por la precisión. El diseño está siempre relacionado con las limitaciones económicas, bien en forma de espacio, tiempo o esfuerzo limitados. Para un número fijo (y limitado) de individuos medidos, se trata de obtener la distribución óptima del número de familias y de descendientes medidos por familia, de forma que se minimice la varianza de muestreo de la regresión o la correlación.

Diseño de regresión descendencia-progenitor

Los errores típicos de las estimaciones de heredabilidad a partir de la regresión de descendencia sobre progenitor fueron descritos por Latter y Robertson en 1960. La varianza de un coeficiente de regresión es la siguiente:

$$\sigma_b^2 = 1/(N-2) [(\sigma_y^2/\sigma_x^2) - b^2]$$

Dado que N , el número total pares de observaciones, suele ser bastante grande y b^2 comparativamente bastante pequeño, podemos aproximar la fórmula por

$$\sigma_b^2 = 1/N (\sigma_y^2/\sigma_x^2) \quad (\text{aprox.})$$

Supongamos ahora que medimos k progenitores ($k = 1$ ó $k = 2$) y n descendientes por familia. Las correspondientes varianzas serían:

$$\begin{aligned} \sigma_x^2 &= 1/k V_p \\ \sigma_y^2 &= 1/n [1+(n-1)t] V_p \end{aligned} \quad (\text{se demostrará más adelante})$$

Sustituyendo, tenemos

$$\sigma_b^2 = \frac{k[1+(n-1)t]}{nN} \quad (\text{aprox.})$$

Cuando sólo medimos un progenitor ($k = 1$), y hay un límite en el número de individuos que pueden medirse (nN fijo), el mínimo se alcanza para $n = 1$, puesto que $(n-1)t = 0$. El diseño más eficiente en estas condiciones es medir un descendiente por familia y el mayor número posible de familias. El error típico del coeficiente de heredabilidad estimado puede deducirse como:

$$\begin{aligned} V(h^2) &= 4\sigma_b^2, \text{ dado que } h^2 = 2b_{DP} \\ \text{e.t.}(h^2) &= 2\sigma_b \approx 2/\sqrt{N} \end{aligned} \quad (\text{aprox.})$$

Cuando medimos los dos progenitores y un descendiente por pareja se tiene:

$$\text{e.t.}(h^2) = \sigma_b = \sqrt{2/N} \quad (\text{aprox.})$$

En general, la regresión sobre el progenitor medio dará mayor precisión que la regresión sobre un progenitor.

Diseños de análisis de hermanos

Consideremos ahora estimaciones obtenidas a partir de la correlación intraclase de familias de medios hermanos o de hermanos carnales. Robertson (1959a) demostró que la varianza de muestreo de la correlación intraclase es, entonces,

$$\sigma_i^2 = \frac{2[1+(n-1)t]^2 (1-t)^2}{n(n-1)(N-1)}$$

Cuando el producto $nN = T$ está limitado por los recursos disponibles, se demuestra que la varianza de muestreo de la correlación intraclase es mínima cuando $n = 1/t$ aproximadamente. Por ello, el tamaño óptimo de la familia depende de la correlación intraclase y, por lo tanto, de la heredabilidad.

Si asumimos que tanto el ambiente común como la varianza de dominancia son despreciables, para hermanos carnales $t = h^2/2$, mientras que $t = h^2/4$ para medios hermanos. Los tamaños de familia que dan el diseño más eficiente son $n = 2/h^2$ para familias de hermanos carnales y $n = 4/h^2$ para familias de medios hermanos. Dado que el conocimiento de la heredabilidad *a priori* será sólo aproximado, no podemos estimar con exactitud el tamaño de la familia. No obstante, se demuestra que en ausencia de conocimiento previo de la heredabilidad, ésta debería estimarse a partir de familias de medios hermanos que tengan de 20 a 30 miembros.

Para calcular la varianza de muestreo de la correlación intraclase en diseño óptimo, es decir, cuando $n = 1/t$, se sustituye este valor en la ecuación anterior. Tras algunas aproximaciones se obtiene que

$$\sigma_i^2 = 8t / T \quad (\text{aprox.})$$

Entonces, para hermanos carnales

$$V(h^2) = 4\sigma_i^2 = 16h^2 / T \quad (\text{aprox.})$$

De forma similar, para medios hermanos se obtiene

$$V(h^2) = 16\sigma_i^2 = 32h^2 / T \quad (\text{aprox.})$$

Como resumen, teniendo en cuenta los conceptos de sesgo y de precisión conjuntamente, en general la estimación más fiable de la heredabilidad es la regresión de descendiente sobre progenitor y, en segundo lugar, la basada en la correlación de medios hermanos, a pesar de que la precisión de esta estimación es inferior a la de hermanos carnales, cuyo error típico se reduce a la mitad.

Poblaciones seleccionadas

La teoría expuesta hasta ahora concerniente a la estimación de la heredabilidad se ha sustentado en la suposición de apareamiento aleatorio. Cuando nos enfrentamos a estimar el coeficiente de heredabilidad en poblaciones de ganado, a menudo éstas se encuentran sometidas a selección. Esta circunstancia genera una fuente adicional de sesgo, que no afecta a la estimación de la heredabilidad mediante la regresión de descendiente sobre progenitor. Por otra parte, se demuestra que la selección causa la reducción de la

varianza entre padres y, consecuentemente, la de la covarianza de los hermanos. El resultado final es una infraestimación del valor de la heredabilidad a partir de la correlación intraclase.

De hecho, los métodos descritos, interesantes para introducir al alumno en el armazón conceptual de la estimación de la heredabilidad a partir de varianzas y covarianzas estimadas en la población, han dado paso a métodos estadísticos para diseños desequilibrados en los que se considera el valor mejorante de cada individuo como un nivel de un factor aleatorio. Este tipo de modelo se conoce como **modelo animal**, y permite obtener estimaciones tanto de componentes genéticos (varianza aditiva) como de la varianza residual, originada por factores no incluidos en el modelo. Estos métodos, de entre los que cabe destacar el de máxima verosimilitud restringida (*REML*), permiten obtener estimaciones insesgadas de los componentes genéticos de la población base, siempre que se incluya la matriz de parentesco entre los animales de la población sometida a selección. Sobre este tema, no obstante, volveremos más adelante cuando estudiemos los métodos de evaluación de reproductores.

Caracteres correlacionados

Vamos a tratar las relaciones existentes entre dos caracteres cuantitativos o métricos. Es bien conocido en el mundo de la producción animal que existe una correlación positiva entre la cantidad de leche y la cantidad de grasa producida por una vaca en una lactación. En cambio, la correlación entre la cantidad de leche y el porcentaje de grasa es negativa. Una correlación negativa entre dos caracteres puede ser, a diferencia del caso anterior, favorable; tal es la relación entre el crecimiento diario de los cerdos y el índice de conversión.

La asociación entre los valores observados de dos caracteres se cuantifica mediante la **correlación fenotípica**. Se calcula a partir de medidas realizadas (valores fenotípicos) de los dos caracteres en cada uno de los individuos de una población o muestra. Si pudiéramos conocer para cada carácter e individuo cuál es su valor genotípico y su desviación ambiental, podríamos calcular las correlaciones entre valores genotípicos y entre desviaciones ambientales. Si pudiésemos conocer los valores mejorantes, las desviaciones de dominancia y las de interacción epistática, podríamos calcular las correlaciones genéticas, la debida a la dominancia y la de interacción, respectivamente.

Correlaciones genéticas y ambientales

Tratar de calcular todas las correlaciones arriba mencionadas sería complejo e inmanejable. A efectos prácticos consideraremos la **correlación genética**, o correlación de los valores mejorantes, y la **correlación ambiental**, que estrictamente hablando no es la correlación de las desviaciones ambientales, sino la correlación de las desviaciones ambientales junto con las desviaciones de origen genético no aditivo.

La principal causa genética de correlación es la pleiotropía, aunque el ligamiento entre *loci* puede ser una causa de correlación transitoria, particularmente en poblaciones derivadas de cruces entre estirpes divergentes. Los genes que incrementan el crecimiento, aumentan a la vez la estatura y el peso, por lo cual tienden a causar correlación entre estos dos últimos caracteres.

El grado de correlación que surge de la pleiotropía expresa la magnitud en que dos caracteres están influidos por los mismos genes. La correlación que resulta es el efecto neto o promedio de todos los genes que segregan y afectan simultáneamente a los dos caracteres. Algunos genes pueden hacer aumentar los dos caracteres, provocando una correlación positiva, mientras que otros aumentan un carácter a la vez que reducen el otro, originando una correlación negativa.

El ambiente es causa de correlación en la medida en que dos caracteres estén influidos por las mismas diferencias de condiciones ambientales. De nuevo, la correlación que resulta de causas ambientales es el efecto medio de todos los factores ambientales que varían.

Descomposición de la correlación fenotípica

La correlación entre dos variables cualesquiera (x, y) es el cociente entre la covarianza y el producto de sus desviaciones típicas. En el caso de una correlación fenotípica tenemos:

$$r_{P_{xy}} = r_P = \frac{\text{cov}_P(x, y)}{\sigma_{P_x} \sigma_{P_y}}$$

por lo que podemos escribir la covarianza fenotípica como

$$\text{cov}_P(x, y) = \text{cov}_P = r_P \sigma_{P_x} \sigma_{P_y}$$

Asimismo, denotando la covarianza entre los valores mejorantes como cov_A y la covarianza entre las desviaciones ambientales, además de los efectos genéticos no aditivos, como cov_E , tenemos:

$$\text{cov}_P = \text{cov}_A + \text{cov}_E$$

Expresando las covarianzas como producto de correlaciones por las desviaciones típicas respectivas se obtiene:

$$r_P \sigma_{P_x} \sigma_{P_y} = r_A \sigma_{A_x} \sigma_{A_y} + r_E \sigma_{E_x} \sigma_{E_y}$$

donde r_A y r_E son las correlaciones genética y ambiental, respectivamente. Dado que $\sigma_A = h\sigma_P$ y que $\sigma_E = e\sigma_P$, donde $e^2 = 1-h^2$, sustituyendo se obtiene

$$r_P \sigma_{P_x} \sigma_{P_y} = r_A h_x \sigma_{P_x} h_y \sigma_{P_y} + r_E e_x \sigma_{P_x} e_y \sigma_{P_y}$$

Dividiendo los dos miembros de la igualdad por el producto $\sigma_{P_x} \sigma_{P_y}$ queda

$$r_P = r_A h_x h_y + r_E e_x e_y$$

Si las heredabilidades de los dos caracteres son bajas, las correlaciones fenotípicas serán fundamentalmente de origen ambiental. Si las heredabilidades son altas, la correlación genética (aditiva) será la más importante en la determinación de r_P , consecuencia de su mayor peso relativo.

Estimación de la correlación genética

La estimación de la correlación genética se basa en el parecido entre parientes, de forma análoga a la estimación de la heredabilidad. En vez de calcular los componentes de varianza a partir de un análisis de la varianza, calculamos los componentes de la covarianza a partir de un análisis de covarianza. Sea una muestra estructurada por familias de padre común, en la que cada individuo se ha medido respecto a dos caracteres, por ejemplo el crecimiento diario (x) y el índice de conversión (y). En la siguiente tabla, el primer subíndice corresponde al progenitor, mientras que el segundo corresponde al individuo en el que se miden dos caracteres.

1	2	3	s
$x_{11} y_{11}$	$x_{21} y_{21}$	$x_{31} y_{31}$	$x_{s1} y_{s1}$
$x_{12} y_{12}$	$x_{22} y_{22}$	$x_{32} y_{32}$	$x_{s2} y_{s2}$
...
$x_{1n} y_{1n}$	$x_{2n} y_{2n}$	$x_{3n} y_{3n}$	$x_{sn} y_{sn}$

A este diseño equilibrado corresponde el siguiente análisis de la covarianza, análogo a un análisis de la varianza:

Fuente de variación	Grados de libertad	Suma de productos (SP)	Producto medio (PM)	Esperanza de los productos medios (EPM)
Entre (B)	$s - 1$	$SP_B = \sum_i(x_i y_i)/n - (x_{..} y_{..})/ns$	$SP_B / (s - 1)$	$cov_w + n cov_B$
Dentro de (W)	$s(n - 1)$	$SP_W = \sum_{ij} x_{ij} y_{ij} - \sum_i(x_i y_i)/n$	$SP_W / (s(n - 1))$	cov_w
Total	$sn - 1$	$SP_P = \sum_{ij} x_{ij} y_{ij} - (x_{..} y_{..})/ns$	$SP_P / (sn - 1)$	

Si la familia es de medios hermanos, el componente de covarianza entre padres, cov_B , estima $1/4 cov_A$, es decir, un cuarto de la covarianza entre los valores genéticos de los dos caracteres. Podemos deducirlo como sigue:

$$cov(x_{11}, y_{12}) = cov_B = cov(1/2 A_x, 1/2 A_y) = 1/4 cov_A$$

Mediante el análisis de varianza aplicado por separado a cada carácter, con la misma estructura de datos se obtienen los componentes de varianza entre padres que se igualan respectivamente a $1/4 \sigma_{A_x}^2$ y $1/4 \sigma_{A_y}^2$. Por lo tanto, la correlación genética se obtiene como

$$r_A = \frac{cov_A}{\sqrt{\sigma_{A_x}^2 \sigma_{A_y}^2}} = \frac{cov_B}{\sqrt{\sigma_{B_x}^2 \sigma_{B_y}^2}}$$

siendo $\sigma_{B_x}^2$ y $\sigma_{B_y}^2$ las varianzas entre grupos de los caracteres x e y , respectivamente.

En el caso de parentesco descendencia-progenitor, calculamos la covarianza respecto a un carácter a partir del producto del valor del progenitor o del progenitor medio y el valor medio de la descendencia. Para estimar la correlación genética entre los dos caracteres calculamos la covarianza cruzada, a partir del producto del valor x de los progenitores y el valor y de la descendencia. Esta covarianza cruzada es igual a la mitad de la covarianza genética entre los dos caracteres, es decir, $1/2 cov_A$, puesto que $cov(P_x, D_y) = cov_{xy} = cov(A_x, 1/2 A_y) = 1/2 cov_A$. La correlación genética viene dada por

$$r_A = \frac{cov_{xy}}{\sqrt{cov_{xx} cov_{yy}}}$$

donde cov_{xy} es la covarianza cruzada (o la media geométrica de covarianzas cruzadas) y cov_{xx} y cov_{yy} son las covarianzas de progenitor y descendencia de cada carácter por separado.

Tabla 3. Correlaciones genéticas entre caracteres económicos del ganado. (Tomado de Pirchner, 1983.)

<u>Ganado vacuno</u>		
Producción lechera	Grasa (%)	-0,07 a -0,67
	Sólidos no grasos (%)	-0,02 a -0,20
	Proteína (%)	-0,10 a -0,50
	Eficiencia alimentaria	+0,80 a +0,95
	Velocidad de ordeño	+0,20 a +0,30
Grasa	Sólidos no grasos (%)	+0,30 a +0,70
	Proteína (%)	+0,40 a +0,70
Ganancia diaria	Rendimiento a la canal (%)	+0,10 a +0,40
	Grasa de la canal (%)	+0,00 a +0,10
<u>Ganado porcino</u>		
Ganancia diaria	Eficiencia alimentaria	+0,50 a +1,00
	Espesor del tocino dorsal	-0,25 a +0,30
	Longitud corporal	-0,50 a +0,10
	Area del <i>longissimus dorsi</i>	-0,10 a +0,40
Espesor del tocino dorsal	Longitud corporal	-0,25 a -0,50
	Eficiencia alimentaria	-0,05 a -0,40
	Area del <i>longissimus dorsi</i>	-0,15 a -0,40
<u>Ganado ovino</u>		
Peso del vellón	Peso del vellón limpio	+0,65 a +0,70
	Longitud de la fibra	0,00 a +0,20
	Peso corporal	-0,10 a 0,00
<u>Aves</u>		
Número de huevos	Peso del huevo	-0,25 a -0,50
	Peso corporal	-0,20 a -0,60
	Edad a la madurez sexual	-0,15 a -0,50
	Incubabilidad (%)	-0,20 a +0,30
	Eficiencia alimentaria	+0,50 a +1,00
Peso del huevo	Peso corporal	+0,20 a +0,60
	Espesor de la cáscara	+0,10 a -0,40
	Eficiencia alimentaria	+0,20 a -0,60
	Incubabilidad (%)	-0,20 a -0,40

Precisión de las estimaciones

Las estimaciones de las correlaciones genéticas están sometidas usualmente a errores de muestreo bastante grandes, por lo que raramente son precisas. Dichas correlaciones están muy influidas por las frecuencias génicas, por lo cual pueden diferir notablemente entre poblaciones. Los valores presentados en la tabla 3 son una buena muestra de ello, pues reflejan la variación de los resultados obtenidos en diferentes poblaciones,

por lo cual deben tomarse como aproximaciones orientativas. A efectos de establecer un modelo de evaluación de reproductores multicarácter será muy conveniente contar con estimaciones precisas de correlaciones genéticas y ambientales procedentes de la misma población que se intenta mejorar.

Según Robertson (1959b), el error típico de una estimación viene dado aproximadamente por la fórmula

$$\sigma(r_A) = \frac{1 - r_A^2}{\sqrt{2 \frac{\sigma(h_x^2) \sigma(h_y^2)}{h_x^2 h_y^2}}}$$

Implicaciones

1. Cuando seleccionamos individuos respecto a un carácter, es interesante conocer en qué medida la mejora de un carácter causará cambios simultáneos en otros caracteres. Sobre esto volveremos en el contexto de respuestas correlacionadas a la selección.
2. La relación existente entre un carácter métrico y su eficacia biológica, *fitness*, es el agente primario que determina las propiedades genéticas de este carácter tanto en una población natural como de ganado, y va a condicionar la respuesta a la selección.
3. La correlación genética expresa el grado en que dos medidas reflejan lo que es genéticamente el mismo carácter, por lo que cobra sentido en la interpretación de la repetibilidad. Se dijo que la repetibilidad tenía una interpretación genética sólo si las diferentes medidas son del mismo carácter genético. Esto es lo mismo que decir que sea 1 la correlación genética entre las medidas. No obstante, en la práctica, la repetibilidad es un concepto útil incluso aunque la correlación genética pueda ser a menudo algo menor que 1.

Evaluación genética de reproductores. I. Un solo carácter

Se cuenta que Robert Bakewell, un inglés pionero en la mejora del ganado por métodos empíricos, tal vez el ganadero más famoso del siglo XVIII, propugnaba “aparear lo mejor con lo mejor”. Este principio, aparentemente simple y directo, dejaba planteadas dos cuestiones: la primera, la definición de qué es lo mejor, lo cual cae de lleno dentro de la discusión de los objetivos de selección; la segunda cuestión es cómo elegir los mejores animales, y es precisamente la respuesta a esta pregunta el objetivo de este capítulo y el siguiente.

La evaluación genética de un animal consiste en realizar una predicción de su valor mejorante, o valor genético aditivo, a partir de información procedente del animal a evaluar (candidato a la selección) y/o de alguno o algunos de sus parientes. La evaluación genética tiene como objetivo ordenar los animales según su valor genético de manera que podamos elegir como reproductores aquellos individuos superiores, es decir, realizar la selección. Implícitamente hemos supuesto que se evaluaban los animales respecto a un único carácter, lo cual constituye el objeto de este capítulo, aunque también podemos evaluar los animales respecto a varios caracteres, como veremos posteriormente.

La información utilizada para evaluar genéticamente los animales es, obviamente, el fenotipo del individuo o de sus parientes. El modelo que hemos visto hasta ahora supone que el valor fenotípico es el resultado de sumar el valor genotípico y la desviación ambiental, es decir,

$$P = G + E$$

Podemos modificar esta ecuación suponiendo que el valor fenotípico es el resultado de sumar el valor mejorante con una nueva desviación, E' , que incluiría, además de los efectos ambientales, la parte no aditiva del genotipo:

$$P = A + E'$$

Si lo que pretendemos es evaluar A a partir de P , tendremos que pensar que en E' podremos encontrar factores que actuando de una manera sistemática podrán distorsionar, –sesgar–, el valor de A predicho. Supongamos, por ejemplo, que tenemos dos vacas clónicas: se da la circunstancia que mientras que la primera de ellas produce leche en un rebaño de nivel de manejo y régimen nutritivo elevados, la segunda la produce en unas condiciones ambientales muy inferiores. Es previsible que la segunda de las vacas dé una producción lechera inferior; si utilizásemos la producción de leche como criterio de evaluación, sin considerar el efecto del rebaño, concluiríamos la superioridad genética de la primera vaca sobre la segunda, lo cual sería un grave error: habríamos realizado una predicción sesgada del valor mejorante de la vaca.

El reconocimiento de este planteamiento llevó a que en la primera mitad de este siglo se comenzaran a implantar los centros o estaciones de control. En éstas se propor-

cionaba a los animales un medio uniforme, por lo que se asumía que las diferencias fenotípicas que mostraban eran consecuencia directa de sus diferencias genéticas. Como ejemplo histórico, podemos citar la prueba de toros lecheros jóvenes a partir de la primera lactación de sus hijas que se implantó en los países nórdicos; en la actualidad, podemos citar los centros de control porcino implantados en prácticamente todos los países, y los centros de evaluación de toros cárnicos jóvenes, bien por sus crecimientos o caracteres de la canal, o bien respecto a caracteres maternos, a través de sus hijas, como sucede en Francia. A efectos de evaluación, las aves, tanto los pollos *broiler* como las gallinas de puesta, se someten también a condiciones de crianza extremadamente homogéneas.

Cuando somos capaces de eliminar las diferencias ambientales sistemáticas y podemos asumir una media común para todos los animales objeto de evaluación, como son los casos anteriormente descritos, podemos utilizar la metodología de **índices de selección**, que estadísticamente tienen propiedades de mejor predicción lineal, o *Best Linear Predictor (BLP)*.

Es fácil suponer que la evaluación de los animales en centros de control es costosa, razón que ha llevado a investigar nuevos métodos que permitan utilizar datos de campo, es decir, datos que recoge habitualmente el ganadero para gestionar técnicamente su granja. La utilización de estos datos de campo en la evaluación de reproductores requiere el conocimiento de ciertos factores ambientales, “fijos” desde un punto de vista estadístico, que tienen un efecto sobre el fenotipo.

Tomando como ejemplo la producción lechera, Minvielle (1990) sistematiza estos factores ambientales como sigue:

Previsibles	Estimables	Aleatorios
Nº de ordeños diarios	Hato o rebaño	Microambiente
Intervalo parto-concepción	Año de parto	
Edad al parto		
Estación de parto		

Los factores “previsibles” son a menudo objeto de corrección o ajuste. Para ello se elaboran tablas con factores de corrección adecuados a cada país o región, las cuales se revisan periódicamente. El efecto de cada hato, en cambio, ya es más difícil de predecir. Así las cosas, podemos plantearnos la evaluación de dos maneras. La primera, la más simple, consistiría en evaluar los animales a partir de datos precorregidos, desconsiderando los efectos del rebaño. Ello nos llevaría a utilizar los índices de selección, pero las predicciones generalmente estarían sesgadas. La segunda alternativa es utilizar un método de evaluación que contemple el efecto del rebaño –al igual que otros efectos ambientales fijos– de manera que obtengamos predicciones insesgadas del valor mejorante de los candidatos a la selección. Este método se conoce como **mejor predicción lineal insesgada, BLUP**, pues éstas son las iniciales de los términos ingleses *Best Linear Unbiased Prediction*. En las páginas que siguen vamos a describir en detalle las metodologías de índices de selección y *BLUP*.

Índices de selección Suele ser habitual el tratar como dos problemas distintos las tareas de (1) encontrar la combinación óptima de registros fenotípicos del propio individuo y de sus parientes respecto a un carácter determinado, y (2) la de combinar en alguna clase de valor agregado de un individuo sus valores fenotípicos propios y/o de sus parientes referentes a dis-

tintos caracteres. El método de índices de selección se desarrolló para resolver el segundo problema. Sin embargo, como ha demostrado Henderson (1963), la teoría de índices de selección abarca ambos problemas, pudiendo extenderse a la evaluación de familias y cruces de líneas.

Los índices de selección se desarrollaron inicialmente para la selección multicarácter en plantas (Smith, 1936), a partir de la noción de función discriminante (Fisher, 1936), siendo posteriormente aplicados a la evaluación del ganado (Hazel, 1943). En este índice inicial a cada candidato a la selección se le asignaba un valor que combinaba varios caracteres, ponderados por su importancia económica relativa.

Para cada candidato, el índice de selección sintetiza la información disponible procedente de varias fuentes o de distintos caracteres en un valor sencillo, único. El índice será la mejor predicción del verdadero valor mejorante del individuo. Tiene las siguientes propiedades:

1. Maximiza r_{AI} , es decir, la correlación entre el verdadero valor mejorante (A) y el índice (I).
2. Maximiza la probabilidad de ordenar correctamente los candidatos según sus valores mejorantes verdaderos.
3. Maximiza el progreso genético (o el valor mejorante promedio de los candidatos) a obtener por selección.
4. Es la mejor predicción lineal: minimiza el cuadrado medio de las diferencias, entre el valor mejorante verdadero y el índice, es decir, $E(I-A)^2 = \text{mínimo}$.

El índice de selección para un candidato α será de la siguiente manera:

$$I_{\alpha} = b_1X_1 + b_2X_2 + \dots + b_nX_n = \sum b_iX_i \quad [1]$$

donde los X_i son los registros (expresados como desviaciones a la media que se supone conocida) del individuo α o de sus parientes y b_i son los factores de ponderación. El problema de construir el índice radica en encontrar los valores apropiados para los coeficientes de regresión parcial b_i . Aunque esto puede hacerse a partir de cualquiera de los cuatro criterios dados anteriormente, basaremos la deducción en el criterio primero.

La correlación entre el valor mejorante (A) y el índice (I) se expresa como sigue:

$$r_{AI} = \frac{\text{cov}(A, I)}{\sqrt{\sigma_A^2 \sigma_I^2}} = \frac{\sigma_{AI}}{\sqrt{\sigma_A^2 \sigma_I^2}}$$

Los b_i deben maximizar r_{AI} . Por comodidad, haremos máximo el logaritmo de r_{AI} , $\log r_{AI}$, que da el mismo resultado final.

$$\log r_{AI} = \log \sigma_{AI} - 1/2 \log \sigma_A^2 - 1/2 \log \sigma_I^2$$

Tomando derivadas parciales con respecto a cada b_i e igualando a cero, obtenemos una serie de ecuaciones simultáneas que al resolverlas dan los b_i . Para cualquier b_i , digamos b_1 , la ecuación es:

$$0 = \delta \log r_{AI} / \delta b_1 = (1/\sigma_{AI}) \cdot (\delta \sigma_{AI} / \delta b_1) - 0 - (1/2 \sigma_I^2) \cdot (\delta \sigma_I^2 / \delta b_1) \quad [2]$$

Podemos simplificar esta ecuación dado que:

$$a) \sigma_{AI} = \text{cov}(A, \sum b_i X_i) = \sum b_i \sigma_{AX_i}, \text{ por lo cual}$$

$$\delta \sigma_{AI} / \delta b_1 = \delta(b_1 \sigma_{AX_1}) / \delta b_1 + \delta(b_2 \sigma_{AX_2}) / \delta b_1 + \dots + \delta(b_n \sigma_{AX_n}) / \delta b_1 = \sigma_{AX_1}$$

$$b) \sigma_I^2 = b_1^2 \sigma_{X_1}^2 + b_2^2 \sigma_{X_2}^2 + \dots + b_n^2 \sigma_{X_n}^2 + 2b_1 b_2 \sigma_{X_1 X_2} \dots, \text{ por lo cual}$$

$$\delta \sigma_I^2 / \delta b_1 = 2b_1 \sigma_{X_1}^2 + 0 + \dots + 0 + 2b_2 \sigma_{X_1 X_2} + \dots + 2b_n \sigma_{X_1 X_n}$$

$$= 2[b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_n \sigma_{X_1 X_n}]$$

puesto que la derivada respecto a b_1 de los dobles productos que no lo contienen es cero.

Sustituyendo $a)$ y $b)$ en la ecuación [2], tenemos

$$0 = \sigma_{AX_1} / \sigma_{AI} - [(b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_n \sigma_{X_1 X_n}) / \sigma_I^2]$$

que multiplicado por σ_I^2 da

$$b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_n \sigma_{X_1 X_n} = \sigma_{AX_1} (\sigma_I^2 / \sigma_{AI})$$

Como σ_I^2 / σ_{AI} es común a la parte derecha de todas las ecuaciones, únicamente tiene efecto escala sobre los valores de los b_i obtenidos, por lo cual le daremos el valor arbitrario 1. Este valor arbitrario, no obstante, implica que aumente una unidad el valor mejorante que se predice por cada unidad que aumenta el índice. Ello se debe a que $\sigma_I^2 / \sigma_{AI} = \sigma_{AI} / \sigma_I^2 = 1 = b_{AI}$, es decir, la regresión de los valores mejorantes sobre los valores del índice.

La forma final de la ecuación b_1 será:

$$b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_n \sigma_{X_1 X_n} = \sigma_{AX_1}$$

Se obtendrán ecuaciones similares para b_2, \dots, b_n , siendo la serie final:

$$\begin{aligned} b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_n \sigma_{X_1 X_n} &= \sigma_{AX_1} \\ b_1 \sigma_{X_2 X_1} + b_2 \sigma_{X_2}^2 + \dots + b_n \sigma_{X_2 X_n} &= \sigma_{AX_2} \\ b_1 \sigma_{X_n X_1} + b_2 \sigma_{X_n X_2} + \dots + b_n \sigma_{X_n}^2 &= \sigma_{AX_n} \end{aligned} \quad [3]$$

Esta es la **forma general de las ecuaciones del índice de selección**. Puede hacerse más simple en el caso de selección para un carácter, y necesita extenderse algo en el caso de selección simultánea de varios caracteres.

En forma matricial se expresa de la siguiente manera:

$$\begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \dots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \dots & \sigma_{X_2 X_n} \\ \dots & \dots & \dots & \dots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \dots & \sigma_{X_n}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} = \begin{bmatrix} \sigma_{AX_1} \\ \sigma_{AX_2} \\ \dots \\ \sigma_{AX_n} \end{bmatrix}$$

Simbólicamente, $\mathbf{P} \cdot \mathbf{b} = \mathbf{c}$, donde \mathbf{P} es la matriz de varianzas-covarianzas fenotípicas, \mathbf{b} es el vector de coeficientes y \mathbf{c} es el vector de covarianzas entre el criterio de selección (X_i) y el objetivo de selección (A).

La eficacia o **precisión de un índice** se define como la correlación entre el valor del índice, predictor, y el verdadero valor mejorante, predictando. Cuanto mayor sea la correlación, mejor será el índice como predictor del valor mejorante; veremos que la precisión será uno de los factores que determinan la respuesta a la selección. Por otra parte, el cálculo de la precisión nos permite comparar la eficacia relativa de distintos índices que combinan diferentes fuentes de información.

$$r_{AI} = \frac{\text{cov}(A, I)}{\sqrt{\sigma_A^2 \sigma_I^2}}$$

$$\begin{aligned} \sigma_I^2 &= \sigma^2(b_1 X_1 + b_2 X_2 + \dots + b_n X_n) \\ &= b_1^2 \sigma_{X_1}^2 + b_2^2 \sigma_{X_2}^2 + \dots + b_n^2 \sigma_{X_n}^2 + 2b_1 b_2 \sigma_{X_1 X_2} + 2b_1 b_3 \sigma_{X_1 X_3} + \dots \\ &\quad + 2b_{n-1} b_n \sigma_{X_{n-1} X_n} \\ &= b_1(b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_n \sigma_{X_1 X_n}) + b_2(b_1 \sigma_{X_2 X_1} + b_2 \sigma_{X_2}^2 + \dots \\ &\quad + b_n \sigma_{X_2 X_n}) + \dots + b_n(b_1 \sigma_{X_n X_1} + b_2 \sigma_{X_n X_2} + \dots + b_n \sigma_{X_n}^2) \end{aligned}$$

por lo que

$$\sigma_I^2 = b_1 \sigma_{AX_1} + b_2 \sigma_{AX_2} + \dots + b_n \sigma_{AX_n}$$

Como hemos supuesto al construir las ecuaciones del índice que $\sigma_I^2 / \sigma_{AI} = 1$, $\sigma_I^2 = \sigma_{AI}$, por lo cual la precisión del índice es

$$r_{AI} = \frac{\sigma_I^2}{\sqrt{\sigma_I^2 \sigma_A^2}} = \frac{\sigma_I^2}{\sigma_I \sigma_A} = \frac{\sigma_I}{\sigma_A}$$

La correlación r_{AI} es una correlación múltiple. Su cuadrado, denominado genéricamente coeficiente de determinación, expresa la fracción de la varianza aditiva que se explica mediante las medidas combinadas en el índice; como contrapartida, la fracción $(1-r_{AI}^2)$ es la fracción de la varianza aditiva que no se tiene en cuenta en el índice.

Ejemplos de índices de selección

Comenzaremos describiendo índices de selección sencillos que sintetizan la información fenotípica del individuo o de alguna clase de parientes, para describir más tarde índices apropiados para la evaluación familiar y después plantear como pueden considerarse dos de ellos conjuntamente para dar lugar a un índice complejo, como es el caso de la evaluación combinada.

Evaluación individual

Parte de la información fenotípica del propio individuo, como por ejemplo la producción de leche ajustada a 305 días o el peso al destete normalizado a 210 días. De la ecuación general de los índices de selección sólo se necesitan los siguientes términos:

$$b_1 \sigma_{X_1}^2 = \sigma_{AX_1}$$

donde $\sigma_{X_1}^2$ es la varianza fenotípica de la fuente de información, en este caso el registro del propio individuo, X_1 , y σ_{AX_1} es la covarianza entre el valor fenotípico del individuo y su valor mejorante; por lo tanto,

$$b_1 = \frac{\sigma_{AX_1}}{\sigma_{X_1}^2} = \frac{\sigma_A^2}{\sigma_P^2} = h^2$$

y el índice es como sigue:

$$I = h^2 X_1$$

donde X_1 es la desviación a la media fenotípica, supuestamente conocida *a priori*.

La precisión del índice requiere los siguientes cálculos:

$$\sigma_I^2 = b_1 \sigma_{AX_1} = h^2 \sigma_A^2$$

$$r_{AI} = \frac{\sigma_I}{\sigma_A} = h$$

Evaluación a partir de n medidas del individuo

Cuando un carácter puede observarse en un individuo dos o más veces a lo largo de su vida productiva, como la producción de leche o el número de lechones nacidos en partos sucesivos, es conveniente utilizar toda la información con el fin de predecir el valor mejorante del individuo con mayor precisión.

De las ecuaciones generales sólo tendremos en cuenta una de las variables, por ejemplo X_1 , que en este caso será la media de n producciones de un mismo animal expresada como desviación a la media.

$$b_1 \sigma_{X_1}^2 = \sigma_{AX_1}$$

$$\begin{aligned} a) \quad \sigma_{X_1}^2 &= \sigma^2 [(x_1 + x_2 + \dots + x_n)/n] \\ &= (\sigma_{x_1}^2 + \sigma_{x_2}^2 + \dots + \sigma_{x_n}^2 + 2\sigma_{x_1x_2} + \dots + 2\sigma_{x_{(n-1)}x_n}) / n^2 \\ &= (n\sigma_{x_i}^2 + n(n-1)\sigma_{x_i x_j}) / n^2 \\ &= (\sigma_{x_i}^2 + (n-1)r\sigma_{x_i}^2) / n \\ &= \sigma_P^2 \frac{1+(n-1)r}{n}, \text{ siendo } r \text{ la repetibilidad del carácter} \end{aligned}$$

$$\begin{aligned} b) \quad \sigma_{AX_1} &= \text{cov} [A, 1/n(x_1 + x_2 + \dots + x_n)] \\ &= [\text{cov}(A, x_1) + \text{cov}(A, x_2) + \dots + \text{cov}(A, x_n)] / n \\ &= (n\sigma_A^2) / n \\ &= \sigma_A^2 \end{aligned}$$

El coeficiente del índice es, entonces,

$$b_1 = \frac{n\sigma_A^2}{\sigma_p^2(1+(n-1)r)}$$

$$= \frac{nh^2}{1+(n-1)r}$$

Para algunos, este coeficiente b_1 sería el equivalente a la heredabilidad de las medias de n medidas tomadas en cada candidato.

El índice queda como sigue:

$$I = \frac{nh^2}{1+(n-1)r} X_1$$

donde X_1 es la desviación de la media de n medidas tomadas en el individuo expresada como desviación a la media de la población.

La precisión del índice se calcularía como sigue:

$$\sigma_i^2 = b_1\sigma_{AX_1}$$

$$\sigma_i^2 = \frac{nh^2}{1+(n-1)r} \sigma_A^2$$

$$r_{AI} = \frac{\sigma_I}{\sigma_A} = \sqrt{\frac{nh^2}{1+(n-1)r}}$$

Evaluación genealógica

Vamos a considerar el índice derivado de la medición de los dos valores fenotípicos de los progenitores del individuo. De la ecuación general extraemos las siguientes ecuaciones:

$$b_1\sigma_{X_1}^2 + b_2\sigma_{X_1X_2} = \sigma_{AX_1}$$

$$b_1\sigma_{X_2X_1} + b_2\sigma_{X_2}^2 = \sigma_{AX_2}$$

donde se suponen conocidos X_1 , el valor fenotípico del padre y X_2 , el valor fenotípico de la madre. A es el valor mejorante del descendiente que se pretende predecir. Suponiendo que los padres no están emparentados, que pertenecen a la misma población (igual varianza σ_p^2) y que tampoco están sometidos a ninguna fuente de covarianza ambiental, las ecuaciones se reducen a:

$$b_1\sigma_p^2 = 1/2\sigma_A^2; \quad b_1 = 1/2h^2$$

$$b_2\sigma_p^2 = 1/2\sigma_A^2; \quad b_2 = 1/2h^2$$

El índice, por lo tanto, es:

$$I = 1/2 h^2 X_1 + 1/2 h^2 X_2$$

donde X_1 y X_2 son las desviaciones de los progenitores respectivos a una media conocida y común.

La precisión del índice queda:

$$\sigma_I^2 = 1/4 h^2 \sigma_A^2 + 1/4 h^2 \sigma_A^2 = 1/2 h^2 \sigma_A^2$$

$$r_{AI} = \sigma_I / \sigma_A = \sqrt{(1/2 h^2)} = \frac{h}{\sqrt{2}}$$

Vemos cómo el valor mejorante esperado (o predicho) de un animal es la semisuma de los valores mejorantes de sus padres. Si uno de los padres es desconocido, la predicción del valor mejorante del individuo es sencillamente la mitad del valor mejorante del progenitor conocido. La extensión de este índice al caso de uno o dos abuelos es directa, pues éstos transmiten sólo la cuarta parte de sus genes al nieto. Debemos observar que a veces se exagera el valor de los índices genealógicos o de pedigrí. Un toro, por ejemplo, no tiene por qué ser bueno porque su abuelo o bisabuelo fuera un gran campeón, puesto que sólo tendrá 1/4 ó 1/8 de sus genes, respectivamente. No hay que olvidar que cuantas más generaciones medien entre el descendiente y el antepasado, menor proporción de genes tendrá de éste.

Evaluación por descendencia (*Progeny test*)

Existen caracteres que no son observables en el candidato a la selección, por lo cual deberemos predecir el valor mejorante de los individuos a partir de observaciones realizadas en alguna clase de parientes. Un ejemplo común es la evaluación de toros lecheros a partir de las lactaciones de sus hijas, mientras que si nos remontamos a principios de este siglo, podemos hablar de la evaluación por descendencia empírica que realizaban en las estaciones danesas de mejora del porcino para conocer el valor genético de los futuros verracos respecto a caracteres de la canal.

Hemos dado una definición de valor mejorante empírico, que era el doble de la desviación de la media de la descendencia a la media general. Esta definición se basaba en el hecho de suponer que el animal, por ejemplo un toro, se apareaba al azar con un elevado número de hembras. En la práctica, el número de hembras con las que se suele aparear un toro, sobre todo en el inicio de la prueba para conocer su potencial lechero, suele ser limitado. En estas circunstancias, debemos pensar en métodos que nos aproximen al máximo el valor mejorante predicho al valor mejorante verdadero, utilizando de la mejor manera la información disponible. El índice apropiado cuando se miden n descendientes, asumiendo ausencia de covarianza ambiental, se deduce como sigue:

$$b_1 \sigma_{X_1}^2 = \sigma_{AX_1}$$

$$a) \sigma_{\bar{x}_1}^2 = \sigma_p^2 \frac{1 + (n-1)t}{n}$$

(por analogía con lo deducido anteriormente respecto a la media de n medidas tomadas en un individuo)

$$= \sigma_p^2 \frac{1+(n-1)1/4h^2}{n}$$

(para familias de medios hermanos)

$$\begin{aligned} b) \sigma_{AX_1} &= \text{cov} [A, (x_1 + x_2 + x_3 + \dots + x_n) / n] \\ &= [r\sigma_A^2 + r\sigma_A^2 + r\sigma_A^2 + \dots + r\sigma_A^2] / n \\ &= nr\sigma_A^2 / n \quad (\text{donde } r \text{ es el coeficiente de parentesco entre} \\ &= 1/2\sigma_A^2 \quad \text{progenitor y descendiente, igual a } 1/2) \end{aligned}$$

El coeficiente del índice es:

$$\begin{aligned} b_1 &= \frac{nr\sigma_A^2}{\sigma_p^2 (1 + (n-1)1/4h^2)} \\ &= \frac{1/2h^2n}{1 + (n-1)1/4h^2} = \frac{2h^2n}{4 + (n-1)h^2} \end{aligned}$$

Por lo cual, el índice es el siguiente:

$$I = \frac{2h^2n}{4+(n-1)h^2} (\bar{x}_i - \bar{x}),$$

donde \bar{x}_i es la media de los descendientes del macho i , y \bar{x} es la media de la población.

La precisión del índice se calcula como sigue:

$$\begin{aligned} \sigma_i^2 &= b_1\sigma_{AX_1} \\ \sigma_i^2 &= \frac{2h^2n}{4 + (n-1)h^2} 1/2\sigma_A^2 = \frac{h^2n\sigma_A^2}{4 + (n-1)h^2} \end{aligned}$$

$$r_{AI} = \sigma_i / \sigma_A = \sqrt{\frac{nh^2}{4 + (n-1)h^2}}$$

Evaluación por colaterales

Se basa en evaluar los reproductores partiendo del conocimiento del fenotipo de sus hermanos o medios hermanos; es decir, a través de la media de sus hermanos evaluamos el valor mejorante de un individuo. Es un índice que ha sido y es muy útil para evaluar animales respecto a caracteres de la canal y de calidad de carne que solamente pueden medirse después del sacrificio. El índice apropiado se deduce de forma similar

al anterior (queda como ejercicio para el alumno). Considerando que los colaterales pueden ser hermanos carnales o medios hermanos, el índice resulta ser:

$$I = \frac{rnh^2}{1+(n-1)t} X_1$$

donde r es el parentesco entre los hermanos y t la correlación intraclase de los mismos, siendo X_1 la desviación de la media del grupo de hermanos a la media general.

También podemos deducir la precisión del índice, que es

$$r_{IA} = hr \sqrt{\frac{n}{1+(n-1)t}}$$

Evaluación de familias

Existen ocasiones en las que no se busca seleccionar un individuo particular sino toda una familia completa, como a menudo se realiza en las aves. Otras veces se pretende realizar una selección del mejor individuo de cada familia, atendiendo a la desviación de su fenotipo a la media familiar.

Comencemos por la **selección familiar**. Al igual que en los casos anteriores, puede deducirse el índice de selección apropiado. Para ello tendremos que considerar que la observación fenotípica es una media de n observaciones, la media familiar, y que el valor mejorante esperado se supone idéntico para cada miembro de la familia. En principio, el valor esperado de cualquier miembro de la familia es la semisuma de los valores mejorantes de los padres; no obstante, existen diferencias entre hermanos debidas al efecto de segregación mendeliana, ϕ , que se asume distribuido con media igual a cero. La ecuación del índice es la siguiente:

$$b_1 \sigma_{X_1}^2 = \sigma_{AX_1}$$

$$a) \sigma_{X_1}^2 = \sigma_P^2 \frac{1+(n-1)t}{n}, \text{ al ser la varianza de una media.}$$

Suponiendo que el valor A corresponda al primer individuo, tenemos

$$\begin{aligned} b) \sigma_{AX_1} &= \text{cov}[A, 1/n(x_1 + x_2 + x_3 + \dots + x_n)] \\ &= (\sigma_A^2 + r\sigma_A^2 + r\sigma_A^2 + \dots + r\sigma_A^2) / n \\ &= (\sigma_A^2 + (n-1)r\sigma_A^2) / n \\ &= \sigma_A^2 \frac{1+(n-1)r}{n} \end{aligned}$$

siendo r el coeficiente de parentesco entre los miembros de la familia.

El coeficiente del índice queda, por lo tanto,

$$b_1 = h^2 \frac{1 + (n-1)r}{1 + (n-1)t}$$

que algunos denominan, por ejemplo Falconer, la heredabilidad de las medias familiares. Se propone al lector demostrar que la precisión del índice es

$$r_{AI} = \frac{h(1+(n-1)r)}{\sqrt{n(1+(n-1)t)}}$$

En el caso de la **evaluación intrafamiliar**, vamos a deducir las varianzas fenotípica y aditiva de las desviaciones intrafamiliares como complemento de las varianzas fenotípica y aditiva estimadas en la evaluación de familias. De esta manera tenemos:

$$\begin{aligned} a) \quad \sigma_{\bar{x}_1}^2 &= \sigma_p^2 \left[1 - \frac{1+(n-1)t}{n} \right] \\ &= \sigma_p^2 \left[\frac{(n-1)(1-t)}{n} \right] \\ b) \quad \sigma_{AX_1} &= \sigma_A^2 \left[1 - \frac{1+(n-1)r}{n} \right] \\ &= \sigma_A^2 \left[\frac{(n-1)(1-r)}{n} \right] \end{aligned}$$

Siendo el coeficiente del índice, por lo tanto,

$$b_1 = h^2 \frac{1-r}{1-t}$$

que, como en el caso anterior, puede entenderse como la heredabilidad de las desviaciones intrafamiliares.

La precisión del índice puede demostrarse que es como sigue:

$$r_{AI} = h(1-r) \sqrt{\frac{n-1}{n(1-t)}}$$

Evaluación combinada

En el ganado porcino se ha venido realizando un control combinado consistente en la evaluación simultánea de un macho y uno o dos hermanos de camada. La finalidad obviamente era aumentar la precisión de la predicción del valor mejorante de los machos candidatos a la selección. Vamos a deducir en lo que sigue un índice adecuado a esta evaluación combinada. De la ecuación general, los términos que necesitamos son los siguientes:

$$\begin{aligned} b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} &= \sigma_{AX_1} \\ b_1 \sigma_{X_2 X_1} + b_2 \sigma_{X_2}^2 &= \sigma_{AX_2} \end{aligned}$$

donde X_1 es el fenotipo del individuo y X_2 la media familiar, expresados como desviación a la media de la población.

Analicemos cada una de las varianzas y covarianzas implicadas:

$$\begin{aligned}
 a) \quad \sigma_{\bar{X}_1}^2 &= \sigma_p^2 \\
 b) \quad \sigma_{X_1 X_2} &= \text{cov}(x_i, \bar{x}_i) \\
 &= \text{cov}\left[x_i, \frac{1}{n}(x_1 + x_2 + \dots + x_i + \dots + x_n)\right] \\
 &= (\sigma_B^2 + \sigma_B^2 + \dots + \sigma_p^2 + \dots + \sigma_B^2) / n \\
 &= ((n-1)\sigma_B^2 + \sigma_p^2) / n \\
 &= ((n-1)t\sigma_p^2 + \sigma_p^2) / n \\
 &= \sigma_p^2 \frac{1+(n-1)t}{n}
 \end{aligned}$$

es decir, la covarianza de un individuo con la media de su familia, cuando participa él mismo en dicha media, es igual a la varianza de la media de n medidas.

$$\begin{aligned}
 c) \quad \sigma_{X_2 X_1} &= \sigma_{X_1 X_2} \\
 d) \quad \sigma_{\bar{X}_2}^2 &= \sigma_p^2 \frac{1+(n-1)t}{n}, \text{ al ser la varianza de una media} \\
 e) \quad \sigma_{AX_1} &= \sigma_A^2 \\
 f) \quad \sigma_{AX_2} &= \sigma_A^2 \frac{1+(n-1)r}{n}
 \end{aligned}$$

(de igual manera que se ha deducido antes en el contexto de la selección familiar).

Tras alguna simplificación algebraica, que se propone al lector como ejercicio, se llega a los siguientes coeficientes:

$$\begin{aligned}
 b_1 &= h^2 \frac{(1-r)}{(1-t)} \\
 b_2 &= h^2 \frac{(r-t)}{(1-t)} \frac{n}{1+(n-1)t}
 \end{aligned}$$

Por lo cual el índice resulta ser

$$I = h^2 \frac{(1-r)}{(1-t)} X_1 + h^2 \frac{(r-t)}{(1-t)} \frac{n}{1+(n-1)t} X_2$$

donde X_1 y X_2 son, respectivamente, los valores fenotípicos del individuo y la media familiar, expresados como desviación a la media de la población. Debemos hacer notar que si el número de hermanos es distinto en cada familia, debemos calcular un índice distinto para cada candidato a la selección.

Igualmente puede demostrarse que la precisión del método es como sigue:

$$r_{AI} = h \sqrt{\frac{(r-t)^2}{(1-t)} \frac{(n-1)}{1+(n-1)t}}$$

Comparación de índices de acuerdo con su precisión

Es fácil suponer que la evaluación genealógica es menos precisa que la evaluación a partir del fenotipo del propio individuo, y que ésta, asimismo, es menos precisa que si evaluamos el animal a partir de *n* medidas. Una forma general de comparar los distintos métodos es a través del número de individuos que debemos medir para igualar la precisión de la evaluación fenotípica o sobre la media de *n* medidas. Minvielle (1990) nos proporciona la siguiente tabla:

<i>h</i> ²	<i>r</i>	Precisión de la evaluación individual		Número de parientes a medir para para obtener la misma precisión					
		Una medida por candidato	Tres medidas por candidato	Medios hermanos		Hermanos carnales		Progenie	
0,1	0,3	0,32	0,43	26	117	5	12	5	9
0,3	0,5	0,55	0,67	imposible		9	51	6	11
0,6	0,8	0,77	0,83	imposible		imposible		9	13

Según esta tabla, podemos decir que midiendo 117 medios hermanos o 12 hermanos carnales o bien nueve descendientes, obtendríamos la misma precisión que evaluando el candidato a partir de la media de tres producciones.

Si sólo se puede hacer una medida sobre el candidato, es fácil suponer que los demás métodos de evaluación permitirán fácilmente obtener una precisión de evaluación equivalente si la *h*² es baja. En cualquier caso, la precisión se verá incrementada realizando una evaluación a partir de la descendencia, aunque este incremento será sólo marginal si la *h*² es elevada.

La precisión no es, sin embargo, el único criterio a tener en cuenta a la hora de decantarnos por un método u otro de evaluación. El hecho de que la producción lechera esté limitada a un sexo, las hembras, obliga a evaluar los machos a través de sus parientes, hijas fundamentalmente. Otro caso es la evaluación respecto a caracteres de la canal que impliquen el sacrificio de los animales, en cuyo caso debemos recurrir a la evaluación por colaterales, hermanos o medios hermanos. La posibilidad práctica de llevar a cabo un determinado método de evaluación puede ser más determinante que su precisión a la hora de elegirlo.

No debemos olvidar, por otra parte, que el criterio final de elección de un método de evaluación es hacer máximo el progreso genético anual, entrando así en juego un nuevo factor, el intervalo generacional. Más adelante, no obstante, volveremos a tratar estos aspectos de optimización en el contexto de la respuesta a la selección.

Intervalos de confianza en la predicción de valores mejorantes

La metodología de índices de selección proporciona la mejor manera de ordenar los animales de acuerdo con su valor genético. El mejorador puede preguntarse en qué medida el valor predicho mediante un índice se acerca al valor mejorante verdadero del animal. Formulada en estos términos, la pregunta no tiene respuesta puesto que nunca

podremos conocer el valor mejorante verdadero. No obstante, podemos calcular un cierto intervalo en el que tengamos una probabilidad determinada (seguridad) de que se encuentra el verdadero valor mejorante.

El intervalo de confianza de un valor mejorante verdadero, A_i , se calcula asumiendo que los valores mejorantes y los del índice siguen una distribución normal bivariada. Dicho intervalo es como sigue:

$$I_i \pm \tau \sqrt{(1-r_{AI}^2) \sigma_A^2}$$

Los valores de τ son los de la siguiente tabla:

Probabilidad de que A caiga dentro del rango	τ
50%	0,67
60%	0,84
68%	1,00
70%	1,04
80%	1,28
90%	1,65
99%	2,58

Si el intervalo elegido es el 80%, $\tau = 1,28$, existe una probabilidad del 10 por ciento de que el verdadero valor mejorante esté por encima del límite superior y asimismo una probabilidad del 10 por ciento de que el verdadero valor mejorante sea menor que el límite inferior.

Mejor predicción lineal insesgada (BLUP)

En el capítulo anterior hemos analizado en detalle la metodología de índices de selección. Hemos visto la necesidad de asumir una media conocida, así como de conocer las varianzas y covarianzas fenotípicas y aditivas; es decir, se hacía necesario el conocimiento de los momentos de primer orden (media) y segundo orden (varianzas y covarianzas) de la población. En el capítulo que iniciamos vamos a describir una metodología que sólo necesita del conocimiento de los momentos de segundo orden (varianzas y covarianzas de los factores aleatorios incluidos en el modelo), y que es capaz de estimar los efectos fijos (efectos ambientales sistemáticos) y predecir los efectos aleatorios (en nuestro caso los valores mejorantes de los candidatos a la selección). Debido a estas propiedades, además de otras que veremos a lo largo del presente capítulo, la metodología *BLUP* se ha venido imponiendo paulatinamente para evaluar reproductores a partir de datos de campo.

La deducción de las propiedades estadísticas, así como de los algoritmos de cálculo se debe a Henderson (1973). Esta deducción es más compleja que la que presentamos para los índices de selección, pues requiere planteamientos de álgebra matricial que exceden los objetivos de este curso. Las propiedades son semejantes a las enunciadas para los índices de selección, si bien se relaja la necesidad de conocer la media.

Existen diversos modelos que presentan propiedades *BLUP*, pero con fines didácticos comenzaremos por el que es conceptualmente más sencillo, que por otra parte admite desarrollos adicionales y, además, es el más utilizado en la evaluación genética de reproductores en la actualidad: se trata del **modelo animal**.

Modelo animal

Supongamos que registramos el peso al nacimiento de terneros (ajustado según el sexo) nacidos en dos rebaños, con resultados que se expresan en la siguiente tabla:

Nº del individuo	Rebaño	Peso al nacimiento (kg)
1	1	43
2	2	41
3	1	45
4	2	38
5	1	40
6	1	42
7	2	39
8	1	43

El valor fenotípico observado de cada uno de los terneros puede describirse según el modelo lineal siguiente:

$$y_{ij} = R_i + a_j + e_{ij}$$

donde

- y_{ij} peso al nacimiento del ternero j que pertenece al rebaño i ,
- R_i efecto del rebaño i ,
- a_j efecto del ternero j , concretamente su valor mejorante,
- e_{ij} error o residuo.

De acuerdo con el modelo matemático descrito, podríamos escribir las siguientes igualdades:

$$\begin{aligned} 43 &= R_1 + a_1 + e_{11} \\ 41 &= R_2 + a_2 + e_{22} \\ 45 &= R_1 + a_3 + e_{13} \\ 38 &= R_2 + a_4 + e_{24} \\ 40 &= R_1 + a_5 + e_{15} \\ 42 &= R_1 + a_6 + e_{16} \\ 39 &= R_2 + a_7 + e_{27} \\ 43 &= R_1 + a_8 + e_{18} \end{aligned}$$

Disponiendo estas ecuaciones en notación matricial podríamos escribir lo siguiente:

$$\begin{bmatrix} 43 \\ 41 \\ 45 \\ 38 \\ 40 \\ 42 \\ 39 \\ 43 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{22} \\ e_{13} \\ e_{24} \\ e_{15} \\ e_{16} \\ e_{27} \\ e_{18} \end{bmatrix}$$

Asumiendo que el efecto rebaño es fijo desde el punto de vista estadístico, mientras que el valor mejorante del animal es un efecto aleatorio, es decir, una variable que pue-

de tomar distintos valores con una probabilidad asociada a cada valor, estaríamos ante un modelo mixto. En notación matricial, podríamos representarlo de la siguiente manera:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

donde:

- \mathbf{y} vector de observaciones,
- \mathbf{X} matriz de incidencia que relaciona las observaciones con los efectos rebaño,
- \mathbf{b} vector de efectos fijos, en nuestro caso rebaño,
- \mathbf{Z} matriz de incidencia que relaciona las observaciones fenotípicas con el valor mejorante de cada animal,
- \mathbf{a} vector de efectos aleatorios, en nuestro caso los valores mejorantes de los animales,
- \mathbf{e} vector de errores o residuos, que recoge la parte de la variabilidad no explicada por el modelo.

Henderson (1973) demostró que dado un modelo mixto como el que acabamos de describir, bajo las condiciones de insesgamiento y varianza del error de predicción mínima, pueden obtenerse estimaciones de los efectos fijos y predicciones insesgadas de los efectos aleatorios mediante el siguiente algoritmo:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

que se conoce con el nombre de **ecuaciones de modelo mixto** o **ecuaciones BLUP**. En estas ecuaciones aparecen dos nuevos componentes, \mathbf{R} y \mathbf{G} , que corresponden a las matrices de varianzas-covarianzas del vector de residuos y del vector de efectos aleatorios, respectivamente.

Si la inversa de la matriz de coeficientes premultiplica la matriz de coeficientes, se obtiene obviamente una matriz de identidad:

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

En relación con estas ecuaciones se demuestra lo siguiente:

$$\begin{aligned} V(\hat{\mathbf{b}}) &= \mathbf{C}_{11} \\ V(\hat{\mathbf{a}}) &= \mathbf{G} - \mathbf{C}_{22} \\ V(\hat{\mathbf{a}} - \mathbf{a}) &= \mathbf{C}_{22} \end{aligned}$$

De estas cantidades, la más interesante es la última: la varianza del error de predicción, es decir, la varianza del vector de diferencias entre el valor mejorante predicho y el real. Se trataría, pues, de una medida de la precisión. Un valor mejorante predicho sería tanto más preciso cuanto menor fuera la varianza del error de predicción asociada.

Como consecuencia de la gran flexibilidad de las ecuaciones de modelo mixto (Henderson, 1974), podemos construir modelos animales de muy distintas formas, dependiendo del número de medidas del animal, de los objetivos del estudio y de que incluyamos o no las relaciones de parentesco entre los animales.

Una observación por animal

Las ecuaciones de modelo mixto pueden simplificarse si asumimos la siguiente estructura de varianzas-covarianzas de los efectos aleatorios \mathbf{a} y \mathbf{e} :

$$V \begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

que es lo mismo que decir que los valores mejorantes y los errores no están correlacionados; igualmente, al asumir una estructura diagonal para la matriz de varianzas-covarianzas entre ellos, se supone su independencia e idéntica distribución, $\mathbf{I}\sigma_a^2$ y $\mathbf{I}\sigma_e^2$, respectivamente. (Obsérvese el cambio de notación. Por razones de fidelidad a la notación original de Henderson, σ_a^2 y σ_e^2 equivaldrían a los valores σ_A^2 y σ_E^2 anteriormente mencionados.)

Bajo este supuesto, las ecuaciones de modelo mixto

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

se simplifican a

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{I}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

ya que

$$\mathbf{G}^{-1} = (\mathbf{I}\sigma_a^2)^{-1} = \mathbf{I}/\sigma_a^2$$

y

$$\mathbf{G}^{-1} \times \mathbf{I}\sigma_e^2 = \mathbf{I}\sigma_e^2/\mathbf{I}\sigma_a^2 = \mathbf{I}(\sigma_e^2/\sigma_a^2) = \mathbf{I}\alpha$$

Si el vector de observaciones estuviera constituido por una única medida o registro tomado en cada animal, por ejemplo el peso al nacimiento o el porcentaje de grasa en la primera lactación, \mathbf{Z} sería igual a una matriz de identidad (\mathbf{I}), y entonces

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I}(1+\alpha) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \end{bmatrix}$$

Despejando $\hat{\mathbf{a}}$ en la segunda ecuación nos da

$$\hat{\mathbf{a}} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) / (1 + \alpha) \quad [4]$$

El valor α puede relacionarse con la heredabilidad (h^2) como sigue

$$h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$$

por lo que

$$\sigma_a^2 = h^2 \sigma_p^2 \text{ y } \sigma_e^2 = (1 - h^2) \sigma_p^2,$$

siendo σ_p^2 la varianza fenotípica. Por lo tanto,

$$\alpha = \sigma_e^2 / \sigma_a^2 = (1 - h^2) / h^2$$

La matriz de coeficientes de las ecuaciones de modelo mixto quedaría:

$$\begin{bmatrix} 5 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1+\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1+\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1+\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1+\alpha & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1+\alpha & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1+\alpha & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1+\alpha & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1+\alpha \end{bmatrix}$$

Suponiendo que la heredabilidad del peso al nacimiento fuera $h^2 = 0,25$, entonces

$$\alpha = (1 - h^2) / h^2 = 0,75 / 0,25 = 3$$

por lo cual obtendríamos las ecuaciones del modelo mixto:

$$\begin{bmatrix} 5 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \hat{R}_1 \\ \hat{R}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 213 \\ 118 \\ 43 \\ 41 \\ 45 \\ 38 \\ 40 \\ 42 \\ 39 \\ 43 \end{bmatrix}$$

La nueva matriz de coeficientes de las ecuaciones de modelo mixto es una matriz no singular, que por lo tanto puede invertirse dando soluciones únicas con propiedades *BLUE* (*Best Linear Unbiased Estimator*) para los efectos fijos y *BLUP* para los efectos aleatorios, en este caso los valores genéticos o mejorantes de los animales. La solución numérica es la siguiente:

$$[42,60 \quad 39,33 \mid 0,100 \quad 0,417 \quad 0,600 \quad -0,333 \quad -0,650 \quad -0,150 \quad -0,083 \quad 0,100]'$$

Analicemos algunas de las soluciones, por ejemplo \hat{R}_1 y \hat{a}_1 :

$$\hat{R}_1 = \frac{213}{5} - \frac{\hat{a}_1 + \hat{a}_3 + \hat{a}_5 + \hat{a}_6 + \hat{a}_8}{5}$$

El efecto rebaño sería la media del rebaño corregida por el valor mejorante medio de los animales del mismo. Por su parte,

$$\hat{a}_1 = (43 - \hat{R}_1) / 4$$

lo cual muestra que el valor mejorante del animal 1 va a proceder de la desviación de su producción al efecto rebaño, que se ha calculado teniendo en cuenta el valor genético de sus contemporáneos, es decir de los animales tratados de igual manera.

Es interesante recordar que, en este caso, $1/4 = 1/(1 + \alpha)$. Se demuestra fácilmente que esta última cantidad equivale a h^2 (la prueba se propone al alumno como ejercicio). Así, la solución \hat{a}_1 , que es la predicción del valor mejorante del animal 1, quedará, por lo tanto como

$$\hat{a}_1 = h^2 (43 - \hat{R}_1)$$

que sería semejante a un índice de selección para la evaluación individual. Esto puede verse también en términos matriciales. Recordemos que, según la ecuación 3, el vector de los valores mejorantes se calcula como $\hat{\mathbf{a}} = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) / (1 + \alpha) = h^2(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$. Si supusiéramos que todos los animales pertenecían a un mismo rebaño, cuyo peso al nacimiento promedio fuese conocido sin error, μ , como es el caso de los índices de selección, podríamos escribir el siguiente modelo:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{I}\mathbf{a} + \mathbf{e}$$

donde $\mathbf{1}$ es un vector columna de unos. Sustituyendo en la fórmula anterior tenemos:

$$\hat{\mathbf{a}} = h^2(\mathbf{y} - \mathbf{1}\mu)$$

que retornando a términos algebraicos corresponde a un índice para la evaluación individual:

$$\hat{a}_1 = h^2 (y - \mu)$$

No obstante, debemos insistir en que para calcular un valor de índice deberíamos conocer la media de la población de referencia previamente, lo cual es difícil en muchos casos. Mediante estas ecuaciones de modelo mixto, la solución de cada animal se ha ajustado automáticamente respecto a la media de su rebaño. Asimismo pueden estimarse los efectos fijos, rebaño en nuestro caso.

Suponer que la matriz de varianzas-covarianzas es una matriz diagonal, con la varianza aditiva como elementos distintos de cero, $\mathbf{G} = \mathbf{I}\sigma_a^2$, es equivalente a asumir que los animales no están emparentados ni sometidos a ninguna fuente de covarianza ambiental, como lo hacíamos al confeccionar un índice para la evaluación individual o fenotípica. El emparentamiento de los animales aporta nueva información que puede acomodarse en las ecuaciones de modelo mixto, permitiendo con ello una predicción más precisa de los efectos aleatorios. Esto va a ser objeto de análisis en los siguientes apartados.

Modelo animal que incluye el parentesco entre los individuos

Muy comúnmente los individuos de un rebaño e incluso de diferentes rebaños están emparentados. Esta nueva información va a permitir predecir los valores mejorantes con toda la información disponible, lo cual se traducirá en la reducción de la varianza del error de predicción en las ecuaciones de modelo mixto. A lo largo de este apartado veremos cómo se incorpora la información genealógica al entramado de las ecuaciones de modelo mixto.

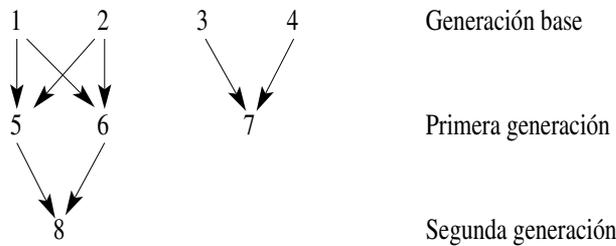
Si \mathbf{a} representa el vector de efectos genéticos aditivos de los animales, la matriz de varianzas-covarianzas será igual al producto de la matriz de parentesco por la varianza aditiva, es decir, $\mathbf{V}(\mathbf{a}) = \mathbf{A}\sigma_a^2$. Recuérdese que la covarianza aditiva entre dos individuos es su coeficiente de parentesco multiplicado por la varianza aditiva del carácter.

Método tabular

El parentesco entre los animales, organizado en una matriz denotada habitualmente como **A**, puede calcularse por un método recurrente conocido como **método tabular** (Emik y Terril, 1949). Debemos partir de una lista de animales cuyos progenitores sean conocidos, como en el siguiente ejemplo:

<u>Animal</u>	<u>Padre</u>	<u>Madre</u>
5	1	2
6	1	2
7	3	4
8	5	6

que corresponde a la siguiente genealogía



Tenemos que ordenar los animales cronológicamente, de manera que ningún animal aparezca en la lista antes que sus padres. Al ser los progenitores de los animales 1, 2, 3 y 4 desconocidos, se asume que los animales no están emparentados.

Con la lista de los animales ordenados cronológicamente hay que construir un cuadro de doble entrada con tantas filas y columnas como animales tengamos. Se debe identificar también cada fila y cada columna con el distintivo del animal y asimismo poner sobre cada identificación de columnas la identificación de los padres del individuo. Los elementos de **A**, la matriz de parentesco, se denotan como a_{ij} , siendo el coeficiente de parentesco entre los animales i y j . De nuevo debemos resaltar un cambio de notación. Mientras que en la escuela de Edimburgo (libro de Falconer), el parentesco se denota con la letra r , en el contexto de la bibliografía relacionada con métodos **BLUP** (procedente en general de Norteamérica), el parentesco habitualmente se denota mediante la letra a , seguida por dos subíndices que identifican los individuos cuyo parentesco se ha calculado, es decir a_{ij} .

Siguiendo ordenadamente los pasos que comentaremos a continuación se llega al siguiente cuadro:

	--	--	--	--	12	12	34	56
	1	2	3	4	5	6	7	8
1	1	0	0	0	0,5	0,5	0	0,5
2	0	1	0	0	0,5	0,5	0	0,5
3	0	0	1	0	0	0	0,5	0
4	0	0	0	1	0	0	0,5	0
5	0,5	0,5	0	0	1	0,5	0	0,75
6	0,5	0,5	0	0	0,5	1	0	0,75
7	0	0	0,5	0,5	0	0	1	0
8	0,5	0,5	0	0	0,75	0,75	0	1,25

Por cada animal cuyos padres se desconocen se escribe un 1 en la diagonal de la tabla y un 0 en los elementos de fuera de la diagonal correspondientes a los animales no emparentados. Al poner un 1 en los elementos diagonales de los animales con padre desconocido queremos decir que el parentesco de un animal no consanguíneo consigo mismo es 1 (se justifica más adelante). Además, se asume que el parentesco entre animales sin padres comunes o cuyos padres son desconocidos es 0.

La siguiente etapa es calcular los parentescos entre el animal 1 y los animales 5, 6, 7 y 8. El parentesco entre 1 y 5 es la media de los parentescos de 1 con los padres de 5, que son 1 y 2. Por ello,

$$a_{15} = \frac{1}{2} (a_{11} + a_{12}) = \frac{1}{2} (1 + 0) = 0,5$$

De forma similar

$$a_{16} = \frac{1}{2} (a_{11} + a_{12}) = \frac{1}{2} (1 + 0) = 0,5$$

$$a_{17} = \frac{1}{2} (a_{13} + a_{14}) = \frac{1}{2} (0 + 0) = 0$$

$$a_{18} = \frac{1}{2} (a_{15} + a_{16}) = \frac{1}{2} (0,5 + 0,5) = 0,5$$

La matriz \mathbf{A} es simétrica, por lo cual $a_{15} = a_{51}$, $a_{16} = a_{61}$, $a_{17} = a_{71}$ y $a_{18} = a_{81}$. Para calcular los parentescos de los animales 2, 3 y 4 procederíamos de forma similar a como lo hemos hecho con el individuo 1.

Ahora debemos calcular el elemento diagonal 5. Recordemos que $a_{ii} = 2f_{ii} = 2(\frac{1}{2}(1 + F_i)) = 1 + F_i$, siendo f la coascendencia y F el coeficiente de consanguinidad, respectivamente. Es decir,

$$a_{55} = 1 + F_5$$

El coeficiente de endogamia es igual a la mitad del parentesco aditivo entre los padres de 5, esto es, $F_5 = \frac{1}{2} a_{12}$. Cuando los padres son desconocidos, se toma como 0 el coeficiente de endogamia, asumiendo que los padres del individuo no estaban emparentados.

Tras computar el elemento diagonal de un animal, se debe calcular el parentesco con los restantes animales hasta completar la fila, usando el mismo procedimiento anteriormente descrito. Tras esto, de forma similar, iniciaríamos el cálculo del siguiente elemento diagonal y así sucesivamente hasta completar la tabla. En nuestro caso es de notar que el único elemento endogámico es 8. Recuérdese que el coeficiente de endogamia de un individuo es la proporción de *loci* del genoma que se ha hecho homogénea, es decir, que los dos alelos del *locus* sean idénticos por descendencia.

Inversa de la matriz de parentesco

Para construir las ecuaciones de modelo mixto de un modelo animal se requiere calcular la inversa de \mathbf{A} , puesto que $\mathbf{G} = \mathbf{V}(\mathbf{a}) = \mathbf{A}\sigma_a^2$ y por tanto $\mathbf{G}^{-1} = \mathbf{A}^{-1}/\sigma_a^2$. Cuando el número de animales es elevado, puede hacerse imposible el cálculo de la inversa de \mathbf{A} . En 1975, Henderson descubrió que conociendo las genealogías de los animales era bastante sencillo calcular la inversa de \mathbf{A} sin formar la misma \mathbf{A} ; incluso era casi más fácil calcular la inversa de \mathbf{A} que crear \mathbf{A} .

Sin entrar en detalles acerca de su deducción, las normas de Henderson para una población no consanguínea se resumen en el algoritmo que sigue:

- I) Partir de valores nulos en todos los elementos de \mathbf{A}^{-1}
- II) Considerar el número de progenitores conocidos
- si se conocen los dos padres, p y q, del individuo i, sumar
 - 2 al elemento (i,i)
 - 1 a los elementos (p,i), (i,p), (i,q) y (q,i)
 - 0,5 a los elementos (p,p), (p,q), (q,p) y (q,q)
 - si sólo se conoce un progenitor, por ejemplo p, sumar
 - 4/3 al elemento (i,i)
 - 2/3 a los elementos (p,i) y (i,p)
 - 1/3 al elemento (p,p)
 - si no se conoce ninguno de los progenitores, sumar
 - 1 al elemento (i,i)

Este algoritmo es una aproximación aceptable para coeficientes de consanguinidad bajos, como por ejemplo en vacuno lechero, que oscilan entre el 0,5 y el 2%. No obstante, Quaas (1976) ideó un algoritmo para calcular la matriz de parentesco cuando los animales son consanguíneos, pero su descripción queda fuera del objetivo de estas notas.

Implicaciones de la matriz de parentesco en la evaluación genética

Hemos comentado anteriormente que la inclusión de la matriz de parentesco permitía aumentar la precisión de las predicciones de los valores mejorantes de los candidatos a la selección, realizadas a través de las ecuaciones de modelo mixto. Ésta, con ser una propiedad deseable, no es la única. En lo que sigue vamos a ver que mediante la inclusión de la matriz de parentesco puede ligarse toda la información disponible, permitiendo la evaluación de los animales a partir de las producciones propias (si las tuvieren) y/o de sus parientes.

Para desarrollar este apartado, continuaremos con el ejemplo iniciado, adaptado de un ejemplo numérico de Kennedy y Sorensen (1988), generalizando los resultados siguiendo la nomenclatura propuesta por Pollak (1988).

Siguiendo las normas de Henderson, la inversa de la matriz de parentesco sería:

$$\mathbf{A}^{-1} = \begin{bmatrix} 2 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 2 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1,5 & 0,5 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0,5 & 1,5 & 0 & 0 & -1 & 0 \\ -1 & -1 & 0 & 0 & 2,5 & 0,5 & 0 & -1 \\ -1 & -1 & 0 & 0 & 0,5 & 2,5 & 0 & -1 \\ 0 & 0 & -1 & -1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 2 \end{bmatrix}$$

Las nuevas ecuaciones de modelo mixto quedarían, por lo tanto, como

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Tomando $h^2 = 0,25$, entonces $\alpha = 3$. Si tenemos en cuenta que en la submatriz $\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha$ los elementos diagonales son 1 ó 0 (según tengan o no registro propio) más $a^{ii}\alpha$, y que los no diagonales son iguales a $a^{ij}\alpha$ (siendo a^{ij} los elementos de la inversa de la matriz de parentesco), en nuestro ejemplo los valores de las ecuaciones de modelo mixto serían:

$$\begin{bmatrix} 5 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 7 & 3 & 0 & 0 & -3 & -3 & 0 & 0 \\ 0 & 1 & 3 & 7 & 0 & 0 & -3 & -3 & 0 & 0 \\ 1 & 0 & 0 & 0 & 5,5 & 1,5 & 0 & 0 & -3 & 0 \\ 0 & 1 & 0 & 0 & 1,5 & 5,5 & 0 & 0 & -3 & 0 \\ 1 & 0 & -3 & -3 & 0 & 0 & 8,5 & 1,5 & 0 & -3 \\ 1 & 0 & -3 & -3 & 0 & 0 & 1,5 & 8,5 & 0 & -3 \\ 0 & 1 & 0 & 0 & -3 & -3 & 0 & 0 & 7 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -3 & -3 & 0 & 7 \end{bmatrix} \begin{bmatrix} \hat{R}_1 \\ \hat{R}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 213 \\ 118 \\ 43 \\ 41 \\ 45 \\ 38 \\ 40 \\ 42 \\ 39 \\ 43 \end{bmatrix}$$

cuyo vector de soluciones es

$$[42,671 \quad 39,444 \mid -0,210 \quad 0,096 \quad 0,529 \quad -0,415 \mid -0,395 \quad -0,109 \quad -0,015 \mid -0,169]'$$

Debemos observar que cuando no hay solape de generaciones y los individuos de la generación base no están emparentados, el promedio de los \hat{a}_i de la generación base es cero; en este caso, $(\hat{a}_1 + \hat{a}_2 + \hat{a}_3 + \hat{a}_4)/4 = 0$. No obstante, el promedio \hat{a}_i de las siguientes generaciones no es cero y refleja los cambios debidos a la deriva genética y/o a la selección; así por ejemplo, en la generación 1, $(\hat{a}_5 + \hat{a}_6 + \hat{a}_7)/3 = -0,173$.

Analizaremos ahora cómo se obtienen las soluciones de algunos de los animales, lo cual nos dará una idea de los efectos implicados. Así, por ejemplo, para el individuo 7, del que se conoce su producción y la de sus padres, su ecuación estaría compuesta por

$$\begin{aligned} \hat{R}_2 - 3\hat{a}_3 - 3\hat{a}_4 + 7\hat{a}_7 &= 39 \\ \hat{a}_7 &= 1/7 [39 - \hat{R}_2 + 3(\hat{a}_3 + \hat{a}_4)] \end{aligned}$$

La ecuación anterior expresada de una forma general sería

$$\hat{a}_{ij} = [1/(1 + 2\alpha)] \{y_{ij} - GC_i\} + [\alpha/(1 + 2\alpha)] \{\hat{a}_{ij}^P + \hat{a}_{ij}^M\}$$

donde GC es el grupo de contemporáneas/os, es decir, animales tratados igual (efecto fijo), y \hat{a}_{ij}^P y \hat{a}_{ij}^M son los valores genéticos de su padre y de su madre, respectivamente. Si hacemos la semisuma de estos dos valores, obtenemos la predicción por índice genealógico del valor mejorante del animal \hat{a}_{ij} . Sustituyendo en nuestra fórmula nos da

$$\begin{aligned} \hat{a}_{ij} &= [1/(1 + 2\alpha)] (\text{desviación de su producción}) + \\ & [2\alpha/(1 + 2\alpha)] (\text{índice genealógico}). \end{aligned}$$

Veamos ahora el animal 3. De él se conoce su compañera/o de apareamiento (consorte), tiene registro propio y, además, se conoce la producción de un descendiente. Su ecuación es

$$\begin{aligned} \hat{R}_1 + 5,5\hat{a}_3 + 1,5\hat{a}_4 - 3\hat{a}_7 &= 45 \\ \hat{a}_3 &= 1/5,5 [45 - \hat{R}_1 - 1,5\hat{a}_4 + 3\hat{a}_7] \end{aligned}$$

que generalizando sería

$$\hat{a}_{ij} = 1/[1 + (1+0,5n)\alpha] \{y_{ij} - GC_i\} + \alpha/[1 + (1 + 0,5n)\alpha] \Sigma\{\hat{a}_{ij}^P - 0,5\hat{a}_{ij}^C\}$$

siendo n el número de descendientes medidos (uno en este caso) y \hat{a}_{ij}^C y \hat{a}_{ij}^D los valores mejorantes del consorte y de su descendiente, respectivamente. Eso es importante puesto que la evaluación de un animal implica un ajuste automático según el valor de su consorte y de su descendencia. El modelo animal ajusta, pues, las desviaciones al apareamiento aleatorio que comúnmente se producen en las poblaciones de ganado, pues es frecuente aparear los mejores animales de una población entre sí (apareamiento clasificado). Esto supone un claro avance respecto a la metodología de índices de selección en la cual se asumía que los animales se apareaban al azar.

Supongamos el caso de que el animal 3 fuera un toro del que se desconoce su propia producción. Es importante darse cuenta de que no habría término GC al no tener su propio registro, por lo que la solución de este toro sería una función del valor mejorante de su descendencia desviada de la mitad del valor mejorante de las vacas:

$$\hat{a}_{ij} = [1/(1 + 0,5n)] \Sigma(\hat{a}_{ij}^D - 0,5\hat{a}_{ij}^C)$$

Si expresamos la suma de las desviaciones como n veces la desviación promedio, el coeficiente se hace

$$n/(1 + 0,5n)$$

que se aproxima a 2 a medida que n aumenta.

El individuo 5 constituye una situación todavía más compleja. De él conocemos la producción propia, la de sus padres, la de su consorte y la de su descendencia. Su ecuación resulta ser

$$\begin{aligned} \hat{R}_1 - 3\hat{a}_1 - 3\hat{a}_2 + 8,5\hat{a}_5 + 1,5\hat{a}_6 - 3\hat{a}_8 &= 40 \\ \hat{a}_5 &= 1/8,5 [40 - \hat{R}_1 + 3(\hat{a}_1 + \hat{a}_2) - 1,5\hat{a}_6 + 3\hat{a}_8] \end{aligned}$$

entonces, generalizando

$$\hat{a}_{ij} = [1/(1 + (2+0,5n)\alpha)] \{ (y_{ij} - GC_i) + \alpha(\hat{a}_{ij}^P + \hat{a}_{ij}^M) + \alpha\Sigma(\hat{a}_{ij}^D - 0,5\hat{a}_{ij}^C) \}$$

debiendo mencionarse que la ponderación de la información propia y de los ascendientes se hace más pequeña a medida que n , el número de descendientes medidos, aumenta:

$$1/[1 + (2 + 0,5n)\alpha] \rightarrow 0, \text{ según } n \rightarrow \infty$$

e igualmente

$$\alpha/[1 + (2 + 0,5n)\alpha] \rightarrow 0, \text{ según } n \rightarrow \infty$$

mientras que

$$\alpha n/[1 + (2 + 0,5n)\alpha] \rightarrow 2, \text{ a medida que } n \rightarrow \infty.$$

Cuando se dispone de la propia producción del animal, la expresión general de la ecuación siempre contiene un término de desviación de la producción al grupo contemporáneo, es decir, $y_{ij} - GC_i$. Esto hace que se ajuste la superioridad (o inferioridad) genética de la competencia, ya que el término GC_i está ajustado por el valor genético medio de todos los animales del grupo, dado que, por ejemplo, $\hat{R}_1 = 213/5 - (\hat{a}_1 + \hat{a}_3 + \hat{a}_5 + \hat{a}_6 + \hat{a}_8)/5$.

Comentarios sobre los efectos fijos

Hemos visto como el *BLUP* tiene unas propiedades muy interesantes, entre las que destaca la ausencia de sesgo en la predicción de los efectos aleatorios, con la condición de que el modelo sea correcto.

¿Qué pasa cuando se ignora en nuestro modelo un factor del medio que tiene un efecto importante sobre las observaciones? En relación con el ejemplo que venimos desarrollando, es bien conocida la influencia de la edad de la madre sobre el peso al nacimiento del ternero, que aumenta a medida que la madre va alcanzando la madurez. La primera consecuencia de esta omisión es una predicción sesgada de los demás efectos, en particular de los valores genéticos (Henderson, 1975); así, imputaríamos, por ejemplo, un menor valor genético a una vaca primeriza que a una adulta. Ello llevaría a errores en la ordenación de los animales, reduciéndose la eficacia global de la selección. Como contrapartida paradójica, la precisión de nuestra evaluación se vería aparentemente sobreestimada.

Si por el contrario, incluimos en el análisis factores ambientales de efecto irrelevante, la predicción de los demás efectos no se ve sesgada, si bien disminuye la precisión de las estimaciones y predicciones. Otra consecuencia negativa será el aumento del riesgo de desconexión de las estimaciones. La desconexión es la ausencia de información que permita una comparación justa entre los animales (Ducrocq, 1992). Entraña una confusión entre efectos. Analicemos la siguiente tabla:

	Desconexión		Conexión	
	Rebaño 1	Rebaño 2	Rebaño 1	Rebaño 2
Toro A	n_A	–	n_{A_1}	n_{A_2}
Toro B	–	n_B	–	n_B

Supongamos que los toros A y B no están emparentados y que tienen hijas en los rebaños 1 y 2. Si el toro A tiene n_A hijas en el rebaño 1 y el toro B n_B hijas en el rebaño 2, no habrá manera de saber en qué medida las diferencias entre las hijas de ambos son debidas a causas genéticas o ambientales: estaremos ante un diseño desconectado. Si por el contrario, el toro A tiene n_{A_1} hijas en el rebaño 1 y n_{A_2} hijas en el rebaño 2, podremos estimar estadísticamente el efecto de cada factor sobre las producciones observadas, ya que las hijas sirven de conexión. Si incluso A y B fueran hermanos o medios hermanos, su emparentamiento aportaría una información –aunque menos rica que en el caso precedente– sobre la parte de las diferencias observadas que es de origen genético. Es decir, las relaciones de parentesco son fuente de conexión.

Modelo animal con medidas repetidas

Este apartado trata del análisis de caracteres que pueden ser medidos más de una vez en un animal, tal como la producción de leche o el número de lechones nacidos vivos en cada parto en el ganado de cerda. El modelo que se asume es el siguiente

$$y = Xb + Za + Zp + e$$

que con respecto a los modelos anteriores introduce los efectos ambientales permanentes, **p**, los cuales, siendo de origen no genético, actúan o influyen durante toda la vida del individuo, como por ejemplo la obstrucción de un pezón en la primera lactación, consecuente a una mamitis.

Este modelo permite utilizar toda la información de que se dispone de un animal. Si el animal no tiene ningún registro, como por ejemplo un toro lechero, la columna de **Z**

que está relacionada con este animal es cero. Así por ejemplo, si tuviéramos un diseño con cuatro animales, tres vacas y un toro, de las cuales la primera tiene tres lactaciones, la segunda dos y la tercera una, la matriz de incidencia \mathbf{Z} quedaría de la siguiente forma:

$$\mathbf{Z} = \begin{array}{c} \begin{array}{cccc} \text{♀}_1 & \text{♀}_2 & \text{♀}_3 & \text{♂} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right] \end{array} \end{array}$$

Dado que $\sigma_y^2 = \sigma_a^2 + \sigma_p^2 + \sigma_e^2$, y que podemos escribir la heredabilidad, h^2 , y la repetibilidad, r , como sigue

$$h^2 = \sigma_a^2 / \sigma_y^2$$

$$r = (\sigma_a^2 + \sigma_p^2) / \sigma_y^2 = (\sigma_y^2 - \sigma_e^2) / \sigma_y^2$$

se obtiene para los efectos genéticos aditivos

$$\alpha_1 = \sigma_y^2 / \sigma_a^2 = (1-r) / h^2$$

y para los valores ambientales permanentes

$$\alpha_2 = \sigma_e^2 / \sigma_p^2$$

donde

$$\sigma_e^2 / \sigma_y^2 = 1-r$$

y

$$\sigma_p^2 / \sigma_y^2 = (r-h^2)$$

por lo cual

$$\sigma_e^2 / \sigma_p^2 = (1-r) / (r-h^2)$$

De esta manera, las ecuaciones del modelo mixto, tras asumir que $\text{cov}(\mathbf{a}, \mathbf{p}') = \mathbf{0}$, $\text{cov}(\mathbf{a}, \mathbf{e}') = \mathbf{0}$ y $\text{cov}(\mathbf{p}, \mathbf{e}') = \mathbf{0}$, quedarían

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}[(1-r)/h^2] & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} + \mathbf{I}[(1-r)/(r-h^2)] \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Debemos hacer notar que todos los elementos de \mathbf{p} que pertenecen a animales sin observaciones son igual a cero, y, como consecuencia, pueden borrarse las ecuaciones y los coeficientes correspondientes a \mathbf{p} de esos animales. De esta forma, la predicción de una futura producción del animal i es $a_i + p_i$, mientras que la predicción de la producción de los animales sin registro y , por lo tanto, sin efecto ambiental permanente sería su valor mejorante.

Resolución de las ecuaciones de modelo mixto

Cuando se realiza una evaluación a partir de datos de campo, el número de ecuaciones que resultan de la aplicación de la metodología de modelo mixto (modelo animal) puede ser muy elevado, haciendo imposible la resolución de dichas ecuaciones por inversión directa de la matriz de coeficientes. La construcción de dicha matriz puede ser inabordable en estas condiciones incluso para los superordenadores más potentes. Como ejemplo, baste decir que la primera evaluación de ganado Holstein mediante un modelo animal, realizada en 1987 en los Estados Unidos, se hizo sobre registros de más de 7,5 millones de vacas. El sistema completo, incluyendo los efectos fijos y los aleatorios, constaba de aproximadamente 23,4 millones de ecuaciones (Wiggans *et al.*, 1988).

Es fácil suponer que dadas estas magnitudes ha sido necesario idear técnicas que de forma indirecta permitan acceder a soluciones idénticas (o muy aproximadas) a las que se obtendrían resolviendo las ecuaciones de modelo mixto. Estas técnicas se engloban dentro del concepto de “iteración sobre datos”, desarrollado a través de diferentes algoritmos. Mediante estas técnicas es posible llegar a soluciones con propiedades de modelo mixto sin necesidad incluso de construir la matriz de coeficientes completa. La descripción de tales técnicas queda, no obstante, fuera del alcance de un curso de introducción a la mejora.

Estimación de componentes de varianza

Los componentes de varianza necesarios para la evaluación *BLUP* deben estimarse a partir del mismo modelo asumido para realizar la evaluación genética. El método más aceptado en el mundo de la mejora genética, debido a sus propiedades estadísticas óptimas, es el de **máxima verosimilitud restringida**, *REML* en iniciales inglesas (*Restricted Maximum Likelihood*). Asumiendo un modelo animal sencillo de la forma

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

pueden estimarse los componentes de varianza de los factores aleatorios, es decir, la varianza aditiva, σ_a^2 , y la residual, σ_e^2 . En la iteración (i+1) del procedimiento de estimación, son los siguientes:

$$\hat{\sigma}_{e(i+1)}^2 = [\mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}_{(i)}'\mathbf{X}'\mathbf{y} - \hat{\mathbf{a}}_{(i)}'\mathbf{Z}'\mathbf{y}] / [N - r(\mathbf{X})]$$

$$\hat{\sigma}_{a(i+1)}^2 = [\hat{\mathbf{a}}_{(i)}'\mathbf{A}^{-1}\hat{\mathbf{a}}_{(i)} + \hat{\sigma}_{e(i)}^2 \text{tr}(\mathbf{A}^{-1}\mathbf{C}_{(i)})] / N$$

donde N es el número de registros o de animales (un registro por animal), los subíndices (i) e (i + 1) nos indican la iteración, siendo r y tr los operadores rango y traza respectivamente. La solución requiere asimismo la obtención de $\mathbf{C}_{(i)}$:

$$\mathbf{C}_{(i)} = (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{A}^{-1}\alpha_{(i)})^{-1}$$

donde

$$\mathbf{M} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$$

Evaluación genética de reproductores. II. Caracteres múltiples

La selección del ganado habitualmente se realiza tras evaluar los animales en relación con varios caracteres. En la medida en que los caracteres se encuentran correlacionados, se puede utilizar dicha información tanto para obtener predicciones de valores mejorantes respecto a varios caracteres a partir de información de los propios caracteres y de otros con ellos correlacionados, como para calcular un valor genético agregado consistente en una combinación de valores mejorantes ponderados por su importancia económica relativa.

En el capítulo que iniciamos vamos a describir los índices que permiten evaluar genéticamente los animales respecto a varios caracteres, pero, a diferencia de lo visto hasta ahora, teniendo en cuenta la información de otros caracteres correlacionados; posteriormente pasaremos a describir la metodología de evaluación por valor genético agregado y, finalmente, la metodología *BLUP* multicarácter.

Evaluación con información de caracteres correlacionados

Como se ha señalado anteriormente, las fuentes de información utilizadas para predecir el valor mejorante de un individuo respecto a un carácter pueden ser el fenotipo observado en el individuo o en algún pariente en relación con dicho carácter junto con el valor fenotípico de otros caracteres medidos en el propio individuo o en parientes del mismo. A efectos de simplificar la descripción del método, no obstante, consideraremos únicamente las mediciones simultáneas realizadas en el propio individuo.

Para cada carácter a evaluar, el índice tendrá la misma forma que la descrita en la selección para un solo carácter, es decir:

$$I_{\alpha} = b_1X_1 + b_2X_2 + \dots + b_nX_n$$

donde los X_i son las observaciones realizadas en cada individuo en relación con n caracteres, expresadas como desviación a la media de la población. Las ecuaciones del índice para el carácter i son semejantes a las que describimos en un capítulo anterior, siendo A_i el valor mejorante del carácter en evaluación:

$$b_1\sigma_{X_1}^2 + b_2\sigma_{X_1X_2} + \dots + b_n\sigma_{X_1X_n} = \sigma_{A_iX_1}$$

$$b_1\sigma_{X_2X_1} + b_2\sigma_{X_2}^2 + \dots + b_n\sigma_{X_2X_n} = \sigma_{A_iX_2}$$

$$b_1\sigma_{X_nX_1} + b_2\sigma_{X_nX_2} + \dots + b_n\sigma_{X_n}^2 = \sigma_{A_iX_n}$$

En este caso, no obstante, las covarianzas fenotípicas y genéticas deberán expresarse en términos de correlaciones genéticas ($r_{A_{ij}}$) y fenotípicas ($r_{P_{ij}}$), heredabilidades (h^2) y varianzas fenotípicas ($\sigma_{P_i}^2$):

$$\begin{aligned}\sigma_{X_i}^2 &= \sigma_{P_i}^2 \\ \sigma_{X_i X_j} &= r_{P_{ij}} \sigma_{P_i} \sigma_{P_j} \\ \sigma_{A_i X_i} &= h_i^2 \sigma_{P_i}^2 \\ \sigma_{A_i X_j} &= r_{A_{ij}} h_i h_j \sigma_{P_i} \sigma_{P_j}\end{aligned}$$

La varianza del índice, σ_I^2 , se calcula como se describió anteriormente, y asimismo la precisión del índice, $r_{AI} = \sigma_I / \sigma_A$.

Predicción del valor genético agregado

Aplicando el procedimiento que acabamos de describir para cada uno de los caracteres a evaluar, podríamos construir sendos índices que proporcionarían predicciones de los valores mejorantes del individuo respecto a distintos caracteres. Ponderando adecuadamente el valor mejorante predicho para cada carácter, según su importancia económica relativa, obtendríamos el **valor genético agregado** o mérito neto, H , mediante la siguiente ecuación:

$$H = a_1 A_1 + a_2 A_2 + \dots + a_n A_n$$

siendo a_i los coeficientes de ponderación económica y A_i los valores mejorantes predichos para cada carácter mediante índices construidos como se acaba de describir. Es fácil entender que el valor genético agregado se medirá en unidades monetarias, puesto que la importancia económica viene referida a la ganancia económica por cada incremento de una unidad del valor mejorante de cada carácter.

Como alternativa a calcular un índice para cada carácter, puede plantearse un índice que sintetiza tanto la información de los distintos caracteres como los coeficientes de ponderación económica. Las ecuaciones generales de los índices se plantearían como sigue:

$$\begin{aligned}b_1 \sigma_{X_1}^2 + b_2 \sigma_{X_1 X_2} + \dots + b_m \sigma_{X_1 X_m} &= \sigma_{HX_1} \\ b_1 \sigma_{X_2 X_1} + b_2 \sigma_{X_2}^2 + \dots + b_m \sigma_{X_2 X_m} &= \sigma_{HX_2} \\ b_1 \sigma_{X_m X_1} + b_2 \sigma_{X_m X_2} + \dots + b_m \sigma_{X_m}^2 &= \sigma_{HX_m}\end{aligned}$$

Si sustituimos H por su valor, tenemos:

$$\begin{aligned}\sigma_{HX_1} &= \text{cov}(X_1, a_1 A_1 + a_2 A_2 + \dots + a_n A_n) = a_1 \sigma_{X_1 A_1} + a_2 \sigma_{X_1 A_2} + \dots + a_n \sigma_{X_1 A_n} \\ \sigma_{HX_2} &= \text{cov}(X_2, a_1 A_1 + a_2 A_2 + \dots + a_n A_n) = a_1 \sigma_{X_2 A_1} + a_2 \sigma_{X_2 A_2} + \dots + a_n \sigma_{X_2 A_n} \\ \sigma_{HX_m} &= \text{cov}(X_m, a_1 A_1 + a_2 A_2 + \dots + a_n A_n) = a_1 \sigma_{X_m A_1} + a_2 \sigma_{X_m A_2} + \dots + a_n \sigma_{X_m A_n}\end{aligned}$$

Matricialmente, podemos resumir estas ecuaciones como sigue:

$$\begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \dots & \sigma_{X_1 X_m} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \dots & \sigma_{X_2 X_m} \\ \sigma_{X_m X_1} & \sigma_{X_m X_2} & \dots & \sigma_{X_m}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix} = \begin{bmatrix} \sigma_{X_1 A_1} & \sigma_{X_1 A_2} & \dots & \sigma_{X_1 A_n} \\ \sigma_{X_2 A_1} & \sigma_{X_2 A_2} & \dots & \sigma_{X_2 A_n} \\ \sigma_{X_m A_1} & \sigma_{X_m A_2} & \dots & \sigma_{X_m A_n} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix}$$

Puede observarse que las variables medidas, X_i , es decir, los criterios de evaluación o criterios de selección, no tienen que coincidir necesariamente con los caracteres a mejorar (objetivos de selección), A_i . La primera matriz, \mathbf{P} ($m \times m$), contiene las varianzas y covarianzas fenotípicas entre las m variables que se miden en cada individuo, mientras que la segunda, \mathbf{G} ($m \times n$), contiene las varianzas y covarianzas (genéticas) entre las variables medidas y los n valores mejorantes de los caracteres incluidos en el valor genético agregado. Completan el sistema el vector de coeficientes del índice, \mathbf{b} , así como el vector de pesos económicos relativos, \mathbf{a} . Podemos escribir simbólicamente este sistema matricial como sigue:

$$\mathbf{P} * \mathbf{b} = \mathbf{G} * \mathbf{a}$$

El índice de selección tendrá la misma forma que los que hemos visto hasta ahora, es decir,

$$I_\alpha = b_1X_1 + b_2X_2 + \dots + b_mX_m$$

Debemos observar, no obstante, que aunque se utilicen mediciones realizadas en el individuo o en alguno/s de sus parientes en relación con diferentes caracteres, los coeficientes del índice, los b_i , se han calculado teniendo en cuenta la información que relaciona las observaciones realizadas en el individuo con los objetivos de selección, así como los coeficientes de ponderación económica. Resulta obvio recordar que la construcción de este tipo de índices requiere el conocimiento preciso de las varianzas y covarianzas fenotípicas de las variables medidas, así como de las varianzas y covarianzas genéticas entre las variables medidas –criterios de evaluación– y los objetivos de selección, idealmente estimadas en la misma población.

Puede demostrarse, por otra parte, que el valor numérico del índice para un individuo concreto obtenido mediante la aplicación de la última fórmula es idéntico al valor genético agregado, H , obtenido a partir de una ponderación, según su importancia económica relativa, de los valores mejorantes de cada individuo predichos mediante la metodología descrita en el apartado anterior, es decir, operando carácter a carácter, pero considerando la información de caracteres correlacionados. Este último índice, por lo tanto, sería un predictor del valor económico global del animal.

La precisión del índice puede calcularse de forma similar a como se realiza en los índices para la evaluación de un solo carácter. Los cálculos son como sigue:

$$\begin{aligned}\sigma_I^2 &= b_1\sigma_{HX_1} + b_2\sigma_{HX_2} + \dots + b_m\sigma_{HX_m} \\ &= \mathbf{b}'\mathbf{Pb}\end{aligned}$$

Por otra parte, tras una deducción que se sugiere realizar al lector, se demuestra que:

$$\begin{aligned}\sigma_H^2 &= \sigma^2 (a_1A_1 + a_2A_2 + \dots + a_nA_n) \\ &= \mathbf{a}'\mathbf{Ca}\end{aligned}$$

siendo \mathbf{C} una matriz $n \times n$ de varianzas y covarianzas genéticas entre los caracteres incluidos en el valor genético agregado.

Finalmente,

$$r_{HI} = \sqrt{\frac{\mathbf{b}'\mathbf{Pb}}{\mathbf{a}'\mathbf{Ca}}}$$

Consideraciones genético-económicas

Un problema que debe resolverse es el número de objetivos y de criterios a considerar para la construcción del índice. Se trata de un problema complejo, puesto que idealmente los animales deberían evaluarse simultáneamente respecto a todos los caracteres que influyen o están relacionados con el sistema productivo, incluyendo estos caracteres en el valor genético agregado. Esta aproximación tiende a desembocar en la definición de un valor genético agregado muy complejo. A modo de ejemplo, en las aves se ha llegado a incluir 15 caracteres en la definición de H . Es fácil imaginar la dificultad que esta aproximación entraña en lo referente a la estimación de los parámetros adecuados.

Como alternativa, podemos entender que el valor económico de un animal depende fundamentalmente de unos pocos caracteres principales, mientras que otros son secundarios. Los primeros son los que deben incluirse en la definición del valor genético agregado, mientras que algunos de los segundos pueden ser incluidos como variables en el índice. Esta aproximación ha permitido, por ejemplo, reducir H en las aves a una combinación de cuatro caracteres (número de pollitos vendidos por gallina, número de huevos vendidos por gallina, calidad media del huevo y peso de la canal de la gallina a la reposición).

Otro ejemplo lo tenemos en el ganado porcino, donde se utiliza en la evaluación en granja un índice que incluye como variables el crecimiento diario entre 35 y 95 kg, así como el espesor del tocino dorsal medido mediante ultrasonidos. Los caracteres incluidos en H son, en cambio, el crecimiento en ese mismo periodo, el índice de conversión (correlacionado con el anterior) y el porcentaje de magro de la canal, correlacionado (negativa aunque favorablemente) con el espesor del tocino dorsal.

El cálculo de los coeficientes de ponderación económica puede realizarse de distintas maneras que no siempre conducen a resultados semejantes. Es método más utilizado tal vez sea el cálculo del beneficio marginal que se obtiene por la mejora de una unidad de un determinado carácter, considerando los demás fijos. La discusión de los métodos de cálculo de dichos coeficientes queda, no obstante, fuera del alcance de estas notas. Cabe señalar, por otra parte, que errores “moderados” –de hasta el 50%– en la estimación de los coeficientes económicos tienen poco efecto sobre la eficacia (precisión) del índice (Pease *et al.*, 1967), aunque cuanto mayores sean los errores, los índices se alejan del óptimo.

Respecto a los parámetros genéticos, Pease *et al.* (1967) han demostrado también que los errores en las heredabilidades tienen un efecto pequeño, siendo más importantes los errores en las correlaciones. Los errores de diferentes parámetros, además, tienen un efecto acumulativo sobre el índice.

Reescalado de los índices

Se ha comentado que los índices multicarácter predicen el valor genético agregado, el cual se expresa en unidades monetarias. Si los X_i se han incluido en el índice como valores absolutos, no desviados a la media, los valores de índice de cada animal oscilarán en torno a una media distinta de cero. Esto puede llevar a confusiones a los ganaderos que prefieren que los animales evaluados simultáneamente tengan valores de índice que oscilen en torno a un valor medio conocido, por ejemplo 100, en una escala cuya desviación típica es 10. El cálculo de los índices en la nueva escala se realiza muy fácilmente, según vamos a ver.

Sean I el valor del índice en la escala antigua;
 I' el valor del índice en la nueva escala;
 $\mu_i = b_1M_1 + b_2M_2 + \dots + b_mM_m$, la media de los valores del índice en la escala antigua, donde M_i es la media de la población estimada para la variable i del índice;

- μ_r la media del índice en la nueva escala;
- σ_l la desviación típica del índice en la escala antigua;
- σ_r la desviación típica en la nueva escala.

Entonces,

$$(I - \mu_l)/\sigma_l = (I' - \mu_r)/\sigma_r$$

Reordenando términos, se obtiene:

$$I' = [(I - \mu_l)/\sigma_l] \sigma_r + \mu_r$$

BLUP multicarácter (modelo animal)

De la misma forma que la teoría de los índices de selección puede adaptarse para la evaluación multicarácter, la metodología de modelo mixto puede abordar asimismo este tipo de análisis (Henderson y Quaas, 1976).

Hasta hace poco tiempo, el coste de computación disuadía a los mejoradores del uso de modelos *BLUP* multicarácter. No obstante, diversas estrategias de iteración sobre datos han cristalizado en la aparición de programas de evaluación como “*PEST*” (Groeneveld y Kovac, 1990) que permiten realizar evaluaciones *BLUP*, tanto multicarácter como de un solo carácter, a un coste de computación razonable.

Deducción de las ecuaciones de modelo mixto

Supongamos el caso más sencillo –dos caracteres– con las siguientes ecuaciones:

$$\begin{aligned} y_1 &= \mathbf{X}_1 \mathbf{b}_1 + \mathbf{Z}_1 \mathbf{a}_1 + \mathbf{e}_1 \\ y_2 &= \mathbf{X}_2 \mathbf{b}_2 + \mathbf{Z}_2 \mathbf{a}_2 + \mathbf{e}_2 \end{aligned}$$

donde los subíndices 1 y 2 denotan los caracteres 1 y 2, respectivamente. En esta disposición los animales se encuentran ordenados dentro de caracteres. Las ecuaciones pueden expresarse en forma matricial como

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

Asumiendo un modelo genético aditivo, la varianza de \mathbf{a} queda

$$\mathbf{V}(\mathbf{a}) = \mathbf{G} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} \\ g_{21}\mathbf{A} & g_{22}\mathbf{A} \end{bmatrix} = \mathbf{G}_0 \otimes \mathbf{A}$$

donde \mathbf{A} es la matriz de parentesco, \otimes es el operador producto directo de matrices, y g_{ij} es el elemento ij de \mathbf{G}_0 , la matriz de varianzas-covarianzas genéticas entre los caracteres.

La varianza del error queda

$$\mathbf{V}(\mathbf{e}) = \mathbf{R} = \begin{bmatrix} r_{11}\mathbf{I} & r_{12}\mathbf{I} \\ r_{12}\mathbf{I} & r_{22}\mathbf{I} \end{bmatrix} = \mathbf{R}_0 \otimes \mathbf{I}$$

donde r_{ij} es el elemento ij de \mathbf{R}_0 , la matriz de varianzas-covarianzas ambientales.

El equivalente a la submatriz $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$ en las ecuaciones de modelo mixto en este caso sería el producto de matrices

$$\begin{bmatrix} \mathbf{X}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' \end{bmatrix} \begin{bmatrix} \mathbf{I}r^{11} & \mathbf{I}r^{12} \\ \mathbf{I}r^{21} & \mathbf{I}r^{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' r^{11} \mathbf{X}_1 & \\ & \mathbf{X}_2' r^{22} \mathbf{X}_2 \end{bmatrix}$$

donde r^{ij} son los elementos ij de \mathbf{R}_0^{-1} . Operando de similar manera con las demás submatrices se obtiene

$$\begin{bmatrix} \mathbf{X}_1' r^{11} \mathbf{X}_1 & \mathbf{X}_1' r^{12} \mathbf{X}_2 & \mathbf{X}_1' r^{11} \mathbf{Z}_1 & \mathbf{X}_1' r^{12} \mathbf{Z}_2 \\ \mathbf{X}_2' r^{21} \mathbf{X}_1 & \mathbf{X}_2' r^{22} \mathbf{X}_2 & \mathbf{X}_2' r^{21} \mathbf{Z}_1 & \mathbf{X}_2' r^{22} \mathbf{Z}_2 \\ \mathbf{Z}_1' r^{11} \mathbf{X}_1 & \mathbf{Z}_1' r^{12} \mathbf{X}_2 & \mathbf{Z}_1' r^{11} \mathbf{Z}_1 + \mathbf{G}^{11} & \mathbf{Z}_1' r^{12} \mathbf{Z}_2 + \mathbf{G}^{12} \\ \mathbf{Z}_2' r^{21} \mathbf{X}_1 & \mathbf{Z}_2' r^{22} \mathbf{X}_2 & \mathbf{Z}_2' r^{21} \mathbf{Z}_1 + \mathbf{G}^{21} & \mathbf{Z}_2' r^{22} \mathbf{Z}_2 + \mathbf{G}^{22} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' (r^{11} \mathbf{y}_1 + r^{12} \mathbf{y}_2) \\ \mathbf{X}_2' (r^{21} \mathbf{y}_1 + r^{22} \mathbf{y}_2) \\ \mathbf{Z}_1' (r^{11} \mathbf{y}_1 + r^{12} \mathbf{y}_2) \\ \mathbf{Z}_2' (r^{21} \mathbf{y}_1 + r^{22} \mathbf{y}_2) \end{bmatrix}$$

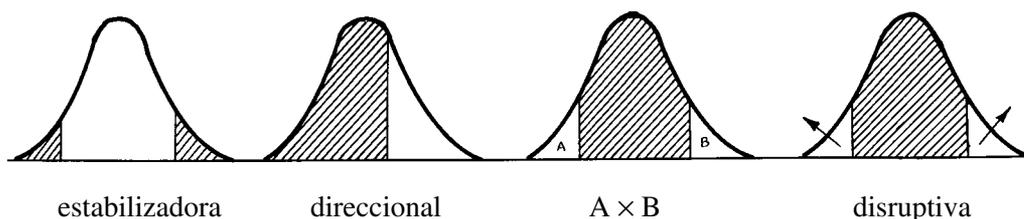
siendo \mathbf{G}^{ij} el producto $\mathbf{A}^{-1}g^{ij}$, donde g^{ij} es el elemento ij de \mathbf{G}_0^{-1} .

Se debe mencionar que mediante la resolución de estas ecuaciones se obtienen los valores mejorantes de los distintos individuos respecto a los dos caracteres. Para obtener el valor genético agregado de un individuo deberemos ponderar dichos valores mejorantes según su importancia económica relativa.

Selección artificial. I. Predicción de la respuesta

La mejora genética del ganado es la consecuencia de la acción secuencial de la selección y el apareamiento entre los mejores individuos. La **selección** implica la reproducción de los mejores individuos, cualquiera que sea la acepción de “mejor”. En capítulos precedentes hemos visto cómo detectar los mejores individuos a partir de su evaluación genética; ahora nos ocuparemos de predecir los resultados del apareamiento de estos individuos en una población cerrada. Los distintos métodos de evaluación genética darán lugar a diferentes métodos de selección que compararemos desde distintas perspectivas.

Para describir la teoría de selección comenzaremos por analizar el caso más sencillo, que corresponde a la selección de individuos respecto a un carácter, sobre la base de sus producciones. Genéricamente, podemos distinguir tres tipos de selección: estabilizadora, disruptiva y direccional, que esquemáticamente recoge la siguiente figura.



En el caso de la **selección estabilizadora** se eligen los fenotipos del entorno de la media de la población. Aunque es común en la naturaleza, en los animales domésticos se practica principalmente en la reproducción de animales de compañía o de exposición. En el ganado rústico, asimismo, suele suceder que las condiciones ambientales requieren un tamaño óptimo de los animales, razón por la cual los ganaderos tienden a eliminar los animales extremos. Una variante de esta selección estabilizadora consiste en seleccionar los extremos, pero apareándolos de forma cruzada –individuos de un extremo con los del otro–, de manera que se mantenga la media pero aumente la varianza.

En la **selección disruptiva** se seleccionan los extremos, pero los animales de cada uno de éstos se aparean entre sí. Resulta difícil encontrar ejemplos adecuados en el ganado, aunque uno aceptable sería la selección de los machos según su “masculinidad” y las hembras según su expresión femenina.

En la **selección direccional** se eligen los animales cuya producción es superior en algún sentido. En el ganado lechero se seleccionarían los animales de mayor producción lechera, mientras que en el ganado porcino se elegirían aquellos animales que tuvieran mayores crecimientos, pero con los menores índices de conversión y espesores de grasa dorsal. En el capítulo que iniciamos nos centraremos exclusivamente en la selección direccional en relación con un solo carácter.

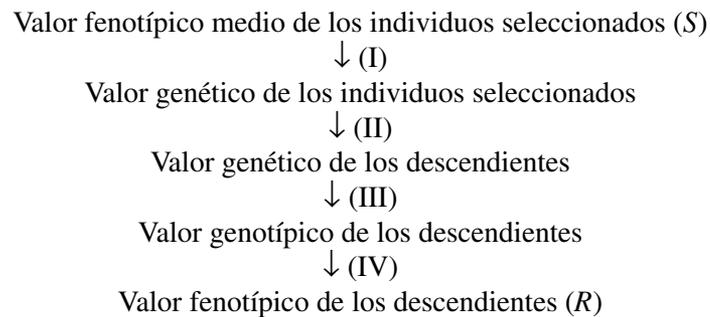
Es razonable pensar –y puede demostrarse formalmente (Falconer, 1983)– que el efecto básico de la selección debe ser cambiar las frecuencias génicas de los alelos favorables que determinan los caracteres métricos. No obstante, estos cambios de las frecuencias no van a ser observables, por lo que tendremos que analizar los cambios de la media de la población.

Efecto de la selección sobre la media: respuesta a la selección

El cambio de la media de la población es la **respuesta a la selección**, que simbolizaremos por R . Se define como la diferencia entre el valor fenotípico medio de la descendencia de los progenitores seleccionados y la media de la generación parental antes de realizar la selección. La respuesta a la selección se denomina con frecuencia también **progreso genético**, simbolizado por ΔG .

El **diferencial de selección** es la superioridad fenotípica media de los progenitores seleccionados y se designa por S ; se define como el valor fenotípico medio de los individuos seleccionados como progenitores expresado como desviación a la media de la población, es decir, el valor fenotípico medio de todos los individuos de la generación parental antes de que se hiciera la selección.

Para deducir la relación entre R y S es conveniente analizar las etapas que presentamos en el siguiente esquema (Ollivier, 1981):



- El paso (I) se hace suponiendo una relación lineal entre P , el valor fenotípico, y A , el valor aditivo. Aplicando la teoría de regresión, que supone una distribución normal bivariada de A y P , el valor mejorante de los individuos seleccionados es el producto de S , superioridad fenotípica, por b_{AP} , es decir,

$$S \cdot b_{AP} = S \cdot (\text{cov}(A,P)/V_P) = S \cdot \sigma_A^2 / \sigma_P^2 = S \cdot h^2$$

- El paso (II) requiere la definición de la forma de apareamiento. En caso de panmixia, hemos visto que el valor mejorante medio de los descendientes es igual al de sus padres; por ello, Sh^2 es el valor mejorante medio de los descendientes. Este planteamiento supone que todos los individuos seleccionados aportan descendencia a la generación siguiente en la misma medida, lo que es equivalente a suponer que toda la selección (artificial) se hace antes de la reproducción y que el carácter seleccionado es independiente del valor selectivo (ausencia de selección natural).
- El paso (III) requiere la ausencia de dominancia y de interacciones epistáticas, pues en este caso, el valor genotípico medio de los descendientes es igual al valor mejorante medio de los reproductores.
- El paso (IV) supone la ausencia de interacción entre el genotipo y el medio. Por lo expuesto, podemos escribir la siguiente ecuación:

$$R = Sh^2$$

Por otra parte, el diferencial de selección puede predecirse *a priori* a partir de la proporción de individuos seleccionados si se cumplen dos condiciones:

1. los valores fenotípicos del carácter que se selecciona se distribuyen normalmente, y
2. la selección es por truncamiento, es decir, ningún animal seleccionado por su valor fenotípico es inferior a cualquiera de los animales rechazados.

El diferencial de selección tipificado S/σ_p se denomina **intensidad de selección**, y se simboliza por i . El diferencial de selección es, por lo tanto,

$$S = i \sigma_p$$

y la respuesta esperada se hace

$$R = i h^2 \sigma_p$$

Como $h = \sigma_A/\sigma_p$, donde σ_A es la desviación típica de los valores mejorantes, podemos reescribir esta ecuación como

$$R = i h \sigma_A$$

Pero dado que la correlación entre el valor fenotípico y el valor mejorante de un individuo es como sigue

$$r_{AP} = \text{cov}(A,P)/(\sigma_A \sigma_p) = \sigma_A^2/(\sigma_A \sigma_p) = \sigma_A/\sigma_p = h,$$

esta última ecuación tiene un ámbito de aplicación más general y veremos que se utiliza para comparar métodos de selección. De hecho, generalizando, h puede entenderse como la correlación entre el criterio de evaluación o criterio de selección, hasta ahora el valor fenotípico, y el objetivo de selección, el valor mejorante. A tal correlación se la conoce como precisión del método de selección. La respuesta esperada queda, pues, en función del producto de la intensidad de selección, de la precisión del método y de la desviación típica aditiva. Simbólicamente podemos escribir:

$$R = i r_{AI} \sigma_A$$

En el contexto de los índices de selección para un solo carácter, la precisión era igual al cociente de la desviación típica del índice a la desviación típica aditiva, es decir, $r_{AI} = \sigma_I / \sigma_A$. Sustituyendo en la fórmula anterior, tenemos:

$$R = i \sigma_I$$

Por otra parte, si z es la altura de la ordenada en el punto de truncamiento (figura 4), por las propiedades de la ley normal reducida (media 0, varianza 1) se deduce que

$$S/\sigma_p = i = z/p$$

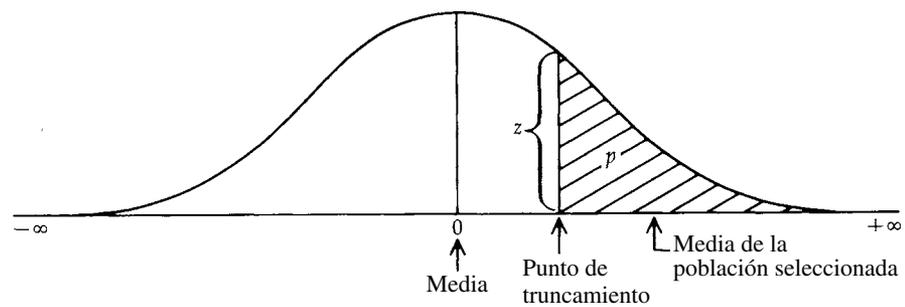


Figura 4. Parámetros relacionados con la respuesta a la selección en una población que sigue una distribución normal reducida. (Figura tomada de Van Vleck *et al.*, 1987.)

Según esas mismas propiedades de la ley normal, la intensidad de selección puede obtenerse a partir de la proporción de individuos seleccionados, p , tal como se describe en la tabla 4. Debe observarse que para obtener el valor de la intensidad de selección correspondiente a una determinada proporción de individuos seleccionados hay que tener en cuenta el tamaño de la población; los valores correspondientes de i varían según estemos ante una población grande, de más de 30 individuos en la práctica, o estemos ante poblaciones pequeñas de 10 o 20 individuos.

Tabla 4. Intensidad de selección según la proporción de individuos seleccionados y el efectivo de la población. (Tomada de Pirchner, 1983.)

Porcentaje seleccionado	Intensidad de selección		
	$n = 10$	$n = 20$	∞
1			2,67
2			2,42
5		1,87	2,06
10	1,54	1,64	1,75
20	1,27	1,33	1,40
30	1,06	1,11	1,16
40	0,89	0,93	0,97
50	0,74	0,77	0,80
60	0,60	0,62	0,64
70	0,45	0,48	0,50
80	0,32	0,34	0,35
90	0,17	0,18	0,19
100	0	0	0

Hemos analizado hasta ahora la superioridad media global de los progenitores utilizados. Los machos y las hembras, no obstante, pueden diferir en la cantidad de selección que se les aplica. Algunos caracteres, como la producción de leche, sólo pueden medirse en un sexo. Si la selección aplicada a machos y hembras difiere, los valores S ó i a utilizar son las medias no ponderadas de los dos sexos, a saber,

$$S = \frac{1}{2} (S_M + S_H)$$

$$i = \frac{1}{2} (i_M + i_H)$$

Progreso genético anual

El objetivo de la selección no es tanto maximizar la respuesta a la selección cada generación, sino que, por consideraciones de tipo económico, debemos maximizar la respuesta por unidad de tiempo, generalmente medido en años. Este planteamiento requiere la definición de un nuevo parámetro, el intervalo generacional. Es fácil sospechar que algunos métodos precisos, como por ejemplo la selección por descendencia, requieren un número elevado de años para evaluar los animales. A modo de ejemplo, la evaluación genética de un toro lechero se conoce cuando éste alcanza una edad de 6 a 7 años. La elevada respuesta generacional que puede obtenerse por este método, sobre todo debida a la alta precisión con que se evalúan los animales, puede verse claramente reducida si medimos la respuesta a la selección anual.

El **intervalo generacional**, que designamos como L , se define como el tiempo que transcurre entre los apareamientos de dos generaciones sucesivas. Cuando las generaciones se solapan –esto es, no son discretas–, el intervalo generacional se mide como la edad media de los padres cuando nace la descendencia seleccionada.

Sea ΔG el progreso genético anual, L_M el intervalo generacional de machos, y L_H el de hembras. En estas circunstancias

$$\Delta G = \frac{\Delta M + \Delta H}{L_M + L_H}$$

Este resultado se debe a Dickerson y Hazel (1944). Para demostrarlo supongamos que M es el valor genético aditivo de los machos elegidos para producir la siguiente generación y H el valor genético aditivo de las madres seleccionadas. Estos machos seleccionados nacieron L_M años antes de producir la descendencia que los sustituye, cuyo valor genético denominamos P . El valor genético medio de los toros nacidos L_M años antes es $P - L_M \Delta G$, suponiendo un progreso genético anual constante. La superioridad de los toros seleccionados sobre esta media es ΔM . Así, $M = P - L_M \Delta G + \Delta M$. Mediante un razonamiento similar, obtenemos que $H = P - L_H \Delta G + \Delta H$. Por otra parte, $P = (M + H)/2$, por lo que sustituyendo

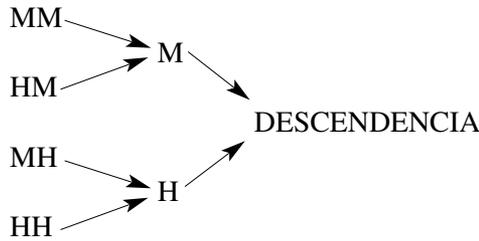
$$P = (M + H)/2 = 1/2(P - L_M \Delta G + \Delta M + P - L_H \Delta G + \Delta H)$$

Sustrayendo P de ambos lados de la ecuación y reordenando obtenemos

$$\Delta G (L_M + L_H) - (\Delta M + \Delta H) = 0, \text{ y, por lo tanto,}$$

$$\Delta G = \frac{\Delta M + \Delta H}{L_M + L_H}$$

Por su parte, Rendel y Robertson (1950) extendieron este método hasta considerar padres de machos (MM), madres de padres (HM), padres de hembras (MH) y madres de hembras (HH) elegidos como abuelos, teniendo cada una de estas vías un intervalo generacional diferente (L_{MM} , L_{HM} , L_{MH} , L_{HH} , respectivamente).



Sean ΔMM , ΔHM , ΔMH e ΔHH las superioridades genéticas respectivas de los abuelos seleccionados sobre la media de su generación. Por un razonamiento similar al anterior tenemos:

$$\begin{aligned} MM &= M - L_{MM} \Delta G + \Delta MM \\ HM &= M - L_{HM} \Delta G + \Delta HM \\ MH &= H - L_{MH} \Delta G + \Delta MH \\ HH &= H - L_{HH} \Delta G + \Delta HH \end{aligned}$$

Sabemos que en la generación parental, $M = (MM + HM)/2$ y $H = (MH + HH)/2$, y que el valor genético aditivo de la descendencia es, por lo tanto,

$$D = (M + H)/2 = (MM + HM + MH + HH)/4$$

sustituyendo,

$$(M + H)/2 = \frac{1}{4} (M - L_{MM} \Delta G + \Delta MM + M - L_{HM} \Delta G + \Delta HM + H - L_{MH} \Delta G + \Delta MH + H - L_{HH} \Delta G + \Delta HH)$$

Tras reordenar términos y sustraer $(M + H)/2$ de ambos lados,

$$\Delta G (L_{MM} + L_{HM} + L_{MH} + L_{HH}) - (\Delta MM + \Delta HM + \Delta MH + \Delta HH) = 0;$$

$$\Delta G = \frac{\Delta MM + \Delta HM + \Delta MH + \Delta HH}{L_{MM} + L_{HM} + L_{MH} + L_{HH}}$$

El progreso genético anual, por lo tanto, es equivalente a la superioridad promedio de los abuelos seleccionados dividida por el intervalo generacional medio de los diferentes abuelos. La expresión que acabamos de deducir, junto con la precedente, permiten comparar el progreso genético esperado en diferentes programas considerando diferencias en intervalos generacionales, intensidades de selección y precisiones de la predicción.

Optimización del progreso genético anual

Hemos visto que la respuesta a la selección anual depende de la intensidad de selección, de la precisión del método de selección, de la varianza aditiva y del intervalo generacional. De estos parámetros, la varianza aditiva a corto plazo se considera a efectos prácticos una propiedad de la población que no cambia en las primeras generaciones de selección. Sin embargo, estudios teóricos (Bulmer, 1971, 1976) han demostrado que la varianza aditiva en la subpoblación de padres seleccionados, V_A' , se ve reducida en una proporción que depende de la heredabilidad, según la siguiente ecuación:

$$V_A' = V_A(1 - h^2k)$$

siendo $k = i(i - x)$, donde x es la desviación del punto de truncamiento a la media de la población, expresada en unidades de desviación típica. En la segunda generación de selección la varianza aditiva se reduce nuevamente, aunque no mucho, no habiendo nuevos cambios importantes en las siguientes generaciones.

El mejorador sólo podrá actuar en principio sobre tres parámetros, la intensidad de selección, la precisión del método y el intervalo generacional. Desafortunadamente, no se conoce una regla general que permita decidir de forma analítica el plan de selección a adoptar. Dichos parámetros están interrelacionados de forma compleja y van a depender de la especie, del carácter y del método o métodos de evaluación, algunos de los cuales no son aplicables en ciertos casos por razones biológicas o técnicas. Dado que cada caso práctico es un caso particular, nos dedicaremos a exponer las bases de un plan de selección.

Valores de los parámetros de respuesta

Los valores de precisión referentes a cada uno de los métodos de selección son los que derivamos en un capítulo anterior donde se describieron los distintos métodos de evaluación de reproductores.

Los valores i y L son parámetros zootécnicos que aproximadamente –y en términos medios– tienen los valores que se expresan en la siguiente tabla, debida a Minvielle (1990).

Especie	L (años)		i y (% seleccionado/generación)			
	Machos	Hembras	Machos		Hembras	
Bovinos	3	8*	5	2,1 (4)	0,64 (60)	
Porcinos	1,5	3*	1,5	2,5 (1,4)	1,4 (20)	
Ovinos	2	4*	3,5	2,1 (4)	0,97 (40)	
Aves	1,5	–	1,5	2,5 (1,4)	1,4 (20)	

* Corresponde a evaluación por descendencia; los demás valores están referidos a selección individual.

Los valores presentados en la tabla pueden variar según las razas, los sistemas de producción y la aplicación o no de nuevas tecnologías reproductivas, como la transferencia de embriones. La intensidad de selección concretamente está limitada por la eficiencia reproductiva de la especie, puesto que es requisito imprescindible para muchos planes de selección mantener al menos el tamaño de la población constante. Como ejemplo ilustrativo, pensemos en una población porcina en la que cada hembra pare 10 lechones, 5 de los cuales son hembras. Si sólo se permite a cada cerda un solo parto, deberíamos sustituirla por una de sus 5 hijas; es decir, la proporción de individuos seleccionados para este sexo sería como máximo el 20%, a cuyo valor corresponde una intensidad de selección de 1,4.

Relaciones entre i , L y r_A

Como hemos dicho anteriormente, las relaciones existentes entre los parámetros que determinan la respuesta a la selección son complejas, puesto que no podemos, en general, actuar sobre alguno de ellos sin que se modifique alguno de los otros. Ello nos obliga a buscar para cada caso particular la combinación óptima entre i , L y r_A .

Relaciones entre i y r_{AI}

La precisión de un método de selección aumenta con el número de medidas por animal o el número de parientes medidos. No obstante, bien se realicen los controles en granja o en centro de control, el número de controles no puede sobrepasar un máximo, que generalmente se fija por limitaciones económicas. Por otra parte, cuando la evaluación de un reproductor implica el sacrificio de algunos colaterales, ello redundará en una disminución efectiva del número de candidatos a la selección y, por tanto, de i .

Al margen de este planteamiento quedarían en la actualidad la selección para caracteres que normalmente se controlan en granja con fines principalmente de gestión técnica de la explotación, con independencia de que después se utilice la información para realizar la evaluación genética de los animales.

Relaciones entre i y L

El progreso genético anual aumenta cuando L , el intervalo generacional, disminuye, es decir, cuando disminuye el tiempo de utilización de los reproductores. No obstante, el número de descendientes por reproductor también disminuye y, consecuentemente, el número de candidatos disponibles para la generación siguiente es menor, al igual que la intensidad de selección. Así, en las especies poco prolíficas nos veremos obligados a alargar el intervalo generacional para alcanzar una intensidad de selección suficiente. En las especies prolíficas, en cambio, podrá obtenerse una intensidad de selección elevada de forma rápida, lo cual reducirá considerablemente los intervalos generacionales.

Relaciones entre L y r_{AI}

Podemos aumentar el progreso genético anual utilizando un método preciso de evaluación de los animales, aunque ello pueda conducir a aumentar el intervalo generacional de forma importante. Por ejemplo, recurrir a la selección por descendencia podrá permitir –en ciertas condiciones– el aumento de la precisión, aunque implicará un aumento del intervalo generacional. En la selección del ganado porcino respecto a caracteres productivos, tradicionalmente se ha considerado que el incremento en precisión no es suficiente para compensar el aumento del intervalo generacional, razón por la que se recomienda utilizar métodos menos precisos, con una rotación rápida del ganado.

La discusión que acabamos de realizar ha mostrado la importancia de evaluar las relaciones entre los factores de respuesta a la selección antes de diseñar cualquier plan de selección. Los razonamientos expuestos han sido de carácter muy general y siempre simplificando la realidad. Para comparar en cada caso concreto diversas alternativas de selección deberemos considerar parámetros técnicos realistas, extraídos del sistema de producción en el que están inmersos los animales.

Comparación general de planes de selección

La comparación de los planes de selección se realiza evaluando el progreso genético que se espera en cada uno de ellos. No obstante, un programa de coste excesivo deberá ser descartado. En la práctica deberemos buscar el óptimo derivado de considerar tanto el progreso genético esperado, como el coste económico.

En un contexto económico, puede actualizarse progreso genético esperado, es decir, transformamos el valor económico del cambio esperado en años sucesivos en la ganancia se haría efectiva hoy en día. La tasa de actualización es $1/(1 + e)^t$, donde e es la tasa de descuento y t indica los años en los que se realizará la ganancia.

Pueden darse algunas recomendaciones generales sobre el tipo de plan de selección a adoptar, según sea el tipo de carácter a mejorar y según sea el origen de la información disponible. Los resumiremos en la tabla 5.

Tabla 5. Recomendaciones generales sobre el plan de selección a adoptar. (Adaptado de Minvielle, 1990.)

Origen de las observaciones (ejemplo)	Heredabilidad	
	Baja ($h^2 < 0,2$)	Alta ($h^2 > 0,4$)
	<i>Selección de los machos</i>	
Sin restricción alguna (crecimiento) o machos solamente (semen)	P o I	I
Hembras solamente (de leche)	P o C	P o G o C
Emparentados solamente (canal)	P	P o C
	<i>Selección de las hembras</i>	
Sin restricción alguna (crecimiento) o hembras solamente (leche)	I o G	I
Machos solamente (semen)	C o G	C o G
Emparentados (canal)	C o G	C

C, selección por colaterales; G, selección por ascendientes; I, selección individual; P, selección sobre descendencia.

En general, siempre que sea posible y la heredabilidad sea bastante elevada, el método de elección es la selección individual. Es un método simple que, a diferencia de otros métodos de evaluación por índices y del *BLUP*, no necesita registrar las genealogías. La utilización de colaterales, y todavía más de descendientes, permite un progreso más rápido, si bien requiere la gestión de un mayor número de datos, aunque con los sistemas informáticos actuales esto no constituye ningún problema. La selección por ascendencia tiene como ventaja el permitir una selección muy temprana, aunque su precisión es siempre inferior a la de la selección individual. Es el método obligatorio para elegir los toros jóvenes lecheros que más tarde serán evaluados mediante una prueba de descendencia.

Acerca de la selección familiar, no reflejada en la comparación mostrada en la tabla 5, cabe decir que es eficaz cuando los efectivos familiares son grandes, la heredabilidad es baja y los efectos del medio ambiente común son inexistentes; como contrapartida presenta el inconveniente de aumentar rápidamente la tasa de consanguinidad de poblaciones pequeñas cerradas.

La selección intrafamiliar, en cambio, posibilita limitar el aumento de la tasa de consanguinidad, aunque su precisión, y por lo tanto la respuesta esperada, son siempre inferiores a las de la selección individual. Otra circunstancia que puede justificar la selección intrafamiliar es la existencia de un elevado efecto ambiental común a los miembros de la misma familia, por ejemplo un efecto materno importante.

Selección artificial. II. Medida de la respuesta

Hasta ahora hemos estudiado la predicción de la respuesta a la selección, es decir, el cambio genético debido a la selección. Debemos preguntarnos en qué medida concuerdan las ganancias esperadas en los procesos de selección con las que se dan de hecho en poblaciones concretas. Este tema ha sido objeto de numerosas investigaciones. La mayoría de ellas se han hecho utilizando animales de laboratorio (ratones e insectos de los géneros *Drosophila* y *Tribolium*) por su elevada prolificidad, su bajo intervalo generacional y su más bajo coste, comparado con las especies de ganado. En otros estudios se ha utilizado la simulación mediante computador para investigar la sensibilidad de la respuesta a la selección a simplificaciones, omisiones, sesgos e imprecisiones de la teoría de la selección.

Medida de la respuesta

Cuando se lleva a cabo un experimento de selección, se observa que las medias de las generaciones varían; es decir, las medias de las generaciones no progresan de una forma regular, sino que fluctúan erráticamente, más o menos violentamente. La mejor medida de la respuesta media por generación se obtiene mediante la pendiente de una recta de regresión ajustada a las medias generacionales, tal como se ilustra en la figura 5.

Las causas de variación son la deriva genética aleatoria, los errores de muestreo en la estimación de las medias de cada generación, las diferencias en el diferencial de selección y los factores ambientales. La variación debida a la deriva genética y a los errores de muestreo puede reducirse solamente aumentando los efectivos medidos y seleccionados. Las otras dos causas vamos a discutir las con más detalle.

Diferencias ambientales

Las diferencias ambientales entre generaciones pueden derivarse de causas climáticas, nutritivas y de manejo en general. La manera obvia de eliminar las fluctuaciones ambientales es manteniendo una población control no seleccionada. Suponiendo que las diferencias ambientales afectan de igual modo a las poblaciones seleccionada y control, la diferencia entre ellas estima la mejora genética realizada mediante selección.

El uso de una población control, no obstante, no siempre mejora la precisión con que se estima la respuesta. Ambas poblaciones están sometidas a la deriva aleatoria y a los errores de muestreo, por lo que la diferencia entre las dos está sometida a la varianza de estas dos causas que es la suma de la varianza de las dos líneas. Por otra parte, si se tiene alguna limitación de espacio, hay que reducir el tamaño de la población de la línea seleccionada. Si la línea seleccionada y el control son en tamaño la mitad de una línea seleccionada de un experimento en que no se use control, la utilización de un control cuadruplica la varianza de muestreo de la respuesta medida como desviaciones al control, con lo cual se dobla el error típico. Esta pérdida de precisión puede contrapesar la ganancia de eliminar diferencias ambientales.

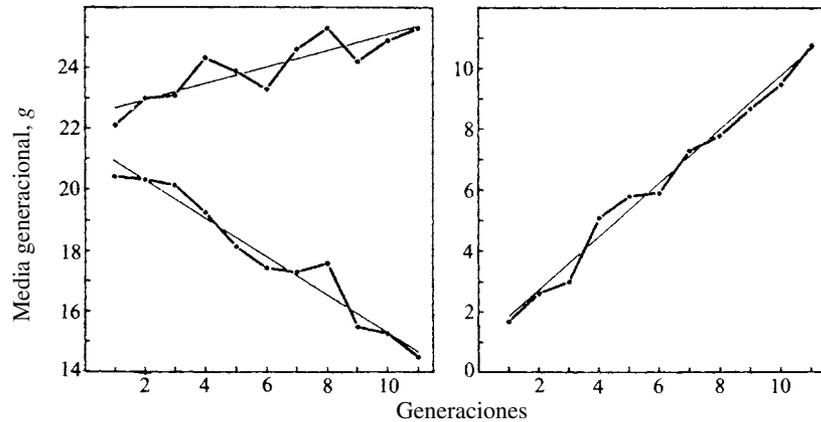


Figura 5. Selección divergente para el peso a las seis semanas en ratones. En la izquierda, las respuestas de las dos líneas se muestran por separado. En la derecha, se presenta la diferencia entre la línea seleccionada a favor y la línea seleccionada en contra. (Tomado de Falconer, 1986.)

Como alternativa puede plantearse un experimento de selección divergente. Cada línea seleccionada actúa como control de la otra, midiéndose la respuesta como divergencia entre las dos líneas. En este caso, en relación con la magnitud de la respuesta, la precisión es la misma que la de una línea que tuviera en total los mismos individuos medidos. Ello se debe a que aunque el error típico de la diferencia entre líneas se doble, tal como acabamos de explicar, también la respuesta se dobla aproximadamente, ya que se seleccionan ambas líneas.

Los cambios aleatorios de ambiente reducen la precisión con que se estima la respuesta, pero no sesgan la estimación. Otra cosa distinta es cuando aparecen tendencias ambientales, es decir, cambios progresivos en el tiempo, pues lo que parece una respuesta a la selección puede deberse realmente a una tendencia ambiental. Esto hace difícil la evaluación de la eficacia de la selección en la mejora del ganado, ya que sin una población control no hay forma segura de asignar cuánta mejora se debe a la selección, y cuánta al manejo.

Utilizando los símbolos habituales, podríamos escribir

$$P_t = A_t + E_t \text{ y } P_{t+1} = A_{t+1} + E_{t+1}$$

donde los subíndices t y $t+1$ indican la generación en que nos encontramos. La diferencia entre dos generaciones consecutivas sería:

$$P_{t+1} - P_t = (A_{t+1} - A_t) + (E_{t+1} - E_t)$$

De forma similar, las diferencias entre ambientes podríamos obtenerlas a partir de las medias de la población control:

$$C_{t+1} - C_t = (A_0 - A_0) + (E_{t+1} - E_t),$$

de modo que sustituyendo se calcularía el progreso genético realizado.

Por otra parte, tanto en experimentos de selección como en programas de mejora llevados a la práctica, podríamos obtener la tendencia genética libre de las influencias ambientales a partir de un modelo animal (*BLUP*) que incluyera la matriz de parentescos

co. La tendencia genética se mediría a través de la evolución de la media de los valores mejorantes de los animales nacidos en el mismo año. Las tendencias ambientales podrían asimismo estimarse analizando la evolución anual de los valores medios de los efectos fijos (rebaño, por ejemplo).

Ponderación del diferencial de selección

Se debe distinguir entre diferencial de selección efectivo y esperado, pues en la práctica la aportación de los reproductores a la siguiente generación no suele ser la misma, dado que siempre están presentes las diferencias de fertilidad. Por ello hay que ponderar la desviación a la media de cada progenitor por su aportación proporcional a la descendencia.

El diferencial de selección esperado es simplemente la desviación fenotípica media de los progenitores seleccionados; el diferencial de selección efectivo, en cambio, es la desviación media ponderada de los padres.

La ponderación del diferencial de selección tiene en cuenta en buena medida los efectos de la selección natural. Si, por ejemplo, los fenotipos más extremos son menos fértiles, la selección natural trabajará en contra de la selección artificial. Mediante la ponderación del diferencial de selección medimos, por lo tanto, los efectos conjuntos de las selecciones natural y artificial.

Heredabilidad realizada

La respuesta a la selección, estimada como cambio promedio por generación, no tiene en cuenta la magnitud de la selección aplicada. Como alternativa podemos plantearnos estimar la respuesta a la selección en función del diferencial de selección efectivo acumulado, tal y como se muestra en el ejemplo de la figura 6. El valor promedio de la relación R/S viene dado por la pendiente de la regresión de la respuesta sobre el diferencial de selección acumulado.

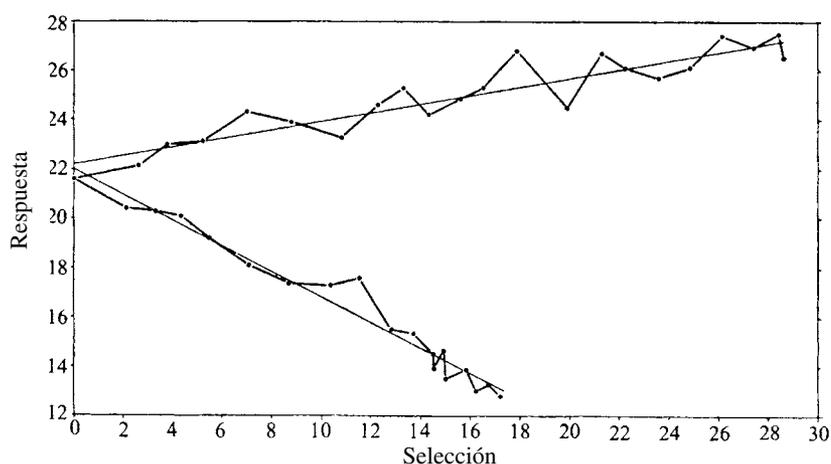


Figura 6. Selección divergente para el peso a las seis semanas en ratones. Las medias generacionales están referidas a los diferenciales de selección acumulados. (Tomado de Falconer, 1986.)

Anteriormente demostramos que en la selección individual $R = h^2S$, siendo, por lo tanto, $h^2 = R/S$. Aplicado al caso que nos ocupa, el cociente R/S , estimado como regresión de la respuesta sobre el diferencial de selección acumulado, se denomina heredabilidad realizada, siendo primariamente una descripción de la respuesta que puede no dar, por varias razones, una estimación válida de la heredabilidad de la población base. Si hay

cambios sistemáticos debidos a las tendencias ambientales o a la depresión endogámica, éstos estarán incluidos en la respuesta a menos que los eliminemos por comparación con una línea control. Los efectos de la deriva genética, en cambio, sólo podrán evaluarse mediante la replicación de la selección.

Por otra parte, como puede verse en la figura 6, la heredabilidad realizada en dos líneas divergentes suele ser distinta. Cada una de ellas es una descripción válida de la respuesta, pero no pueden tomarse como estimaciones válidas de la heredabilidad en la población base. Las posibles causas de la asimetría de la respuesta las comentaremos más adelante.

Respuesta correlacionada a la selección

Hasta ahora nos hemos referido a la respuesta directa, es decir, la de un carácter cuando los animales se seleccionan respecto al mismo. Podríamos plantearnos, no obstante, cuál sería la respuesta en un carácter correlacionado genéticamente, y , cuando seleccionamos respecto a un carácter x , es decir, cuál sería la respuesta si se realizara la selección indirecta de y a través de x . El cambio operado en el carácter y vendrá dado por la regresión del valor genético de y sobre el valor genético de x . Esta regresión es

$$b_{(A)y|x} = \text{cov}_A / \sigma_{A_x}^2 = r_A \sigma_{A_y} / \sigma_{A_x}$$

La respuesta del carácter x , seleccionado directamente, es

$$R_x = i h_x \sigma_{A_x}$$

Por lo tanto, la respuesta correlacionada del carácter y es

$$\begin{aligned} RC_y &= b_{(A)y|x} R_x \\ &= i h_x \sigma_{A_x} r_A \sigma_{A_y} / \sigma_{A_x} \\ &= i h_x r_A \sigma_{A_y} \end{aligned}$$

Dado que podemos expresar $\sigma_{A_y} = h_y \sigma_{P_y}$, la respuesta correlacionada resulta

$$RC_y = r_A i h_x h_y \sigma_{P_y}$$

Si realizamos un experimento de selección doble, es decir, seleccionando en una línea el carácter x y en otra el carácter y , midiendo la respuesta directa y la respuesta correlacionada realizada en ambas líneas podemos obtener una estimación de la correlación genética, r_A . Despejando en la última ecuación se tiene:

$$r_A = RC_y / (i h_x h_y \sigma_{P_y}),$$

De forma similar

$$r_A = RC_x / (i h_y h_x \sigma_{P_x})$$

Multiplicando estas dos últimas expresiones podemos obtener una estimación conjunta del de la correlación genética (de su cuadrado), para lo cual no se necesita conocer ningún parámetro distinto a las respuestas directas y correlacionadas para ambos caracteres.

$$r_A^2 = \frac{RC_x}{R_x} \cdot \frac{RC_y}{R_y}$$

Variabilidad de las respuestas

Si se replican los experimentos de selección, es decir, si se someten contemporáneamente al mismo esquema de selección varias líneas de la misma población base, es habitual que difieran sus respuestas (figura 7). Esta variabilidad será mayor cuanto menor sea el número de individuos seleccionados como progenitores de una nueva generación y menor sea la heredabilidad del carácter seleccionado. A pesar de dicha variabilidad, frecuentemente la respuesta promedio de varias líneas replicadas está en estrecha concordancia con lo esperado (Clayton *et al.*, 1957).

Discordancia entre la respuesta observada y la esperada

A modo de ejemplo, hemos visto que se observó que la respuesta a la selección para el aumento de peso a las seis semanas de los ratones fue menor que la esperada, mientras que la respuesta a la reducción de dicho peso fue superior a la predicha (figura 6). La explicación de tal desacuerdo puede radicar en varios factores, o en una combinación de los mismos.

Variabilidad entre líneas

Si el experimento se realiza solamente con una o unas pocas líneas de selección, el desacuerdo podría ser el resultado de la variabilidad entre líneas, como acabamos de describir. Debido al coste de los factores, muchos de los experimentos de selección que se realizan en el ganado se llevan a cabo sobre una o unas pocas líneas, cada una de ellas de efectivo limitado. En estas circunstancias no es posible eliminar la variabilidad de las líneas como causa de discordancia entre la respuesta esperada y observada.

Selección natural

La respuesta se predice como función de S , la superioridad fenotípica de los padres seleccionados sobre la población. Si el carácter que se selecciona influye sobre la viabilidad, la fertilidad o la longevidad, es decir, si está genéticamente relacionado con la eficacia biológica, los padres seleccionados no tendrán igual número de descendientes, por lo cual no aportarán igual número de genes a la siguiente generación. El diferencial de selección efectivo no será igual al esperado. Como la respuesta observada es una función del diferencial de selección realizado, si hay selección natural, la respuesta real no será igual a la esperada.

La selección natural puede ser antagónica o sinérgica con la selección artificial, dependiendo de la correlación entre el carácter seleccionado y la eficacia biológica. Es razonable pensar que la historia evolutiva de una población la haya llevado a unos niveles casi óptimos, desde el punto de vista de la eficacia biológica, respecto a muchos caracteres. La selección natural opondría resistencia a que intentásemos llevar la población lejos de tal óptimo; la correlación genética del carácter seleccionado con la eficacia biológica sería negativa, o lo que es equivalente, los individuos seleccionados, portadores de los genotipos más extremos, se reproducirían a tasas más bajas siendo el diferencial de selección efectivo menor que el esperado y, por lo tanto, menor también la respuesta.

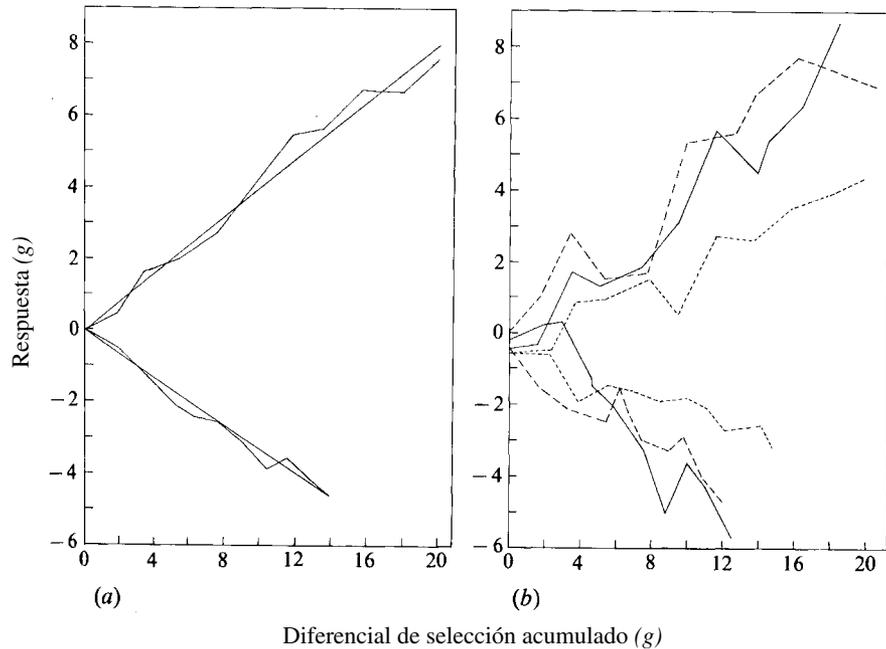


Figura 7. Diez generaciones de selección divergente para el peso a las seis semanas en ratones. Las medias generacionales se presentan en función de los diferenciales de selección acumulados.

a) Población total consistente en seis réplicas seleccionadas en cada generación.

b) Tres de las seis réplicas seleccionadas en cada generación. Cada réplica estuvo constituida por ocho pares de progenitores con endogamia mínima, dando $N_e = 31$.

(Tomado de Falconer, 1986.)

Depresión consanguínea

Las líneas de selección son habitualmente pequeñas, por lo cual es de esperar que la selección deba coexistir simultáneamente con la consanguinidad o endogamia. Si el carácter seleccionado se ve afectado por la depresión endogámica (que analizaremos en un capítulo posterior), se neutralizará una parte o todo el progreso de la selección a favor, mientras que la respuesta a la selección en contra del carácter se verá aumentada. En ambos casos la respuesta a la selección real estará lejos de la esperada. No obstante, en la selección del ganado se intenta evitar generalmente la endogamia, por lo que, en este contexto, los efectos descritos serán poco importantes.

Asimetría de la respuesta

De igual modo que la respuesta a la selección real depende más del diferencial de selección efectivo que del esperado, dicha respuesta a la selección real depende más de la heredabilidad realizada que de la estimada a partir del grado de parecido entre parientes. En un experimento publicado por Falconer en 1954, la heredabilidad del peso de los ratones a las seis semanas, estimada a partir del parecido entre parientes de la población base, fue 0,35, mientras que las heredabilidades realizadas en la selección a favor y en contra fueron, respectivamente, 0,18 y 0,52. Vamos a analizar algunos factores que pueden explicar esta divergencia.

- a) Dominancia direccional. En los numerosos *loci* que afectan al carácter en la población base podría haber una preponderancia de los genes dominantes que influyeran el carácter en una dirección, mientras que los alelos recesivos influirían en la dirección opuesta. Si asumiéramos que los genes estuvieran al principio en frecuencias intermedias, la respuesta a la selección en la dirección influida por los alelos recesivos sería más rápida que en la dirección influida por los alelos dominantes. Ello se debe a que al seleccionar a favor del aumento de la frecuencia de genes con efectos dominantes, no podríamos distinguir entre heterocigotos y homocigotos dominantes, siendo así enmascarado por la dominancia el verdadero valor genético aditivo. La selección, por tanto, sería relativamente imprecisa, situación que no se daría en los recesivos.
- b) Frecuencia génica direccional. Los genes que influyen el carácter en una dirección podrían estar en frecuencias más altas en la población base que los alelos que influyen el carácter en la otra dirección, dependiendo de la historia selectiva previa de la población. En ausencia de dominancia direccional, la frecuencia génica direccional causaría una respuesta a la selección mayor en la dirección de los alelos menos frecuentes (véase la figura 3). Los efectos de dominancia direccional y frecuencia génica direccional no son mutuamente excluyentes ni independientes; pueden interactuar causando muchos modelos de asimetría de la respuesta a la selección, dependiendo de las propiedades genéticas del carácter seleccionado en la población base.
- c) Ventaja del heterocigoto. La selección en una dirección puede favorecer a los genotipos heterocigotos. Pueden estar implicados varios *loci*, pero el razonamiento que ilustra el principio de este fenómeno lo haremos por comodidad refiriéndonos a un solo *locus*. Si un gen con dos alelos influye sobre un carácter según un modelo sobredominante, el heterocigoto será superior a los dos homocigotos. La selección de los individuos superiores favorecería los individuos heterocigotos, que aportarían los dos alelos a la siguiente generación con iguales frecuencias. La población se aproximaría gradualmente al equilibrio de frecuencias génicas en el *locus*, en un valor que sería dependiente del mérito relativo de los dos homocigotos. Una vez que se alcanzase el equilibrio genético, la población seguiría presentando variación genética, pero no habría nuevas respuestas a la selección. La selección realizada en dirección opuesta, en cambio, se vería poco afectada por la superioridad de los heterocigotos, y uno de los alelos se aproximaría rápidamente a la fijación. El efecto consecuente sería la respuesta asimétrica a la selección bidireccional, con heredabilidades realizadas desiguales en las dos direcciones.

Tamaño de población mínimo para que la selección sea efectiva

Si el tamaño efectivo de la población, N_e , es pequeño, por muy bien diseñado que esté un programa de selección, la diferencia entre una pequeña respuesta y una gran respuesta estará determinada principalmente por el azar. Nicholas (1980) ha demostrado que para muchas situaciones prácticas, aceptando coeficientes de variación en la respuesta no superiores al 20%, cuando el programa se evalúa tras t años, y L e i son el intervalo generacional y la intensidad promedio, respectivamente, el tamaño efectivo de la población debería ser como mínimo igual a

$$50L / (i^2 h^2 t).$$

Duración de la respuesta a la selección

Los resultados de experimentos de selección a largo plazo han demostrado que la respuesta por generación no es constante. En ciertas ocasiones se ha visto que la tasa de respuesta decrece con el tiempo hasta que cesa enteramente. Este fue el caso tras 20 a 30 generaciones de selección direccional con respecto a la longitud del tórax en *Drosophila* (Robertson, 1955), así como del peso a las seis semanas de los ratones (Falconer, 1955). No obstante, como contraste, la respuesta en el porcentaje de aceite de un maíz se prolongó más de ochenta generaciones (Dudley, 1977). Las causas de estas divergencias no son bien conocidas. Dos cuestiones son relevantes en este apartado: *a*) la respuesta máxima esperada, y *b*) las propiedades genéticas de una población en la que ha cesado la respuesta, que está en el límite de selección.

Respuesta máxima esperada

Se ha demostrado que la selección es eficaz tanto más tiempo cuanto mayor es el tamaño efectivo de la población, N_e . Con genes aditivos, Robertson (1960) ha demostrado que la respuesta máxima esperada es la siguiente:

$$R_{(\text{máx})} = 2N_e i h^2 \sigma_p$$

siendo $ih^2\sigma_p$ la respuesta a la selección fenotípica en una generación, o bien la respuesta promedio observada en las primeras generaciones.

Cuando la selección se extiende durante un largo período, Hill (1982) ha demostrado que la respuesta puede verse incrementada como consecuencia de mutaciones espontáneas, prolongándose indefinidamente, de manera que no se llegue a alcanzar el límite de selección. Por otra parte, aunque los experimentos de inducir mutaciones por irradiaciones de los órganos genitales en general no han sido exitosos, no debemos olvidar la vía abierta por Mackay (1984), quien observó un incremento sustancial de la varianza y la respuesta usando cruces disgénicos de *Drosophila*, asociado al movimiento de elementos transponibles.

Causas del cese de la respuesta

Hay al menos cuatro razones posibles por las que una población sometida a selección direccional deje de responder, es decir, alcance el límite de selección o meseta de selección:

- a*) Agotamiento de la varianza genética aditiva, debiéndose al ambiente todas las diferencias entre individuos. Si esto fuera la causa de la respuesta en meseta, la heredabilidad estimada en una población en meseta debería ser cero; asimismo, la población no debería responder a la selección inversa. Debido a que no se da habitualmente esta última circunstancia, el agotamiento de la varianza genética aditiva no parece una explicación general del cese de la respuesta.
- b*) Sobredominancia. Como ya hemos mencionado, la acción génica sobredominante podría ser importante para el carácter seleccionado, llevando la ventaja del heterocigoto a un equilibrio intermedio de las frecuencias génicas. En estas condiciones, la población debería responder a la selección inversa con una disminución de la media del carácter seleccionado, aunque estuviera la heredabilidad estimada en el equilibrio próxima a cero, pues toda la varianza genética sería no aditiva.
- c*) Ligamiento. Los genes que afectan al carácter en direcciones opuestas podrían estar ligados. Si este fuera el caso, las ganancias procedentes de la selección de genes en un *locus*, podrían ser anuladas por las pérdidas originadas por la selección concomitante de un alelo indeseado del *locus* ligado. Si esta fuera la causa de la respuesta en meseta, la media de la población no debería cambiar bajo selección rela-

jada, aunque podría responder a la selección inversa. Varias generaciones de selección relajada podrían permitir que la recombinación rompiera los ligamientos desfavorables, permitiendo con ello una respuesta a una nueva selección a favor.

- d) Selección natural. Aunque hubiera todavía varianza genética y la heredabilidad fuera positiva, los efectos favorables de la selección artificial podrían ser contrarrestados por los efectos negativos de la selección natural. Esto podría dar como resultado un diferencial de selección efectivo cero, sin respuestas adicionales a la selección.

Muchos de los intentos realizados para alterar una población y acercarla a las ideas preconcebidas del ganadero respecto a los rendimientos o al tipo, serían una agresión para una población moldeada óptimamente a su ambiente por muchas generaciones de selección natural. Por ello, no debemos sorprendernos de que las poblaciones de ganado se resistan a los cambios que queramos imponerles. Pero también es cierto que mediante la alteración del ambiente –alojamientos, higiene, buena alimentación y otros medios de control– podremos relajar algunas de las presiones de la selección natural, expandiendo en consecuencia la posibilidad de que estas poblaciones respondan a la selección artificial. De hecho, las poblaciones de ganado de distintas especies han venido respondiendo positivamente a la selección realizada en los últimos años, con resultados a menudo espectaculares.

Selección artificial. III. Caracteres múltiples

Cuando analizamos los métodos de evaluación de reproductores, comentamos que los objetivos de selección en la mejora del ganado suelen ser múltiples. En dicho capítulo, se describieron los métodos de evaluación de reproductores mediante índices y modelos *BLUP* que permitían la evaluación y posterior selección de los candidatos a la reproducción de acuerdo con su valor económico global. En el capítulo que iniciamos vamos a describir de nuevo la respuesta esperada global mediante tales métodos y la respuesta correlacionada en cada uno de los objetivos de selección, así como dos nuevos métodos de selección multicarácter, como la selección *en tándem* y la selección por niveles de elección independientes, o más simplemente niveles independientes.

Como consecuencia de la selección simultánea de varios caracteres debemos esperar una reducción de la respuesta esperada a la selección de cada uno de ellos por separado. Si, por ejemplo, concedemos idéntica importancia a cada carácter, la intensidad de selección será proporcional al factor $(p)^{1/n}$, siendo n el número de caracteres y p la fracción total seleccionada.

Selección por índices

El índice de selección para varios caracteres, I_α , fue definido anteriormente como un predictor del valor genético agregado, H , es decir, del valor económico global del individuo. El índice tiene la propiedad de compensar la posible deficiencia del candidato a la selección en algún carácter por la superioridad en otro (u otros) de los objetivos de selección. Simbólicamente, podemos describirlos, respectivamente, como

$$\begin{aligned} I_\alpha &= b_1X_1 + b_2X_2 + \dots + b_mX_m \\ H &= a_1A_1 + a_2A_2 + \dots + a_nA_n \end{aligned}$$

Cabe recordar que la respuesta global a la selección se predecía a partir del conocimiento de la intensidad de selección y la desviación típica del índice, es decir,

$$R = i \sigma_I$$

La varianza del índice se calcula como sigue:

$$\begin{aligned} \sigma_I^2 &= b_1\sigma_{HX_1} + b_2\sigma_{HX_2} + \dots + b_m\sigma_{HX_m} \\ &= \mathbf{b}'\mathbf{Pb} \end{aligned}$$

Respuesta correlacionada en cada objetivo de selección

Supongamos que queremos predecir la respuesta del primer objetivo de selección, A_1 . Podemos hacer para ello una adaptación de la teoría de respuesta correlacionada.

Dado que la respuesta global se predice mediante $R = i \sigma_I$, la respuesta correlacionada en el primer objetivo de selección puede expresarse como

$$R_{A_1} = b_{A_1 I} i \sigma_I = (\text{cov}(A_1, I) / \sigma_I^2) i \sigma_I = i \text{cov}(A_1, I) / \sigma_I,$$

donde

$$\begin{aligned} \text{cov}(A_1, I) &= \text{cov}(A_1, (b_1 X_1 + b_2 X_2 + \dots + b_m X_m)) \\ &= \text{cov}(A_1, b_1 X_1) + \text{cov}(A_1, b_2 X_2) + \dots + \text{cov}(A_1, b_m X_m) \\ &= b_1 \sigma_{A_1 X_1} + b_2 \sigma_{A_1 X_2} + \dots + b_m \sigma_{A_1 X_m} \end{aligned}$$

Puede comprobarse fácilmente que cada una de las covarianzas de A_1, A_2, \dots, A_n con I corresponde, respectivamente, a los elementos del vector $\mathbf{b}'\mathbf{G}$.

Método en *tándem*

El método de selección *en tándem* consiste en hacer la selección de cada carácter de forma alternativa, es decir seleccionando un solo carácter en cada generación. El orden en que cada uno de los caracteres es sometido a selección y el número de generaciones (ciclos) sucesivos de selección aplicados a cada uno influyen de forma compleja en la respuesta global a la selección. La ventaja principal de este método radica en la simplicidad de su realización.

Selección por niveles independientes

La selección por niveles independientes supone la elección de aquellos candidatos que superen unos niveles mínimos prefijados para cada carácter, con independencia de que puedan ser muy superiores en relación con alguno de dichos caracteres. Los diferentes caracteres pueden aparecer secuencialmente en el curso del desarrollo del candidato, como por ejemplo el peso al nacimiento, al destete y al año en los bovinos de carne, o bien pueden tener valores económicos que en un determinado mercado obedecen a una cierta jerarquía, como la producción de leche, la calidad de la leche y el tipo en bovinos lecheros.

Es fácil intuir que la intensidad de selección óptima (es decir, aquella que maximiza el progreso genético global) a aplicar a cada carácter es difícil de determinar. Por ejemplo, si la proporción de animales a seleccionar es $p = 24\%$, con dos caracteres no correlacionados, podríamos elegir entre otras combinaciones el retener el 30 y 80% ó 80 y 30% ó 40 y 60% o 60 y 40% de los animales, respectivamente, para cada carácter. La intensidad de selección óptima será difícil de obtener, aunque a veces se recurre a relacionarla con su importancia económica. Cuando los caracteres no están correlacionados y la importancia relativa de uno de ellos es claramente superior a la de los otros, Young (1961) ha demostrado para el caso de tres caracteres que la estrategia óptima es concentrar la presión en el carácter de más peso económico.

Eficiencia relativa de los tres métodos de selección

La comparación de los tres métodos de selección que acabamos de describir es compleja. Podemos hacerla atendiendo al coste y las posibilidades de aplicación, y asimismo al progreso genético máximo esperado, siendo esta última aproximación la que vamos a presentar. Cabe resaltar que hay que analizar cada situación, pues el óptimo depende de múltiples parámetros. No obstante, se pueden hacer algunas afirmaciones de carácter general, que siguiendo a Turner y Young (1969) resumimos como sigue.

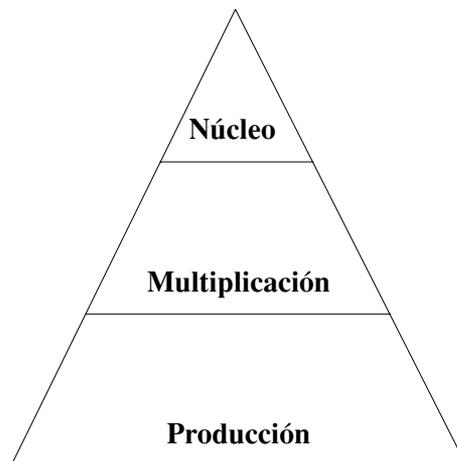
1. En cualquier circunstancia, la selección por índice no es nunca menos eficiente que la selección por niveles independientes, aunque en algunos casos tampoco es superior. Lo mismo puede argumentarse respecto al método de niveles independientes en relación con el método en tándem.
2. La superioridad de los índices sobre los otros dos métodos aumenta a medida que lo hace el número de objetivos de selección, aunque disminuye según aumentan las diferencias en la importancia relativa de los mismos. La superioridad de los índices sobre los niveles independientes disminuye según se incrementa la intensidad de selección, siendo independiente de dicha intensidad su superioridad sobre el método en tándem.
3. La superioridad de los niveles independientes sobre la selección por tándem aumenta con la intensidad de selección y el número de objetivos de selección, aunque disminuye según aumentan las diferencias en importancia económica relativa de los objetivos de selección.
4. La eficiencia relativa de los índices sobre los otros métodos se ve más afectada por la correlación fenotípica entre caracteres cuando los objetivos son de igual importancia, siendo la eficiencia relativa de los índices mayor cuando la correlación fenotípica es baja o negativa. El efecto de la correlación genética es solamente aparente cuando los objetivos son de importancia desigual, cambiando su influencia sobre la eficiencia relativa a medida que lo hacen otros parámetros.

Como conclusión general podemos decir que, de acuerdo con su eficiencia relativa, el orden de mérito sería: índices > niveles independientes > tándem, si bien las diferencias entre dos de ellos consecutivos podrían ser, según los casos, despreciables o incluso nulas.

En la aplicación práctica de la selección, suelen darse simultáneamente dos o más de estos métodos. Por ejemplo, la selección del ganado porcino se realiza atendiendo al valor de un índice para la mejora de caracteres de producción y de calidad de la canal, pero también se exige que los animales presenten un número mínimo de tetinas y, según las razas, que no sean sensibles al estrés. Asimismo, son objeto de desecho aquellos animales que presentan problemas de patas, las hembras que no quedan fecundadas tras dos ciclos estrales o los machos que tienen una libido deficiente o una calidad espermática insuficiente. Por otra parte, debemos traer a colación el distinto énfasis puesto en los caracteres a lo largo del tiempo, como por ejemplo en la prolificidad porcina, que sólo ha sido objeto de selección en los últimos años.

Difusión del progreso genético

Se ha venido suponiendo hasta ahora que todos los individuos de una población eran potenciales candidatos a la selección, de entre los cuales se elegía a los mejor evaluados. Ello requiere el control productivo y genealógico de todos los individuos de la población, así como la informatización y almacenamiento de la ingente información generada. Por razones prácticas, no obstante, esto a menudo es imposible, de manera que la selección se centra en una pequeña parte de la población, en la cual se toman decisiones de selección a partir de predicciones de los valores mejorantes altamente fiables. La consecuencia de esta práctica es la estructuración de las poblaciones en estratos que se configuran de forma piramidal. Una pirámide típica consta de tres estratos: en la cima se encuentra el núcleo, a continuación existe un estrato de multiplicación, y en la base se sitúa el estrato de producción.



Esquemas de mejora de núcleo cerrado

La mejora genética se lleva a cabo únicamente en el núcleo y posteriormente se difunde en sentido descendente. Este planteamiento, conveniente desde el punto de vista de la precisión en la evaluación de los animales y, por lo tanto, de la respuesta a la selección, tiene su coste, que se mide en términos de **retraso** o **desfase genético**, es decir, de diferencia entre el nivel genético de un estrato y el nivel del estrato inmediatamente inferior.

Para calcular el retraso genético vamos a centrarnos de momento en dos estratos: el núcleo y el de multiplicación. Esencialmente, se trata de resolver un problema de migración. Supongamos que cada generación el núcleo de selección experimenta un progreso genético constante igual a ΔG , y que una proporción m (entre 0 y 1) de los individuos del estrato de multiplicación son sustituidos por individuos procedentes del núcleo. Al principio podemos suponer que tanto el núcleo como el estrato de multiplicación son mues-

tras aleatorias de la población no seleccionada y, por tanto, de idéntico valor genético medio, que por simplificar designaremos como 0. La evolución del valor genético medio en sucesivas generaciones es como sigue:

Generación	Valor genético medio en cada generación (A_i)	
	Núcleo	Estrato de multiplicación
0	0	$A_0 = 0$
1	ΔG	$A_1 = 0$
2	$2\Delta G$	$A_2 = (1 - m)A_1 + m\Delta G = m\Delta G$
3	$3\Delta G$	$A_3 = (1 - m)A_2 + m2\Delta G = m(1 - m)\Delta G + 2m\Delta G$ $= \Delta G[m + m + m(1 - m)]$
4	$4\Delta G$	$A_4 = \Delta G[m + m + m(1 - m) + m + m(1 - m) + m(1 - m)^2]$ $= \Delta Gm[1 + 1 + (1 - m) + 1 + (1 - m) + (1 - m)^2]$

Se aprecia que en los sumandos del corchete podemos distinguir tres progresiones geométricas de razón $(1 - m)$, la primera constituida por un término (1) , la segunda por dos $[1 + (1 - m)]$ y la tercera por tres $[1 + (1 - m) + (1 - m)^2]$. Si generalizamos, A_n será la suma de los 1, 2, 3, ..., $n-1$ primeros términos de $(n-1)$ progresiones geométricas idénticas de razón $(1 - m)$. Recordando que la suma de los n términos de una progresión geométrica de razón p es $S_n = [(1 - p^{n+1})/(1 - p)]$, tras una sencilla manipulación algebraica, se llega a la siguiente expresión:

$$A_n = \Delta G\{n - [1 - (1 - m)^n]/m\}$$

El retraso genético con respecto al núcleo de selección es, entonces,

$$n\Delta G - A_n = \Delta G [1 - (1 - m)^n]/m$$

que tiende a $\Delta G/m$ a medida que n aumenta indefinidamente. Como podría esperarse, asintóticamente, el retraso genético es inversamente proporcional a la tasa de migración. En el caso concreto de que el núcleo proporcione los machos reproductores al estrato de multiplicación, $m = 0,5$, por lo cual el retraso genético es igual al doble de la respuesta obtenida en una generación de selección.

El planteamiento que acabamos de hacer es válido para cualesquiera dos estratos contiguos, siempre que el flujo de genes se produzca en sentido descendente. El retraso genético del estrato de producción con respecto al núcleo será el correspondiente a acumular los dos retrasos, núcleo-multiplicación y multiplicación-producción.

Esquemas de mejora de núcleo abierto

A pesar de las diferencias entre el nivel genético medio de los estratos, todavía existe una variabilidad genética importante entre los animales que hace aconsejable transferir algunos animales hacia estratos superiores, es decir, abrir el núcleo. Los cálculos referentes al retraso genético en una situación de núcleo abierto son más complejos que los expuestos anteriormente y no los describiremos en estas notas. No obstante, cabe comentar que en los esquemas de núcleo abierto de ovinos y de ganado vacuno de carne

es posible aumentar la respuesta a la selección (entre un 10 y un 15%), mientras que se reduce a la mitad la tasa de consanguinidad, en relación con un esquema de núcleo cerrado del mismo tamaño global (James, 1977).

Dadas las ventajas expuestas, cabe preguntarse por qué se siguen manteniendo esquemas de mejora de núcleo cerrado en cerdos y aves a pesar de sus desventajas genéticas. Las razones son principalmente de tipo sanitario, pues la transferencia de animales entre estratos comporta unos riesgos muy elevados de introducir una enfermedad en el núcleo, que podría llevar incluso a la desaparición de la población.

Como comentario final, se debe señalar que lo expuesto hasta aquí corresponde a pirámides integradas por animales de una misma raza o población. No obstante, es bastante frecuente que el estrato de producción esté constituido por animales cruzados, si bien la teoría de cruzamientos la veremos en un próximo capítulo.

Sistemas de apareamiento. I. Reproducción consanguínea

En la producción animal se trabaja con poblaciones de efectivos variables, tales como razas, rebaños o líneas familiares. La población experimenta una renovación constante al desaparecer algunos individuos y ser sustituidos por otros nuevos. Si la población fuese muy numerosa, en ausencia de selección, la constitución genética permanecería inalterada al pasar de una generación a la siguiente si los apareamientos se produjeran al azar. Sin embargo, la reproducción panmíctica y la ausencia de selección no suelen ser habituales, pues se presentan variaciones que podemos agrupar en dos tipos:

1. Modificación de las frecuencias genotípicas sin que se produzca un cambio simultáneo en la frecuencia de los genes. Ello puede hacerse aplicando un determinado sistema de apareamiento distinto a realizar las cubriciones al azar: los animales se pueden aparear según su parecido genético (parentesco) o fenotípico.
2. Las frecuencias génicas se modifican por medio de la selección natural o artificial, lo cual determina también un cambio de las frecuencias cigóticas (genotípicas). La selección natural se basa en la viabilidad del individuo y en su capacidad para reproducirse en las condiciones ambientales predominantes. La selección artificial, como hemos visto, es aplicada por el hombre y puede orientarse hacia la mejora de cualquier carácter.

La mejora genética implica una alteración genética de la población de acuerdo con una estrategia prefijada, de manera que la descendencia pueda cubrir mejor las exigencias de un sistema de producción y un mercado determinados. La selección, ya descrita, y el sistema de apareamiento serán los pilares fundamentales de cualquier programa de mejora.

Especies, razas, estirpes, líneas y familias

Como es bien sabido, tanto la producción como la mejora genética del ganado se llevan a cabo mediante animales de distintas poblaciones. Con el fin de definir críticamente los tipos de poblaciones con que normalmente se trabaja, vamos a seguir la aproximación sistemática que hacen Johansson y Rendel (1971).

El reino animal se divide en **especies** sobre la base de los caracteres morfológicos y fisiológicos. El criterio fundamental, no obstante, es la discontinuidad reproductiva, es decir, el hecho de que los animales de dos especies distintas no se reproduzcan o no produzcan descendencia fértil cuando se aparean.

Esta discontinuidad varía en sus causas y en su grado de manifestación. Los asnos y las yeguas pueden aparearse y producir descendientes híbridos enteramente viables, mulas y mulos, aunque generalmente estériles. El yak tibetano y el ganado vacuno también se aparean, dando en una primera generación hembras fértiles y machos estériles. Entre el cebú y el ganado vacuno no existe, en cambio, discontinuidad reproductiva; ello ha permitido la obtención de razas nuevas como la Santa Gertrudis. Debido a sus dife-

rencias morfológicas, el cebú y el ganado vacuno pueden considerarse subespecies de una misma especie. De todo lo expuesto se deduce que, en muchos casos, la línea divisoria entre las especies es confusa.

Dentro de una especie pueden distinguirse diferentes **razas**. Desde un punto de vista biológico, podemos definir la raza como una población de animales que difieren de los de otras poblaciones de la misma especie en determinados caracteres definidos genéticamente. Los rasgos característicos de una raza pueden ser cualitativos –por ejemplo, tipo de pelo, color del pelo o cuernos– o cuantitativos, como tamaño, tipo corporal, producción de leche o contenido graso de la leche.

Es posible producir una raza homocigota para uno o varios caracteres cualitativos, aunque como norma esto no servirá para distinguirla de las restantes razas, ya que diversas razas pueden poseer idénticas características externas. Además, estos caracteres raciales cualitativos poseen una importancia escasa o nula, al menos en el ganado. Los aspectos de la producción, que determinan la importancia económica de la raza, muestran una variación continua, no siendo posible en muchos casos el establecimiento de una línea divisoria clara entre las razas, aun cuando los promedios raciales muestren unas diferencias bastante amplias. En consecuencia, en el sentido mendeliano del término, no existen razas puras de animales domésticos y, además, no parece ser posible producir una.

Cuando utilizamos el término “**raza pura**”, éste se refiere a animales que han sido o pueden ser registrados en el libro genealógico de la raza. Estos animales de raza pura suelen constituir un grupo selecto destinado a producir animales de reposición. Los requisitos de ingreso de un animal en el libro genealógico de la raza varían según la época y el lugar. El concepto de raza es, por lo tanto, más convencional que biológico. No obstante, en la práctica la división de los animales en razas parece justificada debido a que las poblaciones que las integran se han especializado para fines diferentes y para distintas condiciones locales.

Se ha venido denominando reproducción en raza pura la realizada entre los animales de la misma raza, y cruzamiento la reproducción entre animales de distintas razas o de estirpes o líneas de la misma raza. Una raza, frecuentemente, se divide en **estirpes** que difieren debido a su aislamiento reproductivo y/o al hecho de haberse seleccionado respecto a distintos objetivos. Podemos encontrar un ejemplo claro en la raza Landrace, cuyo grado de desarrollo muscular depende de la población de que se trate; un Landrace materno de origen alemán será generalmente más musculado que otro de origen francés o inglés.

Una **línea** consta de un grupo de animales que, como consecuencia de la consanguinidad, se encuentran más íntimamente emparentados entre sí que los restantes individuos de la estirpe o raza. Si una línea alcanza el coeficiente de consanguinidad de 0,375 como mínimo, correspondiente al apareamiento entre hermanos durante dos generaciones, entonces recibe el nombre de línea consanguínea. El término **familia** servirá para designar al conjunto de hermanos carnales o medios hermanos, siendo idéntico el parentesco entre los miembros de la misma familia.

Sistemas de apareamiento

El objetivo que persigue un determinado sistema de apareamiento puede ser el aumento o la disminución del grado de heterocigosis de la descendencia, en comparación con el de los padres, o mantener la homocigosis inalterada en algunos casos. Podemos hacer la siguiente clasificación: apareamiento al azar, apareamiento para aumentar la homocigosis y apareamiento para aumentar la heterocigosis.

1. **Apareamiento al azar** Tiene como objetivo mantener inalterada la constitución genética de la población, como por ejemplo en estudios experimentales, en los que se precisa mantener una población testigo para comparar o medir el efecto de los restantes sistemas de apareamiento o para un experimento de selección.

2. **Apareamiento para aumentar la homocigosis** Supone a menudo efectuar los apareamientos entre individuos de la misma raza. En este apartado, distinguiremos dos prácticas distintas: la consanguinidad y el apareamiento según el parecido fenotípico.

Reproducción consanguínea

Puede definirse como el apareamiento realizado entre animales que están más estrechamente emparentados entre sí que el promedio de todos los individuos que componen la población. La reproducción en poblaciones de efectivo reducido, como es a menudo el caso de poblaciones sometidas a selección, determina automáticamente un cierto grado de consanguinidad. Ésta se mide a través del coeficiente de consanguinidad, F , es decir, la probabilidad de que los dos alelos de un *locus* que porta un individuo sean idénticos por descendencia. Desde un punto de vista práctico, el apareamiento de animales que estén emparentados menos estrechamente que los primos ejerce una influencia tan ligera sobre el grado de consanguinidad que puede despreciarse.

Por otra parte, la consanguinidad puede practicarse de forma sistemática y regular a lo largo de varias generaciones. Podemos distinguir tres tipos principales de consanguinidad sistemática:

1. Consanguinidad estrecha. Es el apareamiento entre hermanos carnales o entre padres e hijos, con el fin de obtener líneas consanguíneas con un grado de homocigosis relativamente alto.
2. Formación de estirpes. Constituye una forma de consanguinidad mucho más suave; a largo plazo, origina un aumento de la homocigosis dentro de las estirpes.
3. Consanguinidad colateral. Se lleva a cabo dentro de una línea de antepasados con la finalidad de aumentar la proporción con que un determinado antepasado masculino o femenino interviene en la constitución genética de la descendencia. El procedimiento más intensivo de consanguinidad colateral consiste en el cruzamiento retrógrado con el mismo antepasado durante varias generaciones sucesivas. No obstante, generalmente se utilizan formas más suaves, como son el apareamiento de una hembra con un abuelo o un tío que poseerá la mitad de los genes de sus abuelos.

Apareamiento según el parecido fenotípico

En su versión más pura, no tiene en cuenta el grado de parentesco entre los individuos; es un apareamiento de “igual con igual”, o asociativo positivo, que mantiene la media inalterada, aunque aumenta la varianza. En la reproducción consanguínea se reúnen los genes idénticos, con independencia de su efecto; cuando apareamos individuos semejantes externamente, se reúnen los genes que tienen un efecto similar con independencia de que sea común o no su origen. En consecuencia, este método de apareamiento ejerce una influencia sobre el grado de homocigosis mucho menor que la consanguinidad. También cabe la posibilidad de que se realice la reproducción de “desigual con desigual” y se obtenga una forma intermedia, aunque este procedimiento conduce, natural-

mente, al incremento de heterocigotos. Se habla en este caso de apareamiento asociativo negativo, que tiende a hacer disminuir la varianza.

3. Apareamiento para aumentar la heterocigosis

El **cruzamiento** es lo opuesto a la reproducción en consanguinidad, ya que el parentesco de los animales que se aparean es menos estrecho que el parentesco medio de las poblaciones a que pertenecen. El cruzamiento puede realizarse entre razas o entre estirpes o líneas de la misma raza. Por el momento no entraremos en su análisis, pues será objeto del siguiente capítulo.

Efectos de la consanguinidad sobre los caracteres métricos

Los efectos perjudiciales de la consanguinidad sobre la tasa reproductiva y el vigor son bien conocidos por los mejoradores. En el apartado que iniciamos consideraremos en primer lugar los cambios que produce la endogamia sobre los caracteres métricos, es decir, la depresión consanguínea, para centrarnos posteriormente en los cambios sobre la varianza.

Depresión consanguínea

La consecuencia más importante de la consanguinidad es la reducción del valor fenotípico medio de los caracteres relacionados con la capacidad reproductiva o la eficiencia fisiológica, fenómeno que se conoce con el nombre de **depresión consanguínea** o depresión endogámica. En general, la consanguinidad tiende a reducir la eficacia biológica. Por ello, los caracteres relacionados con ésta, tales como el tamaño de la camada o la lactación en mamíferos muestran reducción bajo endogamia (tabla 6).

Tabla 6. Efectos de la consanguinidad sobre distintos caracteres productivos, expresado en porcentaje de la media por cada aumento del 1% de consanguinidad. (Adaptado de Pirchner, 1985.)

Especies y carácter	% de la media
Ganado vacuno	
Producción de leche	-0,50
Tasa de concepción (M)	-1,26
Tasa de concepción (D)	-0,95
Viabilidad (D)	-0,45
Peso al año	-0,16
Ganancia posdestete	-0,20
Epermatozoides por eyaculado	-0,50
Ganado porcino	
Nacidos vivos (D)	-0,28
Nacidos vivos (M)	-0,70
Lechones destetados (D)	-0,73
Lechones destetados (M)	-0,46
Ganancia media diaria	-0,42
Espesor del tocino dorsal	-0,20

Especies y carácter	% de la media
Ganado ovino	
Peso al año	-0,22
Producción de leche	-0,26
Peso del vellón	-0,70
Aves	
Incubabilidad (D)	-0,36
Incubabilidad (M)	-0,10
Viabilidad (D)	-0,23
Producción de huevos (%)	-0,38
Tamaño del huevo	0
Tamaño corporal	-0,20

M, consanguinidad de la madre; D, consanguinidad de la descendencia.

Para obtener una explicación formal, supongamos una “población total” constituida por un conjunto de líneas derivadas de una población base. Además, por el momento ignoraremos los efectos que pueda causar la selección sobre la dispersión de las frecuencias génicas. Es sabido que las frecuencias génicas de la población en su conjunto no cambian bajo endogamia, pero sí lo hacen las frecuencias genotípicas. La consanguinidad origina un incremento de las frecuencias de los homocigotos con la consiguiente disminución de la proporción de heterocigotos. Por ello, los cambios en la media deben estar relacionados con esta circunstancia.

Consideremos ahora una población subdividida en muchas líneas con un coeficiente de consanguinidad F . Las frecuencias genotípicas, los valores genotípicos y la media de la población son como sigue:

Genotipo	Frecuencia	Valor	Frecuencia \times valor
A_1A_1	$\bar{p}^2 + \bar{p}\bar{q}F$	$+a$	$\bar{p}^2a + \bar{p}\bar{q}aF$
A_1A_2	$2\bar{p}\bar{q} - 2\bar{p}\bar{q}F$	d	$2\bar{p}\bar{q}a - 2\bar{p}\bar{q}dF$
A_2A_2	$\bar{q}^2 + \bar{p}\bar{q}F$	$-a$	$-\bar{q}^2a - \bar{p}\bar{q}aF$

Si sumamos para todos los genotipos obtenemos el valor medio bajo consanguinidad,

$$\begin{aligned}
 M_F &= a(\bar{p} - \bar{q}) + 2d\bar{p}\bar{q}(1 - F) \\
 &= M_0 - 2d\bar{p}\bar{q}F,
 \end{aligned}$$

donde M_0 es la media de la población antes de la consanguinidad. El cambio de media debido a la endogamia es, por lo tanto $-2d\bar{p}\bar{q}F$. Es decir, un *locus* cooperará al cambio de la media bajo endogamia únicamente si d no es cero, o lo que es lo mismo, si el valor del heterocigoto difiere del intermedio entre los dos homocigotos. Esta conclusión, demostrada sólo para el caso dialélico, es extensible asimismo a *loci* multialélicos. La magnitud del cambio de la media depende también de las frecuencias génicas, de manera que los genes con frecuencias intermedias influyen más en el cambio de la media que los genes con frecuencias extremas.

Considerando ahora todos los *loci* que afectan al carácter, suponiendo que todos los *loci* afecten aditivamente, la media de la población viene dada por la suma de las aportaciones de todos los *loci* que afectan al carácter:

$$\begin{aligned} M_F &= \Sigma a(\bar{p} - \bar{q}) + 2\Sigma d\bar{p}\bar{q}(1 - F) \\ &= M_0 - 2\Sigma d\bar{p}\bar{q}F, \end{aligned}$$

siendo el cambio de la media bajo consanguinidad de $-2\Sigma d\bar{p}\bar{q}F$. Para que se produzca cambio de la media es necesario que se dé dominancia direccional, esto es, que los genes que incrementan el valor del carácter sean dominantes sobre los alelos que lo reducen.

Los datos experimentales corroboran la hipótesis de acción aditiva, pues la depresión consanguínea tiende a ser lineal con respecto a F , lo cual puede interpretarse en el sentido de que la interacción epistática entre *loci* no es de gran importancia. Esta observación, no obstante, es válida solamente en las primeras generaciones, ya que como consecuencia del proceso dispersivo se perderán algunas líneas. Otras líneas, en cambio, podrán presentar incluso una superioridad respecto a los caracteres medidos. La depresión consanguínea, por lo tanto, debe entenderse en términos medios, aunque algunas líneas –raras– puedan ser incluso superiores a la población base, como se ha observado en ratones. A efectos prácticos, no obstante, el efecto de la consanguinidad debe considerarse perjudicial.

Cambios en la varianza bajo consanguinidad

Debido a que los componentes genéticos de la varianza disminuyen a medida que las frecuencias génicas tienden hacia los extremos, la varianza genética dentro de las líneas disminuye. Ello conduce a una redistribución de la varianza genética: mientras que el componente de varianza entre líneas aumenta, el componente dentro de líneas disminuye. En otras palabras, la consanguinidad lleva a una diferenciación genética entre líneas y a la uniformidad genética dentro de líneas.

En realidad no se ha logrado todavía un tratamiento teórico completo de la redistribución de la varianza. En este apartado no intentaremos hacer más que una descripción breve de los aspectos principales, y para ello tendremos que hacer algunas simplificaciones. Lo que sigue va a estar referido a la varianza debida a genes aditivos, no aplicándose estrictamente a genes que presentan dominancia. Las conclusiones que extraeremos, no obstante, nos indicarán el efecto general de la endogamia sobre la varianza, en el caso de endogamia lenta, no siendo aplicable a los casos de apareamientos de hermanos.

Consideremos en primer lugar un solo *locus*. Cuando no hay dominancia, la varianza genotípica en la población base es toda aditiva y puede expresarse como

$$V_G = 2p_0q_0a^2$$

La varianza dentro de cualquier línea es, asimismo,

$$V_G = 2pqa^2$$

donde p y q son las frecuencias génicas de esa línea. La varianza media dentro de las líneas, V_{Gd} , resulta

$$V_{Gd} = 2\bar{p}q a^2$$

donde $\bar{p}q$ es el valor medio de pq en todas las líneas. Pero $2\bar{p}q$ es la frecuencia media de heterocigotos en la población total, que en términos de las frecuencias de la población base puede escribirse como $2p_0q_0(1-F)$. Por lo tanto,

$$\begin{aligned} V_{Gd} &= 2p_0q_0 a^2(1-F) \\ &= V_G(1-F) \end{aligned}$$

siendo esto válido cuando se suman las varianzas correspondientes a todos los *loci*. Por lo tanto, cuando F se aproxima a la unidad, la varianza dentro de líneas se aproxima a cero.

Consideremos ahora la varianza entre líneas, V_{Ge} . Se trata de la varianza de las medias verdaderas de las líneas, que se estimaría en un análisis de varianza como componente entre líneas. Para un *locus* sin dominancia, la media de la población sería

$$\begin{aligned} M &= a(p-q) \\ &= a(1-2q) \end{aligned}$$

Por lo tanto, la varianza entre líneas resultaría

$$\begin{aligned} \sigma_M^2 &= \sigma^2(2aq) \\ &= 4a^2\sigma_q^2 \\ &= 4a^2p_0q_0F \\ &= 2FV_G \end{aligned}$$

siendo esto igualmente válido cuando se suman los efectos de todos los *loci*. La varianza genética queda, por lo tanto, descompuesta como sigue:

Entre líneas	$2FV_G$
Dentro de líneas	$V_G(1-F)$
Total	$V_G(1+F)$

La varianza genética total en el conjunto de la población es la suma del componente dentro de líneas y entre líneas, siendo igual a $(1+F)$ veces la varianza genética original. De esta forma, cuando la consanguinidad es completa, $F = 1$, la varianza genética de la población total se dobla y toda ella aparece como componente entre líneas.

La heredabilidad esperada dentro de cualquier línea será $(1-F)V_G / [(1-F)V_G + V_E]$. Tras algunas deducciones algebraicas que se proponen al alumno como ejercicio, la heredabilidad resulta ser

$$h_t^2 = \frac{h_0^2(1-F_t)}{1-h_0^2F_t}$$

donde h_t^2 y F_t son la heredabilidad dentro de líneas y el coeficiente de consanguinidad en el tiempo t , siendo h_0^2 la heredabilidad de la población base. Es de esperar, por consiguiente, que la heredabilidad decline con la consanguinidad en una población peque-

ña. La fórmula solamente es aplicable a caracteres sin varianza no aditiva y en ausencia de selección. Cuando consideramos genes sometidos a cualquier grado de dominancia, los cambios de varianza bajo consanguinidad dependen de las frecuencias génicas iniciales, siendo imposible obtener una solución general expresada en términos de la varianza genética presente en la población base.

A efectos prácticos, para que los cambios de varianza sean demostrables, éstos deben ser sustanciales. La varianza genotípica raramente constituye la mayor parte de la varianza fenotípica. Por lo tanto, en relación a la varianza fenotípica original, los cambios esperados debidos a la consanguinidad habitualmente son bastante pequeños, por lo cual su contrastación estadística es difícil.

Sistemas de apareamiento. II. Cruzamiento

La diversidad genética existente entre las razas de ganado y sus cruces constituye una parte importante de la variación genética total, que puede usarse en la mejora de la productividad de los animales. A diferencia de la variación intra-raza, las diferencias existentes entre las razas y sus cruces respecto a eficiencia o adaptación, pueden explotarse con rapidez, precisión y gran flexibilidad para producir los genotipos de la descendencia, maternos y paternos, mejor adaptados a un sistema de manejo dado y a las preferencias del mercado. Sin embargo, dichas ganancias no son acumulativas, requiriéndose la selección dentro de las razas existentes para mantener una mejora continuada.

El cruzamiento implica aparear individuos menos emparentados entre sí, es decir con menos genes en común, que el promedio de la población de la que forman parte. Supone, por lo tanto, una desviación del apareamiento aleatorio. El efecto genético primario del cruzamiento es el aumento de la proporción de animales heterocigotos, en comparación con la proporción que existe en una población en la que los apareamientos se producen de forma aleatoria.

A lo largo de este capítulo examinaremos en detalle los efectos genéticos del cruzamiento y describiremos, asimismo, los efectos de éste sobre las medias de los caracteres cuantitativos. Discutiremos el cálculo de la heterosis (vigor híbrido o ventaja de los individuos cruzados), así como ciertas observaciones relativas a la heterosis estimada en las poblaciones de ganado. Finalmente, describiremos y compararemos mediante el modelo de Dickerson diversos sistemas de cruzamiento.

Efectos genéticos y fenotípicos del cruzamiento

El efecto probablemente más importante del cruzamiento, aunque no el único, es la **heterosis**. Dicho término fue acuñado por Shull en 1914 para describir el aumento del vigor que presentaban los individuos cruzados respecto a sus padres, con independencia de la causa que lo originase. A este respecto, las deducciones que siguen tratan de explicar formalmente el origen genético de la heterosis.

Consideremos dos poblaciones (razas o líneas) que se reproducen al azar y que en lo sucesivo denominaremos poblaciones paternas. Dichas poblaciones se cruzan entre sí para producir una F_1 o primera generación cruzada. Veamos, en primer lugar, los efectos de un solo *locus* con dos alelos cuyas frecuencias son p y q en una población y p' y q' en la otra. Sea y la diferencia de las frecuencias génicas entre las dos poblaciones, de manera que $y = p - p' = q' - q$. Las deducciones algebraicas se hacen más fáciles si se expresan las frecuencias génicas p' y q' de la segunda población como $(p - y)$ y $(q + y)$.

Asumamos que los valores genotípicos son iguales en ambas poblaciones y que se ajustan al modelo habitual para un *locus* dialélico. Se trata de encontrar la media de cada población paterna y el progenitor medio, así como la media de la F_1 . Las medias paternas M_{P_1} y M_{P_2} son, respectivamente,

$$M_{P_1} = a(p - q) + 2dpq$$

$$\begin{aligned}
 M_{P_2} &= a(p - y - q - y) + 2d(p - y)(q + y) \\
 &= a(p - q - 2y) + 2d(pq + y(p - q) - y^2)
 \end{aligned}$$

El progenitor medio o valor medio de los progenitores de las poblaciones paternas, M_{P_m} , es, por lo tanto,

$$\begin{aligned}
 M_{P_m} &= 1/2(M_{P_1} + M_{P_2}) \\
 &= a(p - q - y) + d(2pq + y(p - q) - y^2)
 \end{aligned}$$

Cuando se cruzan las dos poblaciones para producir la F_1 , se aparean animales tomados al azar de una población con otros tomados también al azar de la otra población. Esto es equivalente a tomar genes al azar de las dos poblaciones:

		A_1	A_2
		p	q
A_1	$p - y$	$p(p - y)$	$q(p - y)$
A_2	$q + y$	$p(q + y)$	$q(q + y)$

Así, la población queda constituida como sigue:

Genotipos	A_1A_1	A_1A_2	A_2A_2
Frecuencias	$p(p - y)$	$2pq + y(p - q)$	$q(q + y)$

El valor genotípico medio será, por lo tanto,

$$\begin{aligned}
 M_{F_1} &= a(p^2 - py - q^2 - qy) + d(2pq + y(p - q)) \\
 &= a(p - q - y) + d(2pq + y(p - q))
 \end{aligned}$$

Podemos medir la magnitud del grado de heterosis como desviación al progenitor medio, es decir, como desviación a la media de las dos poblaciones paternas. Así, la heterosis de la F_1 es como sigue:

$$h_{F_1} = M_{F_1} - M_P = dy^2$$

Por ello, la existencia de heterosis depende de la dominancia. Los *loci* que no presenten dominancia ($d = 0$), no originan heterosis. Asimismo, la heterosis producida por el cruzamiento de dos líneas o razas concretas depende del cuadrado de las diferencias de las frecuencias génicas entre las dos poblaciones. Si las poblaciones que se han cruzado no difieren en frecuencias génicas, no habrá heterosis; en contrapartida, ésta será máxima cuando un alelo esté fijado en una población y el otro alelo en la otra. Además, mediante una deducción algebraica sencilla, puede demostrarse que el valor y^2 es el aumento de la proporción de heterocigotos de la descendencia F_1 , en relación con la media de las razas paternas. Cuando hay más de dos alelos por *locus*, el cruzamiento produce también un aumento de la proporción de heterocigotos, por lo cual lo dicho hasta ahora sigue siendo válido.

Por otra parte, si imaginamos una población dividida en muchas líneas, y suponemos que tomamos pares de líneas al azar, la diferencia cuadrática media de la frecuencia génica entre líneas será igual al doble de la varianza de la frecuencia génica entre las líneas, es decir, $y^2 = 2\sigma_q^2$. Dado que $\sigma_q^2 = pqF$, $y^2 = 2pqF$ (la demostración se propone al

lector como ejercicio). Es decir, la magnitud media de heterosis en cruzamientos entre pares de líneas tomadas al azar es igual a la depresión consanguínea, pero de signo opuesto.

Consideremos los efectos conjuntos de todos los *loci* en que difieren las poblaciones paternas. Si los valores genotípicos de cada uno de los *loci* combinan aditivamente, podemos expresar la heterosis producida por los efectos conjuntos de todos los *loci* como suma de sus aportaciones individuales. Así, la heterosis de la F_1 es:

$$h_{F_1} = \sum dy^2$$

La heterosis, al igual que la depresión consanguínea, depende de la arquitectura genética del carácter: los caracteres determinados principalmente por *loci* en los que la acción génica es dominante o sobredominante, presentarán más heterosis que aquellos determinados por genes de acción predominantemente aditiva.

Por otra parte, Willham y Pollak (1985) han revisado teóricamente el papel que juega la epistasia como probable causa de heterosis. Concluyen que las deducciones derivadas de un modelo de *locus* simple con dos alelos sin epistasia entre *loci*, que permita la suma de todos los *loci*, son una teoría razonablemente robusta para describir las causas de la heterosis, dado que son relativamente pequeños los coeficientes de los efectos epistáticos.

En la vertiente biológica, no resulta fácil atribuir causas concretas a lo comentado. Se puede explicar la sobredominancia en el plano molecular. Si un *locus* codifica un enzima, los productos de los dos alelos (alozimas) podrían tener diferentes propiedades, tales como actividad enzimática, estabilidad térmica u óptimo respecto a diferentes condiciones ambientales como temperatura o pH. La mezcla de alozimas haría más versátil al heterocigoto que cualquiera de los dos homocigotos con alozimas simples; alternativamente, si los alozimas difiriesen en actividad, la actividad intermedia del heterocigoto sería más favorable que las altas o bajas actividades de los homocigotos.

Por otra parte, aunque no se trata de un fenómeno de carácter general, se ha observado que las líneas consanguíneas son más susceptibles que los animales cruzados a las fuentes de variación ambiental. Como explicación se ha sugerido la posibilidad de que la posesión de alelos diferentes en ciertos *loci* confiere a los animales cruzados una mayor “versatilidad bioquímica”, que los capacita para que ajusten sus mecanismos fisiológicos al ambiente, de manera que presenten el fenotipo óptimo a pesar de valores ambientales subóptimos, lo que se conoce como homeostasis. La menor capacidad homeostática de las líneas endogámicas podría entenderse como una manifestación de la depresión consanguínea. Podríamos suponer dominancia direccional, en el sentido de que los genes que incrementan la capacidad homeostática tenderían a ser dominantes sobre los alelos que la reducen.

Estimación del grado de heterosis

En la práctica, en una primera aproximación, la heterosis se estima como desviación de la media de la descendencia de cruzamientos recíprocos, M_{ij} , a la media de las razas paternas respecto al carácter evaluado. Habitualmente, se expresa como porcentaje:

$$\text{Heterosis (\%)} = \frac{(M_{AB} + M_{BA}) - (M_{AA} + M_{BB})}{(M_{AA} + M_{BB})} \times 100$$

El porcentaje de heterosis que acabamos de definir se denomina heterosis individual y cuantifica el vigor híbrido atribuible al aumento de la heterocigosidad de los individuos cuyos caracteres se miden. Algunos caracteres pueden presentar también

heterosis materna, es decir, superioridad debida a la heterocigosidad de las madres de los animales que se evalúan. Si, por ejemplo, las cerdas cruzadas tuvieran un comportamiento materno mejor que las de las raza pura, ésto podría reflejarse en porcentajes de supervivencia o crecimientos superiores de sus lechones. Si en el sistema de cruzamiento, tanto las madres como los descendientes fueran animales cruzados, podrían acumularse los dos tipos de efectos.

Antes de entrar a describir los sistemas específicos de cruzamiento, haremos algunas observaciones derivadas de los porcentajes de heterosis estimados en el ganado y en algunas especies experimentales. La presentación de datos que apoyen lo que vamos a comentar sería excesivamente prolija, por lo que nos limitaremos a exponer valores de heterosis en el ganado porcino (tabla 7).

1. El porcentaje de heterosis frecuentemente difiere entre cruces recíprocos. La causa más probable e importante es la influencia del ambiente materno. En el ganado vacuno, el peso al nacimiento, por ejemplo, se ve influido por la capacidad uterina de la madre. De igual manera, la producción lechera de ésta influye en el peso al destete.
2. El porcentaje de heterosis depende de las razas que intervengan en el cruzamiento. Como la heterosis depende de las diferencias de frecuencias génicas de los *loci* que determinan el carácter, es razonable esperar diferentes porcentajes de heterosis en diferentes cruzamientos, ya que las diferencias entre las razas surgen de la selección para diferentes objetivos en distintos ambientes y de cambios aleatorios acumulativos, originando ambos procesos cambios en las frecuencias génicas.
3. El porcentaje de heterosis observado depende del ambiente en el que se ha hecho la comparación. Puede haber una interacción genotipo-ambiente, como ha demostrado Orozco (1976), que haga que sea superior la heterosis en los ambientes más desfavorables.

Tabla 7. Porcentajes de heterosis en cruzamientos de razas porcinas. (Tomado de Pirchner, 1983.)

	Cruces simples	Cruces de tres razas
Madre	Raza pura	Cruzada
Descendencia	Cruzada	Cruzada
Caracteres maternos		
Número de nacidos vivos	102	108
Porcentaje de destetados	106	106
Número de destetados	108	116
Peso de la camada al nacimiento	104	110
Peso de la camada al destete	115	125
Caracteres del individuo		
Peso al destete del lechón	106	107
Ganancia media diaria	106	106
Peso a los cinco meses	110	110
Conversión del alimento	103	103
Piezas de primera (%)	100	100

4. Los porcentajes de heterosis difieren entre caracteres. Como la heterosis depende, como hemos dicho, de la arquitectura genética del carácter, será mayor en aquellos en los que predominen los efectos de dominancia. Tales caracteres son los relacionados con la supervivencia, vigor y eficacia reproductiva. En cambio, los caracteres poco importantes para la supervivencia y reproducción, como la tasa de crecimiento, el índice de conversión y el valor de la canal, han mostrado bajos porcentajes de heterosis, puesto que se trata de caracteres en los que hay un predominio de los efectos genéticos aditivos.

En general, los caracteres que presentan elevada heterosis se ven afectados fuertemente por la depresión endogámica, y tienden a ser de baja heredabilidad. En cambio, los caracteres altamente heredables tienden a manifestar una baja heterosis en cruzamiento, así como una baja depresión endogámica.

5. Finalmente, pueden acumularse las heterosis individual y materna en cada uno de los caracteres para dar mayores porcentajes de heterosis respecto la eficiencia total de la producción (véase la segunda columna de la tabla 7).

Sistemas de cruzamiento

Pueden cruzarse animales pertenecientes a distintas razas, estirpes o líneas. En cualquiera de los casos será aplicable lo que vamos a describir; no obstante, con el fin de simplificar la exposición, la referiremos al cruzamiento entre razas.

Cada una de las razas paternas de la descendencia cruzada transmite a ésta la mitad de la superioridad o inferioridad respecto a un carácter, que es atribuible principalmente a los efectos medios de esos genes. El cruzamiento permite al mejorador o ganadero combinar genes de diversas fuentes y crear combinaciones de genes y de caracteres que no existían en ninguna de las poblaciones parentales. La combinación de razas con el fin de maximizar el valor de la descendencia en relación con la productividad total, se conoce con el nombre de **complementariedad**. Dicha complementariedad requiere el uso de razas que contrarrestan parcialmente las deficiencias de la otra. Frecuentemente implica el uso de razas o líneas adaptadas al sistema de producción solamente como líneas maternas y de razas o líneas interesantes para el mercado (producto que se comercializa) como líneas paternas. Por ejemplo, en la producción de animales para carne, la línea materna puede seleccionarse respecto a caracteres reproductivos, mientras que la línea paterna se seleccionará para crecimiento y valor de la canal.

Evaluación de poblaciones

El diseño de programas para contrastar diferencias entre poblaciones respecto a caracteres productivos presenta, sin embargo, más problemas de los que pueda parecer a simple vista y normalmente requiere la evaluación de un gran número de individuos.

El requisito básico es evaluar todos los animales en un ambiente común o eliminando mediante el diseño estadístico todos los efectos ambientales. A la hora de decidir el número de animales a evaluar, lo más sencillo es especificar un valor aceptable del error típico de la media. La varianza de la media de una población en la que se evalúan n individuos de cada una de s familias de hermanos carnales viene dada por

$$V = \frac{1 + 0,5h^2 (n - 1)}{ns} V_p$$

Modelo de Dickerson

Dickerson (1969, 1973) ha expuesto una serie de conceptos y métodos que permiten predecir los efectos conjuntos de la complementariedad de razas y de la heterosis en sistemas de cruzamiento específicos. Estos métodos se resumen en un modelo matemático aditivo que describe las causas de que un grupo de animales de raza pura o cruzados tenga una media fenotípica concreta en relación con un carácter. El modelo asume que el fenotipo se ve sometido a influencias directas (I) e influencias maternas (M). La forma general de la ecuación es la siguiente:

$$\bar{y}_i = \mu + \sum P_j g_j^I + \sum R_k g_k^M + \sum S_{lm} h'_{lm} + \sum T_{np} h^M_{np}$$

donde

- \bar{y}_i es el valor fenotípico medio de los animales del grupo genético i;
 μ es el valor fenotípico promedio de todas las razas y cruces;
 g_j^I son las desviaciones a μ atribuibles al valor genético de los animales de la raza parental j para el carácter en cuestión, siendo P_j la proporción de genes del grupo i que aporta la raza j. Para cualquier grupo, $\sum P = 1$;
 g_k^M es la desviación a μ atribuible al valor genético respecto a efectos maternos de los animales de la raza k, siendo R_k la proporción de genes de las madres del grupo genético i aportado por la raza k. De nuevo $\sum R = 1$;
 h'_{lm} representa la desviación a μ debida a la heterosis de los descendientes del cruce de las razas l y m, siendo h^M_{np} la desviación atribuible a la heterosis de los efectos maternos de hembras procedentes de cruzar las razas n y p;
 S_{lm} es la heterocigosidad potencial del grupo de animales i, atribuibles a las razas l y m, siendo T_{np} la proporción de heterocigosidad potencial en las madres de los animales del grupo genético i que puede atribuirse a las razas n y p. El sumatorio de S y T no necesariamente debe ser la unidad, ni tampoco puede excederla.

Tanto el valor genético individual (g^I) como el materno (g^M), incluyen desviaciones debidas a efectos de genes no alélicos (epistasia), así como de la heterocigosidad media característica de la raza e interacciones de g^I con g^M .

La superioridad de una raza en caracteres controlados maternamente la identifica como probable línea materna. De forma similar, cuando observamos una raza superior en caracteres determinados en gran parte por el componente individual, estamos ante una prometedora raza paterna.

A través del modelo descrito vamos a analizar distintos cruzamientos que podemos agrupar en: a) cruzamientos con finalidad comercial: de dos, tres y cuatro razas, y cruzamiento rotativo; y b) cruzamientos con finalidad genética: de absorción y creación de razas sintéticas.

Cruzamiento de dos razas

Es el cruzamiento más simple y constituye la primera etapa de cruzamientos más complejos. Las aportaciones de los distintos efectos pueden resumirse en la siguiente ecuación:

$$\bar{y}_i = \mu + 1/2(g^I_A + g^I_B) + h'_{AB} + g^M_B \quad (A\sigma \times B\varphi)$$

que supone una superioridad sobre el progenitor medio de

$$\Delta\bar{y} = h'_{AB} + 1/2(g^M_B - g^M_A)$$

Este cruzamiento es útil cuando la raza materna está adaptada al manejo o a ciertas condiciones ambientales, aportando su rusticidad, fertilidad y aptitud materna. Podemos beneficiarnos de la complementariedad y la heterosis individual si cruzamos estas madres con machos de crecimientos elevados y de buena conformación y cualidades cárnicas. La descendencia se destina generalmente al abasto de carne. Tiene el inconveniente de necesitar un reaprovisionamiento continuo de hembras, bien por compra del exterior, bien manteniendo un rebaño especial que asegure la reposición.

Cruzamiento de tres razas

La media esperada de la descendencia cruzada ($C\sigma \times A.B\varphi$), puede predecirse mediante la siguiente expresión:

$$\bar{y} = \mu + \frac{1}{2}g'_C + \frac{1}{4}(g'_A + g'_B) + \frac{1}{2}(h'_{CA} + h'_{CB}) + \frac{1}{2}(g^M_A + g^M_B) + h^M_{AB}$$

La superioridad del cruzamiento de tres razas sobre la media ponderada de las mismas es como sigue:

$$\Delta\bar{y} = \frac{1}{2}(h'_{CA} + h'_{CB}) + h^M_{AB} + \frac{1}{2}(g^M_{A,B} - g^M_C)$$

Teniendo en cuenta premisas estrictamente genéticas, el cruzamiento de tres razas o líneas es óptimo, pues combina las ventajas potenciales de las heterosis individual y materna y la complementariedad. Las razas A y B se eligen en función de su valor respecto a caracteres maternos; los machos, como en el caso anterior, atendiendo a caracteres de crecimiento, eficiencia alimentaria y caracteres de la canal. Los problemas que plantea son tanto de manejo como económicos al tener la necesidad de mantener tres rebaños, a menos que se realice la reposición con animales procedentes del exterior, práctica habitual en el estrato de producción.

Cruzamiento de cuatro razas

Consiste en cruzar dos generaciones F_1 procedentes de cuatro razas paternas distintas ($C.D\sigma \times A.B\varphi$). El modelo que pasamos a exponer predice la media esperada en la descendencia:

$$\bar{y} = \mu + \frac{1}{4}(g'_A + g'_B + g'_C + g'_D) + \frac{1}{4}(h'_{CA} + h'_{CB} + h'_{DA} + h'_{DB}) + \frac{1}{2}(g^M_A + g^M_B) + h^M_{AB}$$

que supone una superioridad sobre la media de las cuatro razas parentales de

$$\Delta\bar{y} = \frac{1}{2}(h'_{CA} + h'_{CB} + h'_{DA} + h'_{DB}) + h^M_{AB} + \frac{1}{2}(g^M_{A,B} - g^M_{C,D})$$

El cruzamiento de cuatro razas utiliza completamente la heterosis individual, materna y paterna, junto con la superioridad de efectos maternos de las dos razas que se cruzaron para dar la línea materna y asimismo la posible complementariedad.

Cruzamiento rotativo

En el cruzamiento rotativo pueden estar implicadas dos, tres, cuatro o incluso más razas. Cuando intervienen dos razas, se denomina cruzamiento alternante. El cruzamiento rotativo de tres razas es el caso que vamos a analizar en detalle, ya que puede servir de modelo para otros cruzamientos rotativos.

Según se aprecia en la tabla 8, la primera generación es equivalente a la de un cruzamiento de dos razas; la segunda, a uno de tres razas. Sin embargo, en vez de sacrificar la descendencia cruzada triple, ésta se mantiene en la explotación y se cruza con la

Tabla 8. Cruzamiento rotativo de tres razas. (Adaptado de Hohenboken, 1985.)

Generación	Raza ♂	Raza o cruce de la madre	Efectos genéticos individuales			Efectos genéticos maternos			Heterosis individual	Heterosis materna
			g'_A	g'_C	g'_B	g^M_A	g^M_C	g^M_B	h^I	h^M
1	A	C	0,5	0,5	0	0	1,0	0	1,0	0
2	B	AC	0,25	0,25	0,5	0,5	0,5	0	1,0	1,0
3	C	B×AC	0,125	0,625	0,25	0,25	0,25	0,5	0,75	1,0
4	A	C×(B×AC)	0,563	0,312	0,125	0,125	0,625	0,25	0,875	0,75
5	B	A×C×(B×AC)	0,281	0,156	0,563	0,563	0,312	0,125	0,875	0,875
6	C	B×A×(C×(B×AC))	0,141	0,578	0,281	0,281	0,156	0,563	0,844	0,875
.
.
.
<i>n</i>	C	BACBAC...	0,143	0,571	0,286	0,286	0,143	0,571	0,857	0,857

raza con la que está más distantemente emparentada. La descendencia nacida de esta tercera generación es portadora de 5/8 de genes de la raza de su padre, 1/4 de la de su abuelo materno y 1/8 de la raza del macho del cruzamiento que dio la primera generación. Los coeficientes de g^I son 0,625, 0,25 y 0,125, respectivamente. Las madres son resultado de un cruzamiento de tres razas, con lo que los coeficientes de g^M son 0,25, 0,25 y 0,5, respectivamente.

Las madres tienen la máxima heterocigosidad potencial, por lo que el coeficiente de h^M es 1. La descendencia, en cambio, no goza de la máxima heterocigosidad. Cuando los genes de la raza C se emparejan con genes originados en las razas B (0,25) o A (0,125), potencialmente se puede producir heterosis; ahora bien, cuando los genes C se emparejan con otros genes C (probabilidad de 0,25), no hay ocasión para nueva heterocigosidad y heterosis. Por ello, el coeficiente de heterosis individual es 0,75, es decir, 1 (potencial) menos 0,25 (no alcanzada). La heterosis potencial materna de la cuarta generación será igual a la heterosis potencial individual de la tercera, y, en general, la heterosis potencial materna en la generación n será igual a la heterosis potencial individual de la generación $n-1$.

En un cruzamiento rotativo de tres razas, los coeficientes se estabilizan con bastante rapidez, pudiendo resumir la ecuación que describe la generación n en los siguientes términos:

La proporción de genes que cada raza aporta a la descendencia, y por lo tanto los coeficientes adecuados de g^I son 0,571 para la raza del padre de la descendencia, 0,286 para la del abuelo materno, y 0,143 para la otra raza que interviene en el cruzamiento. La proporción de genes de las madres, y por lo tanto los coeficientes de g^M , son los de la generación anterior. La proporción de heterosis máxima individual o materna que se alcanza es 0,857 ó 6/7.

La ganancia promedio de este sistema para n razas paternas viene predicha por la expresión siguiente:

$$\Delta\bar{y} = (2^n - 2)/(2^n - 1) [h^I_m + h^M_m]$$

donde h^I_m y h^M_m son las heterosis individual y materna promedio de las combinaciones parentales.

El cruzamiento rotativo de tres razas no aprovecha la gran proporción de heterosis individual o materna que se da en un cruzamiento de tres razas. Sin embargo, una vez que se ha estabilizado, tanto las madres como la descendencia son cruzadas, por lo cual se benefician de la heterosis. La pérdida de $1/7$ de heterosis sería pequeña para muchos caracteres, en comparación con el beneficio de tener una población enteramente cruzada.

El sistema presenta también problemas de manejo, ya que las generaciones se solapan. Las hembras del hato suelen ser de distintas edades y generaciones. Esto obliga a una permanente identificación de las hembras y un cuidadoso programa de apareamientos, que requieren medios materiales y humanos para llevarlos a cabo. Otro problema es la variabilidad fenotípica de la descendencia. Puede complicarse el manejo y también disminuir el valor económico de los animales vendidos, si el mercado prima la uniformidad de un determinado tipo de animales; pero probablemente sea la mayor desventaja el no poder utilizar la complementariedad de razas con la eficacia que permite el cruzamiento a tres vías.

El cruzamiento alternante, denominado popularmente con el término inglés *criss-cross*, es un sistema también eficiente, aunque sólo se alcancen $2/3$ de las heterosis potenciales máximas, tanto para la heterosis individual como para la materna.

Cabe señalar, por último, que los cruzamientos rotativos pueden combinarse con el uso de un macho terminal. En este caso, frecuente en el sistema de producción porcina de los Estados Unidos, una parte de la piara se dedicaría a la producción de hembras híbridas, mientras que la otra parte se cruzaría con un macho terminal, generalmente comprado al exterior, que podría ser de raza pura o incluso cruzado.

Cruzamiento de absorción

Es un método de mejora de razas que consiste en retrocruzamientos repetidos con una determinada raza paterna, tras un cruce F_1 inicial. Primariamente no va dirigido al aprovechamiento de la complementariedad y la heterosis, sino a la implantación de una nueva raza. La proporción de genes de la raza original en la generación de apareamiento n sería $(1/2)^n$, y por lo tanto $1 - (1/2)^n$ la proporción de genes de la raza a la que se dirige la absorción.

Presenta la ventaja de no tener que comprar un nuevo rebaño, particularmente cuando puede utilizarse la inseminación artificial. Como inconveniente, la necesidad de un tiempo relativamente largo para llegar a la sustitución: a menudo se considera que ésta se ha realizado cuando se llega a una proporción de $31/32$ genes de la raza que se quiere implantar, es decir, 5 generaciones.

Razas sintéticas

Generalmente, se desarrolla una nueva raza cuando se considera que las razas existentes no poseen la combinación óptima de caracteres para un determinado ambiente (sistema de producción y/o mercado). El desarrollo de una nueva raza requiere predecir, a partir de la información que se tiene de las razas puras o por intuición, las razas y las proporciones de genes que éstas deben aportar a la nueva población. Una vez hecho el cruzamiento, se convierte la descendencia en una población cerrada que posteriormente se somete a selección en el sentido deseado. A modo de ejemplo, pueden citarse las razas sintéticas de la siguiente tabla:

RAZA SINTÉTICA	RAZAS ORIGINALES
Maine Anjou	Durham × Mancelle
Santa Gertrudis	Brahman × Durham
Île de France	Dishley × Merinos
Columbia	Lincoln × Rambouillet
Corriedale	Merino × Lincoln
Targhee	Rambouillet × Corriedale × Lincoln
F.S.L.	Frisona × Sarda × Lacaune
Salz	Romanov × Rasa aragonesa
Lacombe	Berkshire × Chester × Landrace danés

El valor fenotípico promedio de la generación inicial (F_1) de la nueva raza integrada por dos razas originales puede predecirse según la siguiente ecuación:

$$\bar{y} = \mu + \frac{1}{2}(g'_A + g'_B) + \frac{1}{2}(g^M_A + g^M_B) + h^I_{AB}$$

En una segunda generación, habría que añadir los efectos de la heterosis materna generadas por el cruzamiento. No obstante, existen diversos aspectos potencialmente negativos:

- Cuando los individuos de la población cerrada se aparean entre sí, se perderá una proporción de la heterocigosidad inicial y, por lo tanto, de la heterosis inicial. En general, si N razas contribuyen en igual medida a la fundación de una población de la que se va a derivar una nueva raza, en posteriores generaciones de apareamiento se perderá aproximadamente $1/N$ de la heterocigosidad y heterosis iniciales (Wright, 1969).
- Si la población se ha formado con pocos machos y/o hembras y es, por lo tanto, de tamaño efectivo pequeño, la consanguinidad conducirá a una cierta fijación aleatoria de genes que originará depresión consanguínea en algunos caracteres.
- Si una proporción de la superioridad de cualquiera de las razas originales se ha debido a combinaciones epistáticas, en generaciones siguientes la media de la población cerrada regresará, en la medida en que se rompan por recombinación las combinaciones génicas favorables.

Como contrapartida de estos efectos negativos, pueden crearse también nuevas combinaciones favorables que no existieran en ninguna de las razas paternas. En cualquier caso, se crea nueva variabilidad genética, que en principio deberá aumentar la heredabilidad, aunque no se tienen estimaciones fiables de esta última que lo confirmen. Por otra parte, es posible que el progreso genético dependiente de las nuevas combinaciones génicas y de la nueva variabilidad, pueda compensar los efectos desfavorables de la pérdida de heterosis, depresión endogámica y pérdidas de superioridad epistática de la media de la población; no obstante, carecemos de nuevo de pruebas experimentales concluyentes acerca de la magnitud de estos efectos en las poblaciones de ganado.

Caracteres de umbral

Existen muchos caracteres de interés biológico o importancia económica que varían de una forma discontinua aunque no se heredan de una forma mendeliana sencilla. Ejemplos familiares son la susceptibilidad a ciertas enfermedades, como la displasia de cadera en los perros o la criptorquidia porcina, donde hay dos clases fenotípicas –afectada o no afectada– y tamaño de la camada de los mamíferos más grandes que sólo son portadores de un hijo o a veces de dos o tres. A primera vista parece que los caracteres de este tipo quedan fuera del campo de la genética cuantitativa; sin embargo, cuando se someten a análisis genético, se observa que se heredan de igual forma que los caracteres de variación continua.

Predisposición y umbral

La clave para comprender la herencia de tales caracteres se sustenta en la idea de que el carácter tiene una continuidad subyacente con un **umbral** que impone una discontinuidad en la expresión visible, tal como aparece en la figura 8. Cuando el valor de la variable subyacente está por debajo del nivel de umbral, el individuo presenta el fenotipo “normal”; cuando se sitúa por encima del umbral, el individuo está “afectado”.

Se ha denominado **predisposición** a la variable continua subyacente en el contexto de enfermedades humanas de caracteres de umbral, siendo este término el que usaremos aquí. La variación de la predisposición es de origen genético y ambiental; puede imaginarse como la concentración de alguna sustancia o la tasa de algún proceso de desarrollo o algo que pueda medirse en principio o estudiarse como un carácter métrico de forma ordinaria. Puede estar compuesto por varios procesos fisiológicos o de desarrollo distintos, aunque no sea necesario conocer cómo se combinan para dar la predisposición, o incluso conocer cuales son realmente.

Dos clases, un umbral

Consideremos caracteres que tienen sólo dos clases fenotípicas con un umbral sencillo que las separa. Nos referiremos a las dos clases como normal y afectada. En el plano fenotípico, o escala visible, los individuos pueden tener solamente dos valores posibles que podrían designarse 0 para el normal y 1 para el afectado. Sin embargo los grupos de individuos, tales como familias o la población global, pueden presentar cualquier valor, en forma de proporción o porcentaje de individuos que estén afectados. Esto se conoce como incidencia, aunque en el contexto epidemiológico se denomine habitualmente prevalencia. En lo que sigue, por razones de fidelidad a los trabajos originales, hablaremos de incidencia para referirnos tanto a la proporción de individuos que presentan una determinada enfermedad como a la proporción de individuos resistentes a un fármaco o tratamiento.

La incidencia describe con sencillez la población o grupo, pero la escala de porcentaje en que se expresa varía de acuerdo con la media. Por lo tanto, para análisis genéticos se debe convertir las incidencias en predisposiciones medias. Para hacer esta trans-

formación es necesario suponer que la predisposición se distribuye normalmente, siendo la unidad de predisposición la desviación típica σ . La predisposición media está relacionada entonces con la incidencia mediante la desviada normal x , que es la desviación del umbral a la media en unidades de desviación típica.

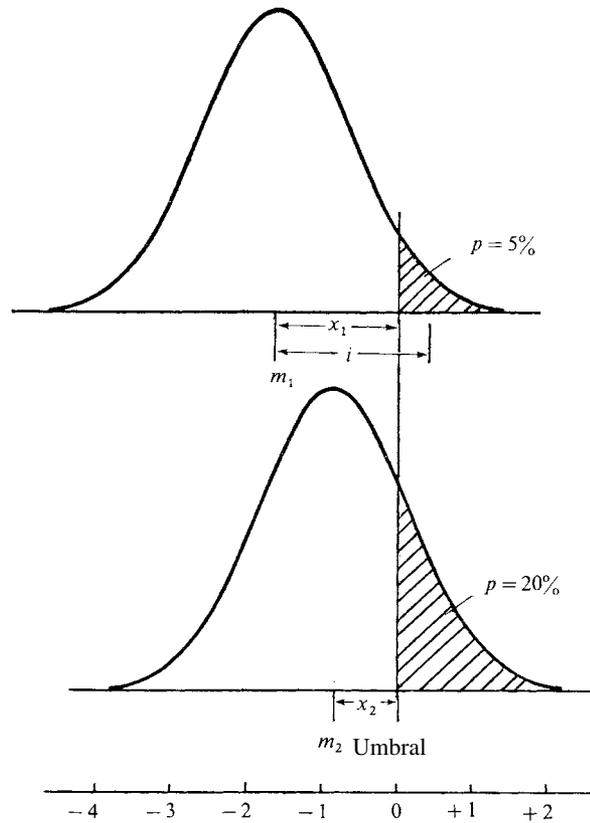


Figura 8. Dos poblaciones con diferentes incidencias, p , de un carácter de umbral y, consecuentemente, predisposiciones medias diferentes. (Tomado de Falconer, 1986.)

Comparación de medias

Consideremos dos poblaciones o grupos con diferentes incidencias, como se muestra en la figura 8. ¿En qué medida difieren en predisposición media? Al comparar diferentes grupos, podemos definir el umbral como fijo; es decir, está en el mismo nivel de predisposición en todos los grupos. La media de cualquier grupo se expresa entonces como una desviación en unidades σ al umbral. En otras palabras, se toma el umbral como origen o punto cero de la escala de predisposición. El grupo superior de la figura 8 tiene una incidencia $p_1 = 0,05$, que da $x_1 = 1,6\sigma_1$. La predisposición media es por ello $m_1 = -1,6\sigma_1$. (Debemos tener cuidado con los signos de las predisposiciones medias. Si el umbral está por encima de la media, ésta será negativa). De forma similar, el grupo inferior de la figura 8, con una incidencia de 0,20, tiene una media de $m_2 = -0,8\sigma_2$. Nótese que las medias se expresan en unidades de desviación típica de sus propias poblaciones. Sin embargo, no podemos avanzar en la comparación de las medias a menos que asumamos que las dos desviaciones típicas son iguales. Si puede aceptarse esto como un supuesto razonable, entonces $\sigma_1 = \sigma_2 = \sigma$, y $m_2 - m_1 = 0,8\sigma$; es decir, las medias de los dos grupos difieren en 0,8 desviaciones típicas de predisposición.

Heredabilidad de la predisposición

Supongamos que la distribución superior de la figura 8 representa una generación parental de la que se seleccionen como progenitores los individuos afectados. Cuando se aparean, producen una descendencia que presenta la distribución inferior. Conociendo la incidencia en la generación parental y en la descendencia, tenemos todo cuanto se necesita para calcular la regresión de la descendencia sobre el progenitor medio respecto a valores de predisposición, y a partir de ésta, la heredabilidad de la predisposición.

Consideremos en primer lugar la respuesta a la selección.¹ Es la diferencia media entre las generaciones paterna y filial, que fue dada anteriormente como $0,8\sigma$ asumiendo igualdad de las varianzas de las dos generaciones. (De hecho, las varianzas no serán del todo iguales, pero el pequeño error introducido lo despreciaremos de momento). Veamos ahora el diferencial de selección. Es la predisposición media de los individuos afectados en la generación paterna expresada como desviación a su media poblacional. La proporción de individuos utilizados como padres puede ser menor que la incidencia: en otras palabras, no todos los individuos afectados pueden utilizarse como progenitores. Pero en la medida en que todos los padres estén afectados, la media de los utilizados se espera que sea la misma que la media de todos los individuos afectados. La media de los individuos afectados en unidades de desviación típica es equivalente a la intensidad de selección, i , que corresponde a la incidencia como la proporción seleccionada. Con una incidencia de $p = 0,05$, la intensidad de selección es $i = 2,1$, y el diferencial de selección es, por lo tanto, $S = 2,1\sigma$.

La regresión de los valores de la descendencia sobre el progenitor medio es la razón de la respuesta al diferencial de selección, y $R/S = 0,8\sigma / 2,1\sigma = 0,38$. Finalmente, si no existe parecido ambiental entre descendiente y sus progenitores, la heredabilidad de la predisposición es igual a la regresión de la descendencia sobre los valores del progenitor medio.

El cálculo de la regresión y heredabilidad explicadas anteriormente con referencia a progenitores y descendientes puede aplicarse a cualquier clase de parentesco. Supongamos que la distribución superior de la figura 8 represente una población y que la distribución inferior representa cualquier clase de parientes, p.e., hermanos completos de individuos afectados. Entonces la regresión calculada es la de un individuo sobre su pariente. Como las varianzas de los dos grupos –la población y sus parientes– son aproximadamente iguales, la regresión es aproximadamente igual a la correlación. Por ello, con las incidencias de la figura 8 la correlación de hermanos completos en relación con la predisposición sería de 0,38. En ausencia de parecido debido a ambiente común y de dominancia, la heredabilidad se estimaría como dos veces la correlación de hermanos completos, ó 0,76.

Los cálculos explicados más arriba pueden resumirse en las siguientes fórmulas. La correlación de predisposición entre parientes de cualquier clase especificada viene dada por

$$t = \frac{(m_R - m_P)}{i} = \frac{(x_P - x_R)}{i}$$

donde los subíndices P y R hacen referencia a la población y a los parientes, respectivamente, m es la media como desviación al umbral, x es la desviada normal del umbral

1. El razonamiento que sigue cobra aparentemente más sentido en el contexto de selección para la resistencia a un agente infeccioso o a un fármaco.

a la media, e i es la desviación media de individuos afectados a la media de la población. Los signos corresponden a una escala que asigne una mayor predisposición a los individuos afectados que a los normales. La heredabilidad se obtiene, entonces, a partir de la correlación como

$$h^2 = t/r$$

siendo r el coeficiente de parentesco.

El error introducido al asumir igualdad de varianzas en los parientes de individuos afectados y en la población global conduce a que la correlación estimada anteriormente sea demasiado baja, en torno al 5-10 por ciento. La siguiente fórmula modificada (Reich *et al.*, 1972) tiene en cuenta la desigualdad de las varianzas

$$t = \frac{x - x_R \sqrt{1 - (x^2 - x_R^2) \left(1 - \frac{x}{i}\right)}}{1 + x_R^2 (1 - x)}$$

donde las x e i sin subíndices se refieren a la población, y x_R lo hace los parientes; además, se toma el signo de la raíz cuadrada que haga oscilar a t entre 0 y 1.

El cálculo de la heredabilidad de la predisposición se ha aplicado ampliamente en el estudio de la herencia de enfermedades humanas. Los datos se recogen preguntando a pacientes de una enfermedad concreta sobre el estado de sus parientes en relación con la enfermedad. La correlación respecto a la predisposición puede calcularse según hemos visto. La interpretación de la correlación en términos de heredabilidad, no obstante, se ve sometida a la incertidumbre sobre parecidos debidos al ambiente común.

Las estimaciones obtenidas para la heredabilidad de la predisposición asumiendo ausencia de parecido ambiental, oscilan desde el 85% de la esquizofrenia al 35 por ciento de las enfermedades cardíacas congénitas (Emery, 1976). El conocimiento de la heredabilidad es útil en el consejo genético para calcular los riesgos de recurrencia en familias, pues permite que se combine correctamente toda la información sobre la familia. El riesgo de recurrencia se define como la probabilidad de aparición de individuos afectados entre los descendientes de un determinado apareamiento.

Adecuación del modelo de predisposición

La definición de predisposición como variable continua distribuida normalmente implica que la distribución de la predisposición es unimodal. Si en realidad fuese bimodal o multimodal, los cálculos de correlaciones y heredabilidades serían, entonces, inválidos. Una distribución bimodal o trimodal puede surgir de dos maneras principalmente; la primera, si existe un gen sencillo cuyo efecto sobre la predisposición sea suficientemente grande en relación con la variación residual, y la segunda, si hubiera un factor ambiental de efecto grande en relación con el resto de la variación. Tal factor ambiental que afecte la predisposición a la enfermedad puede ser la exposición a un agente patógeno. Por ello, los análisis genéticos en términos de predisposición son válidos sólo si la predisposición es multifactorial, lo que significa que hay muchas causas de variación, todas de efectos relativamente pequeños, estando el control genético regulado por genes de más de uno o pocos *loci*.

No hay manera de conocer por adelantado si se cumplen los requisitos que validan el análisis genético, puesto que la predisposición no puede medirse para ver si su distribución es unimodal. No obstante, se puede ver si los resultados obtenidos son razo-

nables y consistentes. Por ejemplo, una heredabilidad que excediese el 100 por ciento sería obviamente inaceptable, y sugeriría un gen principal sencillo. También en ausencia de parecido debido al ambiente común, la heredabilidad estimada a partir de diferentes clases de parientes debería ser la misma. En general, los resultados obtenidos han sido razonablemente consistentes y no han dado ninguna razón de peso para dudar de la adecuación del modelo de predisposición.

Selección para caracteres de umbral

La aplicación de la selección a un carácter de umbral no implica las dificultades teóricas del análisis genético. Tiene una cierta importancia práctica en relación con la reducción de la incidencia de anomalías o defectos, y asimismo en relación con el cambio de la respuesta de los animales a los tratamientos, bien sea para aumentar o reducir la resistencia a los fármacos.

Consideraremos la selección individual aplicada a un carácter con dos clases visibles. La respuesta depende del diferencial de selección. Pero el diferencial de selección no depende primariamente de la proporción seleccionada, como en un carácter de variación continua, sino de la incidencia, puesto que dentro de una misma clase fenotípica no podemos discriminar entre los individuos que poseen una alta o baja predisposición. Los individuos seleccionados son una muestra aleatoria de la clase deseada, con independencia que seleccionemos todos los individuos de dicha clase o sólo una parte de ella. En el caso de la resistencia a fármacos o a alguna enfermedad, el seleccionar una proporción menor que la incidencia no tiene ninguna ventaja.

Si la proporción que hay que seleccionar es mayor que la incidencia (caso de animales que, p.e., han resistido al ataque de un determinado agente patógeno), nos veremos obligados a usar algunos individuos de la clase no deseada. Su predisposición promedio estará por debajo de la media de la población, por lo que el uso de individuos no deseados como progenitores aplicará algo de selección negativa. La media de la clase no deseada se calcula fácilmente como $-ip/(1 - p)$, donde i es la media de la clase deseada cuya incidencia es p .

Aunque no vamos a entrar en una descripción formal, se demuestra que en los caracteres de umbral la respuesta esperada es superior en la selección familiar que en la individual. Es decir, sería más eficaz la eliminación de todos los individuos de las familias en las que al menos uno de los miembros estuviera afectado, que la eliminación solamente de los individuos afectados. Esta superioridad de la selección familiar puede justificarse en el hecho de que entre los parientes próximos a los individuos afectados el componente hereditario de la predisposición deba ser mayor, de manera que al eliminarlos (mantenerlos en el caso de resistencia) el cambio de la media de la predisposición sea más acusado al serlo uno de sus componentes, el genético.

Referencias bibliográficas

- BULMER, M.G., 1971. "The effect of selection on genetic variability". *Am. Nat.* 105, 201-211.
- BULMER, M.G. 1976. "The effects of selection on genetic variability: a simulation study". *Genet. Res.*, 28, 101-117.
- CLAYTON, G.A.; MORRIS, J.A.; ROBERTSON A. 1957. "An experimental check on quantitative genetic theory. I. Short-term responses to selection". *J. Genet.*, 55, 131-151.
- CROW, J.F.; KIMURA M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, Nueva York.
- DICKERSON, G.E. 1969. "Experimental approaches in utilising breed resources". *Anim. Breed. Abstr.*, 37, 191-202.
- DICKERSON, G.E. 1973. "Inbreeding and heterosis in animals". *Proc. An. Breeding and Genetics Symp. in Honor of Dr. J.L. Lush*. ASAS y ADSA, Champaign, IL., 54-77.
- DICKERSON, G.E.; HAZEL L.N. 1944. "Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock". *J. Agric. Res.* 69 (12), 459-476.
- DUCROCQ, V. 1992. "Les bases de la génétique quantitative. Du modèle génétique au modèle statistique". *INRA Prod. Anim., hors série "Eléments de génétique quantitative et application aux populations animales"*, 75-81.
- DUDLEY, J.W. 1977. "76 Generations of selection for oil and protein percentage in maize". En: *Proceedings of the International Conference on Quantitative Genetics*, E. Pollak; O. KEMPTHORNE; T.B. BAILEY, Jr. (eds.) Iowa State University Press, Ames, pp. 459-473.
- EMERY, A.E.H. 1976. *Methodology in Medical Genetics*. Churchill Livingstone, Edinburgh.
- EMIK, L.O.; TERRIL, C.E. 1949. "Systematic procedures for calculating inbreeding coefficients". *J. Hered.* 40, 51-55.
- FALCONER, D.S. 1954. "Asymmetrical responses in selection experiments". *Symposium on Genetics of Population Structure*. International Union of Biological Science, Nápoles, Series B, N° 15, pp. 16-41.
- FALCONER, D.S. 1955. "Patterns of response in selection experiments with mice". *Cold Spring Harbor Symposium on Quantitative Biology*, 20, 178-96.
- FALCONER, D.S. 1983. *Introduction to Quantitative Genetics*. 2nd. ed., Longman, Londres y Nueva York.
- FALCONER, D.S. 1986. *Introducción a la genética cuantitativa*. 2ª ed., CECSA, México.
- FISHER, R.A. 1918. "The correlation between relatives on the supposition of Mendelian inheritance". *Trans. Roy. Soc. Edinburgh*, 52, 399-433.
- FISHER, R.A. 1936. "The use of multiple measurements in taxonomic problems". *Ann. Eugen.*, 7, 179-189.
- GROENEVELD, E.; KOVAC, M. 1990. "A generalized computing procedure for setting up and solving mixed linear models". *J. Dairy Sci.*, 73, 513-531.
- HALDANE, J.B.S. 1932. *The Causes of Evolution*. Longmans Green, Londres.
- HAZEL, L.N. 1943. "The genetic basis for constructing selection indexes". *Genetics*, 28, 476-490.

- HENDERSON, C.R. 1963. "Selection index and expected genetic advance". En: *Statistical Genetics and Plant Breeding*, W.D. HANSON; H.F. ROBINSON (eds.) National Academy of Sciences – National Research Council, Publ. N° 982, Washington, pp. 141-163.
- HENDERSON, C.R. 1973. "Sire evaluation and genetic trends". *Proc. An. Breeding and Genetics Symp. in Honor of Dr. J.L. Lush*. ASAS and ADSA, Champaign, IL., pp. 10-41.
- HENDERSON, C.R. 1974. "General flexibility of linear model techniques for sire evaluation". *J. Dairy Sci.*, 57, 963-972.
- HENDERSON, C.R. 1975. "Comparison of alternative sire evaluation methods". *J. Anim. Sci.*, 43, 760-770.
- HENDERSON, C.R.; QUAAS R.L. 1976. "Multi-trait selection using relatives' records". *J. Anim. Sci.*, 43, 1188-1197.
- HILL, W.G. 1982. "Rates of change in quantitative traits from fixation of new mutations". *Proc. Natnl. Acad. Sci.*, 79, 142-145.
- HOHENBOKEN, W.D. 1985. "Genetic structure of populations. 2. Matings among distantly related individuals". En: *General and Quantitative Genetics*, A.B. CHAPMAN (ed.) Elsevier, Amsterdam, pp. 251-274.
- JAMES, J.W. 1977. "Open nucleus breeding systems". *Anim. Prod.* 24, 287-305.
- JOHANNSEN, W. 1903. *Über erblichkeit in populationen und in reinen linien*. G. Fischer, Jena.
- JOHANSSON, I.; RENDEL, J. 1971. *Genética y mejora animal*. Acribia, Zaragoza.
- KEMPTHORNE, O. 1955. "The correlations between relatives in random mating populations". *Cold Spring Harbor Symposium in Quantitative Biology*, 20, 60-75.
- KENNEDY, B.W.; SORENSEN, D. 1988. Properties of mixed-model methods for prediction of genetic merit. En: *Proceedings of the 2nd International Conference on Quantitative Genetics*, B.S. WEIR, E.J. EISEN, M.M. GOODMAN, G. NAMKOONG (eds.) Sinauer, Sunderland, Massachusetts, pp. 91-103.
- LATTER, B.D.H.; ROBERTSON, A. 1960. "Experimental design in the estimation of heritability by regression methods". *Biometrics*, 16, 348-53.
- MACKAY, T.F.C. 1984. "Jumping genes meet abdominal bristles: Hybrid dysgenesis induced quantitative variation in *Drosophila melanogaster*". *Genet. Res.*, 44, 231-237.
- MATHER, K. 1943. "Polygenic inheritance and natural selection". *Biol. Rev.*, 18, 32-64.
- MINVIELLE, F. 1990. *Principes d'amélioration génétique des animaux domestiques*. INRA, París.
- NICHOLAS, F.W. 1980. "Size of population required for artificial selection". *Genet. Res.*, 35, 85-105.
- NILSSON-EHLE, H. 1909. "Kreuzungsuntersuchungen an Hafer und Weizen". *Lunds Univ. Aarskr. N. F. Alf., Ser. 2, Vol. 5, N° 2*, 1-122.
- OLLIVIER, L. 1981. *Éléments de génétique quantitative*. INRA y Masson, París.
- OROZCO, F. 1976. "A dynamic study of genotype-environment interaction with egg laying of *Tribolium castaneum*". *Heredity*, 37, 157-171.
- PEASE, A.H.R.; COOK, G.L.; GRIEG, M.; CUTHBERTSON, A. 1967. "Combined Testing". *Report DA 188 of the Industry Development Authority*, Hitchin, Herts, England.
- PIRCHNER, F. 1983. *Population Genetics in Animal Breeding*. Plenum Press, New York.
- PIRCHNER, F. 1985. "Genetic structure of populations. 1. Closed populations or matings among related individuals". En: *General and Quantitative Genetics*, A.B. CHAPMAN (ed.) Elsevier, Amsterdam, pp. 227-250.
- POLLAK, E.J. 1988. *Current Genetic Prediction Systems Used in the Beef Cattle Industry*. Mimeo.
- QUAAS, R.L. 1976. "Computing the diagonal elements and inverse of a large numerator relationship matrix". *Biometrics*, 32, 949-953.

- REICH, T.; JAMES, J.E.; MORRIS, C.A. 1972. "The use of multiple thresholds in determining the mode of transmission of semicontinuous traits". *Ann. Hum. Genet.*, 36, 163-184.
- RENDEL, J.M.; ROBERTSON, A. 1959. "Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle". *J. Genet.*, 50, 1-8.
- ROBERTSON, A. 1959a. "Experimental design in the evaluation of genetic parameters". *Biometrics*, 15, 219-26.
- ROBERTSON, A. 1959b. "The sampling variance of the genetic correlation coefficient". *Biometrics*, 15, 469-85.
- ROBERTSON, A. 1960. "A theory of limits in artificial selection". *Proc. Roy. Soc. London, B.*, 153, 234-249.
- SCHAEFFER, L.R., MAO, I.L. 1987. *Notes on Linear Model Theory and Variance Component Estimation*. Mimeo.
- SMITH, H. 1936. "A discriminant function for plant selection". *Ann. Eugen.*, 7, 240-250.
- TURNER, H.N.; YOUNG, S.S.Y. 1969. *Quantitative Genetics in Sheep Breeding*. Macmillan, Melbourne, Australia.
- VAN VLECK, L.D.; POLLAK, E.J.; OLTENACU, E.A.B. 1987. *Genetics for the Animal Sciences*. W.H. Freeman and Company, Nueva York.
- WIGGANS, G.R.; MISZTAL, I.; VAN VLECK, L.D. 1988. "Implementation of an animal model for genetic evaluation in the United States". *J. Dairy Sci.*, 71 (supl. 2), 1330-1337.
- WILLHAM, R.L.; POLLAK, E. 1985. "Theory of heterosis". *J. Dairy Sci.*, 68, 2411-2418.
- WRIGHT, S. 1921. "Systems of mating". *Genetics*, 6, 111-178.
- WRIGHT, S. 1969. *Evolution and the Genetics of Populations. Vol. 2, The Theory of Gene Frequencies*. University of Chicago Press, Chicago.
- YOUNG, S.S.Y. 1961. "A further examination of the relative efficiency of three methods of selection for genetic gains under less restricted conditions". *Genet. Res.*, 2, 106-121.
- YULE, G.U. 1906. "On the theory of inheritance of quantitative compound characters on the basis of Mendel's Laws, a preliminary note". *Report of the 3rd International Conference in Genetics*, 140-142.

Anexos

Anexo I. Propiedades de la esperanza

- 1) $E(ax) = a E(x)$
- 2) $E(x+y) = E(x) + E(y)$
- 3) $E(xy) = E(x) E(y)$ (x e y independientes)

- 4) $\sigma^2(x) = E(x^2) - E(x)^2$
- 5) $\sigma^2(ax) = a^2\sigma^2(x)$
- 6) $\sigma^2(x+y) = \sigma^2(x) + \sigma^2(y)$ (x e y no correlacionadas)
- 7) $\sigma^2(x+y) = \sigma^2(x) + \sigma^2(y) + 2\text{cov}(x,y)$
- 8) $\sigma^2(x-y) = \sigma^2(x) + \sigma^2(y) - 2\text{cov}(x,y)$

- 9) $\text{cov}(x,y) = E(xy) - E(x)E(y)$
- 10) $\text{cov}(x,ay) = a \text{cov}(x,y)$
- 11) $\text{cov}(ax,y) = a \text{cov}(x,y)$
- 12) $\text{cov}(x,y+z) = \text{cov}(x,y) + \text{cov}(x,z)$
- 13) $\text{cov}(y+z,x) = \text{cov}(y,x) + \text{cov}(z,x)$
- 14) $\text{cov}(x,x) = \sigma^2(x)$

Nota: De una manera intuitiva y poco formalista, podemos definir la **esperanza** como el valor medio de la variable observada cuando se repite infinitas veces la experiencia.

Anexo 2. Análisis de la varianza

Descomposición de la varianza en un modelo jerárquico¹

El caso más sencillo de diseño jerárquico, correspondiente a un diseño equilibrado, es decir, con igual número de observaciones por clase, será el que tiene el siguiente modelo matemático:

$$y_{ij} = \mu + T_i + \varepsilon_{j(i)} \quad \text{cuando} \quad \begin{array}{l} \text{i) } 1 \rightarrow t \\ \text{j) } 1 \rightarrow n \end{array}$$

donde μ es la media general, T_i es el efecto del toro o grupo i —fijo desde el punto de vista estadístico— y $\varepsilon_{j(i)}$ es el efecto del individuo j dentro del grupo i . La desviación de una observación cualquiera respecto de la media general se puede considerar descompuesta en dos sumandos cuya expresión es:

$$\begin{aligned} y_{ij} - \bar{y}_{..} &= (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}_{..}) \\ (y_{ij} - \bar{y}_{..})^2 &= [(y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}_{..})]^2 \\ (y_{ij} - \bar{y}_{..})^2 &= (y_{ij} - \bar{y}_i)^2 + (\bar{y}_i - \bar{y}_{..})^2 + 2(\bar{y}_i - \bar{y}_{..})(y_{ij} - \bar{y}_i) \end{aligned}$$

Teniendo en cuenta que al sumar para j un término en el que no existe j , lo que hacemos es repetir dicho término tantas veces como unidades indique el valor de esta letra; la expresión de la suma de todas las diferencias cuadráticas para i, j sería:

$$\sum_{ij} (y_{ij} - \bar{y}_{..})^2 = \sum_{ij} (y_{ij} - \bar{y}_i)^2 + n \sum_i (\bar{y}_i - \bar{y}_{..})^2 + 2 \sum_{ij} P \quad (I)$$

El último sumando sería:

$$2 \sum_{ij} (\bar{y}_i - \bar{y}_{..}) (y_{ij} - \bar{y}_i) = 2 \sum_i [(\bar{y}_i - \bar{y}_{..}) \sum_j (y_{ij} - \bar{y}_i)]$$

El sumatorio:

$$\sum_j (y_{ij} - \bar{y}_i) = y_i - n \bar{y}_i = y_i - y_i = 0$$

es nulo, al igual que el $\sum_i (\bar{y}_i - \bar{y}_{..})$, luego toda la expresión del doble producto es cero.

El análisis de varianza se basa en la expresión (I). La variabilidad general se descompone en variabilidad debida a las diferencias entre toros o grupos (*between*) y variabilidad residual (dentro de grupos o *within*). Para la mecánica del análisis de varianza no se calcula esta serie de diferencias elevadas al cuadrado, sino que éstas se transforman en unas expresiones más manejables para el cálculo. El segundo término del segundo miembro sería:

$$\begin{aligned} n \sum_i (\bar{y}_i - \bar{y}_{..})^2 &= n \sum_i (y_i/n - y_{..}/tn)^2 \\ &= 1/n \sum_i (y_i^2 - 2y_i y_{..}/t + y_{..}^2/t^2) \\ &= 1/n (\sum_i y_i^2 - (2y_{..}/t) \sum_i y_i + t y_{..}^2/t^2) \end{aligned}$$

1. Tomado de Cunningham, E.P., 1969. *Animal Breeding Theory*. Landbruksforlaget, Universitetsforlaget, Vollebekk, Oslo, y Orozco F., 1975. *Estadística. Curso superior de horticultura*. I.A.M.Z. Mimeo.

$$\begin{aligned}
 &= 1/n(\sum_i y_i^2 - 2y_{..}^2/t + y_{..}^2/t) \\
 &= 1/n\sum_i y_i^2 - y_{..}^2/tn \qquad \text{(II) S.C. entre grupos, } SC_B
 \end{aligned}$$

El primer término del segundo miembro se convertiría en:

$$\begin{aligned}
 \sum_{ij} (y_{ij} - \bar{y}_i)^2 &= \sum_{ij} (y_{ij} - y_i/n)^2 \\
 &= \sum_{ij} (y_{ij}^2 - 2y_{ij}y_i/n + y_i^2/n^2) \\
 &= \sum_{ij} y_{ij}^2 - 2 \sum_i y_i (\sum_j y_{ij})/n + n/n^2 \sum_i y_i^2 \\
 &= \sum_{ij} y_{ij}^2 - 2 \sum_i y_i^2/n + 1 \sum_i y_i^2/n \\
 &= \sum_{ij} y_{ij}^2 - 1/n \sum_i y_i^2 \qquad \text{(III) S.C. dentro de grupos, } SC_W
 \end{aligned}$$

El primer miembro de la igualdad es:

$$\begin{aligned}
 \sum_{ij} (y_{ij} - \bar{y}_{..})^2 &= \sum_{ij} (y_{ij} - y_{..}/tn)^2 \\
 &= \sum_{ij} (y_{ij}^2 - 2y_{ij}y_{..}/tn + y_{..}^2/t^2n^2) \\
 &= \sum_{ij} y_{ij}^2 - (2/tn)y_{..}^2 + tny_{..}^2/t^2n^2 \\
 &= \sum_{ij} y_{ij}^2 - y_{..}^2/tn \qquad \text{(IV) S.C. total, } SC_T
 \end{aligned}$$

En las expresiones últimas se verifica que:

$$\text{S.C. total} = \text{S.C. entre grupos} + \text{S.C. dentro de grupos}$$

Esta igualdad se expresa a continuación en las siguientes ecuaciones algebraicas:

$1/n\sum_i y_i^2 - y_{..}^2/tn$	Entre grupos
$\sum_{ij} y_{ij}^2 - 1/n \sum_i y_i^2$	Dentro de grupos o error
$\sum_{ij} y_{ij}^2 - y_{..}^2/tn$	Total

En la composición de estas expresiones, –(II), (III) y (IV)– se ve que hay una lógica: cada sumando elevado al cuadrado está dividido por el número de elementos que han formado esta suma; por ejemplo, si han intervenido n números, se dividirá por n ; si estuviera compuesta por tn individuos (sumados), habría que dividir por tn , etc. Es importante comprender esta lógica, porque en la práctica es la que nos va a ayudar a la realización material de cualquier análisis de varianza, puesto que la descomposición algebraica no se efectúa normalmente, suponiéndose conocida, aplicando ya directamente las fórmulas II, III y IV.

Puede demostrarse que SC_W , SC_T y SC_B se distribuyen como χ^2_{tn-t} , χ^2_{tn-1} y χ^2_{t-1} respectivamente. El término grados de libertad (g.l.) hace, pues, referencia a los parámetros de las respectivas χ^2 . Los grados de libertad se rigen bajo un mismo esquema que las sumas de cuadrados de las fórmulas anteriores. En el componente entre grupos se observa que el primer término es un sumatorio que está compuesto por t sumas y el

segundo por una sola, entonces sus g.l. son $t - 1$. Respecto al término del error (dentro de grupos), y siguiendo el mismo criterio, el primer sumatorio está compuesto por tn sumas y el segundo tiene t , sus g.l. son $tn - t$. De una forma análoga se realizan los g.l. del “total”. Esquematizando todo esto en un cuadro tenemos:

Fuente de variación	g.l.	Suma de cuadrados
Entre grupos	$t - 1$	$1/n \sum_i y_i^2 - y_{..}^2/tn$
Dentro de grupos (error)	$tn - t$	$\sum_{ij} y_{ij}^2 - 1/n \sum_i y_i^2$
Total	$tn - 1$	$\sum_{ij} y_{ij}^2 - y_{..}^2/tn$

Esperanza de los cuadrados medios en un modelo jerárquico

Si definimos el efecto del grupo y el residuo como aleatorios, tenemos

$$E(T_i) = 0; \quad E(T_i^2) = \sigma_B^2; \quad E(T_i T_k) = 0, \quad \text{siendo } \sigma_B^2 \text{ la varianza entre grupos}$$

$$E(\varepsilon_{ij}) = 0; \quad E(\varepsilon_{j(i)}^2) = \sigma_W^2; \quad E(\varepsilon_{j(i)} \varepsilon_{j'(i)}) = E(\varepsilon_{j(i)} \varepsilon_{j'(i')}) = 0, \\ \text{donde } \sigma_W^2 \text{ es la varianza dentro de grupos}$$

Los componentes de varianza σ_W^2 y σ_B^2 pueden estimarse igualando los cuadrados medios calculados (S.C./g.l.) a sus valores esperados. Obsérvese que las dos sumas de cuadrados están compuestas solamente por tres cantidades. En primer lugar obtendremos las esperanzas de esas cantidades y, después, las combinaremos para dar lugar a las esperanzas de los cuadrados medios deseadas.

$$E[\sum_{ij} y_{ij}^2] = E\sum_{ij} [\mu + T_i + \varepsilon_{j(i)}]^2 \\ = E\sum_{ij} [\mu^2 + T_i^2 + \varepsilon_{j(i)}^2 + 2\mu T_i + 2\mu \varepsilon_{j(i)} + 2T_i \varepsilon_{j(i)}] \\ = tn\mu^2 + tn\sigma_B^2 + tn\sigma_W^2 + 0 + 0 + 0$$

$$E[y_{..}^2/tn] = 1/tn E[\sum_{ij} (\mu + T_i + \varepsilon_{j(i)})^2] \\ = 1/tn E[tn\mu + \sum_i nT_i + \sum_{ij} \varepsilon_{j(i)}]^2 \\ = 1/tn E[t^2 n^2 \mu^2 + (\sum_i nT_i)^2 + (\sum_{ij} \varepsilon_{j(i)})^2 + 2tn\mu \sum_i nT_i + 2tn\mu \sum_{ij} \varepsilon_{j(i)} + \\ 2(\sum_i nT_i)(\sum_{ij} \varepsilon_{j(i)})] \\ = tn\mu^2 + n\sigma_B^2 + \sigma_W^2 + 0 + 0 + 0$$

$$E[1/n \sum_i y_i^2] = E(1/n) \sum_i [\sum_j (\mu + T_i + \varepsilon_{j(i)})^2] \\ = E(1/n) \sum_i [n\mu + nT_i + \sum_j \varepsilon_{j(i)}]^2 \\ = E(1/n) \sum_i [n^2 \mu^2 + n^2 T_i^2 + (\sum_j \varepsilon_{j(i)})^2 + 2n^2 \mu T_i + 2n\mu \sum_j \varepsilon_{j(i)} + \\ 2nT_i \sum_j \varepsilon_{j(i)}] \\ = (1/n) \sum_i [n^2 \mu^2 + n\sigma_B^2 + n\sigma_W^2 + 0 + 0 + 0] \\ = tn\mu^2 + tn\sigma_B^2 + t\sigma_W^2$$

El valor esperado de la suma de cuadrados residual o dentro de grupos es, entonces,

$$\begin{aligned} E(SC_W) &= E\{\sum_{ij} y_{ij}^2 - 1/n \sum_i y_i^2\} \\ &= (tn\mu^2 + tn\sigma_B^2 + tn\sigma_W^2) - (tn\mu^2 + tn\sigma_B^2 + t\sigma_W^2) \\ &= (tn-t)\sigma_W^2 \end{aligned}$$

El cuadrado medio correspondiente tiene como esperanza

$$E(CM_W) = E(SC_W) / (tn-t) = \sigma_W^2$$

La suma de cuadrados de tratamientos tiene como valor esperado

$$\begin{aligned} E(SC_B) &= E\{1/n \sum_i y_i^2 - y_{..}^2/tn\} \\ &= (tn\mu^2 + tn\sigma_B^2 + t\sigma_W^2) - (tn\mu^2 + n\sigma_B^2 + \sigma_W^2) \\ &= n(t-1)\sigma_B^2 + (t-1)\sigma_W^2 \end{aligned}$$

Entonces, la esperanza del cuadrado medio de tratamientos es

$$E(CM_B) = E(SC_B) / (t-1) = [n(t-1)]/(t-1) \sigma_B^2 + \sigma_W^2 = n\sigma_B^2 + \sigma_W^2$$

Nótese que con un análisis de este tipo los T_i no se estiman. La hipótesis habitual de $H_0 = \sigma_B^2 = 0$ (ausencia de variabilidad debida a las diferencias entre grupos) se contrasta mediante la prueba F, $F = CM_B / CM_W \sim (n\sigma_B^2 + \sigma_W^2) / \sigma_W^2 \geq 1$, con lo cual, si F es grande, es un indicativo de que $\sigma_B^2 > 0$.

Por otra parte, la estimación de componentes de varianza computando los cuadrados medios, derivando sus valores esperados e igualando los dos puede utilizarse para diseños más complicados, tanto equilibrados como desequilibrados. El método utilizado da estimaciones insesgadas de varianza mínima en diseños equilibrados. En el caso de subclases desiguales (diseños desequilibrados), las estimaciones serán también insesgadas si el modelo incluye solamente efectos aleatorios incorrelados entre sí. Se trata del “método I” de Henderson (1953)². Caso de que el modelo utilizado contenga tanto factores fijos como aleatorios, sería mejor utilizar el “método II” o el “método III” del mismo autor.

2. Henderson, C.R., 1953. “Estimation of variance and covariance components”. *Biometrics*, 9, 226–252.

Anexo 3. Equivalencia de la correlación entre clases e intraclase

La fórmula convencional para la correlación entre dos variables aleatorias es:

$$r = \frac{\text{cov}(x, y)}{\sqrt{\sigma^2(x) \sigma^2(y)}}$$

Retomemos el modelo de análisis de la varianza que hemos descrito anteriormente, habiendo realizado en cada vaca dos medidas. Dichas medidas, realizadas en dos lactaciones distintas, serían:

$$y_{ij} = \mu + V_i + e_{ij}$$

$$y_{ij'} = \mu + V_i + e_{ij'}$$

Nuestra intención es estimar la correlación intraclase, $r_I = \sigma_B^2 / (\sigma_B^2 + \sigma_W^2)$. Para ello partiremos de la correlación entre clases, que puede expresarse como sigue:

$$r = \frac{\text{cov}(y_{ij}, y_{ij'})}{\sqrt{\sigma^2(y_{ij}) \sigma^2(y_{ij'})}}$$

$$= \frac{\text{cov}(u + V_i + e_{ij}, u + V_i + e_{ij'})}{\sqrt{\sigma^2(u + V_i + e_{ij}) \sigma^2(u + V_i + e_{ij'})}}$$

El numerador, es decir, la covarianza entre las dos medidas, se desglosa en:

$$\begin{aligned} \text{cov}(y_{ij}, y_{ij'}) &= \text{cov}(\mu, \mu) + \text{cov}(\mu, V_i) + \text{cov}(\mu, e_{ij}) + \text{cov}(V_i, \mu) + \text{cov}(V_i, V_i) \\ &\quad + \text{cov}(V_i, e_{ij'}) + \text{cov}(e_{ij}, \mu) + \text{cov}(e_{ij}, V_i) + \text{cov}(e_{ij}, e_{ij'}) \end{aligned}$$

Según las normas expresadas en el anexo 1, tenemos:

$$\text{cov}(\mu, \mu) = 0, \text{ al ser dos constantes}$$

$$\text{cov}(\mu, V_i) = 0, \text{ covarianza de una constante con una variable}$$

$$\text{cov}(V_i, V_i) = \sigma_{V_i}^2$$

$$\text{cov}(V_i, e_{ij'}) = \text{cov}(e_{ij}, V_i) = 0, \text{ pues asumimos independencia entre los efectos}$$

$$\text{cov}(e_{ij}, e_{ij'}) = 0, \text{ pues se asume independencia entre los errores}$$

y, por lo tanto,

$$\text{cov}(y_{ij}, y_{ij'}) = \sigma_{V_i}^2 = \sigma_B^2, \text{ es decir, el componente entre vacas.}$$

En un contexto más general, debe resaltarse que la covarianza entre dos observaciones de un mismo grupo en un análisis de la varianza es igual a la varianza entre grupos.

El denominador de la anterior fracción, $\sqrt{\sigma^2(\mu + V_i + e_{ij}) \sigma^2(\mu + V_i + e_{ij'})}$, se reduce a:

$$= \sqrt{(\sigma_{V_i}^2 + \sigma_{e_{ij}}^2)(\sigma_{V_i}^2 + \sigma_{e_{ij'}}^2)}$$

ya que la varianza de una constante (μ) es cero y las covarianzas también lo son. Por otra parte, dado que se asume que todos los errores tienen una distribución común, $\sigma_{e_{ij}}^2 = \sigma_{e_{ij'}}^2 = \sigma_W^2$. En resumen, dadas las asunciones realizadas, tenemos:

$$\begin{aligned}r &= \frac{\sigma_{V_i}^2}{\sigma_{V_i}^2 + \sigma_{\epsilon_{ij}}^2} \\ &= \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \\ &= r_I\end{aligned}$$

Anexo 4. Nociones de álgebra matricial¹

Definiciones

Una matriz es una disposición rectangular de números o símbolos que representan números. De esta forma, la matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = (a_{ij})$$

tiene dos filas y tres columnas, es decir, es una matriz de 2×3 . Nótese que las matrices se expresan con una letra mayúscula en negrilla. Un vector, representado por una letra minúscula en negrilla, es una matriz $1 \times n$ ó $n \times 1$, es decir, una matriz de una sola fila o una sola columna. Se representaría, por ejemplo,

$$\mathbf{b}' = [1 \quad 2 \quad 3] = [b_1 \quad b_2 \quad b_3] = (b_i)$$

Un escalar es una matriz 1×1 o, lo que es igual, un número.

Algunas matrices que nos vamos a encontrar son cuadradas y simétricas. Así, la matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

es simétrica puesto que

$$a_{12} = a_{21}, \quad a_{13} = a_{31} \quad \text{y} \quad a_{23} = a_{32}$$

Los elementos a_{11} , a_{22} y a_{33} forman la diagonal principal de la matriz. Una matriz diagonal es una matriz cuadrada cuyos elementos son cero a excepción de los de la diagonal principal:

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Una matriz diagonal particularmente interesante es la matriz identidad, que es una matriz cuyos elementos diagonales son unos:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

La traspuesta de una matriz es otra matriz que se obtiene intercambiando las filas por las columnas en la matriz original. Se representa añadiendo una prima al símbolo de la matriz original

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \mathbf{A}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

1. Adaptado de Schaeffer, L.R.; Mao, I.L. 1987. *Notes on Linear Model Theory and Variance Component Estimation*. University of Guelph. Mimeo.

Operaciones aritméticas

La traza de una matriz \mathbf{A} , $\text{tr}(\mathbf{A})$, que debe ser cuadrada, es la suma de los elementos diagonales de la matriz.

Adición y sustracción de matrices

Para sumar o sustraer matrices, éstas deben ser de las mismas dimensiones (mismo orden), es decir, deben tener el mismo número de filas y de columnas. Sólo se necesita sumar o restar los elementos correspondientes de las dos matrices. Por ejemplo,

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 2 \\ 3 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 5 \\ 7 & 7 & 9 \end{bmatrix}$$

y de forma similar

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} - \begin{bmatrix} 3 & 1 & 2 \\ 3 & 2 & 3 \end{bmatrix} = \begin{bmatrix} -2 & 1 & 1 \\ 1 & 3 & 3 \end{bmatrix}$$

Multiplicación de matrices

La multiplicación de matrices es un poco más complicada. El número de columnas de la primera matriz debe ser igual al número de filas de la segunda. El elemento ij de la matriz producto se obtiene sumando los productos de los correspondientes elementos de la fila i de la primera matriz por los de la columna j de la segunda. La multiplicación de $\mathbf{A} \times \mathbf{B}$, dos matrices de 3×2 y 2×2 , respectivamente, produce

$$\begin{aligned} \mathbf{A} \times \mathbf{B} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} \end{bmatrix} \end{aligned}$$

que puede resumirse como

$$\begin{array}{ccc} \mathbf{A} & \times & \mathbf{B} & = & \mathbf{C} \\ 3 \times 2 & & 2 \times 2 & & 3 \times 2 \end{array}$$

Debe observarse que las dos dimensiones internas del producto deben ser siempre iguales, mientras que las dos externas dan las dimensiones de la matriz producto. Además, en multiplicación de matrices, el orden de la multiplicación es importante; es decir, \mathbf{AB} usualmente no será igual a \mathbf{BA} . Concretamente, para matrices no cuadradas, puede no ser posible el producto $\mathbf{B} \times \mathbf{A}$ cuando, en cambio, lo sea el $\mathbf{A} \times \mathbf{B}$. Pero incluso cuando la premultiplicación y la posmultiplicación son posibles, los resultados suelen ser diferentes.

La multiplicación de un escalar por una matriz es un caso especial. La matriz producto es el resultado de multiplicar el escalar por cada uno de los elementos de la matriz original.

La traspuesta de un producto de matrices es el producto de las traspuestas originales pero en orden inverso, es decir,

$$(\mathbf{A} \mathbf{B} \mathbf{C})' = \mathbf{C}' \mathbf{B}' \mathbf{A}'$$

Producto directo de matrices (producto de Kronecker)

Supongamos que \mathbf{B} es una matriz de orden $m \times n$ y que \mathbf{A} es una matriz 2×2 ; entonces, el producto directo es

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{bmatrix}$$

Inversión de matrices

La inversión de matrices es el equivalente a la división, pero en el campo del álgebra matricial. Por definición, la inversa de la matriz cuadrada \mathbf{A} , denotada por \mathbf{A}^{-1} , es una matriz que, pre o posmultiplicada por la matriz original, nos da la matriz identidad. Esto es,

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \text{ y } \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

El primer paso para invertir una matriz es calcular su determinante, que debe ser distinto de cero. El determinante es un escalar que puede calcularse como sigue. Supongamos una matriz \mathbf{A} tal que

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Entonces, el determinante es $|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$. Para una matriz 3×3 , el determinante puede reducirse a una serie de determinantes de matrices 2×2 . Por ejemplo,

$$\mathbf{B} = \begin{bmatrix} 6 & -1 & 2 \\ 3 & 4 & -5 \\ 1 & 0 & -2 \end{bmatrix}$$

entonces

$$\begin{aligned} |\mathbf{B}| &= 6 \begin{vmatrix} 4 & -5 \\ 0 & -2 \end{vmatrix} - 1(-1) \begin{vmatrix} 3 & -5 \\ 1 & -2 \end{vmatrix} + 2 \begin{vmatrix} 3 & 4 \\ 1 & 0 \end{vmatrix} \\ &= 6(-8) + 1(-1) + 2(-4) \end{aligned}$$

La expresión general para un determinante es

$$|\mathbf{B}| = \sum_{j=1}^n (-1)^{i+j} b_{ij} |\mathbf{M}_{ij}|$$

donde \mathbf{B} es de orden n , y \mathbf{M}_{ij} es una submatriz menor de \mathbf{B} resultante de eliminar la fila i y la columna j de \mathbf{B} . Para calcular el determinante de \mathbf{B} puede utilizarse cualquier fila o columna, pues el resultado sería el mismo.

Si el determinante es distinto de cero, pasar a la siguiente etapa, que es encontrar la matriz de menores con signo (cofactores). Sustituir entonces el elemento ij de \mathbf{B} por

$$(-1)^{i+j} |M_{ij}|$$

A la matriz de cofactores se la conoce como matriz adjunta. A partir de la matriz \mathbf{B} se obtienen los siguientes menores:

<u>i</u>	<u>j</u>	<u>signo</u>	<u>menor</u>	=	
1	1	+1	$\begin{vmatrix} 4 & -5 \\ 0 & -2 \end{vmatrix}$	=	-8
1	2	-1	$\begin{vmatrix} 3 & -5 \\ 1 & -2 \end{vmatrix}$	=	-1
1	3	+1	$\begin{vmatrix} 3 & 4 \\ 1 & 0 \end{vmatrix}$	=	-4
2	1	-1	$\begin{vmatrix} -1 & 2 \\ 0 & -2 \end{vmatrix}$	=	2
2	2	+1			-14
2	3	-1			1
3	1	+1			-3
3	2	-1			-36
3	3	+1			27

Combinando los menores con signo en la matriz adjunta se obtiene \mathbf{M}_B ,

$$\mathbf{M}_B = \begin{bmatrix} -8 & 1 & -4 \\ -2 & -14 & -1 \\ -3 & 36 & 27 \end{bmatrix}$$

siendo la inversa de \mathbf{B}

$$\begin{aligned} \mathbf{B}^{-1} &= |\mathbf{B}|^{-1} \mathbf{M}_B' \\ &= 1/(-57) \begin{bmatrix} -8 & -2 & -3 \\ 1 & -14 & 36 \\ -4 & -1 & 27 \end{bmatrix} \end{aligned}$$

La importancia de que el determinante no sea cero es obvia, ya que hay que dividir todos los términos de \mathbf{M}_B por el determinante. Si el determinante es cero, la inversa no está definida. A una matriz cuadrada con determinante no cero se le conoce como no singular. Sólo las matrices no singulares tienen inversa. Una matriz con determinante igual a cero se denomina singular y no tiene inversa.

La inversa de una matriz inversa es la matriz original. Es decir, $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$. De manera similar a la traspuesta de un producto de matrices, la inversa del producto de matrices no singulares es el producto de las inversas, pero en orden inverso, esto es

$$\begin{aligned} (\mathbf{A} \mathbf{B} \mathbf{C})^{-1} &= \mathbf{C}^{-1} \mathbf{B}^{-1} \mathbf{A}^{-1} \quad \text{y} \\ \mathbf{C}^{-1} \mathbf{B}^{-1} \mathbf{A}^{-1} \mathbf{A} \mathbf{B} \mathbf{C} &= \mathbf{I} \end{aligned}$$

Matriz de varianzas-covarianzas de un vector

La varianza de una variable aleatoria (escalar) se define como

$$V(x) = E(x - E(x))^2 = E(x^2) - E(x)E(x) = \sigma_x^2$$

Para dos variables aleatorias, x_1 y x_2 , su covarianza se define como

$$\text{Cov}(x_1, x_2) = E(x_1 x_2) - E(x_1) E(x_2) = \sigma_{x_1 x_2}$$

Las definiciones precedentes pueden extenderse a un vector de variables aleatorias, \mathbf{x} , para dar lugar a una matriz de varianzas-covarianzas de orden igual a la longitud de \mathbf{x} .

$$V(\mathbf{x}) = E(\mathbf{x}\mathbf{x}') - E(\mathbf{x}) E(\mathbf{x}')$$

por ejemplo, si $\mathbf{x} = [x_1 \ x_2 \ x_3]$

$$\begin{aligned} V(\mathbf{x}) &= E \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} [x_1 \ x_2 \ x_3] - E \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} E [x_1 \ x_2 \ x_3] \\ &= \begin{bmatrix} E(x_1^2) & E(x_1 x_2) & E(x_1 x_3) \\ E(x_2 x_1) & E(x_2^2) & E(x_2 x_3) \\ E(x_3 x_1) & E(x_3 x_2) & E(x_3^2) \end{bmatrix} - \begin{bmatrix} \mu_1^2 & \mu_1 \mu_2 & \mu_1 \mu_3 \\ \mu_2 \mu_1 & \mu_2^2 & \mu_2 \mu_3 \\ \mu_3 \mu_1 & \mu_3 \mu_2 & \mu_3^2 \end{bmatrix} \end{aligned}$$

Por tanto,

$$V(\mathbf{x}) = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} & \sigma_{x_1 x_3} \\ \sigma_{x_2 x_1} & \sigma_{x_2}^2 & \sigma_{x_2 x_3} \\ \sigma_{x_3 x_1} & \sigma_{x_3 x_2} & \sigma_{x_3}^2 \end{bmatrix} = \mathbf{V}$$

La matriz de varianzas-covarianzas de un vector aleatorio contiene las varianzas en la diagonal, siendo las covarianzas los demás elementos. Estas matrices son generalmente simétricas y no singulares.

La matriz de varianzas-covarianzas de una serie de funciones lineales de un vector aleatorio se deduce como sigue:

$$\begin{aligned} V(\mathbf{Ax}) &= E(\mathbf{Ax}\mathbf{x}'\mathbf{A}') - E(\mathbf{Ax})E(\mathbf{x}'\mathbf{A}') \\ &= \mathbf{A} E(\mathbf{x}\mathbf{x}') \mathbf{A}' - \mathbf{A} E(\mathbf{x})E(\mathbf{x}')\mathbf{A}' \\ &= \mathbf{A} [E(\mathbf{x}\mathbf{x}') - E(\mathbf{x})E(\mathbf{x}')] \mathbf{A}' \\ &= \mathbf{A} V(\mathbf{x}) \mathbf{A}' \end{aligned}$$

De forma similar, se deduce que, dadas dos funciones de \mathbf{x} , por ejemplo \mathbf{Ax} y \mathbf{Bx} , la covarianza entre las mismas es la siguiente:

$$\text{cov}(\mathbf{Ax}, \mathbf{Bx}) = \mathbf{A} V(\mathbf{x}) \mathbf{B}'$$

