

## Proposal of a mixed linear and a mixed generalized model for the analysis of an experiment in rumen microbiology

### Propuesta del modelo lineal mixto y generalizado mixto para el análisis de un experimento de la microbiología del rumen

Magaly Herrera Villafranca<sup>1</sup>, Juana Galindo Blanco<sup>1</sup>, C. Padilla Corrales<sup>1</sup>,  
Caridad W. Guerra Bustillo<sup>2†</sup>, Yolaine Medina Mesa<sup>1</sup> and Lucia Sarduy García<sup>1</sup>

<sup>1</sup>Instituto de Ciencia Animal, Apartado Postal 24, San José de las Lajas, Mayabeque, Cuba.

<sup>2</sup>Universidad Agraria de La Habana (UNAH) Fructuoso Rodríguez Pérez, Carretera Tapaste y Autopista Nacional, km 23 1/2. San José de Las Lajas, Mayabeque, Cuba.

Magaly Herrera Villafranca: <https://orcid.org/0000-0002-2641-1815>

Juana Galindo Blanco: <https://orcid.org/0000-0001-8639-4693>

César Padilla Corrales: <https://orcid.org/000-0002-6794-8694>

Email: [mvillafranca@ica.co.cu](mailto:mvillafranca@ica.co.cu)

The objective of this study was to propose the mixed generalized and mixed linear models for the analysis of an experiment in rumen microbiology. For developing this research, data from a study developed in the Department of Biophysiological Sciences of the Institute of Animal Science were used. The effect of different origins and/or varieties of *Moringa oleifera* on the ruminal microbial population was evaluated. A completely randomized design was applied, associated with a simple variance analysis model, with a 6x3 factorial arrangement. Eighteen treatments were established, which were related to the origin or varieties of *Moringa oleifera* and three times, with six repetitions each. Theoretical assumptions of the analysis of variance for the original variables homogeneity and normality of errors were verified. When they were not fulfilled, the mixed generalized linear model was used as an alternative to the analysis, and if not, the mixed linear model, with the help of GLIMMIX and MIXED procedure of SAS. In both models, treatment, hour and interaction treatment per hour were considered as fixed effects, and nested repetition within hours was considered as random. Results showed that the mean square of the error was low, when mixed procedures were used. Standard errors also decreased, which contributes to greater precision in results. From this perspective, these models are proposed for the analysis of related variables and counting experiments in the ruminal microbial population.

Key words: *GLIMMIX*, *analysis of variance assumptions*, *nested effect*

Parametric analysis of variance is the most widely used statistical method in data analysis, developed by Fisher in the 1920s. However, it is necessary to comply with the theoretical assumptions for its use. Some of them state that errors are normally and independently distributed, that their variances are homogeneous, and to consider the attachment capacity of the model. When any of these theoretical assumptions fail, the use of other analysis methods is suggested, such as mixed linear (MIXED) and mixed generalized (GLIMMIX) models.

Mixed models, according to Dicozskiy and Pedroza (2017), are a proposal for advanced statistical modeling, which allow improving the quality of analysis of fixed

El trabajo tiene como objetivo proponer el modelo lineal mixto y generalizado mixto para el análisis de un experimento de la microbiología del rumen. Para el desarrollo de la investigación, se usaron datos de un estudio desarrollado en el Departamento de Ciencias Biofisiológicas del Instituto de Ciencia Animal. Se evaluó el efecto de diferentes procedencias y/o variedades de *Moringa oleifera* en la población microbiana ruminal. Se aplicó un diseño completamente aleatorizado, asociado a un modelo de análisis de varianza simple, con arreglo factorial 6x3. Se establecieron 18 tratamientos, que estuvieron relacionados con las procedencias o variedades de *Moringa oleifera* y tres horarios, cada uno con seis repeticiones. Se verificaron los supuestos teóricos del análisis de varianza para las variables originales homogeneidad y normalidad de los errores. Cuando se incumplieron, se utilizó como alternativa de análisis el modelo lineal generalizado mixto y, en caso contrario, el lineal mixto, con ayuda del procedimiento GLIMMIX y MIXED del SAS. En ambos modelos, se consideraron como efectos fijos tratamiento, hora e interacción tratamiento por hora, y como aleatorio, la repetición anidada dentro de horas. Los resultados mostraron que el cuadrado medio del error fue menor, cuando se utilizaron los procedimientos mixtos. También disminuyeron los errores estándar, lo que contribuye a mayor precisión en los resultados. Desde esta perspectiva, se proponen estos modelos para el análisis de variables relacionadas y experimentos de conteo en la población microbiana del rumen.

Palabras clave: *GLIMMIX*, *supuestos del análisis de varianza*, *efecto anidado*.

El análisis de varianza paramétrico es el método estadístico más difundido en el análisis de datos, desarrollado por Fisher en la década de los años 20 del siglo pasado. Sin embargo, para su utilización es necesario el cumplimiento de los supuestos teóricos. Entre ellos, que los errores se distribuyan de forma normal e independiente, que sus varianzas sean homogéneas, además de considerar la aditividad del modelo. Cuando alguno de estos supuestos teóricos falla, se sugiere el empleo de otros métodos de análisis, entre los que se encuentran los modelos lineales mixtos (MIXED) y generalizados mixtos (GLMMIX).

Los modelos mixtos, según Dicozskiy y Pedroza (2017), son una propuesta de modelación estadística avanzada, que permiten mejorar la calidad del análisis de los factores fijos

and random factors, by modeling random variability and error correlation. They are very useful for unbalanced data analysis, which are data with some type of hierarchical structure. Therefore, they allow to estimate variability among groups and that of effects nested within groups.

Nelder and Wedderburn (1972) group different statistical models, which they released as generalized linear (MLGnz), which constitute an extension of classical general linear (MLG). These models can be applied to distributions of normal, binomial, Poisson, gamma type, among others (Mandujano *et al.* 2016, Díaz *et al.* 2017 and Monterubbianesi 2017).

Wang *et al.* (2015) state that data measured in agricultural research does not satisfy the premises of general linear models, so that mixed generalized linear models provide an analysis that does not necessarily require normal distribution of variables, by allowing these to be fitted to an exponential family distribution.

These models have been widely disseminated in social sciences, psychology, and medical sciences. However, in agriculture, they have had little application, without taking into account that, on many occasions, situations are involved in which it is difficult to use the MLG in the analysis of variance and regression. This is because analyzed variables do not meet the assumptions of normality, variance homogeneity and independence of errors, so these models can be proposed as an alternative analysis.

Therefore, the objective was to propose the mixed generalized linear model in the analysis of an experiment in rumen microbiology.

### Materials and Methods

For the research, data from an experiment developed in the Department of Biophysiological Sciences of the Institute of Animal Science was used. This study aimed to evaluate the effect of different varieties of *Moringa oleifera* and *Cynodon nlemfuensis* (star grass) on the ruminal microbial population, for which the chemical variables total bacteria and isovaleric acid were measured. The experiment consisted of a completely randomized design, with a 6 x 3 factorial arrangement. The factors were the six grass varieties and the three hours, with six repetitions each. Measurements were not performed on the same experimental unit. The statistical models used were the following:

Mixed generalized linear model:

$$E(y) = \mu = g^{-1}(X\beta)$$

Where:

$E(y)$  = expected value of response variable (total bacteria counting and isovaleric acid)

$X\beta$ : linear predictor (linear combination of a  $\beta$  unknown parameter)

$g$ : link function, which belongs to a member of exponential families of probability distributions.

y factores aleatorios, al modelar la variabilidad aleatoria y la correlación de los errores. Son muy útiles en el análisis de datos desbalanceados, datos con algún tipo de estructura jerárquica. Por tanto, permiten estimar la variabilidad entre grupos y la de los efectos anidados dentro de grupos.

Nelder y Wedderburn (1972) agrupan diferentes modelos estadísticos, los que dieron a conocer como lineales generalizados (MLGnz), que constituyen una extensión de los lineales generales clásicos (MLG). Estos modelos se pueden aplicar a distribuciones de tipo normal, binomial, Poisson, gamma, entre otras (Mandujano *et al.* 2016, Díaz *et al.* 2017 y Monterubbianesi 2017).

Wang *et al.* (2015) plantean que los datos que se miden en las investigaciones agrícolas no satisfacen las premisas de los modelos lineales generales, por lo que los modelos lineales generalizados mixtos proporcionan una vía de análisis que no requiere necesariamente distribución normal de las variables, al posibilitar que estas se puedan ajustar a una distribución de la familia exponencial.

Estos modelos han sido muy difundidos en las ciencias sociales, en la psicología y en las ciencias médicas. Sin embargo, en las agropecuarias han tenido poca aplicación, sin tener en cuenta que, en muchas ocasiones, están involucradas situaciones en las que es difícil utilizar el MLG en los análisis de varianza y regresión. Y esto es porque las variables que se analizan no cumplen los supuestos de normalidad, homogeneidad de varianza e independencia de los errores, por lo que se pueden proponer estos modelos como alternativa de análisis.

A partir de lo antes expuesto, se identificó como objetivo proponer el modelo lineal generalizado mixto en el análisis de un experimento de la microbiología del rumen.

### Materiales y Métodos

Para la investigación se utilizaron los datos de un experimento desarrollado en el Departamento de Ciencias Biofisiológicas, del Instituto de Ciencia Animal. El presente estudio tuvo como objetivo evaluar el efecto de diferentes variedades de *Moringa oleifera* y *Cynodon nleumfuensis* (pasto estrella) en la población microbiana ruminal, para lo que se midieron las variables bacterias totales y ácido isovalérico. El experimento se conformó en un diseño completamente aleatorizado, con arreglo factorial 6 x 3. Los factores fueron las seis variedades de pastos y los tres horarios, con seis repeticiones cada uno. Las mediciones no se realizaron sobre la misma unidad experimental. Los modelos estadísticos que se utilizaron fueron los siguientes:

Modelo lineal generalizado mixto:

$$E(y) = \mu = g^{-1}(X\beta)$$

Donde:

$E(y)$  = valor esperado de la variable respuesta (conteo de bacterias totales y ácido isovalérico)

$X\beta$ : predictor lineal (combinación lineal de un parámetro desconocido  $\beta$ ).

$g$ : función de enlace, que pertenece a un miembro de familias exponenciales de distribuciones de probabilidad.

Mixed linear model:

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij}$$

Where:

$y_{ijk}$  = response variable

$\mu$  = general mean for all observations

$\alpha_i$  = fixed effect of the i-th grass ( $i = 1, \dots, 6$ )

$\beta_j$  = fixed effect of the j-th hour ( $j = 1, \dots, 3$ )

$(\alpha\beta)_{ij}$  = fixed effect of the i-th grass in interaction with the fixed effect of the j-th hour ( $ij = 1, \dots, 18$ )

$e_{ik}$  = random error associated with all observations

The theoretical assumptions of the analysis of variance for the original variables were verified. For variance homogeneity of treatments, Levene (1960) test was used. Normality of errors was evaluated using Shapiro-Wilk (1965) test. In this analysis, the variable total bacteria did not comply with both assumptions, and after transformation, its fulfillment did not improve. The original isovaleric acid variable did meet these assumptions, so it was not necessary to perform data transformation.

For the variable that did not meet the theoretical assumptions of analysis of variance, mixed generalized linear model was applied as an analysis alternative, using GLIMMIX procedure. When theoretical assumptions of the analysis of variance were fulfilled, mixed linear model was used, with the help of the PROC MIXED, both from SAS. In the statistical analyzes, treatments, hours and interaction treatments per hours were considered as fixed effects. Nested repetition within hours was considered as a random effect. For total bacteria variable, normal, Poisson, lognormal, and gamma distributions were tested, the latter being the best fit, with log link function.

Toeplitz (Toep) variance-covariance structures, variance component (VC), composite symmetry (CS), autoregressive of order 1 (AR [1]) and unstructured (UN) were tested. To select the one with the best fit to the data, information criteria [Akaike (AIC), corrected Akaike (AICC) and Bayesian (BIC)] were used, which was considered the smallest value. For mean comparison, fixed range test was used (Kramer 1956). Data was analyzed with SAS (2013) statistical package, version 9.3.

## Results and Discussion

Table 1 shows the analysis of theoretical assumptions normality of errors and variance homogeneity for the analyzed variables. It was observed that, for total bacteria, probability values in both tests were lower than 0.05, so these assumptions are not fulfilled. However, this value was higher than 0.05 for isovaleric acid. This shows the fulfillment of base hypotheses that support the analysis of variance.

Steel and Torrie (1996) and Peña (1994) point out that normal distribution of errors has little influence on ANOVA to compare means, since this technique is robust to error deviations. However, they argue that

Modelo lineal mixto.

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij}$$

Donde:

$y_{ijk}$  = variable respuesta

$\mu$  = media general a todas las observaciones

$\alpha_i$  = efecto fijo del i-ésimo pasto ( $i=1, \dots, 6$ )

$\beta_j$  = efecto fijo del j-ésima hora ( $j=1, \dots, 3$ )

$(\alpha\beta)_{ij}$  = efecto fijo del i-ésimo pasto en interacción con el efecto fijo del j-ésima hora ( $ij=1, \dots, 18$ )

$e_{ik}$  = error aleatorio asociado a todas las observaciones

Se verificaron los supuestos teóricos del análisis de varianza para las variables originales. Para la homogeneidad de varianza de los tratamientos, se usó la dócima de Levene (1960). La normalidad de los errores se evaluó mediante la dócima de Shapiro-Wilk (1965). En este análisis, la variable bacterias totales incumplió con ambos supuestos, y después de transformada no mejoró su cumplimiento. La variable original ácido isovalérico sí cumplió con estos supuestos, por lo que no fue necesario realizar la transformación de los datos.

Para la variable que incumplió con los supuestos teóricos de análisis de varianza, se aplicó como alternativa de análisis el modelo lineal generalizado mixto, con ayuda del procedimiento GLIMMIX. Cuando se cumplieron los supuestos teóricos del análisis de varianza, se utilizó el modelo lineal mixto, con ayuda del PROC MIXED, ambos del SAS. En los análisis estadísticos se consideraron como efectos fijos los tratamientos, las horas y la interacción tratamientos por horas. Como efecto aleatorio se consideró la repetición anidada dentro de horas. Para la variable bacterias totales, se probaron las distribuciones normal, Poisson, lognormal y gamma, siendo esta última la de mejor ajuste, con función de enlace log.

Se probaron las estructuras de varianza-covarianza Toeplitz (Toep), componente de varianza (VC), simetría compuesta (CS), autoregresiva de orden 1 (AR[1]) y no estructurada (UN). Para seleccionar la de mejor ajuste a los datos, se utilizaron los criterios de información [Akaike (AIC), Akaike corregido (AICC) y Bayesiano (BIC)], que se consideró el valor más pequeño. Para la comparación de medias, se utilizó la dócima de rango fijo (Kramer 1956). Los datos se analizaron con el paquete estadístico SAS (2013), versión 9.3.

## Resultados y Discusión

En la tabla 1 se muestra el análisis de los supuestos teóricos normalidad de los errores y homogeneidad de varianza para las variables que se analizaron. Se observó que para las bacterias totales, los valores de probabilidad en ambas dócimas fueron menores que 0.05, por lo que se incumplen dichos supuestos. Sin embargo, para el ácido isovalérico este valor fue superior a 0.05. Esto evidencia el cumplimiento de las hipótesis de bases que sustentan el análisis de varianza.

Steel y Torrie (1992) y Peña (1994) señalan que la distribución normal de los errores tiene poca influencia en el ANOVA para comparar medias, ya que esta técnica

Table 1. Fulfillment of ANOVA theoretical assumptions, for total bacteria and isovaleric acid variables

Variabes	ANOVA theoretical assumptions	Statistical tests	P Value
Total bacteria, 10 <sup>11</sup> CFU/mL	Variance homogeneity	Levene	0.0266
	Normality of errors	Shapiro-Wilk	0.0303
Isovaleric acid, mmol/L	Variance homogeneity	Levene	0.3513
	Normality of errors	Shapiro-Wilk	0.2033

CFU: colony forming units

the lack of normality can affect other assumptions, such as the variance homogeneity, and this happens especially when the number of observations of groups are very different. Nevertheless, when variance components are analyzed, normality can affect the analysis result.

According to Gutiérrez and de la Vara (2012), variance homogeneity is an assumption that relates the residues of treatments, and offers an overview of the possible equality between them. For its analysis, Levene, Bartlett, Hartley, and other tests were used. However, Levene test is the most robust in the absence of normality.

When analyzing variables under study, it was observed that the total bacteria did not meet the variance homogeneity of residuals. Peña *et al.* (2015) state that, according to the nature of this type of variable, the use of classical statistical methods is not recommended because, in some cases, homogeneity assumption is not met.

It is necessary to verify the fulfillment of the theoretical assumptions of classical statistical methods before starting the statistical analysis for this type of research, since, according to results, selection of the appropriate statistical method is defined. The use of these statistical models also avoids all inconveniences that may affect the expected results. In addition, this type of model does not require fulfillment of these assumptions, and these are no longer a problem for data analysis.

Table 2 shows the analysis of variance and covariance structures in order to select the best fit model. For this, information criteria were considered. For total bacteria variable, the lowest value was obtained with that of variance components (VC), and for isovaleric acid, with the autoregressive of order one (AR (1)). However, composite symmetry (CS), unstructured (UN) and Toeplitz structures did not achieve convergence, and did not fit to the analyzed data. For this reason, the results for these structures are not reported. However, Gómez (2019) states that, for selecting the structure with the best fit to data, the one with the lowest values in the information criteria should be taken into account.

Valdivieso (2013) states that, to model covariance structures, data is available, in which the sample variance-covariances of the observed variables estimate

es robusta ante las desviaciones de los errores. Sin embargo, plantean que la falta de normalidad puede afectar otros supuestos, como la homogeneidad de varianza, y esto sucede sobre todo, cuando el número de observaciones de los grupos son muy diferentes. No obstante, cuando se analizan componentes de varianza, la normalidad sí puede afectar el resultado del análisis.

Según Gutiérrez y de la Vara (2012), la homogeneidad de varianza es un supuesto que relaciona los residuos de los tratamientos, y ofrece una visión general de la posible igualdad entre ellos. Para su análisis se utilizan las dójimas de Levene, Bartlett, Hartley, entre otras. Sin embargo, la de Levene es más robusta ante la falta de normalidad.

Al realizar el análisis para las variables en estudio, se observó que las bacterias totales no cumplieron con la homogeneidad de varianza de los residuos. Peña *et al.* (2015) plantean que, de acuerdo con la naturaleza de este tipo de variable, no se recomienda el uso de los métodos estadísticos clásicos, debido a que en algunos casos el supuesto de la homogeneidad no se cumple.

Resulta necesario verificar el cumplimiento de los supuestos teóricos de los métodos estadísticos clásicos antes de iniciar el análisis estadístico para este tipo de investigación, pues de acuerdo con los resultados que se obtienen se define la selección del método estadístico apropiado. El empleo de estos modelos estadísticos también evita todos los inconvenientes que pueden incidir en los resultados esperados. Además, este tipo de modelo no requiere del cumplimiento de dichos supuestos, y estos dejan de ser un problema para el análisis de los datos.

La tabla 2 muestra el análisis de las estructuras de varianza y covarianza con el objetivo de seleccionar el modelo de mejor ajuste. Para ello se consideraron los criterios de información. Para la variable bacterias totales, el valor más pequeño se obtuvo con la de componentes de la varianza (VC), y para el ácido isoalérico la autoregresiva de orden uno (AR(1)). Sin embargo, las estructuras simetría compuesta (CS), no estructurada (UN) y la Toeplitz, no lograron la convergencia, y no se ajustaron a los datos analizados. Es por ello que no se informan los resultados para esas estructuras. Sin embargo, Gómez (2019) plantea que para la selección de la estructura de mejor ajuste a los datos se debe tener en cuenta aquella que presenta los valores más pequeños en los criterios de información.

Valdivieso (2013) plantea que para modelar las estructuras de covarianza se dispone de datos, en los que las varianzas-covarianzas muestrales de las variables

Table 2. Variance-covariance structure for total bacteria and isovaleric acid

Variables	Covariance structures	Information criteria		
		AIC	AICC	BIC
Total bacteria, 10 <sup>11</sup> CFU/mL	Toep	775.93	815.11	807.98
	VC	742.77	752.77	760.58
	CS	-	-	-
	AR(1)	744.77	755.90	763.47
	UN	-	-	-
Isovaleric acid, mmol/L	Toep	-	-	-
	VC	250.50	260.20	268.30
	CS	-	-	-
	AR(1)	249.10	259.80	267.80
	UN	-	-	-

CFU: colony forming units

the model parameters and their errors. Liscano and Ortiz (2017) report that if a structure that fits data is suspected, its use leads to a more efficient inference and estimation.

In the results of the table of analysis of variance, it is shown that mean square of the error was low, when mixed procedures were used. This could be because, when the effects are nested within the analysis, treatment variability decreases and better estimates are obtained (table 3). Hernández *et al.* (2003) refer that, when speaking of nested structure, and data is grouped into experimental units of different order, each with specific properties, according to the considered grouping level, it is necessary to eliminate this effect so that it does not affect the estimation of results.

observadas estiman los parámetros del modelo y sus errores. Liscano y Ortiz (2017) informan que si se sospecha la existencia de una estructura que se adecue a los datos, su utilización conduce a una estimación e inferencia más eficiente.

En los resultados de la tabla del análisis de varianza, se muestra que el cuadrado medio del error fue menor, cuando se usaron los procedimientos mixtos. Esto puede estar dado porque, cuando se anidan los efectos dentro del análisis, disminuye la variabilidad de los tratamientos y se logran mejores estimaciones (tabla 3). Hernández *et al.* (2003) refieren que, cuando se habla de estructura anidada, y los datos están agrupados en unidades experimentales de diverso orden, cada una con propiedades específicas según el nivel de agrupamiento

Table 3. Results of mean square and error probability type I in the interaction for both analyses

Variables	Statistical analysis	Mean square of the error	Probability value Type I
Total bacteria, 10 <sup>11</sup> CFU/mL	ANAVA	0.3712	<0.0001
	GLMMIX	0.2719	<0.0001
Isovaleric acid mmol/L	ANAVA	0.4951	0.4046
	MIXED	0.3824	0.2122

CFU: colony forming units

Mixed generalized linear models and generalized mixed additive models are used for modelling nested data and spatial and temporal correlation structures in counting data or binomial data. Additive mixed-effect models and mixed-effect models are useful for nested data (also called panel data or hierarchical data), repeated measurements, and temporally and spatially correlated data (Zuur *et al.* 2009).

Table 4 shows interaction results for the classical analysis of variance and the mixed generalized linear model. In both cases, interaction was significant. However, standard error was lower when this last was used. The analysis showed that the mixed generalized linear model, in some of the cases, was more conservative

considerado, es necesario eliminar ese efecto para que no incida en la estimación de los resultados.

Los modelos lineales generalizados mixtos y los mixtos aditivos generalizados se usan para modelar los datos anidados y estructuras de correlación espacial y temporal en datos de conteo o datos binomiales. Los de efectos mixtos aditivos y los modelos de efectos mixtos son útiles para datos anidados (llamados también datos de panel o datos jerárquicos), mediciones repetidas y datos correlacionados, temporal y espacialmente (Zuur *et al.* 2009).

En la tabla 4 se muestran los resultados de la interacción para el análisis de varianza clásico y del modelo lineal generalizado mixto. En ambos casos, la interacción fue significativa. Sin embargo, el error

Table 4. Results of the statistical analysis with both methods, for total bacteria variable

Variable	Statistical analysis	Treatment	Hour			SE Signf.
			1	2	3	
Total viable bacteria, 10 <sup>11</sup> CFU/mL	ANAVA	Star grass	2.80 <sup>abcde</sup> (18.71)	2.29 <sup>abcdef</sup> (11.71)	1.18 <sup>f</sup> (4.71)	±0.31 P<0.0001
		Superganius	1.96 <sup>bcdef</sup> (8.04)	1.70 <sup>cdef</sup> (5.54)	2.49 <sup>abcdef</sup> (16.54)	
		Tunera	3.04 <sup>abcd</sup> (26.21)	2.57 <sup>abcdef</sup> (16.71)	2.22 <sup>abcdef</sup> (10.04)	
		Camerún	3.64 <sup>a</sup> (43.21)	3.17 <sup>abc</sup> (24.71)	1.46 <sup>ef</sup> (7.04)	
		Paraguayaya	2.51 <sup>abcdef</sup> (13.04)	3.41 <sup>ab</sup> (31.71)	1.59 <sup>ef</sup> (7.21)	
		Planin	2.59 <sup>abcdef</sup> (17.21)	3.09 <sup>abcd</sup> (23.21)	2.84 <sup>abcde</sup> (19.71)	
	GLMMIX	Star grass	2.93 <sup>abcde</sup> (18.71)	2.43 <sup>bcdef</sup> (11.71)	1.55 <sup>f</sup> (4.71)	±0.24 P<0.0001
		Superganius	2.08 <sup>cdef</sup> (8.04)	1.71 <sup>ef</sup> (5.54)	2.81 <sup>abcde</sup> (16.55)	
		Tunera	3.27 <sup>abc</sup> (26.20)	2.82 <sup>abcde</sup> (16.33)	2.31 <sup>bcdef</sup> (10.04)	
		Camerún	3.77 <sup>a</sup> (43.23)	3.21 <sup>abc</sup> (24.71)	1.95 <sup>def</sup> (7.03)	
		Paraguayaya	2.57 <sup>abcdef</sup> (13.04)	3.46 <sup>ab</sup> (31.72)	1.98 <sup>def</sup> (7.21)	
		Planin	2.85 <sup>abcde</sup> (17.21)	3.14 <sup>abcd</sup> (23.20)	2.98 <sup>abcd</sup> (19.72)	

CFU: colony forming units

in finding similar groups.

When comparing both models, some of the treatment mean values that correspond to the mixed generalized linear model had a slight increase. This could be related to the adjustment of the link function, selected according to the distribution followed by the variable, so means are estimated by the effect of this link function.

When analyzing the isovaleric acid variable, it was observed that interaction between the main effects was not significant. Therefore, the main effects were reported (tables 5 and 6). In the effect of varieties, the standard error for the mixed procedure was slightly lower than the classical analysis of variance, although for both, no significant differences were found among treatments (table 5).

Table 6 reports the effect of hours. In both methods, standard errors presented similar results, and no significant differences were found among times.

estándar resultó menor cuando se empleó este último. En el análisis se evidenció que el modelo lineal generalizado mixto, en algunos de los casos, fue más conservador para encontrar grupos similares.

Al comparar ambos modelos, algunos de los valores de las medias de los tratamientos que corresponden al modelo lineal generalizado mixto tuvieron ligero incremento. Esto se pudiera relacionar con el ajuste de la función de enlace que se selecciona de acuerdo con la distribución que sigue la variable, por lo que las medias se estiman por el efecto de dicha función de enlace.

Al analizar la variable ácido isovalérico, se observó que la interacción entre los efectos principales fue no significativa, por lo que se informaron los efectos principales (tabla 5 y 6). En el efecto de las variedades, el error estándar para el procedimiento mixto fue ligeramente menor con respecto al análisis de varianza clásico, aunque para ambos no se encontraron diferencias significativas entre los tratamientos (tabla 5).

Table 5. Results of the statistical analysis with both methods for isovaleric acid, according to treatments

Statistical analysis	Treatments Variable	Star grass	Super ganius	Tunera	Camerún	Paraguayaya	Planin	SE Signf.
ANAVA	Isovaleric acid, mmol/L	2.01	1.89	1.45	1.89	1.60	1.83	±0.17 P=0.0693
MIXED		2.01	1.89	1.45	1.89	1.60	1.83	±0.15 P=0.0825

Table 6. Results of the statistical analysis with both methods for isovaleric acid, according hours

Statistical analysis	Variable	Hours			SE and Signif
		1	2	3	
ANAVA	Isovaleric acid mmol/L	1.73	1.87	1.73	±0.12 P=0.6046
MIXED		1.73	1.87	1.73	±0.12 P=0.5469

Therefore, this type of analysis can be proposed for research related to rumen microbiology experiments, as long as an adequate statistical analysis is carried out, justifying the use of these methods.

According to Gómez *et al.* (2012) and Dicovskiy and Pedroza (2017), mixed models are a proposal for advanced statistical modeling, which allow improving the quality of the analysis of fixed and random factors, by modeling random variability and error correlation. These models are very useful in the analysis of unbalanced data, or of data with some type of hierarchical or grouping structure.

From the results of this research, it is concluded that mixed models improve accuracy and precision of analysis results. The mean square of the smallest error is obtained when using mixed procedures, and standard errors decrease with respect to classical analysis of variance. From this perspective, these models are proposed for the analysis of variables related to counting experiments in the rumen microbial population.

En la tabla 6 se informa el efecto de las horas. En ambos métodos, los errores estándar presentaron resultados similares, y no se encontraron diferencias significativas entre los horarios. Por tanto, se puede proponer este tipo de análisis para investigaciones relacionadas con experimentos de microbiología del rumen, siempre y cuando se lleve a cabo un análisis estadístico adecuado, que justifique la utilización de estos métodos.

Según Gómez *et al.* (2012) y Dicovskiy y Pedroza (2017), los modelos mixtos son una propuesta de modelación estadística avanzada, que permiten mejorar la calidad del análisis de los factores fijos y factores aleatorios, al modelar la variabilidad aleatoria y la correlación de los errores. Son modelos muy útiles en el análisis de datos desbalanceados, o de datos con algún tipo de estructura jerárquica o de agrupación.

A partir de los resultados de esta investigación, se concluye que los modelos mixtos mejoran la exactitud y precisión de los resultados del análisis. Se obtiene cuadrado medio del error más pequeño, cuando se utilizan los procedimientos mixtos, y los errores estándar disminuyen con respecto al análisis de varianza clásico. Desde esta perspectiva, se proponen estos modelos para el análisis de variables relacionadas con experimentos de conteo en la población microbiana del rumen.

## References

- Díaz, E.J., Bermúdez, D. & Pineda, W. 2017. Estimación de un modelo lineal generalizado mixto para datos de conteo con exceso de ceros. Diploma Thesis. Facultad de Estadística, Universidad Santo Tomás, Bogotá, Colombia
- Dicovskiy, L.M. & Pedroza, M.E. 2017. "General and Mixed linear models in the characterization of the qualification variable, agroindustrial engineering, uni-north". *Nexo Revista Científica*, 30(2): 84-95, ISSN: 1995-9516
- Gómez, S., Torres, V., García, Y. & Navarro, J.A. 2012. "Statistical procedures most used in the analysis of measures repeated in time in the agricultural sector". *Cuban Journal of Agricultural Science*, 46(1): 1-7, ISSN: 2079-3480
- Gómez, S. 2019. Contribución estadística para el análisis de medidas repetidas en el tiempo en el sector agropecuario. PhD Thesis. Departamento de Biomatemática, Instituto de Ciencia Animal, Mayabeque. Cuba
- Gutiérrez, H. & de la Vara, R. 2012. *Análisis y diseño de experimentos*. 3rd Ed. Ed. Mc Graw-Hill Latinoamericana Editores S.A de C.V, México D.F., México, ISBN: 978-607-15-0725-9
- Hernández, M.V., Colmenares F. & Martínez R. 2003. "Modelos jerárquicos "por piezas" en el análisis de la relación entre discontinuidad conductual y discontinuidad en procesos subyacentes". *Revista Anales de Psicología*, 19(1): 159-171, ISSN: 0212-9728
- Kramer, C.Y. 1956. "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications". *Biometrics*, 12(3): 307-310, ISSN: 0006-341X, DOI: 10.2307/3001469
- Levene, H. 1960. Robust tests for the equality of variance. *Contributions to Probability and Statistics*. 1st Ed. Ed. Stanford University Press, Palo Alto, California, USA, p. 278-292.
- Liscano, J.M. & Ortíz, A. F. 2017. Modelos mixtos para datos composicionales: Una aplicación con resultados electorales en Colombia. Diploma Thesis. Facultad de Estadística, Universidad Santo Tomás, Bogotá, Colombia
- Mandujano, S., Kéry, M. & Royle, J. A. 2016. "Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS". *Revista Mexicana de Biodiversidad*, 88(2): 485-486, ISSN: 2007-8706, DOI: <http://dx.doi.org/10.1016/j.rmb.2017.03.028>
- Monterubbiansi, M.G. 2017. Evaluación de alternativas para el análisis estadístico y de aspectos del diseño en ensayos de

- larga duración para estudios agronómicos. PhD Thesis. Departament de Producció Vegetal i Ciència Forestal , Universitat de Lleida, Catalunya, España, p. 193
- Nelder, J.A. & Wedderburn, R.W.M. 1972. "Generalized linear models". Journal of the Royal Statistical Society: Series A (General), 135(3): 370–384, ISSN: 1467-985X, DOI: <https://doi.org/10.2307/2344614>
- Peña, S. 1994. Estadística. Modelos y métodos: 2. Modelos lineales y series temporales. 4th Ed. Ed. Alianza, S.A., Madrid, España, p. 745, ISBN: 84-206-8110-5
- Peña, J.A., Rosales, Y. & Giampaolo, O. 2015. "Estudio del crecimiento bacteriano. Enfoque de análisis de datos con medidas repetidas". Revista de la Facultad de Farmacia, 57(2): 8-17, ISSN: 2244-8845
- SAS Institute Inc. 2013. Statistical Analysis Software SAS/STAT®, version 9.1.3, Cary, N.C., USA, Available: <<http://www.sas.com/en us/software/analytics/stat.html#>>.
- Shapiro, S. & Wilk, B. 1965. "An analysis of variance test for normality (complete samples)". Biometrika, 52(2): 591-611, ISSN: 1464-3510, DOI: <http://dx.doi.org/10.2307/2333709>
- Steel, R.G. & Torrie, I.H. 1996. Bioestadística: principios y procedimientos. 2nd Ed. Ed. McGraw-Hill Interamericana SA., México D.F., México, p. 622, ISBN: 968-451-495-6
- Valdivieso C.E. 2013. "Efecto de los métodos de estimación en las modelaciones de estructuras de covarianzas sobre un modelo estructural de evaluación del servicio de clases". Revista Comunicaciones en Estadística, 6(1): 21-43, ISSN: 2027-3355, DOI: <https://doi.org/10.15332/s2027-3355.2018.0001.03>
- Wang, T., He, P., Ahn, K.W., Wang, X., Ghosh, S. & Laud, P. 2015. "A re-formulation of generalized linear mixed models to fit family data in genetic association studies". Frontiers in Genetics, 6: 120, ISSN: 1664-8021, DOI: <https://doi.org/10.3389/fgene.2015.00120>
- Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. 2009. Mixed Effects Models and Extensions in Ecology with R. Ed. Springer Science & Business Media, New York, USA, ISBN: 978-0-387-87457-9, DOI: <https://doi.org/10.1007/978-0-387-87458-6>

**Received: June 12, 2019**

**Accepted: January 6, 2020**