

## ¿Es razonable un residuo simétrico?

L. Varona, N. Ibáñez, R.N. Pena, J.L. Noguera, R. Quintanilla

Àrea de Producció Animal. Centre UdL-IRTA. Av. Rovira Roure 191. 25198. Lleida.

E-mail: Luis.Varona@irta.es

### Resumen

Los modelos estadísticos habituales en la evaluación genética de reproductores asumen distribuciones simétricas para todos los parámetros aleatorios. En los últimos años se ha producido un desarrollo teórico de distribuciones asimétricas. El objetivo de este trabajo es estudiar la adecuación de las distribuciones asimétricas para modelizar el carácter tamaño de camada en porcino. Para ello, se han analizado 2.072 datos de número de lechones nacidos vivos, producidos por 657 cerdas y con una genealogía disponible de 765 triada de individuo-padre-madre. En todos los análisis se incluyeron en el modelo los efectos orden de parto y rebaño-año-estación como efectos sistemáticos, y los efectos genético aditivo y ambiental permanente como efectos aleatorios con distribución normal. Para los efectos residuales se consideraron tres distribuciones diferentes: 1) Normal simétrica, 2) Normal asimétrica y 3) *t* asimétrica. Los tres modelos estadísticos se compararon mediante un Factor de Bayes. El modelo más adecuado correspondió a la distribución Normal asimétrica (Modelo 2). La media posterior de la heredabilidad fue 0.063, con una desviación estándar posterior de 0.028. La distribución de residuos obtenida fue fuertemente asimétrica, indicando que las fuentes de variación no controladas por el modelo tuvieron, en su mayor parte, una influencia negativa sobre la prolificidad.

**Palabras clave:** Evaluación genética, inferencia bayesiana, distribuciones asimétricas, prolificidad

### Summary

#### ¿Why residuals are symmetric in genetic evaluation?

Statistical models for genetic evaluation make use of Gaussian or *t* residual distributions. However, some new statistical development allows using asymmetric distribution for the residual not controlled by the model. We have analysed a data set of litter size on pigs consisting of 2072 data for number of piglets born alive from 657 sows, with a total pedigree of 765 individual-sire-dam triplets. The model includes order of parity and herd-year-season systematic effects, Gaussian additive genetic effects and Gaussian permanent environmental effects. In addition, we use three different distributions for the residuals. 1) Symmetric Gaussian distribution, 2) Asymmetric Gaussian distribution and 3) Asymmetric *t* distribution. The three statistical models were compared using a Bayes Factor. The most suitable model corresponds to the asymmetric Gaussian distribution (Model 2). The posterior mean of heritability was 0.063, with a posterior standard deviation of 0.028. The resulting distributions of the residuals are strongly asymmetric, indicating that the sources of variation not controlled by the model have mostly a negative influence on the prolificacy. The asymmetry parameter can be understood as a measure of sensibility to negative environmental influences on the phenotype.

**Key words:** Genetic evaluation, bayesian inference, asymmetric distributions, prolificacy

## Introducción

Los modelos habituales de evaluación de reproductores asumen generalmente un modelo donde los efectos desconocidos se distribuyen mediante una distribución normal (Henderson, 1984). Esta asunción implica que los efectos no controlados por el modelo son muchos y de pequeña magnitud, y convergen hacia la distribución normal por el teorema central del límite. Strandén y Gianola (1999) propusieron sustituir la distribución normal por una distribución más robusta como la distribución **t**, que permite la presencia de valores más extremos y evita el sesgo provocado por el tratamiento preferencial.

Tras generaciones de selección, las poblaciones productivas de las especies domésticas han incrementado su potencial genético considerablemente. Por lo tanto, bajo condiciones ambientales adecuadas, producen rendimientos que resultaban impensables antes de la selección. Dado su patrimonio genético, es difícil imaginar un fuerte incremento productivo causado por efectos no contemplados en el modelo, pero por el contrario resulta sencillo pensar en múltiples causas que conlleven una reducción drástica del rendimiento (patologías, olas de calor o de frío, estrés, etc.). Este razonamiento nos sugiere que un modelo que contemple residuos de naturaleza asimétrica posiblemente pueda ser más apropiado como modelo de evaluación de reproductores.

En los últimos años se han desarrollados varios procedimientos de modelización de distribuciones no simétricas (Fernández et al., 1998; Sahu et al., 2003; Jara y Quintana, 2005). En este trabajo se presenta una evaluación de modelos con residuos asimétricos con datos de prolificidad en porcino.

## Material y métodos

Se han utilizado 2.072 datos de prolificidad (número de lechones nacidos vivos) producidos por 657 cerdas de una población comercial de selección con una media fenotípica de 8.59. Se utilizó una genealogía consistente en 765 triadas individuo-padre-madre. Los datos fueron distribuidos en 6 órdenes de parto (1º, 2º, 3º, 4º, 5º y 6º o superior), y en 43 efectos rebaño-año-estación.

Los datos se analizaron utilizando el siguiente modelo animal:

$$y = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{Z}\mathbf{p} + \mathbf{e}$$

Donde se incluyeron los efectos orden de parto y rebaño-año-estación como efectos sistemáticos ( $\beta$ ), además de los efectos genético aditivo ( $\mathbf{u}$ ) y ambiental permanente ( $\mathbf{p}$ ). Las distribuciones *a priori* de los efectos sistemáticos se asumieron uniformes acotadas, y las distribuciones de los efectos genético aditivo y ambiental permanente se asumieron normales multivariantes.

Además, se contemplaron tres distribuciones distintas para el residuo:

1) Distribución normal simétrica:

$$y \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{I}\sigma_e^2)$$

2) Distribución normal asimétrica:

$$y \sim SN(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{I}\sigma_e^2, \lambda)$$

3) Distribución **t** asimétrica:

$$y \sim ST_v(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{I}\sigma_e^2, \lambda)$$

Los modelos se compararon mediante un Factor de Bayes para modelos jerárquicos (García-Cortés et al., 2001; Varona et al., 2001).

## Resultados

Los resultados de la comparación de modelos se presentan en la tabla 1. Puede obser-

vase que el Factor de Bayes entre los modelos que asumieron distribuciones asimétricas en el residuo (modelos 2 y 3) y el modelo habitual de evaluación de reproductores (modelo 1) tomó valores superiores a  $10^{44}$ .

El factor de Bayes entre los modelos con residuos normales asimétricos (modelo 2) y t asimétricos (modelo 3) no fue relevante, aunque favoreció levemente al modelo normal asimétrico.

Tabla 1. Factor de Bayes entre los tres modelos (a) comparados con distinta distribución para el efecto residual

Table 1. Bayes Factor comparison between the three models (a) with different residual distributions

	Modelo 1	Modelo 2	Modelo 3
Modelo 1	1.00	$4.89 \times 10^{-43}$	$7.41 \times 10^{-43}$
Modelo 2	$2.04 \times 10^{44}$	1.00	1.51
Modelo 3	$1.35 \times 10^{44}$	0.66	1.00

(a) Modelo 1: distribución normal simétrica; Modelo 2: distribución normal asimétrica; Modelo 3: distribución t asimétrica.

La tabla 2 muestra un resumen de las distribuciones marginales posteriores de los parámetros de dispersión bajo el modelo 2. La heredabilidad posterior tomó valores entorno al 6%. La media posterior del parámetro de asimetría ( $\lambda$ ) fue considerablemente negativa (-2.81), con una desviación posterior de 0.2. La distribución de los efectos residuales condicionados a las medias

posteriores de la varianza residual ( $\sigma_e^2$ ) y el parámetro de asimetría ( $\lambda$ ) se presenta en la figura 1. En dicha figura se aprecia que la distribución obtenida para los residuos fue fuertemente asimétrica, indicando que la fuentes de variación no controladas por el modelo tuvieron, en su mayor parte, una influencia negativa sobre la prolificidad.

Tabla 2. Medias Posterior (MP) y Desviaciones Estándar Posterior (DSP) de los parámetros de dispersión (a) bajo el modelo con distribución normal asimétrica para el residuo (modelo 2)  
Table 2. Posterior means (MP) and Posterior Standard Deviations of the variance components (a) under the model with a Gaussian Asymmetry Distribution

	$\sigma_a^2$	$\sigma_p^2$	$\sigma_e^2$	$h^2$	$\lambda$
MP	0.425	0.622	2.876	0.063	-2.808
DSP	0.196	0.212	0.366	0.028	0.200

(a)  $\sigma_a^2$ : varianza genética aditiva;  $\sigma_p^2$ : varianza ambiental permanente;  $\sigma_e^2$ : varianza residual;  $\lambda$ : parámetro de asimetría de la distribución de los efectos residuales;  $h^2$ : heredabilidad.

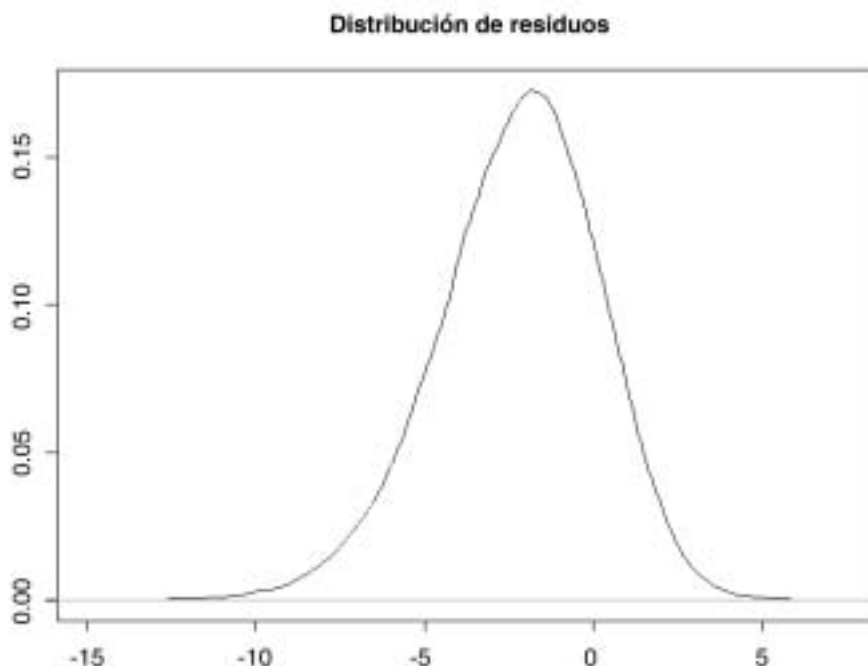


Figura 1. Distribución de los efectos residuales condicionados a las medias posteriores de la varianza residual ( $\sigma_e^2$ ) y el parámetro de asimetría ( $\lambda$ ).

Figure 1. Residual Distribution conditioned to the posterior means of the residual variance ( $\sigma_e^2$ ) and the asymmetry parameter ( $\lambda$ ).

## Discusión

Los resultados obtenidos en el presente trabajo muestran que, en la población analizada, los modelos con distribuciones asimétricas son más apropiados para la evaluación genética del carácter tamaño de camada. La comparación de los dos modelos con distribución asimétrica para los residuos no permite concluir de un modo taxativo cual es el más adecuado. De todas formas, sí parece que el modelo con distribución normal asimétrica, menos parametrizado, sería ligeramente más probable.

La distribución posterior obtenida corrobora la hipótesis previamente planteada sobre la distribución asimétrica negativa de los

residuos, los cuales recogerían aquellos factores no sistemáticos y de origen no genético que comportan una reducción de las producciones, como pudiera ser una eventual patología o una situación de estrés. En este sentido, la aplicación de modelos con residuos asimétricos en procesos de evaluación genética evitaría la penalización de aquellos reproductores que hayan podido sufrir una eventualidad de origen ambiental que determine un menor rendimiento, así como la penalización de sus parientes, al permitir residuos de mayor magnitud en sentido negativo.

Los resultados obtenidos para la heredabilidad (tabla 2) fueron similares a los obtenidos en otros trabajos sobre prolificidad en

porcino (Noguera et al., 2002), quedando comprobado que el uso de una distribución asimétrica influye en la predicción de los valores mejorantes pero no modifica sustancialmente los resultados en términos de heredabilidad.

Además, el modelo con distribuciones asimétricas ofrece información adicional. Por una parte, las estimas de los efectos sistemáticos están referidas a la potencialidad productiva en esa combinación ambiental. Por otra parte, el parámetro de asimetría se puede interpretar como la pérdida de producción debida a causas ambientales direccionales.

Señalaremos, finalmente, que una posible extensión del modelo consiste en la introducción de una jerarquía en el parámetro de asimetría ( $\lambda$ ), permitiendo que éste varíe entre efectos sistemáticos e incluso entre individuos. Esta modelización permitirá obtener valores mejorantes para la sensibilidad de los individuos ante las influencias ambientales direccionales.

## Bibliografía

- Fernández C, Steel MFJ, 1998. On bayesian modelling of fat tails and skewness. *JASA* 93: 359-371.
- García-Cortés LA, Cabrillo C, Moreno C, Varona L, 2001. Hypothesis testing for the genetic background of quantitative traits. *Genetics Selection Evolution* 33: 3-16.
- Henderson CR, 1984. *Applications of Linear Models in Animal Breeding*. University of Guelph.
- Jara A, Quintana F, 2005. Linear effects mixed models with skew-elliptical distributions: A Bayesian approach. Technical Report. y Quintana, 2005.
- Noguera JL, Varona L, Babot D, Estany J, 2002. Multivariate analysis of litter size for multiple parities with production traits in pigs. *J. Anim. Sci.* 80: 2540-2547.
- Sahu SK, Dey DK, Branco MD, 2003. A new class of multivariate skew distributions with applications to bayesian regression models. *The Canadian Journal of Statistics* 31: 129-150. et al., 2003.
- Stranden I, Gianola D, Mixed effects linear models with t-distributions for quantitative genetics analysis: a Bayesian approach. *Genetics Selection Evolution* 31: 25-42.
- Varona L, García-Cortés LA, Perez-Enciso M, 2001. Bayes Factor for detection of Quantitative Trait Loci. *Genetics Selection Evolution* 33: 133-152.

(Aceptado para publicación el 2 de mayo de 2006)