

## Integral evaluation of indicators in models of parametric and non-parametric analysis of variance. Use of the categorical principal component

### Valoración integral de indicadores en los modelos de análisis de varianza paramétrico y no paramétrico. Uso del componente principal categórico

Magaly Herrera<sup>1</sup>, Caridad Walkiria Guerra<sup>2</sup> and Yolaine Medina<sup>1</sup>

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

<sup>2</sup>*Centro Universitario Municipal de Güines, Calle 86, No.7312, Güines, Mayabeque, Cuba*

*Email: mvillafranca@ica.co.cu*

In order to establish possible relations among statistical indicators in models of parametric and non-parametric analysis of variance, belonging to completely randomized and random block experimental designs, the categorical principal component analysis was used because they are quantitative and qualitative. The models of analysis of variance of simple classification, with 16 experiments, were selected, as well as those of double classification with five experiments. An amount of 100 discrete and categorical variables were analyzed. A matrix of data was designed using the indicators of completely randomized designs and the test of Kruskal-Wallis, which is its non-parametric homologue, the model of random blocks with its non-parametric homologue, and the test of Friedman. The categorical principal component analysis showed adequate reliability and a variability percentage explained with 0.94. The indicators with more importance in the first dimension are related to probability of type I error and power, showing absolute values close to one and allowing to determine their contribution in this study. The results evidenced the existing relations among the analyzed statistical indicators, from the high degree of positive correlation over 0.90 among the values of probability of type I error in the F test of Fisher (with and without transformation) and its non-parametric homologue test, as well as the high negative correlations, existing between around 0.8 and 0.93 of them with values of power (with or without data transformation). It is necessary to continue the analysis for different data distributions and sample sizes.

**Key words:** *statistical indicators, models of simple and double variance analysis, categorical principal component analysis*

There are some indicators in the analysis of variance that allow to evaluate quality and rigor of this procedure. Some of them are probability of type I error, test power, sample size and fulfillment of assumptions. In this last, data should distribute normally and have homogeneous variances. However, many times they are not taken into account, even though they provide a valuable information on effectiveness of the used analysis procedures.

De Calzadilla (1999), Vásquez (2011), De Calzadilla *et al.* (2002) studied and determined relations among indicators related to the analysis of variance. However, there is no information in the literature about an analysis that integrates the results of parametric and non-parametric analysis of variance.

In order to carry out this integrating analysis, a

Para establecer posibles relaciones entre indicadores estadísticos en los modelos de análisis de varianza paramétrico y no paramétrico, correspondientes a los diseños completamente aleatorizados y de bloques al azar, se utilizó el análisis de componentes principales categóricos por ser estos cuantitativos y cualitativos. Se seleccionaron los modelos de análisis de varianza de clasificación simple, con dieciséis experimentos, y doble con cinco. Se analizaron en total 100 variables de tipos discretas y categóricas. Con los indicadores de los diseños completamente aleatorizados y su homólogo no paramétrico, la dócima de Kruskal-Wallis, y el de modelo de bloques al azar con su homólogo no paramétrico, la dócima de Friedman, se conformó una matriz de datos. El análisis de componentes principales categóricos mostró adecuada fiabilidad y porcentaje de variabilidad explicada con 0.94. Los indicadores con mayor peso en la primera dimensión se encuentran relacionados con la probabilidad de error tipo I y la potencia, que presentaron valores absolutos cercanos a uno, que permiten determinar el aporte de los mismos en el estudio. Los resultados evidenciaron las relaciones que existen entre los indicadores estadísticos analizados, a partir del alto grado de correlación positiva por encima de 0.90 entre los valores de probabilidad de error tipo I en la dócima F de Fisher (sin transformación y con ella) y la dócima homóloga no paramétrica, así como de las altas correlaciones negativas que existen entre 0.8 y 0.93 aproximadamente de estos con los valores de potencia (sin y con transformación de los datos). Se considera oportuno continuar el análisis para diferentes distribuciones de datos y tamaños de muestras.

**Palabras clave:** *indicadores estadísticos, modelos de análisis de varianza simple y doble, análisis de componentes principales categóricos.*

En el análisis de varianza existen ciertos indicadores que permiten valorar la calidad y rigor de este procedimiento. Entre estos se encuentran la probabilidad error tipo I, la potencia de la dócima, el tamaño de muestra y el cumplimiento de los supuestos. En este último, los datos se deben distribuir aproximadamente normal y tener varianzas homogéneas. Sin embargo, en muchas ocasiones estos no se tienen en cuenta, a pesar de que aportan valiosa información acerca de la efectividad de los procedimientos de análisis empleados.

De Calzadilla (1999), Vásquez (2011), De Calzadilla *et al.* (2002) estudiaron y determinaron relaciones entre los indicadores vinculados con el análisis de varianza. Sin embargo, en la literatura científica no se informa un análisis en el que se integren los resultados del análisis de varianza paramétrico y no paramétrico.

multivariate statistical method is needed, which will allow to establish possible relations among these indicators, and evaluate its contribution to the research. It is suggested the use of categorical principal component analysis (CATPCA), which allows to analyze qualitative and quantitative variables.

The CATPCA is similar to the standard principal component analysis (PCA). It is used with the same objective but, unlike this last, the CATPCA allows to scale the variables at different measuring levels and also allows non-linear relations among them (Molina and de los Monteros 2010).

Like its homologue for continuous variables (principal component analysis), this method may be considered as an exploratory technique for dimension reduction (Navarro *et al.* 2010, Vázquez 2012). The application of this method allows to integrate statistical indicators from ANAVA due to their characteristics.

Therefore, the objective of this study was to perform an integral analysis of a group of statistical indicators associated to the models of parametric and non-parametric analysis of variance, and to establish possible relations among them through the procedure of categorical principal components.

### Materials and Methods

The information selected for the database was processed from 2003 and 2011 by the Department of Biomathematics from the Institute of Animal Science, located in San José de las Lajas, Mayabeque province. Data belonged to researches developed by specialists from different departments.

The selected researches were related to 16 experiments that used the completely randomized design (CRD), and some others related to five experiments that applied the random block design (RBD), summing 100 discrete and categorical variables.

In the selected variables, the theoretical assumptions of the analysis of variance and normality of errors were analyzed. For that purpose, the test of Shapiro and Wilk (1965) was used, as well as the test of Levene (1960) for analyzing homogeneity of variance. Both were applied to the original variables, after the use of  $\sqrt{X}$ ,  $\sqrt{X+0.375}$ , arcsine ( $\sqrt{p}$ ) and Log X transformations for variables of counting and percentage, expressed in a logarithmic scale, respectively.

The parametric analysis of variance was used for CRD and its non-parametric homologue, as well as the Kruskal-Wallis test for RBD and its non-parametric homologue, and the test of Friedman. Both analysis were compared and the statistical results were evaluated. A matrix of data was produced with the obtained information. The lines were the result of the different selected researches, related to the used designs. The columns were the statistical indicators and those of the experimental design that are later mentioned. The matrix

Para realizar este análisis integrador, se necesita un método estadístico multivariado, que permita establecer posibles relaciones entre estos indicadores, y valorar su aporte en la investigación que se realice. Se recomienda la utilización del análisis de componentes principales categórico (CATPCA), que permite analizar variables cualitativas y cuantitativas.

El CATPCA es similar al análisis de componente principal (ACP) estándar. Se utiliza con el mismo objetivo, pero a diferencia de este, el CATPCA es un método que admite escalar las variables a diferentes niveles de medidas y permite relaciones no lineales entre ellas (Molina y de los Monteros 2010).

Al igual que su homólogo para variables continuas (análisis de componentes principales), este método se puede considerar como una técnica exploratoria de reducción de las dimensiones (Navarro *et al.* 2010, Vázquez 2012). La aplicación de este método permite integrar los indicadores estadísticos del ANAVA por las características que ellos presentan.

Es por ello que el objetivo de este trabajo fue realizar un análisis integral de un grupo de indicadores estadísticos asociados a los modelos de análisis de varianza paramétrico y no paramétrico, y establecer posibles relaciones entre ellos mediante el procedimiento de componente principal categórico.

### Materiales y Métodos

Se seleccionó la información de las bases de datos que se procesaron en el período 2003-2011 por el departamento de Biomatemática del Instituto de Ciencia Animal, ubicado en el municipio San José de las Lajas, provincia Mayabeque. Los datos correspondieron a investigaciones que desarrollaron especialistas de los distintos departamentos.

Se seleccionaron las investigaciones relacionadas con dieciséis experimentos que aplicaron el diseño completamente aleatorizado (DCA), balanceado, y otras con cinco experimentos de bloques al azar (DBA), para 100 variables en total, discretas y categóricas.

A las variables seleccionadas se les analizaron los supuestos teóricos del análisis de varianza y normalidad de los errores. Para ello se utilizó la dócima de Shapiro y Wilk (1965) y la homogeneidad de varianza, que se analizó mediante la dócima de Levene (1960). Ambas se aplicaron a las variables originales, después del uso de las transformaciones  $\sqrt{X}$ ,  $\sqrt{X+0.375}$ , arcoseno ( $\sqrt{p}$ ) y Log X, para variables de conteo y porciento, expresadas en escala logarítmica, respectivamente.

Se utilizó el análisis de varianza paramétrico para el DCA y su homólogo no paramétrico, así como la dócima de Kruskal-Wallis y el DBA, con su homólogo no paramétrico, la dócima de Friedman. Ambos análisis se compararon y se valoraron los resultados estadísticos. Con la información obtenida, se confeccionó una matriz de datos. Las filas fueron los resultados de las diferentes

had a 100 x12 dimension:

- Type of design (design)
- Type of experiment (code)
- Fulfillment of the assumptions without transformation (Cumpl S/T)
- Fulfillment of the assumptions with transformation (Cumpl T)
- P value of Fisher F test without transformation (Valp S/T)
- P value of Fisher F test with transformation (ValpT)
- P value of non-parametric test (Valp NP)
- Power value for Fisher F test without transformation (Pot S/T)
- Power value for Fisher F test with transformation (Pot T)
- Sample size (SS)
- Amount of treatments (No. Tto)
- Distribution (Distcod)

The CAPTCA was used for performing an integral evaluation of statistical indicators and establish its possible relations, as well as to represent the information through a Biplot graph. The alpha index of Cronbach was used for measuring reliability of the method with a scale of values proposed by Hair *et al.* (1999).

For the analysis of theoretical assumptions of the model of analysis of variance, Statistica (StatSoft 2007) program was used. For the parametric and non-parametric analysis of variance, the statistical package Infostat (Di Rienzo *et al.* 2012) was applied, as well as the SPSS statistical package, 19.0 version (IBM Corporation 2010), for categorical principal component analysis.

### Results and Discussion

Table 1 shows a summary of the results of CATPCA for the analyzed indicators. Two dimensions were obtained, which explain 60.8 % of variability of the original information. The Cronbach's alpha coefficient showed a total reliability level of 0.94, which is considered as excellent, according to the scale proposed by Hair *et al.* (1999). These authors indicated that the used method is adequate when the values of this coefficient are between 0.60 and 0.70. Dimension one is emphasized because it explains 43.13 % of the total variance, with a coefficient of 0.88. This evidences that the indicators represented in this dimension show a good level of reliability.

investigaciones seleccionadas, relacionadas con los diseños empleados. Las columnas fueron los indicadores estadísticos y del diseño experimental que se relacionan posteriormente. La matriz tuvo una dimensión de 100 x12:

- Tipo de diseño (diseño)
- Tipo de experimento (código)
- Cumplimiento de los supuestos sin transformación (Cumpl S/T)
- Cumplimiento de los supuestos con transformación (Cumpl T)
- Valor p de la dócima F de Fisher sin transformación (Valp S/T)
- Valor p de la dócima F de Fisher con transformación (ValpT)
- Valor p de la dócima no paramétrica (Valp NP)
- Valor de Potencia para dócima F de Fisher sin transformar (Pot S/T)
- Valor de Potencia para dócima F de Fisher con transformación (Pot T)
- Tamaño de muestra (TM)
- Número de tratamientos (No. Tto)
- Distribución (Distcod)

El análisis de componentes principales categóricos (CAPTCA) se empleó para realizar una valoración integral de los indicadores estadísticos y establecer sus posibles relaciones, así como para representar la información mediante el gráfico Biplot. Se utilizó el índice alfa de Cronbach para medir la fiabilidad del método con la escala de valores propuesta por Hair *et al.* (1999).

Para el análisis de los supuestos teóricos del modelo de análisis de varianza, se empleó el programa Statistica (StatSoft 2007). Para el análisis de varianza paramétrico y no paramétrico, el estadístico Infostat (Di Rienzo *et al.* 2012), y para el de componentes principales categóricos (CAPTCA) se aplicó el paquete estadístico SPSS, versión 19.0 (IBM Corporation 2010).

### Resultados y Discusión

La tabla 1 muestra un resumen de los resultados del CATPCA para los indicadores analizados. Se obtuvieron dos dimensiones, que explican 60.8 % de variabilidad de la información original. El coeficiente alfa de Cronbach mostró un nivel de fiabilidad total de 0.94, que se considera excelente, según la escala propuesta por Hair *et al.* (1999). Estos autores indican que cuando los valores de este coeficiente están entre 0.60 y 0.70, el método empleado es adecuado. Se destaca la dimensión uno, que explica 43.13 % de la varianza total, con coeficiente de

Table 1. Results of the Alpha of Cronbach coefficient and explained variance for the analysis of CRD and RBD

Dimension	Cronbach's alpha	Explained variance	
		Total (seigenvalues)	% of variance
1	0.880	5.176	43.130
2	0.576	2.120	17.671
Total	0.941	7.296	60.800

Table 2 shows the most important indicators of each dimension. In the first one, there is a list of those related to probability of type 1 error and power, which present positive and negative values close to one. They allow to determine the contribution of those indicators in the study, which represent one of the aspects to be considered by researchers in the process of experimentation. In dimension two, indicators showing a superior contribution to information are those related to amount of treatments, design (CRD and RBD) and type of experiment, which present values over 0.60.

The most notable indicators in dimension one show correlations over 0.80 of absolute value. Those from dimension two show positive correlations over 0.60 (table 3).

It should be highlighted that high negative correlations among probability values of type 1 error and powers, showing an inverse relation, which means a high probability of finding significant differences or rejecting the null hypothesis when it is false, evidenced that with inferior values of  $\alpha$ , there were superior values of power and vice versa. This result agrees with that stated by Gómez (2005).

Figure 1 shows a graphic representation of the analyzed statistical indicators that were divided into four groups. The first group includes indicators associated with probability values of type 1 error, Fisher F test, ANAVA test for variables (with and without transformation) and the non-parametric homologues Kruskal- Wallis and Friedman (PNP value). The second group contains criteria related to the fulfillment of theoretical assumptions of ANAVA. The third group includes power values (with and without transformation) and the fourth group, which is different from the rest, contains the indicators related to type of experiment and experimental design.

It is considered that the grouping of probability

0.88. Esto evidencia que los indicadores representados en esta dimensión presentan un buen nivel de fiabilidad.

En la tabla 2 se muestran los indicadores con mayor peso en cada dimensión. En la primera se encuentran los relacionados con la probabilidad de error tipo I y la potencia, que presentan valores positivos y negativos cercanos a uno. Estos permiten determinar el aporte de los indicadores en el estudio, que representan uno de los aspectos a considerar por los investigadores en el proceso de experimentación. En la dimensión dos, los indicadores que realizan mayor contribución a la información son aquellos relacionados con el número de tratamientos, diseño (DCA y DBA) y tipo de experimento, que muestran valores por encima de 0.60.

Los indicadores que sobresalen en la dimensión uno presentan correlaciones por encima de 0.80 en valor absoluto. Los de la dimensión dos, muestran correlaciones positivas por encima de 0.60 (tabla 3).

Se debe destacar que las altas correlaciones negativas entre los valores de probabilidad de error tipo I y las potencias, que manifiestan relación inversa, que significa alta probabilidad de encontrar diferencias significativas en gran magnitud o, lo que es lo mismo, rechazar la hipótesis nula cuando es falsa, evidenciaron que a valores pequeños de  $\alpha$  se encontraron valores altos de potencias y viceversa. Este resultado concuerda con lo que plantea Gómez (2005).

La figura 1 muestra la representación gráfica de los indicadores estadísticos analizados que formaron cuatro grupos. En el primero, se agrupan los indicadores asociados con los valores de probabilidad de error tipo I, de la F de Fisher de la dócima del ANAVA, para las variables (con transformación y sin ella) y los homólogos no paramétricos Kruskal- Wallis y Friedman (valor p NP). En el segundo grupo, se ubican los criterios relacionados con el cumplimiento de los supuestos teóricos de ANAVA. En el tercero, los valores de la potencia (con transformación y sin ella) y en el cuatro, más alejado del

Table 2. Results of the most important statistical indicators for the completely randomized and random block designs

Indicators	Dimension	
	1	2
Type of experiment	-0.351	<b>-0.673</b>
Fulfillment of the assumptions without transformation (Cumpl S/T)	-0.027	0.372
Fulfillment of the assumptions with transformation (Cumpl T)	-0.179	0.329
Probability of type 1 error without transformation (PValue S/T)	<b>0.923</b>	-0.186
Probability of type 1 error with transformation (PValue T)	<b>0.952</b>	-0.152
Non-parametric probability of type 1 error	<b>0.942</b>	-0.068
Power without transformation (Pow S/T)	<b>-0.938</b>	0.061
Power with transformation (Pow T)	<b>-0.905</b>	0.122
Sample size (SS)	0.581	-0.173
Amount of treatments	0.236	<b>0.764</b>
Design	0.505	<b>0.730</b>
Distribution	-0.157	0.439

Table 3. Matrix of correlations of transformed variables per dimensions for the completely randomized and random block design

Indicators	Type of experiment	Cumpl S/T	Cumpl C/T	P value S/T	P value C/T	P value NP	Power S/T	Power C/T	T. M	Amount of treatments	Design	Distcod
Typ of experiment (code)												
	1.000											
Cumpl S/T	-0.120	1.000										
Cumpl C/T	-0.001	<b>0.614</b>	1.000									
PValue S/T	-0.186	-0.056	-0.109	1.000								
PValue C/T	-0.219	-0.008	-0.122	<b>0.947</b>	1.000							
PValue NP	-0.274	-0.017	-0.064	<b>0.896</b>	<b>0.929</b>	1.000						
Power S/T	0.224	-0.009	0.091	<b>-0.877</b>	<b>-0.918</b>	<b>-0.875</b>	1.000					
Power C/T	0.175	-0.088	0.135	<b>-0.856</b>	<b>-0.870</b>	<b>-0.853</b>	<b>0.861</b>	1.000				
T. M.	-0.128	-0.175	-0.378	0.428	0.460	.453	-0.486	-0.461	1.000			
Amount of treatments	-0.413	0.071	-0.007	0.013	0.075	.161	-0.170	-0.096	0.033	1.000		
Design	<b>-0.631</b>	0.061	-0.025	0.294	0.322	.374	-0.385	-0.304	0.291	<b>0.731</b>	1.000	
Distcod	-0.177	0.010	0.074	-0.146	-0.199	-0.193	0.117	0.127	-0.142	0.165	0.138	1.000

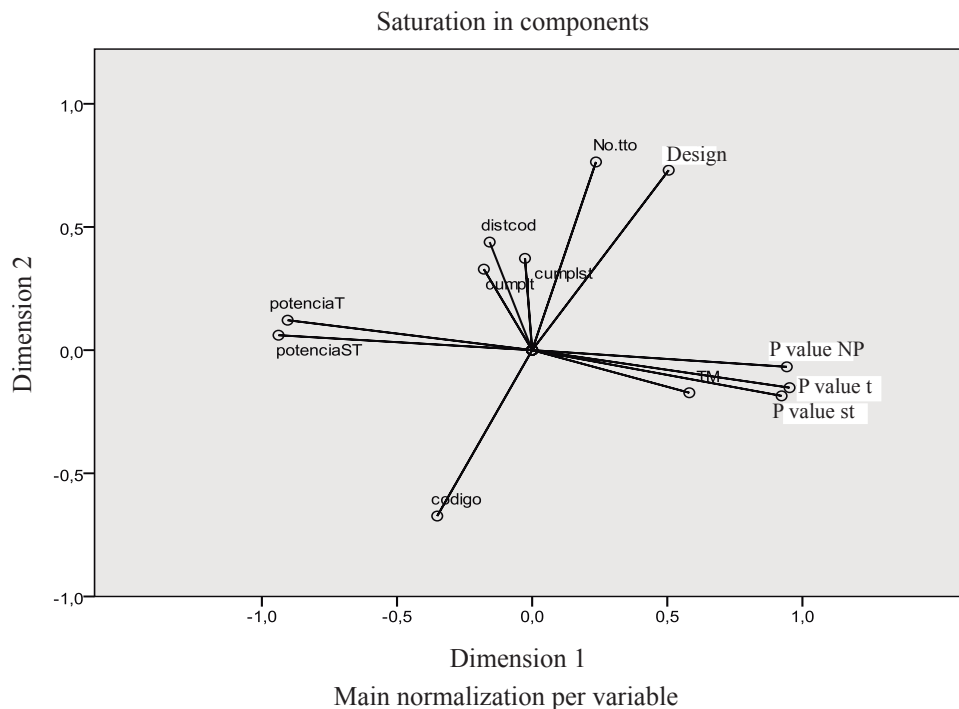


Figure 1. Biplot of analyzed statistical indicators

values of parametric and non-parametric tests is possible because the Fisher F test is characterized by being strong in front of heterogeneity of variance, mainly when it works with the same number of observations per treatments (Steel and Torrie 1992, Peña 1994), as in the case of the analyzed designs. In addition, it presents high power with a low probability of making a type 1 error, manifested in the high negative correlations between powers and probability values of type 1 error (with or without data transformation).

The high negative correlations between power and probability of type 1 error, for the non-parametric test, are due to this last, regarding the parametric one (under the assumption of normal distribution), shows high relative asymptotic efficiency (RAE), which is 95.5%. This corresponds to the criterion of power-efficiency, stated by Siegel (1970), Siegel and Castellan (1995), de Calzadilla (1999), who suggested that, in order to achieve similar results regarding rejection of null hypothesis ( $H_0$ ), the non-parametric test should have a sample size of 100 observations and the parametric one, around 95.

Although sample size is not importantly associated with the analyzed indicators, it is an aspect that should be considered due to its importance for researches. Torres and Cobo (2015) state that nowadays there are no studies on this subject or it is not deeply studied.

The obtained results evidenced the relations among the analyzed statistical indicators, from the high degree of positive correlation among the probability values of type 1 error in the Fisher F test (with or without transformation) and the non-parametric homologue test,

resto, se ubican los indicadores relacionados con el tipo de experimento y diseño experimental.

Se considera que la agrupación de los valores de probabilidad de las dójimas paramétricas y no paramétricas se debe a que la dójima F de Fisher se caracteriza por ser robusta ante la heterogeneidad de varianza, sobre todo cuando se trabaja con igual número de observaciones por tratamientos (Steel y Torrie 1992, Peña 1994), como es el caso de los diseños analizados. Además, se presenta alta potencia conjuntamente con baja probabilidad de cometer error tipo I, manifestada en las altas correlaciones negativas entre las potencias y los valores de probabilidad de error tipo I (con transformación de datos y sin ella).

Las altas correlaciones negativas entre la potencia y la probabilidad de error tipo I, para la dójima no paramétrica, se deben a que esta última con respecto a la paramétrica (bajo el supuesto de distribución normal) presenta alta eficiencia asintótica relativa (ARE), que es de 95.5%. Esto se corresponde con el criterio de potencia-eficiencia planteado por Siegel (1970), Siegel y Castellan (1995), de Calzadilla (1999), quienes sugieren que para lograr similares resultados, en cuanto al rechazo de la hipótesis nula ( $H_0$ ), la dójima no paramétrica debe tener tamaño de muestra de 100 observaciones y la paramétrica, 95 aproximadamente.

Aunque el tamaño de muestra no se asocia de manera importante con los indicadores analizados, es un aspecto que se debe considerar por su importancia para las investigaciones. Torres y Cobo (2015) refieren que en la actualidad no se dispone de trabajos sobre el tema, y no lo tratan con profundidad.

Los resultados obtenidos evidenciaron las relaciones que existen entre los indicadores estadísticos analizados, a partir del alto grado de correlación positiva entre los

as well as the high negative correlations of them with power values (with and without data transformation). This supposes the use of the categorical principal component analysis for an integral evaluation of statistical indicators of models of parametric and non-parametric analysis of variance. It is necessary to continue the study for different data distributions and sample sizes.

valores de probabilidad de error tipo I en la d6cima F de Fisher (con transformaci6n y sin ella) y la d6cima hom6loga no param6trica, as6 como las altas correlaciones negativas de estos con los valores de potencia (con transformaci6n de los datos y sin ella). Esto presupone el empleo del an6lisis de componentes principales categoricos para la valoraci6n integral de los indicadores estadisticos de los modelos de an6lisis de varianza param6trico y no param6trico. Se considera oportuno continuar el estudio para diferentes distribuciones de datos y tamaos de muestras.

## References

- De Calzadilla, J. 1999. Procedimientos de la Estadística no paramétrica. Aplicaciones en las Ciencias Agropecuarias. M.Sc. Thesis, Instituto Nacional de Ciencias Agrícolas, Cuba.
- De Calzadilla, J., Guerra, W. & Torres, V. 2002. "Use and misuse of mathematical transformations. Applications in models of analysis of variance". Cuban Journal of Agricultural Science, 36 (2): 101–104, ISSN: 2079-3480.
- Di Rienzo, J. A., Casanoves, F., Balzarini, M. G., González, L., Tablada, M. & Robledo, C. W. 2012. InfoStat. version 2012, [Windows], Universidad Nacional de Córdoba, Argentina: Grupo InfoStat, Available: <<http://www.infostat.com.ar/>>.
- Gómez, V. M. Á. 2005. Inferencia estadística. Díaz de Santos, 544 p., ISBN: 978-84-7978-687-8, Available: <[https://books.google.com/cu/books?hl=es&lr=&id=MlgqjIRE\\_MUC&oi=fnd&pg=PA207&dq=G%C3%B3mez+2005+Inferencia+estad%C3%ADstica&ots=Vz0rDN8OC2&sig=8y\\_gzxQFQklmbfOtRTxsr6RBuVU&redir\\_esc=y#v=onepage&q=G%C3%B3mez%202005%20Inferencia%20estad%C3%ADstica&f=false](https://books.google.com/cu/books?hl=es&lr=&id=MlgqjIRE_MUC&oi=fnd&pg=PA207&dq=G%C3%B3mez+2005+Inferencia+estad%C3%ADstica&ots=Vz0rDN8OC2&sig=8y_gzxQFQklmbfOtRTxsr6RBuVU&redir_esc=y#v=onepage&q=G%C3%B3mez%202005%20Inferencia%20estad%C3%ADstica&f=false)>, [Consulted: April 19, 2016].
- Hair, J., Anderson, R., Tatham, R. & Llack, W. 1999. Análisis multivariante de datos. Prentice E. & Cano D. (trans.), Edición: 5 ed., Madrid: PRENTICE HALL, 832 p., ISBN: 978-84-8322-035-1, Available: <<https://www.amazon.es/An%C3%A1lisis-multivariante-datos-Joseph-Hair/dp/8483220350>>, [Consulted: April 19, 2016].
- IBM Corporation 2010. IBM SPSS Statistics. version 19, [Windows], Multiplataforma, U.S: IBM Corporation, Available: <<http://www.ibm.com>>.
- Levene, H. 1960. "Robust tests for the equality of variance". In: Olkin I., Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Stanford University Press, pp. 278–292, ISBN: 978-0-8047-0596-7, Available: <[https://books.google.com/cu/books?hl=es&lr=&id=ZUSsAAAAIAAJ&oi=fnd&pg=PA3&dq=Contributions+to+Probability+and+Statistics&ots=GchNfCzOVQ&sig=-W3BeBmuWhTzGRCKktOdQjq0Di4&redir\\_esc=y#v=onepage&q=Contributions%20to%20Probability%20and%20Statistics&f=false](https://books.google.com/cu/books?hl=es&lr=&id=ZUSsAAAAIAAJ&oi=fnd&pg=PA3&dq=Contributions+to+Probability+and+Statistics&ots=GchNfCzOVQ&sig=-W3BeBmuWhTzGRCKktOdQjq0Di4&redir_esc=y#v=onepage&q=Contributions%20to%20Probability%20and%20Statistics&f=false)>, [Consulted: April 19, 2016].
- Molina, M. Ó. & de los Monteros, P. E. 2010. "Rotación en análisis de componentes principales categorico: un caso práctico". Metodología de Encuestas, 12 (1): 63–88, ISSN: 15757803.
- Navarro, C. J. M., Casas, C. G. M. & González, R. E. 2010. "Análisis de Componentes Principales y Análisis de Regresión para Datos Categóricos. Aplicación en la Hipertensión Arterial". Revista de Matemática: Teoría y Aplicaciones, 17 (2): 199–230, ISSN: 2215-3373, DOI: 10.15517/rmta.v17i2.2128.
- Peña, S. de R. 1994. Estadística: Modelos y Métodos. 2nd ed., Madrid: Alianza Universal Textos, 745 p.
- Shapiro, S. S. & Wilk, M. B. 1965. "An Analysis of Variance Test for Normality (Complete Samples)". Biometrika, 52 (3/4): 591–611, ISSN: 0006-3444, DOI: 10.2307/2333709.
- Siegel, S. 1970. Diseño experimental no paramétrico aplicado a las ciencias de la conducta. Available: <<http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=bac.xis&method=post&formato=2&cantidad=1&expresion=mfn=032983>>, [Consulted: April 19, 2016].
- Siegel, S. & Castellan, N. 1995. Estadística no paramétrica aplicada a las ciencias de la conducta. 4th ed., México: Trillas, 432 p., ISBN: 968-24-5101-9, Available: <<http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=CAMZA.xis&method=post&formato=2&cantidad=1&expresion=mfn=014608>>, [Consulted: April 19, 2016].
- StatSoft 2007. STATISTICA (data analysis software system). version 8.0, [Windows], US: StatSoft, Inc., Available: <<http://www.statsoft.com>>.
- Steel, R. G. & Torrie, I. H. 1992. Bioestadística: principios y procedimientos. México: McGraw-Hill Interamericana, 228 p.
- Torres, V. & Cobo, R. 2015. "Applied Mathematics in researches from the Instituto de Ciencia Animal. Fifty years of experience". Cuban Journal of Agricultural Science, 49 (2): 117–125, ISSN: 2079-3480.
- Vásquez, R. E. 2011. Contribución al tratamiento estadístico de datos con distribución Binomial en el Modelo de Análisis de Varianza. Ph.D. Thesis, Instituto Nacional de Ciencias Agrícolas, Cuba.
- Vázquez, Y. 2012. Modelación Estadístico-Matemática con variables mixtas para el estudio de la sostenibilidad social en una empresa ganadera bovina. Ph.D. Thesis, Instituto Nacional de Ciencias Agrícolas, Cuba.

