

Statistical procedure for the analysis of experiments with repeated measures over time in the agricultural and livestock field

Procedimiento estadístico para el análisis de experimentos con medidas repetidas en el tiempo en la esfera agropecuaria.

Sarai Gómez, Verena Torres, Yoleisy García, Magaly Herrera, Yolaine Medina and R. Rodríguez

Instituto de Ciencia Animal

Carretera Central, km. 47 ½. San José de las Lajas. Mayabeque. Cuba. CP: 32 700.

Email: sgomez@ica.co.cu

The objective of this study was the proposal of a statistical analysis methodology that will guide the researcher by making repeated measurements over time in the same experimental unit, through a case study with legumes as a substrate in the production of *in vitro* gas in the agricultural and livestock field. The variable *in vitro* gas production was analyzed. Pearson correlation matrix was calculated, values superior to 0.82 were obtained and the existence of association among sampling days was determined. The sphericity criterion was confirmed by means of the Mauchly statistic and, in front of its failure to fulfill, the fit of the degrees of freedom was made. In the same way, normality assumption was verified ($P < 0.0100$) and when it was not fulfilled, a mixed generalized linear model was used for analyzing the variants of Poisson, Gamma, Binomial, Normal and Normal Log, to determine the distribution that followed the data, which in this case was Gamma. Toeplitz variance-covariance structure was selected as the one that best fits the model based on the lower values of information criteria. The verification of theoretical assumptions necessary for repeated measures defined the model to be used. The use of a mixed generalized linear model increased the accuracy of results by properly estimating the variance-covariance structures and allowed to analyze unbalanced data. A work methodology is proposed for data processing with repeated measures over time.

Keywords: *information criteria, covariance structures, correlation matrix*

With the continuous development of research and the search for new statistical analysis strategies that provide greater precision and accuracy in obtaining results, attention has been focused on determining which is the most appropriate for the analysis of data from experiments with repeated measures at different times in the same experimental unit.

In the agricultural and livestock field, more and more experiments with these characteristics are carried out, since repeated measurements in the same experimental unit over time are cheaper than the use of a different experimental unit for each measurement over time, less experimental units are required, sample size and costs are reduced, test power and accuracy in estimating trends over time are improved. If this type of analysis is properly applied, it emphasizes the validity of statistical conclusions, because it has greater accuracy in the estimation of parameters of the analysis model (Kuehl 2000).

The design with repeated measures over time was

El objetivo de este trabajo fue la propuesta de una metodología de análisis estadístico que sirva de guía al investigador al realizar mediciones repetidas en el tiempo en la misma unidad experimental, a través de un estudio de caso con leguminosas como sustrato en la producción de gas *in vitro* en la esfera agropecuaria. Se analizó la variable producción de gas *in vitro*. Se calculó la matriz de correlación de Pearson, se obtuvieron valores superiores a 0.82 y se determinó la existencia de asociación entre los días de muestreo. Se comprobó el criterio de esfericidad mediante el estadístico de Mauchly y ante su incumplimiento se efectuó el ajuste de los grados de libertad. De igual forma, se verificó el supuesto de la normalidad ($P < 0.0100$) y al no cumplirse se empleó un Modelo Lineal Generalizado Mixto donde se analizaron las variantes de Poisson, Gamma, Log Normal, Normal y Binomial para determinar la distribución que siguieron los datos, que en este caso fue la Gamma. Se seleccionó la estructura de varianza - covarianza Toeplitz como la de mejor ajuste al modelo a partir de los menores valores de los criterios de información. La comprobación de los supuestos teóricos necesarios para medidas repetidas definió el modelo a emplear. La utilización del Modelo Lineal Generalizado Mixto incrementó la precisión de los resultados al estimar de forma adecuada las estructuras de varianza-covarianza y permitió analizar datos desbalanceados. Se propone una metodología de trabajo para el procesamiento de datos con medidas repetidas en el tiempo.

Palabras clave: *criterios de información, estructuras de covarianzas, matriz de correlación.*

Con el continuo desarrollo de las investigaciones y la búsqueda de nuevas estrategias de análisis estadístico que brinden mayor precisión y exactitud al obtener los resultados, se ha centrado la atención en determinar cuál es la más apropiada para el análisis de datos provenientes de experimentos con medidas repetidas en diferentes momentos en el tiempo en la misma unidad experimental.

En la esfera agropecuaria se realizan cada vez más experimentos con estas características, pues las mediciones repetidas en la misma unidad experimental a través del tiempo son más económicas que el uso de una unidad experimental diferente para cada medición en el tiempo, se requieren menos unidades experimentales, se reduce el tamaño de la muestra y los costos, mejora la potencia de prueba y la precisión en la estimación de las tendencias en el tiempo. Si este tipo de análisis se aplica adecuadamente, acentúa la validez de las conclusiones estadísticas, ya que posee mayor exactitud en la estimación de los parámetros del modelo de análisis (Kuehl 2000).

El diseño con medidas repetidas en el tiempo se

studied with the analysis of univariate and multivariate variance (ANOVA and MANOVA), respectively (Fernández *et al.* 1996) and other authors used the mixed linear models and the mixed generalized linear models for the advantages that they present with respect to the traditional ones (Balzarini and Macchiavelli 2005 and Vallejo *et al.* 2010).

The statistical procedure with mixed models allows to analyze correctly and efficiently the data of experiments with repeated measures, through the modeling of the structure of variance-covariance matrix that consider the correlations between repeated measures and the presence of heterogeneous variances to make more precise inferences.

The objective of this study was the proposal of a methodology for statistical analysis to guide the researcher by using experiments with repeated measures over time in the same experimental unit. It is presented through a case study with legumes as a substrate in the production of *in vitro* gas in the agricultural and livestock field.

Materials and methods

Experimental procedure. The information of an experiment belonging to the Department of Physiology of the Institute of Animal Science was carried out with repeated measures at different times in time in the same experimental units, using the *in vitro* gas production technique.

Three shrub legumes were evaluated: *Acacia cornigera* (Acacia), *Albizia lebbekoides* (Albizia) and *Leucaena leucocephala* (Leucaena). Samples were collected from fully established plants in an Arboretum of the Institute of Animal Science (San José de las Lajas, Mayabeque, Cuba) in a typical red ferralitic soil (Hernández *et al.* 2015), without fertilization or irrigation. Leaves and small stems (smaller than 5 mm) of legumes were manually collected simulating the browsing of the animals at 1.5 m height. The plant material was dried in a forced air oven at 60 °C for 72 h. Subsequently, it was ground in a hammer mill, at a particle size of 1 mm. The plant material was properly preserved in sealed nylon bags and sent to the University of Zaragoza (Spain) for further chemical analysis and *in vitro* evaluations (Rodríguez *et al.* 2014).

The variable studied was *in vitro* gas production (mL g⁻¹ OMinc) measured at 2, 4, 6, 8, 10, 12, 16 and 24 hours, at which point the fermentation was stopped after measuring the gas.

Statistical analysis. To determine the existence of an association among sampling schedules, the Pearson correlation matrix was obtained. The sphericity assumption was calculated through the Mauchly statistic (Pérez and Medrano 2010 and Acosta and Sánchez 2015). Before the breach of this, the fit of the degrees of freedom was performed by means of the epsilon of Greenhouse and Geisser (1959) and Huynh

estudió mediante el análisis de la varianza univariado y multivariado (ANOVA y MANOVA), respectivamente (Fernández *et al.* 1996) y otros autores utilizaron los Modelos Lineales Mixtos y los Modelos Lineales Generalizados Mixtos por las ventajas que presentan con respecto a los tradicionales (Balzarini y Macchiavelli 2005 y Vallejo *et al.* 2010).

El procedimiento estadístico con modelos mixtos permite analizar correcta y eficientemente los datos de experimentos con medidas repetidas, a través del modelaje de la estructura de la matriz de varianzas - covarianzas que consideren las correlaciones entre medidas repetidas y la presencia de varianzas heterogéneas para realizar inferencias más precisas.

El objetivo de este trabajo fue la propuesta de una metodología para el análisis estadístico que sirva de guía al investigador al utilizar experimentos con medidas repetidas en el tiempo en la misma unidad experimental. Se presenta a través de un estudio de caso con leguminosas como sustrato en la producción de gas *in vitro* en la esfera agropecuaria.

Materiales y Métodos

Procedimiento experimental. Se empleó la información de un experimento perteneciente al departamento de Fisiología del Instituto de Ciencia Animal realizado con medidas repetidas en diferentes momentos en el tiempo en las mismas unidades experimentales, mediante la técnica de producción de gas *in vitro*.

Se evaluaron tres leguminosas arbustivas: *Acacia cornigera* (Acacia), *Albizia lebbekoides* (Albizia) y *Leucaena leucocephala* (Leucaena). Las muestras se recolectaron de plantas plenamente establecidas en un Arboretum del Instituto de Ciencia Animal (San José de las Lajas, Mayabeque, Cuba) en suelo ferralítico rojo típico (Hernández *et al.* 2015), sin fertilización ni riego. Las hojas y tallos pequeños (menores a 5 mm) de las leguminosas se recolectaron manualmente simulando el ramoneo de los animales a 1.5 m de altura. El material vegetal se secó en estufa de aire forzado a 60 °C durante 72 h. Posteriormente, se molió en molino de martillo, a tamaño de partícula de 1 mm. El material vegetal se conservó adecuadamente en bolsas de nailon selladas y se envió a la Universidad de Zaragoza (España) para posteriores análisis químicos y evaluaciones *in vitro* (Rodríguez *et al.* 2014).

La variable estudiada fue la producción de gas *in vitro* (mL g⁻¹ MOinc) medida a las 2, 4, 6, 8, 10, 12, 16 y 24 horas, momento en que se detuvo la fermentación después de medir el gas.

Análisis estadístico. Para determinar la existencia de asociación entre los horarios de muestreo, se obtuvo la matriz de correlación de Pearson. Se calculó el supuesto de esfericidad a través del estadístico de Mauchly (Pérez y Medrano 2010 y Acosta y Sánchez 2015). Ante el incumplimiento de este, se realizó el ajuste de los

and Feldt (1976). Compliance with the assumption of normality was verified by the tests of Shapiro and Wilk (1965) and Kolmogorov –Smirnov modified by Lilliefors (1967).

In order to obtain estimates with lower bias and lower variance of model parameters, the variance-covariance structures were examined: Unstructured (UN), Toeplitz (TOEP), AutoRegressive of order 1 (Ar (1)), Composite Symmetry (CS) and Components of Variance (CV). These were selected from the smallest values of the information criteria: Akaike (AIC), Corrected Akaike (AICC) and Bayesian (BIC).

Parameters were estimated by the Maximum Restricted Likelihood method and means were compared with the multiple comparison test of Tukey, modified by Kramer, with a significance level for $P < 0.05$ (Tukey 1956).

The estimation method was the approach of Laplace contained in the GLIMMIX procedure of SAS (Gualdrón 2009 and Vallejo *et al.* 2014). Data processing was performed with SAS (2013) statistical package, version 9.3.

To determine the distribution followed by data, SAS Proc Severity was used and Poisson (Logarithmic), Gamma (Reciprocal), Normal Log (Log), Normal (Identity) and Binomial (Logit) distributions were analyzed with their corresponding bonding functions. The expression for the mixed generalized linear model was the following:

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

Where:

y_{ijk} = response variable

μ = intercept or common mean

α_i = fix effect of the i -th treatment ($i=1, \dots, n$)

β_j = fix effect of the j -th time ($j= 1, \dots, n$)

$(\alpha\beta)_{ij}$ = fix effect of the i -th treatment in interaction with the fix effect of the j -th time ($ij=1, \dots, n$)

b_k = random effect of the k -th experimental unit ($k= 1, \dots, n$)

e_{ijk} = random error associated to all observations

Results and Discussion

Table 1 shows the correlation coefficients for the variable *in vitro* gas production where values superior to 0.82 were obtained, which evidences the existence of high correlation throughout the experiment, determined by the proximity in time among sampling schedules. Therefore, the assumption of error independence was not fulfilled. From the hour ten, correlation coefficients reached values of 1.00. At this time, the gas production already expressed its maximum value and begins a stability phase in the process.

Another necessary assumption in repeated measures over time is sphericity, which requires the variances of differences between all pairs of observations to be equal (Caleja *et al.* 2015). Table 2

grados de libertad mediante el ϵ de Greenhouse y Geisser (1959) y Huynh y Feldt (1976). Se comprobó el cumplimiento del supuesto de la normalidad, mediante las d $\acute{o$ cimas de Shapiro y Wilk (1965) y Kolmogorov –Smirnov modificada por Lilliefors (1967).

Con el fin de obtener estimaciones con menor sesgo y menor varianza de los parámetros del modelo, se examinaron las estructuras de varianza-covarianza: No estructurada (UN), Toeplitz (TOEP), Autorregresiva de orden 1 (Ar (1)), Simetría compuesta (CS) y Componentes de varianza (CV). Estas se seleccionaron a partir de los valores más pequeños de los criterios de información: Akaike (AIC), Akaike corregido (AICC) y Bayesiano (BIC).

Los parámetros se estimaron por el método de Máxima Verosimilitud Restringida y para la comparación de medias se utilizó la d $\acute{o$ cima de comparación múltiple de Tukey modificado por Kramer con un nivel de significación para $P < 0.05$ (Tukey 1958).

El método de estimación fue el de aproximación de Laplace contenido en el procedimiento GLIMMIX del SAS (Gualdrón 2009 y Vallejo *et al.* 2014). El procesamiento de los datos se realizó con el paquete estadístico SAS (2013), versión 9.3.

Para determinar la distribución que siguieron los datos se empleó el Proc Severity del SAS y se analizaron las distribuciones Poisson (Logarítmica), Gamma (Recíproca), Log Normal (Log), Normal (Identidad) y Binomial (Logit) con sus correspondientes funciones de enlace.

La expresión del Modelo Lineal Generalizado Mixto es:

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

Donde:

y_{ijk} = variable respuesta

μ = media común o intercepto

α_i = efecto fijo del i -ésimo tratamiento ($i=1, \dots, n$)

β_j = efecto fijo del j -ésimo tiempo ($j= 1, \dots, n$)

$(\alpha\beta)_{ij}$ = efecto fijo del i -ésimo tratamiento en interacción con el efecto fijo del j -ésimo tiempo ($ij=1, \dots, n$)

b_k = efecto aleatorio de la k -ésima unidad experimental ($k= 1, \dots, n$)

e_{ijk} = error aleatorio asociado a todas las observaciones

Resultados y Discusión

En la tabla 1 se muestran los coeficientes de correlación para la variable producción de gas *in vitro* donde se obtuvieron valores superiores a 0.82, lo que evidencia la existencia de alta correlación durante todo el experimento, determinado por la cercanía en el tiempo entre los horarios de muestreo. Por lo tanto, no se cumplió con el supuesto de independencia de los errores. A partir de la hora diez los coeficientes de correlación alcanzan valores de 1.00, en este momento ya la producción de gas expresó su máximo valor y comienza una fase de estabilidad en el proceso.

Otro de los supuestos necesarios en medidas repetidas en el tiempo es la esfericidad, que requiere que las varianzas de las diferencias entre todos los

Table 1. Correlation coefficients for the experiment of *in vitro* gas production

	H2	H4	H6	H8	H10	H12	H16	H24
H2	1.00							
H4	0.95	1.00						
H6	0.91	0.99	1.00					
H8	0.82	0.93	0.96	1.00				
H10	0.85	0.95	0.99	0.96	1.00			
H12	0.85	0.96	0.99	0.97	1.00	1.00		
H16	0.85	0.95	0.99	0.97	1.00	1.00	1.00	
H24	0.85	0.95	0.99	0.97	0.99	0.99	1.00	1.00

H: times

shows the results of the calculation of W statistic of Mauchly (1940) and the correction factor (epsilon) with $P=0.001$, which led to reject the hypothesis that variance-covariance matrix is spherical. That is, variances were not homogeneous (Kirk 1982) and it was necessary to adjust the degrees of freedom by means of the Greenhouse-Geisser and Huynh-Feldt epsilon.

pares de observaciones sean iguales (Caleja *et al.* 2015). En la tabla 2 se presentan los resultados del cálculo del estadístico W de Mauchly (1940) y el factor de corrección (épsilon) con $P=0.001$, lo que llevó al rechazo de la hipótesis que plantea que la matriz de varianza-covarianza es esférica. Es decir, las varianzas no fueron homogéneas (Kirk 1982) y fue necesario realizar el ajuste de los grados de libertad mediante el épsilon de

Table 2. Sphericity test of Mauchly and Epsilon correction for the experiment of *in vitro* gas production

Variable	W of Mauchly	Aprox. χ^2	FD	Value of P	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Inferior limit
<i>In vitro</i> PGas	0.00	398	27	0.001	0.15	0.15	0.14

PGas: gas production

Table 3 shows the traditional technique that accompanies the analysis of univariate variance of fit of degrees of freedom of variability explained by times, and the one attributed to the term of error through their reduction, where it was tried to compensate the positive bias of F test when the assumption of homogeneity of variances was not fulfilled. With the aim of an approach to sphericity assumption, a reduction in the degrees of freedom was performed, with values of 2.11 and 2.79, although it was observed that, in all cases, the same value of F was obtained (the uncorrected and the three corrected). This led to the same conclusion, since the level of significance was lower than 0.05 which allowed to reject the hypothesis of equality of means (Frías and García 1996).

Greenhouse-Geisser y Huynh-Feldt.

La tabla 3 muestra la técnica tradicional que acompaña al análisis de varianza univariado de ajuste de los grados de libertad tanto de la variabilidad explicada por los horarios, como la atribuida al término de error a través de la reducción de los mismos, donde se trató de compensar el sesgo positivo de la prueba F cuando no se cumplió la asunción de homogeneidad de varianzas. Con el objetivo de un acercamiento al supuesto de esfericidad, se realizó una reducción de los grados de libertad con valores de 2.11 y 2.79, aunque se pudo observar que en todos los casos se obtuvo el mismo valor de la F (la no corregida y las tres corregidas) esto condujo a la misma conclusión, ya que el nivel de significación fue menor que 0,05 lo que permitió rechazar la hipótesis de igualdad de medias (Pascual *et al.* 1996).

Tabla 3. Fitting of degrees of freedom for the experiment of *in vitro* gas production

	Origin	Type III of square sum	FD	Mean square	F	Signif.
Schedules	Assumed sphericity	28710.17	7	4101.45	1453.17	0.00
	Greenhouse-Geisser	28710.17	2.11	13591.30	1453.17	0.00
	Huynh-Feldt	28710.17	2.79	10271.34	1453.17	0.00
	Inferior limit	28710.17	1.00	28710.17	1453.17	0.00
Error (times)	Assumed sphericity	296.35	105.00	2.82		
	Greenhouse-Geisser	296.35	31.69	9.35		
	Huynh-Feldt	296.35	41.93	7.07		
	Inferior limit	296.35	15.00	19.80		

Results of statistic tests for checking normality assumption, with $P < 0.0100$, appear in table 4. For the variable *in vitro* gas production, this hypothesis was rejected, residues did not approximate to a normal distribution and this allowed the selection of the mixed generalized linear model as an alternative analysis for non-fulfilling this assumption.

Los resultados de las d cimas estad sticas para la comprobaci n del supuesto de la normalidad, con $P < 0.0100$, se muestran en la tabla 4. Para la variable producci n de gas *in vitro* se rechaz  dicha hip tesis los residuos no se aproximaron a una distribuci n normal y esto permiti  la selecci n del Modelo Lineal Generalizado Mixto como una alternativa de an lisis

Table 4. Normality test for the experiment of *in vitro* gas production

Variable	<i>In vitro</i> PGas
Statistical test	Value of P
Shapiro-Wilk	0.0000
Kolmogoro-Smirnov	<0.0100

Table 5 shows the variance-covariance structures and information criteria studied. UN, TOEP and CS structures showed the same performance, so that any of them could be selected. The analysis was based on the lower values of the obtained information criteria, and the same residual value. However, TOEP was selected after studying the statements by Fern ndez *et al.* (1996) and Vallejo *et al.* (2010), who expressed that observations recorded from the same subject, in addition to being positive and gradually correlated, show a variance-covariance matrix among repeated measures that have a TOEP structure. That means that the closest scores have a higher correlation.

ante el incumplimiento de este supuesto.

En la tabla 5 se exponen las estructuras de varianza-covarianza y los criterios de informaci n estudiados. Las estructuras UN, TOEP y CS, mostraron igual comportamiento, por lo que cualquiera de ellas pudo ser seleccionada, el an lisis se bas  en los menores valores de los criterios de informaci n obtenidos, e igual valor del residual. Sin embargo, se seleccion  la TOEP despu s de estudiar lo expuesto por Fern ndez *et al.* (1996) y Vallejo *et al.* (2010), quienes plantearon que las observaciones registradas desde un mismo sujeto, adem s de estar positiva y gradualmente correlacionadas, presentan una matriz de varianza-covarianza entre las medidas repetidas

Table 5. Variance-covariance structure and information criteria for *in vitro* gas production in legumes experiment

Information criteria	Variance-covariance structures				
	UN	TOEP	Ar(1)	VC	CS
AIC	751.84	751.84	753.84	753.84	751.84
AICC	763.84	763.84	766.88	766.88	763.84
BIC	746.43	746.43	748.22	748.22	746.43
Residual	0.01				

Table 6 shows that, for the variable accumulated gas production, there was interaction among the factors treatment and sampling times ($P = 0.0024$). With the application of this model, it was observed that *in vitro* gas production for Acacia and Leucaena, showed no differences between them at any sampling time and their values were always superior to those of Albizia. It is appreciated that the highest *in vitro* gas productions are reached at 24 hours. Acacia and Leucaena showed a similar performance, as well as acacia at 16 hours. On the other hand, the lowest *in vitro* gas production was obtained with albizia at the beginning of fermentation.

que tienen una estructura TOEP; o sea, las puntuaciones m s pr ximas presentan una correlaci n m s elevada.

En la tabla 6 se observ  que para la variable producci n de gas acumulada hubo interacci n entre los factores tratamientos y horarios de muestreo ($P=0,0024$). Con la aplicaci n de este modelo se observ  que la producci n de gas *in vitro* para Acacia y Leucaena, no mostr  diferencias entre ellos en ning n horario de muestreo y siempre fueron mayores a la de Albizia. Se aprecia que las mayores producciones de gas *in vitro* se alcanzan a las 24 horas, mostrando un comportamiento similar la Acacia y la Leucaena as  mismo la Acacia a las 16 horas. Por otra parte, la menor producci n de gas *in vitro* se obtuvo con la Albizia al inicio de la fermentaci n.

From the results, a methodological proposal is made and the steps to be followed in research in the agricultural and livestock field in which experiments with repeated measures over time in the same experimental unit are

A partir de los resultados obtenidos se realiza una propuesta metodol gica y se describen, con mayor, precisi n los pasos a seguir en investigaciones en la esfera

Table 6. Means of interaction between treatment and sampling times in the accumulated

Times (h)	Treatment			SE± Sign.
	Acacia	Albizia	Leucaena	
2	2.15 ^m (8.57)	1.54 ⁿ (4.67)	2.23 ^m (9.26)	
4	2.96 ^{ijk} (19.23)	2.23 ^m (9.31)	2.93 ^{ijk} (18.74)	
6	3.37 ^{fg} (29.05)	2.55 ^l (12.82)	3.25 ^{fgh} (25.75)	
8	3.63 ^{de} (37.87)	2.76 ^{kl} (15.80)	3.42 ^{ef} (30.58)	±0.0494
10	3.83 ^{cd} (45.90)	2.91 ^{jk} (18.45)	3.61 ^{de} (37.03)	P=0.0024
12	3.96 ^{bc} (52.67)	3.03 ^{hij} (20.79)	3.75 ^{cd} (42.37)	
16	4.13 ^{ab} (61.99)	3.16 ^{ghi} (23.58)	3.90 ^{bc} (49.36)	
24	4.28 ^a (72.18)	3.31 ^{fg} (27.29)	4.07 ^{ab} (58.76)	

a,b,c,d,e,f,g,h,i,j,k,l,m,n Different letters indicate significant differences for P<0.05

() original means

evaluated and described with greater precision:

1. Calculate the Pearson correlation matrix to determine the degree of association among sampling times

2. Analyze the fulfillment of the sphericity condition using the Mauchly test and, otherwise, apply the correction factor.

- Mauchly proves that variance-covariance matrix is spherical or not, and if it is not, it increases the probability of committing type I error. Therefore, it is necessary to correct the degrees of freedom through the epsilon of Huynh-Feldt and Greenhouse-Heisser

3. Analyze the theoretical assumption of normality with the tests of Kolmogorov-Smirnov and Shapiro-Wilk.

4. Examine several variance-covariance structures to obtain estimations with lower bias and lower variance of the model parameters.

- Unstructured (UN)
- Toeplitz (TOEP)
- Autoregressive (AR1)
- Variance components (VC)
- Composite symmetry (CS)

5. Obtaining the information criteria that help to select the most appropriate variance-covariance structure.

- Akaike (AIC)
- Corrected Akaike (AICC)
- Bayesian (BIC)

6. For the best fit of the model, choose the lowest values of the information criteria, to obtain the most appropriate variance-covariance structure.

7. Define the model to be used for each particular

agropecuaria en que se evalúen experimentos con medidas repetidas en el tiempo en la misma unidad experimental:

1. Calcular la matriz de correlación de Pearson para determinar el grado de asociación entre los tiempos de muestreo.

2. Analizar el cumplimiento de la condición de esfericidad mediante la prueba de Mauchly y en caso contrario aplicar el factor de corrección.

- Mauchly prueba que la matriz de varianza-covarianza es esférica o no, si no lo es, aumenta la probabilidad de cometer error de tipo I, por lo tanto es necesario corregir los grados de libertad mediante el epsilon de Huynh-Feldt y Greenhouse-Heisser.

3. Analizar el supuesto teórico de la normalidad mediante las dójimas de Kolmogorov Smirnov y Shapiro Wilk.

4. Examinar varias estructuras de varianza-covarianza para obtener estimaciones con menor sesgo y menor varianza de los parámetros del modelo.

- No estructurada (UN)
- Toeplitz (TOEP)
- Autorregresiva (AR1)
- Componentes de varianza (VC)
- Simetría compuesta (SC)

5. Obtención de los criterios de información que ayuden a seleccionar la estructura de varianza-covarianza más adecuada.

- Akaike (AIC)
- Akaike corregido (AICC)
- Bayesiano (BIC)

6. Para el mejor ajuste del modelo, escoger los menores valores de los criterios de información, para obtener la estructura de varianza-covarianza más adecuada.

situation:

a) If the assumption of normality was met, the Mixed Linear Model will be used.

b) If the assumption of normality was not met, try the variants of Poisson, Gamma, Normal Log, Normal and Binomial distributions with their respective Logarithmic, Identity and Logistic link functions. To use the mixed generalized linear model.

Conclusions

The breach of the assumption of normality of residues from the used tests defined the use of the mixed generalized linear model as an alternative of analysis in experiments with repeated measures over time in the agricultural sector. The information criteria allowed obtaining the optimal structure of the variance-covariance matrix. A work methodology is proposed for processing data with these characteristics.

7. Definir el modelo a emplear para cada situación en particular:

a) Si se cumplió con el supuesto de la normalidad se utilizará el Modelo Lineal Mixto.

b) Si no se cumplió con el supuesto de normalidad probar las variantes de las distribuciones de Poisson, Gamma, Log Normal, Normal y Binomial con sus respectivas funciones de enlace Logarítmico, Identidad y Logístico. Para utilizar el Modelo Lineal Generalizado Mixto.

Conclusiones

El incumplimiento del supuesto de normalidad de los residuos a partir de las dójimas utilizadas definió el empleo del Modelo Lineal Generalizado Mixto como alternativa de análisis en experimentos con medidas repetidas en el tiempo en el sector agropecuario. Los criterios de información permitieron la obtención de la estructura óptima de la matriz de varianza-covarianza. Se propone una metodología de trabajo para el procesamiento de datos con estas características.

References

- Acosta, M.M. & Sánchez, J. P.2015. Desempeño psicométrico de dos escalas de autoeficacia e intereses profesionales en una muestra de estudiantes de secundaria. *CES Psicología*. 8 (2):156-170. Available: <http://www.redalyc.org/articulo.oa?id=423542417009>.
- Balzarini, M. & Macchiavelli, R, 2005. Aplicaciones de Modelo Lineal Mixto en agricultura y forestería. Curso Internacional Aplicaciones de Modelo Lineal Mixto en Agricultura y Foresteria. CATIE, Turrialba, Costa Rica, Mimeo, p.189.
- Caleja, C., Barros, L., Antonio, A.L., Ciric, A., Barreira, J.C.M., Sokovic, M., Oliveira, B.P.P., Santos-Buelga, C. & Ferreira, I.C.F.R. 2015. Development of a functional dairy food: Exploring bioactive and preservation effects of chamomile (*Matricaria recutita* L.). *Journal of Functional Foods*. 16: 114–124. ISSN: 1756-4646. Available: <http://dx.doi.org/10.1016/j.jff.2015.04.033>.
- Fernández, P., Menéndez, I.A., Vallejo, G. & Herrero, J. 1996. Comparación de la potencia y robustez del AMVAR con dependencia serial en el error, cuando diferentes asunciones distribucionales son violadas. *Departamento de Psicología, Universidad de Oviedo. Psicothema* 4(1):277-290 ISSN:0214-9915.
- Greenhouse, S. & Geisser, S. 1959. On methods in the analysis of profile data. *Psychometrika*. 24(2): 95-112. Online ISSN: 1860-0980. Available: <https://doi.org/10.1007/BF02289823>
- Gualdrón, J.C. 2009. Influencia de los criterios de selección AIC Y BIC para la selección del modelo de evolución y la reconstrucción del análisis bayesiano. Available: <<http://tux.uis.edu.co/labsist/docencia/finales/final2009-1/2050158-20070.pdf>. [Consulted: June 20, 2018].
- Hernández, J.A., Pérez, J.J., Bosch, I.D. & Castro, S.N. 2015. Clasificación de los suelos de Cuba 2015. Mayabeque, Cuba. Ediciones INCA. 93 p. ISBN: 978-959-7023-77-7.
- Huynh, H. & Feldt, L.S. 1976. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. *J. Educ. Stat.* 1(1), 69–82. doi:10.3102/10769986001001069. Available: <https://journals.sagepub.com/>
- Kirk, R. 1982. *Experimental design: Procedures for the behavioral sciences*. 2nd Edition. Brooks Cole Publishing Company. California. p. 55. ISBN-10: 081850286X
- Kuehl, R.O. 2000. *Diseño de experimentos, Principios estadísticos de diseño y análisis de investigación*. Segunda edición. Ed, Thomson Learning. Universidad de Arizona, Arizona, USA. p. 492-519. ISBN-0-534-36834-4. Available: <http://www.thomsonlearning.com.mx>.
- Lilliefors, H. 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *J. Am. Stat. Assoc.* 62(318): 399-402. DOI: 10.1080/01621459.1967.10482916. Available: Stable URL: <http://www.jstor.org/stable/2283970> http://academicos.fciencias.unam.mx/wp-content/uploads/sites/91/2015/04/Lilliefors_normality_ks.pdf.
- Mauchly, J. 1940. Significance test of sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics*. 11(2): 204-209. Available: <https://www.jstor.org/stable/2235878>
- Pascual, J., Frías, D. & García, F. 1996. *Manual de psicología experimental. Metodología de investigación*. Libro Entero. Primera Edición. Editorial Ariel, S.A. Barcelona. p. 139-145. ISBN: 84-344-0868-6. Available: https://www.academia.edu/23242604/Manual_de_psicologia_experimental_metodologia_de_investigacion.
- Pérez, E. & Medrano, L. 2010. Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento (RACC)*. 2(1): 58-66. ISSN-e 1852-4206. Available: www.psyche.unc.edu.ar/racc.
- Rodríguez, R., de la Fuente, G., Gómez, S. & Fondevila, M. 2014. Biological effect of tannins from different vegetal origin

- on microbial and fermentation trait *in vitro*. Anim. Prod. Sci. 54 (8): 1039-1046, ISSN: 1836-0939. Available: <http://dx.doi.org/10.1071/AN13045>.
- SAS. 2013. Sistema de análisis estadístico. Universidad de Nebraska. Versión 9.3.
- Shapiro, S. & Wilk, B. 1965. An analysis of variance test for normality (complete samples). Biometrika. 52(3/4): 591-611. doi:10.1093/biomet/52.3-4.591. JSTOR 2333709. MR 0205384. p. 593.
- Tukey, J.W. 1958. Bias and confidence in not quite large samples. The Annals of Mathematical Statistics. 29(2):614-623. Available: doi:10.1214/aoms/1177706647. <https://projecteuclid.org/euclid.aoms/1177706647>
- Vallejo, G., Arnau, J., Bono, R., Fernández, P. & Tuero, E. 2010. Selección de modelos anidados para datos longitudinales usando criterios de información y la estrategia de ajuste condicional. Psicothema. 22 (2):323-333. ISSN: 0214 - 9915
- Vallejo, G., Tuero, E., Núñez, J.C. & Rosario, P. 2014. Performance evaluation of recent information criteria for selecting multilevel models in Behavioral and Social Sciences. International Journal of Clinical and Health Psychology. 14(1): 48-57. ISSN: 1697-2600. Available: <https://www.redalyc.org/pdf/337/33729172006.pdf>

Received: July 5, 2019