

## OBJETIVOS E INDICES DE SELECCION

### Descripción y definiciones

Debido a su importancia económica, es frecuente en planes de mejoramiento genético por selección que varios caracteres deban ser considerados simultáneamente (e.g., ritmo de crecimiento y precocidad sexual en cerdos; peso vivo, tasa reproductiva, y peso de vellón en ovinos; peso al destete y conformación en bovinos).

Una forma de tratar esta situación consiste en formular un **objetivo de selección** ( $V$ ) que combine adecuadamente los caracteres a mejorar:

$$V = \sum_{j=1}^m v_j (y_j - \bar{y}_j) = v_1 (y_1 - \bar{y}_1) + v_2 (y_2 - \bar{y}_2) + \dots + v_m (y_m - \bar{y}_m) \quad [1]$$

En la expresión anterior, los  $y_i$  representan, en general, valores reproductivos de los  $m$  caracteres a mejorar, expresados como desvíos respecto de sus medias poblacionales, y ponderados por coeficientes ( $v_j$ ) que cuantifican su importancia económica relativa.

Para aplicar esta estrategia de mejoramiento, se debería calcular el valor de  $V$  para cada uno de los  $k$  candidatos a selección y elegir como futuros padres a los de mayor  $V_k$ . El cálculo de los  $v_j$  está sujeto a diversas consideraciones (e.g., estadísticas vitales, costos, número de veces que se expresa cada característica) pero es factible. En contraste, los  $y_j$  no son variables observables y, por lo tanto, los  $V_k$  no pueden ser evaluados directamente. El recurso habitual es derivar un estimador; en este caso, un **índice** o **criterio de selección**.

Algebraicamente, un índice de selección ( $I$ ) es una combinación lineal del tipo:

$$I_k = \sum_{i=1}^n b_i (x_{ik} - \bar{x}_i) = b_1 (x_{1k} - \bar{x}_1) + b_2 (x_{2k} - \bar{x}_2) + \dots + b_n (x_{nk} - \bar{x}_n) \quad [2]$$

en la que  $n$  fuentes de información ( $x_i$ ), habitualmente expresadas como desvíos, son agregadas en una suma ponderada. El valor de índice para el  $k$ -ésimo individuo es la suma, ponderada por los  $b_i$ , de los desvíos correspondientes a cada fuente de información considerada. Estas fuentes de información son caracteres (no necesariamente los mismos incluidos en el objetivo) medidos en los propios candidatos a selección o en alguno de sus parientes. Estos caracteres deberían ser, idealmente, fáciles de medir a bajo costo y, preferentemente, antes que los individuos alcancen su edad reproductiva.

Para que sea esperable mejorar el objetivo cuando se seleccione en base al índice obviamente se requerirá que éste satisfaga algún criterio que garantice una estrecha

relación entre V e I. Los siguientes cuatro criterios tienen esa característica y, en realidad, puede demostrarse que son equivalentes:

- maximización de  $r_{VI}$ , la correlación entre objetivo (V) e índice (I);
- maximización de la probabilidad de un ordenamiento correcto de los candidatos en base a su valor para mejorar el objetivo (V);
- maximización del progreso genético por selección (cuando V corresponde a valor genético aditivo);
- minimización del promedio de las diferencias al cuadrado,  $E(I-V)^2$ , entre valores de V y valores de I (i.e., el llamado criterio de ‘mínimos cuadrados’).

## Derivación de la expresión general de un índice de selección

Formular un índice de selección requiere derivar expresiones generales para los  $b_i$  ya que las variables en [2] son observables. La derivación que sigue se basa en el primero de los cuatro criterios citados más arriba: maximización de  $r_{VI}$ .

El problema es encontrar valores para los  $b_i$  tales que maximicen:

$$r_{VI} = \frac{\sigma_{VI}}{\sqrt{\sigma_V^2 \times \sigma_I^2}} \quad [3]$$

donde  $\sigma_{VI}$  indica la covarianza entre V e I.

Maximizar  $\log r_{VI}$  equivale a maximizar  $r_{VI}$  y permite expresar [3] en forma lineal:

$$\log r_{VI} = \log \sigma_{VI} - 1/2 \log \sigma_V^2 - 1/2 \log \sigma_I^2$$

Tomando derivadas parciales con respecto a cada  $b_i$  e igualándolas a cero, se obtiene un conjunto de ecuaciones simultáneas que puede resolverse para valores apropiados de  $b_i$ . Para cualquier  $b_i$ , e.g.  $b_1$ , la ecuación correspondiente será:

$$\frac{\partial \log r_{VI}}{\partial b_1} = \frac{1}{\sigma_{VI}} \frac{\partial \sigma_{VI}}{\partial b_1} - 0 - \frac{1}{2\sigma_I^2} \frac{\partial \sigma_I^2}{\partial b_1} = 0 \quad [4]$$

Note que la varianza de V no es función de los  $b_i$  (es constante con respecto a éstos) y, por lo tanto, la derivada parcial de la varianza de V con respecto a cualquier  $b_i$  será igual a cero. Si se representan los desvíos como  $x_i$  (sólo para simplificar la notación) y considerando que:

$$\sigma_{VI} = \text{cov}(V, \sum_{i=1}^n b_i x_i) = \sum_{i=1}^n b_i \sigma_{Vx_i}$$

y que,

$$\sigma_I^2 = b_1^2 \sigma_{x_1}^2 + b_2^2 \sigma_{x_2}^2 + \dots + 2b_1 b_2 \sigma_{x_1 x_2} + \dots$$

entonces,

$$\frac{\partial \sigma_{VI}}{\partial b_1} = \frac{\partial (b_1 \sigma_{Vx_1})}{\partial b_1} + \frac{\partial (b_2 \sigma_{Vx_2})}{\partial b_1} + \dots + \frac{\partial (b_n \sigma_{Vx_n})}{\partial b_1} = \sigma_{Vx_1}$$

y

$$\begin{aligned} \frac{\partial \sigma_I^2}{\partial b_1} &= 2b_1 \sigma_{x_1}^2 + 0 + 2b_2 \sigma_{x_1 x_2} + \dots + 0 + \dots + 2b_n \sigma_{x_1 x_n} \\ \frac{\partial \sigma_I^2}{\partial b_1} &= 2 (b_1 \sigma_{x_1}^2 + b_2 \sigma_{x_1 x_2} + \dots + b_n \sigma_{x_1 x_n}) \end{aligned}$$

Sustituyendo estos resultados en [4]:

$$\frac{\sigma_{Vx_1}}{\sigma_{VI}} - \frac{b_1 \sigma_{x_1}^2 + b_2 \sigma_{x_1 x_2} + \dots + b_n \sigma_{x_1 x_n}}{\sigma_I^2} = 0$$

Trasponiendo términos y multiplicando por la varianza del índice:

$$b_1 \sigma_{x_1}^2 + b_2 \sigma_{x_1 x_2} + \dots + b_n \sigma_{x_1 x_n} = \sigma_{Vx_1} \left( \frac{\sigma_I^2}{\sigma_{VI}} \right)$$

Esa es la ecuación correspondiente a  $b_1$ , pero como el cociente:

$$\frac{\sigma_I^2}{\sigma_{VI}}$$

será común a todas las ecuaciones de los  $b_i$  y solamente producirá un cambio de escala, se le puede adjudicar un valor arbitrario igual a 1, quedando la ecuación para  $b_1$ :

$$b_1 \sigma_{x_1}^2 + b_2 \sigma_{x_1 x_2} + \dots + b_n \sigma_{x_1 x_n} = \sigma_{Vx_1}$$



equivalente a seleccionar una escala para los  $b_i$  tal que la varianza del índice iguale a su covarianza con el objetivo:

$$\sigma_{VI} = \sigma_I^2$$

## Correlación entre índice (I) y objetivo (V)

La expresión general para esta correlación es :

$$r_{VI} = \frac{\sigma_{VI}}{(\sigma_V^2 \times \sigma_I^2)^{1/2}} = \frac{\sigma_{VI}}{\sigma_V \cdot \sigma_I}$$

pero como

$$\sigma_{VI} = \sigma_I^2$$

entonces

$$r_{VI} = \frac{\sigma_I}{\sigma_V} = \sqrt{\frac{\sigma_{VI}}{\sigma_V^2}}$$

y como

$$\sigma_{VI} = \sum_{i=1}^n b_i \sigma_{Vx_i}$$

entonces

$$\hat{r}_{VI} = \sqrt{\frac{\sum_{i=1}^n \hat{b}_i \sigma_{Vx_i}}{\sigma_V^2}}$$

La correlación entre V e I (también llamada **exactitud**) es una medida de la confianza asociada a cada tipo de índice de selección e influye en el progreso genético esperado.

Se tratan a continuación una serie de casos de formulación de índices de complejidad creciente. La situación más simple imaginable corresponde al uso de una única fuente de información, medida en los propios candidatos a selección, y coincidente con la única variable que integra el objetivo (e.g., selección para mejorar el peso al destete en bovinos [objetivo], utilizando una única medición de peso al destete [criterio] obtenida sobre cada candidato a selección). En el otro extremo de complejidad, se considerarán objetivos de selección constituidos por más de una característica, ponderadas económicamente, y criterios de selección integrados por varias variables, medidas en los propios candidatos a selección o en algún tipo de pariente.

## Índices para mejorar una única característica

**Información : un único registro del propio fenotipo en cada candidato a selección**

Se define como objetivo de selección:  $V = \sigma_{10}$  (valor genético-aditivo o reproductivo)

La relación entre la información ( $x$ , e.g. peso al destete) y el valor está dada por:

$$x = \sigma_{10} + R$$

donde  $R$  indica un término residual que engloba la variación que no es de origen genético aditivo (i.e.  $D, I, Ep$ , etc.).

Como se mide una sola característica, la ecuación del índice para el  $i$ ésimo candidato [2] se reduce a:

$$I_k = \sum_{i=1}^n b_i (x_{ik} - \bar{x}_i) = b(x_k - \bar{x})$$

donde  $x$  representa una medida fenotípica previamente corregida (i.e., estandarizada por todos los factores que afectan sistemáticamente la expresión de la característica; e.g. corrección del peso al destete en base a edad de la madre).

Con una sola fuente de información, el sistema de ecuaciones a resolver para los  $b_i$  [5] se reduce a

$$b \cdot \sigma_x^2 = \sigma_{Vx}$$

de donde

$$\hat{b} = \frac{\sigma_{Vx}}{\sigma_x^2}$$

El denominador de esa expresión es la varianza de la información (i.e., la varianza fenotípica, e.g. la varianza fenotípica del peso al destete) mientras que el numerador es una covarianza teórica para la que se requiere una expresión analítica.

$$\sigma_{Vx} = \text{cov}[\sigma_{10}, (\sigma_{10} + R)] = \text{cov}(\sigma_{10}, \sigma_{10}) + \text{cov}(\sigma_{10}, R)$$

Si se asume que no existe correlación entre efectos aditivos con efectos genéticos no aditivos o con el ambiente, el último término de la expresión anterior será igual a 0. La expresión se reduce entonces a

$$\sigma_{Vx} = \sigma_{10}^2$$

La covarianza genética entre individuos está ponderada por el grado de parentesco existente entre ellos:

$$\text{cov}(X, Y) = a_{XY} \cdot \sigma_{10}^2 + a_{XY} \cdot \sigma_{01}^2 + \dots$$

donde  $a_{XY}$  indica el parentesco entre los individuos  $X$  e  $Y$ . Si sólo se toman en cuenta los efectos aditivos, el primero es el único término relevante en la expresión anterior,

quedando los restantes incluidos en el término residual. Como se trata de información del propio individuo,  $a_{XY}$  será igual a 1 en la expresión anterior.

Obtenida una expresión analítica para la covarianza entre valor e información, se puede derivar una expresión más simple para el valor estimado de  $b$

$$\hat{b} = \frac{\sigma_{Vx}}{\sigma_x^2} = \frac{\sigma_{10}^2}{\sigma_x^2} = h^2$$

la ecuación del índice para el  $k$ -ésimo candidato será entonces:

$$I_k = b (x_k - \bar{x}) = h^2 (x_k - \bar{x})$$

y la expresión correspondiente a la exactitud:

$$f_{VI} = \left[ \frac{\sum_{i=1}^n \hat{b}_i \sigma_{Vx_i}}{\sigma_V^2} \right]^{1/2} = \left[ \frac{h^2 \sigma_{10}^2}{\sigma_{10}^2} \right]^{1/2} = h$$

### **Información: más de un registro del propio fenotipo en cada candidato**

Como en el caso anterior,  $V = \sigma_{10}$

La información a considerar es el promedio de las mediciones fenotípicas realizadas sobre cada candidato ( $k$ ). El índice será entonces:

$$I_k = b (\bar{x}_k - \bar{x})$$

La doble barra sobre la segunda  $x$  indica promedio general de toda la información a través de registros e individuos.

El sistema de ecuaciones puede formularse entonces como

$$b \cdot \sigma_x^2 = \sigma_{Vx}$$

de donde:

$$\hat{b} = \frac{\sigma_{V\bar{x}}}{\sigma_{\bar{x}}^2} = \frac{\text{cov}[V, (\frac{1}{n} \sum_{i=1}^n x_i)]}{\frac{1}{n^2} [n\sigma_{x_i}^2 + n(n-1)\text{cov}(x_i, x_j)]}$$

donde  $\text{cov}(x_i, x_j)$  representa la covarianza entre mediciones.

La repetibilidad ( $r$ ) puede indicarse como la relación:

$$r = \frac{\text{cov}(x_i, x_j)}{\sigma_x^2}$$

entonces,

$$\hat{b} = \frac{\text{cov}[V, (\frac{1}{n} \sum_{i=1}^n x_i)]}{(\frac{\sigma_{x_i}^2}{n}) + [\frac{(n-1)}{n}] \cdot r \cdot \sigma_{x_i}^2} = \frac{\frac{1}{n} (\sum_{i=1}^n \text{cov}(V, x_i))}{(\frac{\sigma_{x_i}^2}{n}) \cdot [1 + (n-1) \cdot r]}$$

$$\hat{b} = \frac{\frac{1}{n} \sum_{i=1}^n \text{cov}[(\sigma_{10} + R), \sigma_{10}]}{(\frac{\sigma_{x_i}^2}{n}) \cdot [1 + (n-1) \cdot r]} = \frac{\frac{1}{n} (n \sigma_{10}^2)}{(\frac{\sigma_{x_i}^2}{n}) \cdot [1 + (n-1) \cdot r]}$$

de donde finalmente:

$$\hat{b} = \frac{nh^2}{[1 + (n-1) \cdot r]}$$

y la exactitud será:

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

que se reduce a la del caso anterior para  $n = 1$ .

Si en lugar de considerar  $V = \sigma_{10}$  se definiera  $V = G + Ep$  donde  $G$  representa todos los efectos genéticos (aditivos y no aditivos) y  $Ep$  indica ambiente permanente se obtendría la siguiente expresión para una estimación de  $b$ :

$$\hat{b} = \frac{nr}{[1 + (n-1) \cdot r]}$$

y la exactitud sería igual a  $b^{1/2}$ .

En este caso se estaría ordenando los candidatos en base a registros ya disponibles con el objetivo de seleccionar los de mayor producción futura durante su vida útil. Se trata de estimar **capacidad probable de producción** en lugar de valor genético aditivo como en el resto de los casos. En esta situación el objetivo contiene tanto componentes genéticos como ambientales.

**Información: un único registro de un antecesor, sin información del candidato**

$V = \sigma_{10D}$  = valor reproductivo del descendiente (D)

Información:  $x_A$  = un registro fenotípico en un antecesor (A)

La ecuación a plantear será:

$$b \cdot \sigma_A^2 = \sigma_{Vx_A}$$

para resolver la cual se requiere una expresión analítica del componente del lado derecho de la expresión:

$$\sigma_{Vx_A} = \text{cov}[\sigma_{10D}, (\sigma_{10A} + R_A)] = \text{cov}(\sigma_{10D}, \sigma_{10A}) + \text{cov}(\sigma_{10D}, R_A) = a_{AD} \sigma_{10}^2$$

donde  $a_{AD}$  corresponde al parentesco entre antecesor y descendiente. De esta forma,

$$\hat{b} = a_{AD} \left( \frac{\sigma_{10}^2}{\sigma_x^2} \right)$$

de donde

$$\hat{b} = a_{AD} h^2 \quad \text{y} \quad \hat{r}_{VI} = a_{AD} h$$

En la Tabla 1 se indican valores de b para diferentes heredabilidades y tipos de

Tabla 1. Valores del coeficiente de ponderación en la ecuación de índice para distintas heredabilidades, cuando se usa como fuente de información un único registro de un antecesor.

Heredabilidad	Antecesor		
	Padre/ madre	Abuelo/ a	Bisabuelo/ a
0.20	0.10	0.05	0.025
0.40	0.20	0.10	0.050
0.60	0.30	0.15	0.075
0.80	0.40	0.20	0.100

antecesoros. Note que los valores de  $b$  y, por ende, los de  $r_{VI}$  se incrementan en forma proporcional a la heredabilidad y disminuyen conforme lo hace el grado de parentesco.

Para los casos simples ya considerados, el *ordenamiento de candidatos* será el mismo independientemente de los valores estimados para  $b$  ya que este coeficiente actúa simplemente como una constante multiplicativa. En contraste, la *dispersión de los valores de índice alrededor de la media* será proporcional al valor de  $b$ . En el caso extremo de un  $b$  cercano a 1, los valores de índice tendrían una dispersión casi idéntica a la de los valores fenotípicos mientras que en el caso opuesto ( $b$  cercano a 0) se verificaría una concentración de valores estimados de índice alrededor del valor medio. Estas tendencias indican que la máxima capacidad de discriminación entre candidatos se verifica, como sería de esperar, cuando  $h^2$  (o  $r$  en el caso de capacidad probable de producción) tiene un valor relativamente elevado.

**Información: un registro de cada progenitor, sin información del candidato**

$V = \sigma_{10}$  = valor reproductivo del candidato

$x_P$  = información del padre (un registro)

$x_M$  = información de la madre (un registro)

Ecuaciones:

$$\begin{aligned} b_P \sigma_{x_P}^2 + b_M \sigma_{x_P x_M} &= \sigma_{V x_P} = \sigma_{10}^2 / 2 \\ b_P \sigma_{x_M x_P} + b_M \sigma_{x_M}^2 &= \sigma_{V x_M} = \sigma_{10}^2 / 2 \end{aligned}$$

Expresión analítica de las covarianzas requeridas:

$$\begin{aligned} \sigma_{x_P x_M} &= \text{cov} [(\sigma_{10_P} + R_P), (\sigma_{10_M} + R_M)] \\ \sigma_{x_P x_M} &= \text{cov} (\sigma_{10_P}, \sigma_{10_M}) + \text{cov} (\sigma_{10_P}, R_M) + \text{cov} (\sigma_{10_M}, R_P) + \text{cov} (R_P, R_M) \\ \sigma_{x_P x_M} &= a_{PM} \sigma_{10}^2 \end{aligned}$$

y si los progenitores no son parientes,

$$\sigma_{x_P x_M} = 0$$

Las ecuaciones quedan:

$$\begin{aligned} b_P \sigma_{x_P}^2 &= \sigma_{10}^2 / 2 \\ b_M \sigma_{x_M}^2 &= \sigma_{10}^2 / 2 \end{aligned}$$

de donde:

$$\hat{b}_P = \hat{b}_M = h^2 / 2$$

Para el k-ésimo candidato, el índice tendrá la forma:

$$I_k = \hat{b}_P(x_{P_k} - \bar{x}_P) + \hat{b}_M(x_{M_k} - \bar{x}_M)$$

y la exactitud será:

$$\hat{r}_{VI} = \left[ \frac{(h^2/2)(\sigma_{10}^2/2) + (h^2/2)(\sigma_{10}^2/2)}{\sigma_{10}^2} \right]^{1/2} = \left[ \frac{h^2}{2} \right]^{1/2} \cong 0.71 h$$

pero podrá incrementarse si se dispone de más de un registro de los progenitores.

**Información: un registro de un progenitor y un registro del candidato**

$V = \sigma_{10}$  = valor reproductivo del candidato

$x_H$  = información de los candidatos

$x_P$  = información de los progenitores de los candidatos

El índice tendrá la forma:

$$I_k = \hat{b}_H(x_{H_k} - \bar{x}_H) + \hat{b}_P(x_{P_k} - \bar{x}_P)$$

Ecuaciones:

$$\begin{aligned} b_H \sigma_{x_H}^2 + b_P \sigma_{x_H x_P} &= \sigma_{V x_H} = \sigma_{10}^2 \\ b_H \sigma_{x_P x_H} + b_P \sigma_{x_P}^2 &= \sigma_{V x_P} = a_{HP} \sigma_{10}^2 = \sigma_{10}^2 / 2 \end{aligned}$$

Expresión analítica de las covarianzas:

$$\begin{aligned} \sigma_{x_H x_P} &= \text{cov}[(\sigma_{10_H} + R_H), (\sigma_{10_P} + R_P)] \\ \sigma_{x_H x_P} &= \text{cov}(\sigma_{10_H}, \sigma_{10_P}) + \text{cov}(\sigma_{10_H}, R_P) + \text{cov}(\sigma_{10_P}, R_H) + \text{cov}(R_H, R_P) \\ \sigma_{x_H x_P} &= a_{HP} \sigma_{10}^2 = \sigma_{10}^2 / 2 \end{aligned}$$

de donde las ecuaciones quedan:

$$\begin{aligned} b_H \sigma_{x_H}^2 + b_P \sigma_{10}^2 / 2 &= \sigma_{10}^2 \\ b_H \frac{\sigma_{10}^2}{2} + b_P \sigma_{x_P}^2 &= \sigma_{10}^2 / 2 \end{aligned}$$

con soluciones para los  $b_i$ :

$$\hat{b}_H = \frac{h^2 - \left(\frac{h^2}{2}\right)^2}{1 - \left(\frac{h^2}{2}\right)^2}$$

$$\hat{b}_P = \frac{h^2(1-h^2)/2}{1 - \left(\frac{h^2}{2}\right)^2}$$

y la  $r_{VI}$ :

$$r_{VI} = \sqrt{\frac{5h^2 - 2h^4}{4 - h^4}}$$

Note que la información del progenitor y la del candidato resultan ponderadas por distintos coeficientes.

Antes de considerar la utilización de registros de progenie y parientes colaterales, es conveniente recordar cuáles son las causas del parecido entre parientes.

En general, se identifican dos causas:

- *genotipo*, por compartir parte de su constitución genética y
- *ambiente*, porque los parientes suelen estar sujetos a similares condiciones ambientales.

Para medir las similitudes debidas a genotipo y ambiente, se define un estadístico ( $t$ ) denominado **correlación intraclase**. Esta correlación incluye las dos causas del parecido entre parientes; la primera (genotipo) está contemplada como heredabilidad ponderada por el grado de parentesco y la segunda (ambiente) está representada por la correlación ambiental ( $c^2$ ).

En general, entonces: 
$$t = a_{ij} h^2 + c^2$$

Para una familia de medio hermanos:

$$t = (1/4) h^2 + c^2$$

Para una familia de hermanos completos:

$$t = (1/2) h^2 + c^2$$

Es habitual considerar que la correlación ambiental vale 0 y entonces  $t$  se reduce a la heredabilidad ponderada por el grado de parentesco. Sin embargo, se debe tener presente que en algunos casos (ej.: hermanos completos o medio hermanos maternos) la correlación ambiental puede tener un valor significativo que debe ser estimado e incluido en los cálculos.

**Información: un registro de cada uno de  $n$  hijos/as del candidato**

$$V = \sigma_{10}$$

Las fuentes de información son los promedios de los registros de  $n_k$  hijos del  $k$ -ésimo candidato (un registro por hijo/a). Este tipo de índice es el utilizado en lo que habitualmente se conoce como **prueba de progenie**.

El índice se construirá en base a los desvíos de los promedios de los hijos/as, con respecto a la media general:

$$I_k = b_k (\bar{x}_k - \bar{x})$$

La ecuación a utilizar en la derivación será:

$$b_k \cdot \sigma_{\bar{x}}^2 = \sigma_{V\bar{x}}$$

de donde puede deducirse que:

$$\hat{b}_k = \frac{n_k h^2}{2[1 + (n_k - 1)t]}$$

y, por lo tanto:

$$\hat{f}_{V I_k} = \sqrt{\frac{n_k h^2}{4[1 + (n_k - 1)t]}}$$

En este caso, a cada candidato le corresponden un  $b_k$  y una exactitud particulares que están relacionadas al número de hijos/as que aportan información para la prueba. Es el único caso en el que  $b$  puede valer más de uno. El valor de  $t$  dependerá de que los hijos/as constituyan familias de medio hermanos o hermanos completos y del valor que se asigne a  $c^2$ . En el caso más común (familias de medio hermanos/as paternos y  $c^2 = 0$ ), las expresiones anteriores se reducen a:

$$\hat{b}_k = \frac{2n_k h^2}{4 + (n_k - 1)h^2}$$

$$\hat{f}_{V I_k} = \sqrt{\frac{n_k h^2}{4 + (n_k - 1)h^2}}$$

En la Figura 1 puede observarse cómo varía  $b_k$  de acuerdo a la heredabilidad del carácter y al número de hijos/as. Note que el valor de  $b_k$  tiende a 2 a medida que se incrementa la cantidad de información y la velocidad con la que lo hace depende de  $h^2$ .

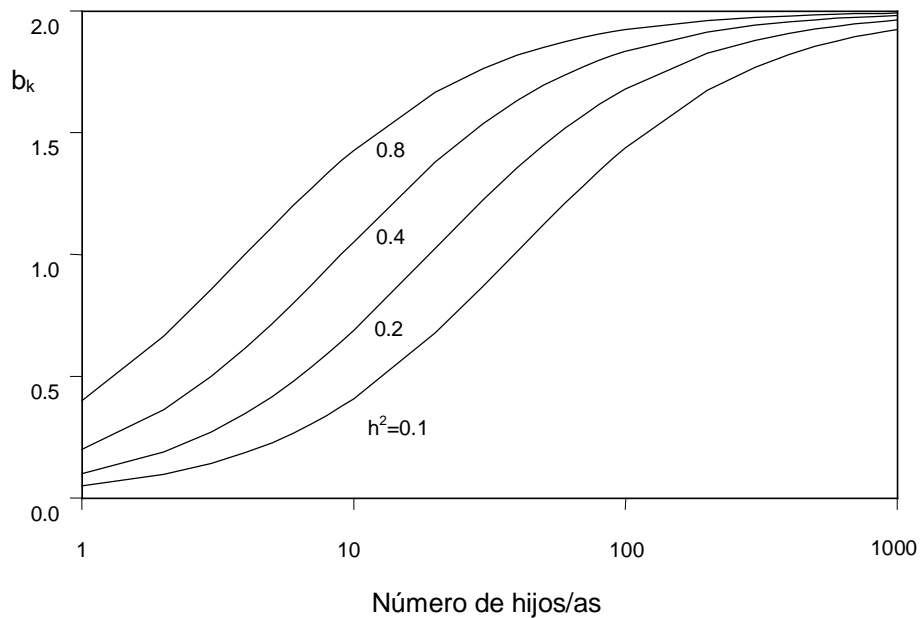


Figura 1. Valores de  $b_k$  en función del número de hijos/as y la heredabilidad (note la escala logarítmica).

**Información: un registro de cada uno de  $n$  hermanos o medio hermanos del candidato**

$$V = \sigma_{10}$$

Las fuentes de información son los promedios de las mediciones de  $n$  hermanos o medio hermanos.

El índice es de la forma:

$$I_k = b (\bar{x}_k - \bar{x})$$

mientras que:

$$\hat{b} = \frac{a_{ij} n h^2}{1 + (n-1)t}$$

$$\hat{f}_{VI} = \sqrt{a_{ij} \cdot \hat{b}}$$

donde  $a_{ij}$  indica relación de parentesco y será igual a  $1/2$  en el caso de hermanos completos o  $1/4$  en el caso de medio hermanos.

Cuando puede asumirse que  $c^2$  tiene un valor despreciable, las expresiones correspondientes a  $b$ , para información de  $n$  medio hermanos (MH) o hermanos completos (HC), quedan:

$$\hat{b}_{MH} = \frac{nh^2}{4 + (n-1)h^2}$$

$$\hat{b}_{HC} = \frac{nh^2}{2 + (n-1)h^2}$$

## Indices para mejorar más de una característica

Para considerar objetivos constituidos por más de una característica es conveniente utilizar álgebra matricial. La información básica será:

$\mathbf{x} = [x_{1k} \ x_{2k} \ \dots \ x_{nk}]$  un vector de longitud  $n$  constituido por registros fenotípicos de las  $n$  variables o fuentes de información que contendrá el índice, medidas en el  $k$ -ésimo candidato o en alguno de sus parientes y expresadas como desvíos.

$\mathbf{y} = [y_{1k} \ y_{2k} \ \dots \ y_{mk}]$  un vector de longitud  $m$  constituido por valores reproductivos del individuo  $k$  para cada uno de los  $m$  caracteres incluidos en el objetivo y expresados como desvíos.

Si se integran estas variables en único vector particionado  $[\mathbf{x}|\mathbf{y}]'$ :

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \\ \hline y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \dots & y_{mp} \end{bmatrix}$$

$n$  fuentes de información

$m$  caracteres a mejorar

$(n+m) \times p$

$p$  individuos

La matriz de varianzas y covarianzas de este vector, calculada aplicando al vector el operador varianza, será:



$$I = \mathbf{b}'\mathbf{x}$$

y su varianza será:

$$\sigma_I^2 = \hat{\mathbf{b}}'\mathbf{P}\hat{\mathbf{b}}$$

La función objetivo tendrá la forma:  $V = \mathbf{v}'\mathbf{y}$

y su varianza:

$$\sigma_V^2 = \mathbf{v}'\mathbf{C}\mathbf{v}$$

La expresión general para las ecuaciones a resolver será:

$$\begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_nx_1} & \sigma_{x_nx_2} & \cdots & \sigma_{x_n}^2 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_n \end{bmatrix} = \begin{bmatrix} \sigma_{y_1y_1} & \sigma_{y_1y_2} & \cdots & \sigma_{y_1y_m} \\ \sigma_{y_2y_1} & \sigma_{y_2y_2} & \cdots & \sigma_{y_2y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{y_ny_1} & \sigma_{y_ny_2} & \cdots & \sigma_{y_ny_m} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_m \end{bmatrix}$$

**P**                      **b̂**                      **G**                      **v**

que es la versión matricial del sistema de ecuaciones indicado en [5] y tiene soluciones:

$$\hat{\mathbf{b}} = \mathbf{P}^{-1}\mathbf{G}\mathbf{v}$$

La covarianza entre I y V será:

$$\sigma_{VI} = \hat{\mathbf{b}}'\mathbf{G}\mathbf{v} = \hat{\mathbf{b}}'\mathbf{P}\hat{\mathbf{b}}$$

y la  $r_{VI}$ :

$$\hat{r}_{VI} = \frac{\sigma_{VI}}{\sigma_V^2} = \frac{\hat{\mathbf{b}}'\mathbf{P}\hat{\mathbf{b}}}{\mathbf{v}'\mathbf{C}\mathbf{v}}$$

El valor relativo de cada fuente de información (VR) en el índice puede calcularse como:

$$VR_i = 100 - 100 \cdot \sqrt{\frac{\hat{\mathbf{b}}'\mathbf{P}\hat{\mathbf{b}} - [\hat{b}_i^2 / (\mathbf{P}^{-1})_{ii}]}{\hat{\mathbf{b}}'\mathbf{P}\hat{\mathbf{b}}}}$$

donde el valor relativo de la  $i$ ésima variable ( $VR_i$ ) se cuantifica como la reducción porcentual en progreso genético esperado que se verificaría si dicha variable fuera retirada del índice y  $(\mathbf{P}^{-1})_{ii}$  representa el elemento de la diagonal principal correspondiente a  $b_i$  en la inversa de  $\mathbf{P}$ .



$$[0.333 \quad 1.0 \quad 2.0] \begin{bmatrix} 302.5891 & 13.5093 & 0.0 \\ 13.5093 & 2.2391 & 0.0 \\ 0.0 & 0.0 & 15.2677 \end{bmatrix} \begin{bmatrix} 0.333 \\ 1.0 \\ 2.0 \end{bmatrix} = 105.8609$$

La correlación entre índice y objetivo será:

$$\hat{r}_{VI} = \frac{\sigma_I^2}{\sigma_V^2} = \frac{16.3646}{105.8609} = 0.3932$$

A título de ejemplo, el valor relativo como criterio de selección del peso del candidato a los 180 días será:

$$VR_1 = 100 - 100 \cdot \sqrt{\frac{16.3646 - [0.1350^2/0.0016]}{16.3646}} = 45.59\%$$

Si esa variable fuera removida del índice, el progreso genético del objetivo se reduciría en un 45.59 %. Los valores correspondientes a  $x_2$  y  $x_3$  son 2.33 y 7.93 %, respectivamente.

A continuación se transcribe una salida de computación correspondiente al ejemplo anterior obtenida utilizando el programa SELIND de Cunningham & Mahon (1977). Los resultados difieren ligeramente de los anteriores debido a diferencias en los valores de las varianzas y covarianzas utilizados.

\* SELECTION INDEX - INPUT DATA \* 3 VARIATES 3 TRAITS

- 
- 1 PESO A LOS 180 DIAS
  - 2 TIPIFICACION
  - 3 PRODUCTIVIDAD

\*\* P-MATRIX \*\*

* J *	* K *	* COVARIANCE *
1	1	1015.05957
1	2	93.50660
1	3	-7.43230
2	2	22.84840
2	3	-3.76340
3	3	94.47840

1 PESO 180 DIAS	0.33333
2 TIPIFICACION	1.00000
3 PROD. DE LA MADRE	2.00000

**\*\* G-MATRIX \*\***

* J *	* K *	* COVARIANCE *
1	1	302.58929
1	2	13.50930
1	3	0.00000
2	1	13.50930
2	2	2.23910
2	3	0.00000
3	1	0.00000
3	2	0.00000
3	3	7.63390

**\*\* C-MATRIX \*\***

* J *	* K *	* COVARIANCE *
1	1	302.58911
1	2	13.50930
1	3	0.00000
2	2	2.23910
2	3	0.00000
3	3	15.26770

\*\*\*\*\*

**SINGLE STAGE SELECTION**

\*\*\*\*\*

**\*SELECTION INDEX - MAIN INDEX\***

-----

*VARIATE*	*B-VALUE*	*VALUE OF VARIATE* (1)
1 PESO 180 DIAS	0.1352	45.6154
2 TIPIFICACION	-0.2312	2.3331
3 PRODUCTIVIDAD	0.1630	7.9149

VARIANCE OF INDEX: 16.3896

STD DEV OF INDEX: 4.0484 (2)

VARIANCE OF AGG GENOT: 105.9364

STD DEV OF AGG GENOT: 10.2925

CORRELATION OF INDEX AND AGG GENOTYPE: 0.3933

REGRESSION OF EACH TRAIT ON INDEX, CORRELATION OF EACH TRAIT WITH INDEX,  
PERCENT OF ECONOMIC GAINS. (3)

*TRAIT*	*REGRESSION*	*CORRELATION*	*PCT. OF GAIN*
1 PESO PROP. 180 DIAS	2.3050	0.5364	76.8311
2 TIPIFICACION PROP	0.0798	0.2160	7.9824
3 PROD. DE LA MADRE	0.0759	0.0787	15.1865

-----SUBINDEX FOR TRAIT: **PESO 180 DIAS**

VARIANCE OF SUBINDEX: 104.7070 CORRELATION OF SUBINDEX AND TRAIT: 0.5882

*VARIATE*	*B-VALUE*	*VALUE OF VARIATE* (1)
1 PESO A LOS 180 DIAS	0.3912	72.2894
2 TIPIFICACION	-1.0112	7.1579
3 PRODUCTIVIDAD	-0.0095	0.0040

-----SUBINDEX FOR TRAIT: **TIPIFICACION**

VARIANCE OF SUBINDEX: 0.2503 CORRELATION OF SUBINDEX AND TRAIT: 0.3344

*VARIATE*	*B-VALUE*	*VALUE OF VARIATE* (1)
1 PESO A LOS 180 DIAS	0.0068	6.0696
2 TIPIFICACION	0.0706	15.2298
3 PRODUCTIVIDAD	0.0033	0.2103

-----SUBINDEX FOR TRAIT: **PRODUCTIVIDAD**

VARIANCE OF SUBINDEX: 0.6216 CORRELATION OF SUBINDEX AND TRAIT: 0.2018

*VARIATE*	*B-VALUE*	*VALUE OF VARIATE* (1)
1 PESO A LOS 180 DIAS	-0.0010	0.0535
2 TIPIFICACION	0.0176	0.3532
3 PRODUCTIVIDAD	0.0814	99.9690

(1) VALUE OF EACH VARIATE IN THE INDEX = PERCENT REDUCTION IN RATE OF OVERALL GENETIC GAIN IF THAT VARIATE IS OMITTED.

(2) THIS IS THE VALUE, IN ECONOMIC UNITS, OF THE GENETIC GAIN IN AGREGATE GENOTYPE ACHIEVED BY ONE STANDARD DEVIATION OF SELECTION ON THE INDEX.

(3) THESE FIGURES ARE THE PERCENTAGES OF TOTAL GAIN (2) ACCOUNTED FOR BY GAIN IN EACH TRAIT.

## Bibliografía

- Brascamp, E.W. 1984. Selection indices with constraints. *Anim. Breed. Abs.* 52: 645-654.
- Cunningham, E.P. 1975. Multi-stage index selection. *Theor. Appl. Gen.* 46: 55-61.
- Cunningham, E.P. and Mahon, G.A.T. 1977. SELIND – A FORTRAN computer program for genetic selection indexes. User's Guide. An Foras Taluntais, Dunsinea, Caatleknock, Dublin and Dublin University. Ireland.
- INTA. 1998. Evaluación genética de carneros Merino en el Centro de Prueba INTA – Pilcaniyeu. Informe No. 4.
- Ponzoni, R.W. 1979. Objectives and selection criteria for Australian Merino sheep. *Proc. Aust. Assoc. Anim. Breeding and Genetics* 1: 320-336.
- Ponzoni, R.W. 1981. Studies on selection objectives and indices for Australian Merino sheep. Animal Industry Division Technical Information Circular No. 56. Department of Agriculture. South Australia.
- Ponzoni, R.W. 1988. Accounting for both income and expense in the development of breeding objectives. *Proc. Aust. Assoc. Anim. Breeding and Genetics* 7: 55-66.
- Ponzoni, R.W. and Newman, S. 1989. Developing breeding objectives for Australian beef cattle production. *Anim. Prod.* 49:35-47.
- Ponzoni, R.W. 1988. Updating of economic values and phenotypic and genetic parameters in WOOLPLAN. *Wool Tech. and Sheep. Breed.* July 1988 issue. pp. 70-75.
- Simm, G., Smith, C. and Prescott, J.H.D. 1986. Selection indices to improve the efficiency of lean meat production in cattle. *Anim. Prod.* 42: 183-193.