

# ESTUDIO DE ALGUNAS HIPOTESIS DE LOS ANALISIS DE VARIANZA APLICADOS A LA PRODUCCION LACTEA OVINA

Por J. A. Carriedo (1)  
F. San Primitivo (1)

## INTRODUCCION

En el análisis de varianza, para la realización de pruebas e interferencias estadísticas, frecuentemente se requiere que se cumplan determinadas hipótesis, como la de homoscedasticidad o igualdad de varianzas residuales y distribución normal de la variable aleatoria. En este estudio tratamos de contrastar estas hipótesis para la producción láctea ovina.

En relación al contraste de homoscedasticidad, contemplamos varios modelos matemáticos, que se diferencian en los factores de variación que incluyen.

Por último, se considera la metodología estadístico informática que permite, por un lado, discutir los supuestos de los modelos matemáticos, y por otro lado llevar a cabo los distintos tipos de análisis de varianza.

## MATERIAL Y METODOS

Hemos utilizado datos procedentes de tres rebaños de ovejas de raza Churra, pertenecientes a las Diputaciones Provinciales de Burgos y Palencia. Las variables contempladas han sido las producciones lácteas en cada uno de los cuatro primeros partos o lactaciones, normalizadas a 150 días de lactación, siguiendo para ello el método de Fleischmann. Los datos genealógicos se han verificado utilizando como marcadores genéticos los polimorfismos bioquímicos hemoglobina y transferrina.

---

(1) Cátedra de Genética y Mejora Animal

En la tabla n.º 1 figuran, para cada lactación, el número de observaciones consideradas y la media de producción láctea.

**TABLA 1**  
**Número de observaciones en cada número de parto o lactación, y media de producción láctea normalizada a 150 días**

Número de lactación o parto	Número de observaciones o efectivo	Media de producción por lactación (en litros)
1	1.379	109,0
2	975	122,6
3	730	129,4
4	394	131,7
1, 2, 3, y 4	3.482	119,7

En este trabajo, se ha contemplado un tipo de estructura de datos que permitiera estudiar factores de variación genéticos y ambientales. Considerando la información disponible en nuestro estudio (tabla n.º 1), las inferencias relativas a los parámetros genéticos de la producción láctea en la primera lactación serán las que tengan un mayor grado de precisión estadística, en relación a las estimaciones de los parámetros genéticos de la producción láctea de las lactaciones segunda a cuarta.

Este aspecto sobre el tamaño muestral creemos que se deberá considerar dado que, además, la distribución entre las distintas familias presenta un fuerte grado de desequilibrio, siendo este desequilibrio en la estructura de los datos aplicable también para los restantes factores de variación, y en especial para el factor año de parto.

El contraste de hipótesis de homoscedasticidad se ha referido a la homogeneidad de varianzas residuales, contemplándose seis tipos de análisis de varianza, correspondientes a los siguientes modelos lineales:

$$1.- y_{ijklm} = \mu + R_i + A_{j(i)} + M_k + RM_{ik} + S_{l(i)} + e_{ijklm}.$$

$$2.- y_{ijlmn} = \mu + R_i + A_{j(i)} + S_{l(i)} + T_m + e_{ijlmn}.$$

$$3.- y_{ijklm} = \mu + R_i + A_{j(i)} + M_k + RM_{ik} + T_l + e_{ijklm}.$$

$$4.- y_{ijkm} = \mu + R_i + A_{j(i)} + S_{k(i)} + e_{ijkm}.$$

$$5.- y_{ijkm} = \mu + R_i + A_{j(i)} + T_k + e_{ijkm}.$$

$$6.- y_{ijkm} = \mu + R_i + A_{j(i)} + M_k + RM_{ik} + e_{ijkm}.$$

Los factores rebaño, tipo de parto y estación o época de parto (R, T y M, respectivamente), fueron incluidos en los modelos como factores de variación principales. El año de parto y el factor semental (A y S, respectivamente), se consideraron ambos subordinados.

dos al rebaño. También se ha contemplado la interacción de la época de parto por el rebaño (RM).

Los factores de variación que se incluyeron en los modelos matemáticos se mostraron como estadísticamente significativos, cuando se analizaron siguiendo el método de mínimos cuadrados (LS).

Hemos seguido el test de Bartlett, con subclases desiguales, para contrastar la hipótesis de homogeneidad de varianzas residuales<sup>6</sup>.

En relación a la función de distribución probabilística de la producción láctea ovina, se han obtenido las estimaciones puntuales del sesgo y de la curtosis, así como sus correspondientes errores típicos, determinándose si diferencian significativamente del valor «cero» correspondiente a la función de distribución normal.

Las estimaciones del sesgo y de la curtosis se han obtenido llevando a cabo tres tipos de análisis (n.º 1, 2 y 3). El n.º 1 se caracterizó porque no se excluyó ningún dato. En los análisis n.º 2 y 3 se consideraron únicamente aquellas lactaciones con producciones superiores a 50 y 70 litros, respectivamente.

Hemos realizado programas informáticos para llevar a cabo el tratamiento de archi-

**TABLA 2**  
Estimaciones del sesgo obtenidas para cada número de parto o lactación, en los tres tipos de análisis estadísticos llevados a cabo

Número de lactación o parto	Tipo de análisis		
	1	2	3
1	0,47±0,07 ***	0,64±0,07 ***	0,88±0,07 ***
2	0,25±0,08 ***	0,38±0,08 ***	0,57±0,08 ***
3	0,18±0,09 *	0,31±0,09 ***	0,47±0,09 ***
4	0,09±0,12	0,14±0,12	0,30±0,13 *
1, 2, 3, y 4	0,32±0,04 ***	0,45±0,04 ***	0,63±0,04 ***

Nivel de significación estadística: \* P < 0,05 ;  
\*\* P < 0,01 ; \*\*\* P < 0,001 .

vos y el tratamiento estadístico, programas HARV, NOR y HOM, en los Centros de Cálculo del C.S.I.C. y del I.N.I.A. de Madrid.

## RESULTADOS

Se han obtenido las estimaciones del sesgo y de la curtosis, así como su grado de significación estadística, de la función de distribución probabilística de la producción láctea ovina, contemplándose 15 situaciones. Estas situaciones resultan de aplicar tres tipos de análisis, denominados 1, 2 y 3, a los datos de cada una de las cuatro primeras lactaciones y al conjunto de todos los datos (Material y Métodos).

En la tabla n.º 2, como cabría esperar se observa que al disminuir el número de datos introducidos en el análisis, aumentan las estimaciones puntuales del sesgo (tabla n.º 2), debido a la mayor proporción de datos desechados en una de las colas de la función de distribución.

**TABLA 3**

Estimaciones de la curtosis obtenidas para cada número de parto o lactación, en los tres tipos de análisis estadísticos llevados a cabo

Número de lactación o parto	Tipo de análisis		
	1	2	3
1	3,34 ± 0,13 *	3,46 ± 0,13 ***	3,80 ± 0,14 ***
2	2,85 ± 0,16	2,85 ± 0,16	2,97 ± 0,16
3	2,81 ± 0,18	2,72 ± 0,18	2,73 ± 0,19
4	2,66 ± 0,25	2,61 ± 0,25	2,65 ± 0,25
1,2,3 y 4	2,95 ± 0,08	2,95 ± 0,08	3,10 ± 0,09

Nivel de significación estadística: \* P < 0,05 ;  
\*\* P < 0,01 ; \*\*\* P < 0,001 .

Para las 15 situaciones señaladas, relativas a los tres tipos de análisis y lactaciones, se ha estimado la curtosis. En la tabla n.º 3 figura, para cada uno de los tipos de análisis y para cada lactación, la estimación puntual de la curtosis, su error típico y el grado de significación estadística.

A partir de estos resultados y considerando las cuatro primeras lactaciones, se puede admitir que el tipo de función de distribución es mesocúrtica.

Como hemos indicado en el apartado de Material y Métodos, para analizar la homoscedasticidad se han considerado seis tipos de modelos matemáticos. En la tabla n.º 4 indicamos, para cada uno de estos modelos y por lactación, el valor «chi-cuadrado» del test de Bartlett con su grado de significación estadística.

**TABLA 4**

**Valores del «chi-cuadrado» del test de Bartlett, y significación estadística, obtenidos para cada número de parto o lactación, en los seis tipos de modelos matemáticos seguidos**

Número de modelo matemático	Número de lactación			
	1	2	3	4
1	157 ***	52	49	19
2	146 ***	81 *	42	21
3	160 ***	93 **	52	38
4	145 ***	81 *	49	24
5	102 ***	62 **	36	32
6	159 ***	78 *	47	48

Nivel de significación estadística: \*  $P < 0,05$  ;  
\*\*  $P < 0,01$  ; \*\*\*  $P < 0,001$  .

Por lo tanto, para la producción láctea ovina en el primer parto o lactación no se cumple el supuesto de homoscedasticidad, siendo las diferencias entre las varianzas residuales altamente significativas.

Mediante los programas HOM y NORM, se obtiene información aplicable a estudios de simulación<sup>1</sup>. Asimismo, el programa HARV es de utilidad para la aplicación de los programas LSML76 y LSMLMW de Harvey<sup>3,4</sup>.

## DISCUSION

Las estimaciones puntuales del sesgo (tabla n.º 2), disminuyen siempre al aumentar el número de lactación. Esta tendencia creemos que es explicable considerando el «estado de desarrollo» del animal, lo que se manifiesta además en un aumento de producción láctea conforme lo hace el número de lactación (tabla n.º 1).

Desde el punto de vista genético, la asimetría positiva que presenta la función de distribución de la producción láctea ovina, puede interpretarse admitiendo que el valor óptimo biológico es algo inferior al valor medio. Falconer<sup>2</sup>, indica que las funciones de distribución del tipo considerado por nosotros son sesgadas.

En relación a la heterogeneidad de varianzas residuales encontradas para la producción láctea ovina, Van Vleck<sup>7</sup> ha llegado a este mismo tipo de conclusión al estudiar la producción láctea del ganado vacuno. Falconer asimismo, indica que frecuentemente no se cumple el supuesto de homoscedasticidad para caracteres productivos.

En cuanto a los resultados obtenidos sobre la función de distribución de la producción láctea ovina, creemos que, si bien el sesgo resulta ser estadísticamente significativo, su estimación puntual no es lo suficientemente alta como para descartar la realización de pruebas de hipótesis sobre factores genéticos y ambientales mediante el uso de importantes métodos de estimación paramétrica, como el LS<sup>3</sup>, ML<sup>3,4</sup> y BLUP<sup>4,5</sup>.

Asimismo, y desde el punto de vista biológico, la heterogeneidad de varianzas residuales no creemos que sea un argumento suficiente para la no realización de inferencias o pruebas de hipótesis sobre los factores de variación genéticos y ambientales siguiendo los métodos de estimación paramétrica anteriormente mencionados, aunque incida negativamente sobre el grado de rigurosidad matemática de las inferencias.

Desde el punto de vista de la rigurosidad en la metodología estadística, podría ser de interés estudiar comparativamente otros métodos de estimación paramétrica, como el de mínimos cuadrados y BLUP con transformación logarítmica<sup>2,3,4,5</sup>. En un posterior trabajo justificamos y realizamos estos estudios comparativos, basándonos en las hipótesis analizadas en el presente estudio.

## AGRADECIMIENTOS

Agradecemos al personal del Centro de Proceso de Datos de la Universidad de León su colaboración en la instalación y puesta a punto de los programas LSML76 y LSMLMW de Harvey.

## RESUMEN

Se ha discutido la validez de algunas de las hipótesis asumidas en diferentes modelos de análisis de varianza para el carácter producción láctea ovina.

Para ello se han considerado 3482 lactaciones de tres rebaños de raza Churra, figurando en los registros los factores de variación: rebaño, semental o progenitor masculino y estación, año, tipo y número de parto.

La función de distribución probabilística de la variable mostró un sesgo estadísticamente significativo, siendo la estimación de  $0,45 \pm 0,04$  la más apropiada considerándose los 15 análisis afectados.

Para las varianzas residuales relativas a los modelos matemáticos contemplados no se cumplió la hipótesis de homoscedasticidad, siendo las diferencias estadísticamente significativas.

## STUDY OF SOME HYPOTHESIS ASSUMED IN THE ANALYSIS OF VARIANCE FOR THE MILK PRODUCTION OF DAIRY EWES

### SUMMARY

We have discussed the validity of some assumptions underlying the analysis of variance, for different statistical models, including the following variation factors: sire, flock, and season, year, number, and type of lambing. For this we have considered 3482 records of milk production in three flock of Churra breed sheep.

The milk production distribution showed statistically significant bias, the value estimation for 15 analysis being:  $0,45 \pm 0,04$ .

The Bartlett's test of heterogeneity of residual error variances was highly significant which is indicative of its non-homogeneity.

### BIBLIOGRAFIA

- 1) DAVIES, R. G. (1971).- *Computer Programing in Quantitative Biology*. Academic Press, London. pp. 394-400.
- 2) FALCONER, D. S. (1970).- *Introducción a la Genética Cuantitativa*. Ed. C.E.C.S.A., México. pp. 347-356 y 133-138.
- 3) HARVEY, W. R. (1977).- *User's guide for LSML76*. The Ohio State University, Columbus, U.S.A. Mimeografía.
- 4) HARVEY, W. R. (1985).- *User's guide for LSMLMW and PARMCARD*. The Ohio State University, Columbus, U.S.A. Mimeografía.
- 5) HENDERSON, C. R. (1974).- General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.*, 57: 963-972.
- 6) SNEDECOR, G. W. y COCHRAN, W. G. (1975).- *Métodos Estadísticos*. Ed. Compañía Editorial Continental, S. A., México-España. pp. 116-118 y 366-368.
- 7) VAN VLECK, L. D. (1968).- Variation of milk records within paternal-sib groups. *J. Dairy Sci.*, 51: 1.465-1.470.