

Selección por reducción de varianza

N. Ibáñez*, ***, D. Sorensen**, A. Blasco***

* Àrea de Producció Animal - Centre UdL-IRTA, 25198 Lleida, España.

** Danish Institute of Agricultural Sciences, Department of Animal Breeding and Genetics, PB50, 8830 Tjele, Denmark.

*** Departamento de Ciencia Animal, Universidad Politécnica de Valencia. P.O. Box 22012, 46071 Valencia, España.

Resumen

Se ha derivado una fórmula aproximada para predecir la respuesta de la selección por reducción de varianza del tamaño de camada en conejos con un modelo que postula que una parte de la varianza ambiental esta controlada genéticamente. Esta fórmula ha sido evaluada mediante simulaciones MCMC. La semejanza entre la respuesta predicha y la simulada en diez generaciones de selección es muy alta cuando la intensidad de selección es baja o moderada, pero menor cuando la intensidad de selección aumenta. También se presenta una evaluación de la potencia estadística para detectar la respuesta por reducción de varianza en tres diseños de selección utilizados para el tamaño de camada en conejo. La potencia aumenta con la intensidad de selección para los tres diseños y con la duración del experimento. La selección divergente es el diseño más eficiente para detectar la respuesta, seguido de una línea seleccionada y del diseño con una línea seleccionada y otra control. Para la selección divergente, se obtiene una potencia del 80% después de 6 generaciones con presiones de selección alta e intermedia.

Palabras clave: Varianzas heterogéneas, Respuesta a la selección, Conejos

Summary

Selection for reduced variance

An approximate formula to predict response to selection for reduced variance of litter size in rabbits is derived under a model postulating that the environmental variance is partly under genetic control. The formula is evaluated using computer simulations. The agreement between predicted and observed results over a period of ten generations of selection is excellent if selection intensity is low to moderate, and less so, as selection increases. An evaluation of the statistical power to detect response to selection for reduced variance under three designs mimicking a prospective experiment involving litter size in rabbits is also reported. Power increases with selection intensity for all three designs and with the duration of the experiment. Divergent selection is the most efficient design followed by a single selected line and then by the design involving a selected and a control line. For divergent selection, power approaches 80% after 6 generations of selection at intermediate and high selection intensities.

Key words: Heterogeneous variances, Response to selection, Rabbits

Introducción

En producción animal, la selección se ha centrado principalmente en aumentar la media de un carácter, como el caso del tamaño de camada en especies prolíficas (cerdos y conejos). Sin embargo, debido a la baja heredabilidad del carácter, la selección ha tenido poco éxito, muy inferior al esperado. Reducir la varianza ambiental del tamaño de camada implicaría aumentar su heredabilidad y con ello sus posibilidades de éxito en la selección. Además es bien conocido que la elevada varianza del tamaño de camada puede tener importantes consecuencias económicas derivadas de la escasa viabilidad de camadas muy numerosas y de las dificultades de manejo de camadas de tamaño muy distinto. Sería deseable que una coneja produjera altos tamaños de camada y que estos fueran similares en todos sus partos. Recientemente hay trabajos que muestran la existencia de mecanismos moleculares afectando la variación fenotípica (Queitsch, C., Sangster, T.A., Lindquist, S., 2002). También, hay trabajos en genética cuantitativa indicando que parte de la variación ambiental puede estar bajo control genético (Sorensen y Waagepetersen 2003 y Ros et al., 2004). Estos estudios son importantes porque abren la posibilidad de controlar la variación, fenotípica o ambiental, de un carácter a través de la selección. En el caso del tamaño de camada, la mayor parte de la variación fenotípica es ambiental, siendo la varianza genética una parte inferior al 10% del total, y en muchas ocasiones inferior al 5% de la variación total.

En este trabajo se presenta una fórmula aproximada para predecir la respuesta a la selección por variación fenotípica y se testa su precisión usando simulación MCMC. También se discuten diferentes diseños de un experimento de selección para reducir la

varianza, teniendo en cuenta la potencia del test en cada uno de los diseños.

Material y métodos

Modelo genético

San Cristobal-Gaudy et al. (1998) propusieron un modelo en el que la varianza residual tiene estructura genética. Se asume que la distribución de muestreo de los datos es Gaussiana, con forma

$$y | \mathbf{b}, \mathbf{a}, \mathbf{a}^*, \mathbf{p}, \mathbf{p}^* \sim N(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p}, \exp(\mathbf{X}'\mathbf{b}^* + \mathbf{Z}'\mathbf{a}^* + \mathbf{W}'\mathbf{p}^*)) \quad (1)$$

Donde los vectores \mathbf{b} y \mathbf{b}^* contienen efectos ambientales (en este caso, efecto generación), y \mathbf{X} , \mathbf{Z} , \mathbf{W} son matrices de incidencia conocidas. Los efectos aditivos \mathbf{a} y \mathbf{a}^* se asumen distribuidos normalmente

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{a}^* \end{bmatrix} | G \sim N(0; G \otimes A), \quad G = \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{a^*} \\ \rho\sigma_a\sigma_{a^*} & \sigma_{a^*}^2 \end{bmatrix}, \quad (2)$$

donde \mathbf{A} es la matriz de relaciones genéticas, σ_a^2 es la varianza aditiva genética de la media del carácter, $\sigma_{a^*}^2$ es la varianza genética aditiva de la varianza ambiental del carácter, ρ es el coeficiente de correlación genética y \otimes corresponde al producto de Kronecker. Los vectores \mathbf{p} y \mathbf{p}^* contienen los efectos ambientales permanentes para la media y la varianza ambiental del carácter respectivamente, y se asumen independientemente con distribuciones,

$$\begin{aligned} \mathbf{p} | \sigma_p^2 &\sim N(0, I_p \sigma_p^2), \\ \mathbf{p}^* | \sigma_{p^*}^2 &\sim N(0, I_{p^*} \sigma_{p^*}^2) \end{aligned} \quad (3)$$

donde I_p es la matriz con tamaño igual al número de hembras con camadas y σ_p^2 y $\sigma_{p^*}^2$ son las varianzas de los ambientes permanentes. Más detalles pueden encontrarse en Sorensen y Waagepetersen (2003).

Predicción de la respuesta genética a la selección

Se presupone que el objetivo de la selección es reducir la varianza ambiental. El criterio de selección empírico es la varianza de datos de una hembra, en este caso la varianza del número de gazapos entre los partos de una hembra. La selección direccional se describe usando la siguiente función lineal de la "fitness"

$$w(y) = 1 - sS^2 \tag{4}$$

donde $S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$ es el criterio de selección, s es una pequeña cantidad que describe la intensidad de la selección artificial y n es el número de datos por hembra. La "fitness" del genotipo (\mathbf{a}, \mathbf{a}^*) se define como

$$w(\mathbf{a}, \mathbf{a}^*) = E[w(y) | \mathbf{a}, \mathbf{a}^*] = 1 - sE(S^2 | \mathbf{a}, \mathbf{a}^*) = 1 - s \exp(\mathbf{b}^* + \mathbf{a}^* + \mathbf{p}^*)$$

La respuesta a la selección de \mathbf{a}^* en un ambiente determinado (condicionado en \mathbf{b}^* y a los componentes de varianza) es,

$$R_{a^*} = \iint \mathbf{a}^* \frac{w(\mathbf{a}, \mathbf{a}^*)}{\bar{w}} p(\mathbf{a}, \mathbf{a}^*) d\mathbf{a} d\mathbf{a}^* = -\frac{s}{\bar{w}} \sigma_{a^*}^2 \exp\left(Xb^* + \frac{\sigma_{a^*}^2}{2} + \frac{\sigma_{p^*}^2}{2}\right) - \frac{s}{\bar{w}} \sigma_{a^*}^2 [(1-h^2)\sigma^2 - \sigma_{p^*}^2]$$

donde $h^2 = \sigma_a^2 / \sigma^2$ y $\sigma^2 = \sigma_a^2 + \exp(Xb^* + \frac{\sigma_{a^*}^2}{2} + \frac{\sigma_{p^*}^2}{2})$. La varianza (condicional) del criterio de selección S^2 es

$$Var(S^2) = [(1-h^2)\sigma^2 - \sigma_{p^*}^2] \left[\left(1 + \frac{2}{n-1}\right) \exp(\sigma_{a^*}^2) - 1 \right]$$

Escribiendo la intensidad de selección $i = \frac{s}{w} \sqrt{Var(S^2)}$ (San Cristobal-Gaudy et al., 1998), se deriva, que la respuesta en \mathbf{a}^* debida a la selección basada en el criterio S^2 viene dada por la siguiente expresión

$$R_{a^*} = -i \frac{\sigma_{a^*}^2}{\sqrt{\left(1 + \frac{2}{n-1}\right) \exp(\sigma_{a^*}^2 + \sigma_{p^*}^2) - 1}} \tag{5}$$

La respuesta predicha a tiempo t es aproximadamente tRa^* .

Predicción de la respuesta correlacionada de la media

La respuesta correlacionada de \mathbf{a} , dado \mathbf{b}^* y los componentes de varianza es

$$CR_a = \iint \mathbf{a} \frac{w(\mathbf{a}, \mathbf{a}^*)}{\bar{w}} p(\mathbf{a}, \mathbf{a}^*) d\mathbf{a} d\mathbf{a}^* = -\frac{s}{\bar{w}} \frac{\sigma_a}{\sigma_{a^*}} \rho \sigma_{a^*}^2 \exp\left(Xb^* + \frac{\sigma_{a^*}^2}{2} + \frac{\sigma_{p^*}^2}{2}\right) = R_{a^*} \frac{\sigma_a}{\sigma_{a^*}} \rho$$

La predicción de la respuesta correlacionada a tiempo t es aproximadamente tCR_a .

Estudio de simulación

Se realizó un experimento de simulación con el objetivo de testar la precisión de las formulas expuestas anteriormente, además de estudiar la potencia del test para detectar la respuesta a la selección por reducción de varianza en tres diseños diferentes. Estos tres diseños tienen el mismo número de animales y corresponden a una sola línea seleccionada, a una línea seleccionada y una línea control y a dos líneas seleccionadas una hacia arriba y otra hacia abajo. En cada uno de los tres diseños se aplicó tres intensidades de selección: baja (Proporción de seleccionados $P = 50\%$), intermedia ($P = 33\%$) y alta ($P = 15\%$). La selección fue aplicada en un solo sexo. La función lineal de la "fitness" (4), es una aproximación del modelo de selección utilizado en simulación, por lo que se espera un mejor comportamiento para bajas intensidades de selección.

Los datos fueron generados por el modelo especificado en (1), (2), y (3). Los valores genéticos aditivos de la descendencia fueron generados usando las reglas mendelianas estándar bajo el modelo infinitesimal y teniendo en cuenta el coeficiente de consanguinidad.

El número de animales usado por generación en los tres diseños fue de 200, (40 machos y 160 hembras). Este tamaño corresponde al espacio disponible en una granja experimental de la Universidad de Orihuela, donde un experimento de selección para reducir la varianza del tamaño de camada en conejo se va a realizar. Cada una de las 160 hembras producen tres datos y el índice S^2 es calculado para cada hembra. La descendencia para crear la siguiente generación fue tomada de hembras (madres) con los menores valores de S^2 .

Los parámetros utilizados en la simulación de los datos proceden de un análisis Bayesia-

no de datos de tamaño de camada en conejo aplicando el modelo descrito por (1), (2) y (3). Estos datos provienen de un experimento de selección de 20 generaciones. Las estimas de las medias posteriores de los parámetros relevantes fueron las siguientes; $\mathbf{b}=10$, $\mathbf{b}^*=2.25$, $\sigma_a^2=1.28$, $\sigma_{a^*}^2=0.14$, $\rho=-0.75$, $\sigma_p^2=0.7$ and $\sigma_{p^*}^2=0.15$.

La respuesta a la selección (verdadera) para la varianza de un carácter en una generación dada, se definió como la media de los valores genéticos aditivos \mathbf{a}^* pertenecientes a cada generación. De manera similar, la respuesta correlacionada a la selección (verdadera) de la media corresponde a la media de los valores de \mathbf{a} .

Potencia del test

La potencia del test se calculo a través de una estima Monte Carlo, obtenida como el porcentaje de replicas en cada generación en las que la respuesta a la selección fue "detectada". Esto implica simular datos bajo la hipótesis nula (una muestra aleatoria de machos y hembras con respecto S^2 y fijar el nivel de significación al 5% de error de tipo I) y bajo la hipótesis alternativa (la selección basada en S^2).

Resultados

Ejemplos de las respuestas aditivas predichas y observadas (verdaderas) se dan en la figura. Las respuestas observadas (medias de 200 replicas) corresponden a las líneas continuas, y las respuestas predichas a las líneas con puntos remarcados. La figura muestra claramente que para intensidades de selección relativamente bajas o medias ($P = 50\%$ o $P = 33\%$) la respuestas predichas y observadas son similares (las dos líneas se sobreponen). Cuando la intensidad de selec-

ción es relativamente alta ($P = 15\%$), las predicciones son menos precisas. Los valores de la potencia del test para detectar la respuesta a la selección por reducción de varianza (respuesta en S^2) en los tres diseños y para las tres presiones de selección se muestran en la tabla 1. La potencia aumenta con la intensidad de selección y con la duración del experimento de selección para los tres diseños. La selección divergente (D3) es la más eficiente seguida del diseño con solo

una línea de selección (D1) y del diseño con dos líneas una seleccionada y otra control. Para D3, la potencia es del 80% después de 6 generaciones con presiones de selección alta e intermedia. Para D1, las potencias son aceptables a partir de la 7 generación y con intensidad de selección alta. La potencia para detectar respuesta en D2 es de 82%, después de 10 generaciones y solo en el caso de aplicar una alta intensidad de selección.

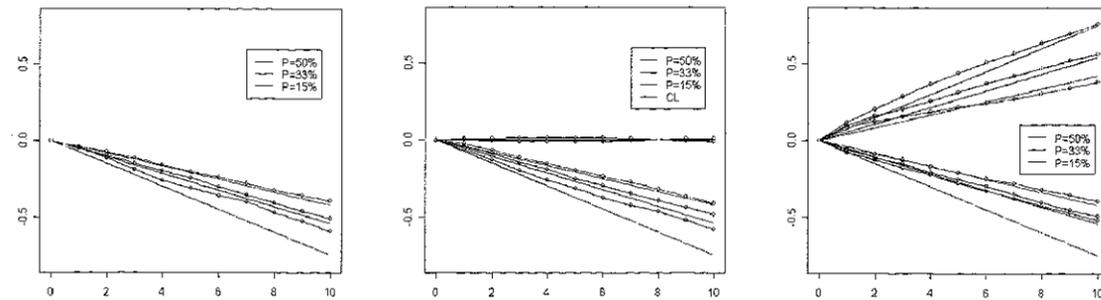


Figura 1. Respuesta a la selección para diferentes diseños de selección e intensidad de selección.
Figure 1. Response to selection for different selection design and selection intensity.

Tabla 1. Potencia del test
Table 1. Power of the test

T	P = 50%			P = 33%			P = 15%		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
1	0.01	0.03	0.04	0.02	0.03	0.06	0.04	0.02	0.06
3	0.05	0.07	0.23	0.04	0.09	0.30	0.13	0.13	0.39
5	0.08	0.13	0.54	0.16	0.22	0.65	0.22	0.28	0.71
7	0.21	0.21	0.67	0.41	0.28	0.80	0.44	0.38	0.91
10	0.45	0.37	0.83	0.62	0.46	0.96	0.77	0.49	1.00

D1, D2 y D3 corresponden a un experimento con una línea de selección, a un experimento con una línea de selección y otra control, y a un experimento de selección divergente, respectivamente.

Conclusiones

Se desarrollo una expresión para predecir la respuesta a la selección por varianza y se estudio su validez usando simulación. Las respuestas predichas son satisfactorias cuando la intensidad de selección no es muy alta. También se estudio la potencia de test para detectar la respuesta a la selección por la varianza, usando una estructura familiar determinada y usando tres diseños de selección. En este estudio se muestra que al menos son necesarias 7 generaciones de selección para que la probabilidad de la detección de la respuesta sea aceptable. La selección divergente es el mejor diseño, seguido del diseño de una sola línea de selección.

Si se mantuviera la media estable mientras se intenta controlar la varianza, se necesitaría realizar un experimento de mayor tamaño para obtener una adecuada potencia. Desde un punto de vista de genética cuantitativa, es deseable entender la evolución de σ_a^2 y de $\sigma_{a^*}^2$ bajo el modelo presentado, así

como la evolución de S^2 . Actualmente estamos trabajando en el desarrollo de estos temas.

Bibliografía

- Ros M, Sorensen D, Waagepetersen R, Dupont-Nivet M, SanCristobal M, Bonnet JC, Mallard RJ, 2004. Evidence for genetic control of adult weight plasticity in the snail *helix aspersa*. *Genetics*, 168: 2089-2097.
- Queitsch C, Sangster TA, Lindquist S, 2002. Hsp90 as a capacitor of phenotypic variation. *Nature*. 417: 618-624.
- SanCristobal M, Elsen JM, Bodin L, Chevalet C, 1998. Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genet. Sel. Evol.* 30: 423-451.
- Sorensen D, Waagepetersen R, 2003. Normal linear models with genetically structured residual variance heterogeneity: a case study. *Genet. Res.* 82: 207-222.

(Aceptado para publicación el 2 de mayo de 2006)