



MODELOS LINEALES GENERALES Y MIXTOS EN LA CARACTERIZACIÓN DE LA VARIABLE CALIFICACIÓN, INGENIERÍA AGROINDUSTRIAL, UNI-NORTE

GENERAL AND MIXED LINEAR MODELS IN THE CHARACTERIZATION OF THE QUALIFICATION VARIABLE, AGROINDUSTRIAL ENGINEERING, UNI-NORTH

L. M. Dicovski Riobóo¹, M. E. Pedroza Pacheco²

¹ Universidad Nacional de Ingeniería, Sede Regional Norte, Estelí, Nicaragua

² Universidad Nacional Autónoma de Nicaragua, Managua, Nicaragua.

* luis.dicovski@norte.uni.edu.ni

(recibido/received: 12-Septiembre-2017; aceptado/accepted: 27-Noviembre-2017)

RESUMEN

Los modelos mixtos son una propuesta de modelación estadística avanzada, que permiten mejorar la calidad del análisis de los factores fijos y factores aleatorios, modelando la variabilidad aleatoria y la correlación de los errores, siendo muy útiles en el análisis de datos desbalanceado, datos con pseudoréplicas, o datos con algún tipo de estructura jerárquica o de agrupación. En esta investigación, con el software InfoStat se realizó una aplicación de los modelos generales y mixtos, sobre la variable “Calificación de los estudiantes Aprobados”, de Ingeniería Agroindustrial, con datos de 11 años. Se respondió preguntas estadísticas sobre el comportamiento del rendimiento académico, en función del Año académico y del Género de los estudiantes. Al modelar el Año Académico, el Género y la interacción como factores fijos, en relación con la nota de los estudiantes aprobados, se determinó que la inclusión de los factores aleatorios: Estudiante y Docente, mejoraron los: AIC, BIC del modelo, la normalidad y homocedasticidad de los residuos, mejorándose así la calidad del análisis estadístico. El modelo con factores mixtos, encontró diferencias significativas en las notas de los estudiantes aprobados por los factores: “Año Académico” y “Género”. Sin embargo, no se determinó efecto de interacción entre ambos factores.

Palabras claves: Modelos mixtos, AIC, BIC, calificaciones.

ABSTRACT

The mixed models are a proposal of advanced statistical modeling, that allow to improve the quality of the analysis of the fixed factors and random factors, modeling the random variability and the correlation of the errors, being very useful in the analysis of unbalanced data, data with pseudo replica, or data with some kind of hierarchical structure or grouping. In this research, with the InfoStat software, an application of the general and mixed models was carried out, on the variable "qualification of the Approved Students", of Agroindustrial Engineering, with data of 11 years. Statistical questions about academic performance behavior were answered, based on Year Academic of the students' and the Gender. In modeling the Academic Year, Gender and interaction as fixed factors, in relation to the grade of the approved students, it was determined that the inclusion of random factors: Student and Teacher, had improved: AIC, BIC model, normality and homoscedasticity of the residues, thus to get better the

quality of the statistical analysis. The model with mixed factors, found significant differences in the qualifications of the students approved by the factors: "Academic Year" and "Gender". However, no interaction effect between the two factors was determined.

Keywords: Mixed models, AIC, BIC, qualifications.

1. INTRODUCCIÓN

“La humanidad asiste al alba de una nueva época” (Souza Silva, 1999). Se vive la “Sociedad del Conocimiento, la Informatización y los Aprendizajes”, que define un nuevo contexto de ruptura de paradigmas, de nuevos y dinámicos aprendizajes a nivel mundial, regional, nacional y local, contexto en el que ocurren una serie de procesos de transformación socio-económica, educativa, tecnológica, etc. El Modelo I+D+i, caracteriza esta nueva era por el predominio de siete mega tendencias, dos de ellas se destacan en este artículo: *a) cambio de paradigmas* y *b) bioestadística*, siendo ambas de carácter transversal holísticas, sistémicas y antropocéntricas (Pedroza, 2015).

En particular en la aplicación de la bioestadística, en los últimos veinte años, se ha observado en cambio de la escala de los experimentos y los muestreos que generan datos. Los métodos estadísticos clásicos no permiten abordar la problemática compleja resultante de este nuevo escenario. Esto plantea un *cambio en el paradigma del análisis de datos*, que ha demandado el desarrollo de modelos estadísticos más flexibles, para garantizar un mejor control estadístico en la calidad de los resultados. *La bioestadística moderna* permite conocer en profundidad nuevos modelos conocidos como Modelos Mixtos para el análisis de datos, que facilitan modelar la diversidad de organización de las unidades experimentales y la estructura de los tratamientos, permitiendo aumentar la precisión de las estimaciones, ampliar el espacio de inferencia y favorecer la comprensión de la estructura de los datos, aumentando una mayor validez y confiabilidad de los resultados (Casanoves, 2016).

Para interpretar el rendimiento académico, se debe considerar que este es un fenómeno multicausal, que involucra múltiples factores, espacios y tiempos, que intervienen en el proceso de enseñanza-aprendizaje. Entre estas causas coexisten determinantes personales, determinantes sociales y determinantes institucionales (Garbanzo Vargas, 2007). En los centros universitarios se acumulan datos académicos poco aprovechados, y ahí hay un campo muy útil para los profesionales de la educación donde se podrán hacer análisis interesantes para la propia institución y para otros profesores (Morales Vallejos, 2002). Esta información puede dar pautas para comprender mejor el comportamiento de los estudiantes en su rendimiento académico.

Dentro de los datos académicos, están los resultados de las evaluaciones. Estas tienen un contenido social, ya que permiten, dentro de unos límites, saber que quienes van a ejercer la profesión dominan los conocimientos, tienen las habilidades y disponen de las actitudes que permiten asegurar que el ejercicio de la práctica profesional será bien realizado (Santos Guerra, 2014). Garbanzo Vargas vincula los conceptos de rendimiento académico y calidad educativa cuando dice que, “el rendimiento académico del estudiantado universitario constituye un factor imprescindible en el abordaje del tema de la calidad de la educación superior, debido a que es un indicador que permite una aproximación a la realidad educativa” (Garbanzo Vargas, 2007).

Si bien el rendimiento del estudiante universitario debería ser explicado a partir de sus procesos de evaluación; la simple medición y/o evaluación de los rendimientos alcanzados por los alumnos no provee por sí misma todas las pautas necesarias para la acción destinada al mejoramiento de la calidad educativa. Sin embargo, unas de las variables más empleadas o consideradas por los docentes e investigadores para aproximarse al rendimiento académico son: las calificaciones. (Navarro, 2003). Esto conlleva a afirmar que una de las dimensiones más importantes en el proceso de enseñanza aprendizaje la constituye el “rendimiento académico del alumno”. El rendimiento académico de un estudiante

universitario es la resultante de una multiplicidad de factores, que van desde los personales, hasta los dependientes de la institución y sus docentes (Ruiz *et al.*, 2010).

Sin embargo, las calificaciones en diferentes escalas y en forma de puntaje, es lo que generalmente se registra y sirve para que un estudiante apruebe las asignaturas de su carrera. También la intensidad de su valor da una aproximación de lo bueno que ha sido el estudiante en aprobar su pensum académico. Finalmente, un buen puntaje de promedio general de carrera abre las puertas a becas de posgrado o mejores trabajos.

En lo que se refiere a la forma de calificar en la Universidad Nacional de Ingeniería, UNI, el Reglamento del Régimen Académico, en su Título 3, Evaluación del Aprendizaje, Arto 25, define que la escala de calificación de las asignaturas estará comprendida en el rango de cero a cien. También facilita una escala cualitativa expresada de la siguiente manera: 0 – 59 Reprobado, 60 – 69 Regular, 70 – 79 Bueno, 80 – 89 Muy Bueno, 90 – 100 Excelente (UNI, 2006). Si un estudiante termina de cursar la asignatura con un promedio final menor a 60 puntos, este tiene la posibilidad de presentarse a dos exámenes finales, llamados de Convocatoria, sino no se presenta a estos exámenes, tendrá según el reglamento, una nota final de 0 puntos.

Por otra parte, el uso masivo de la computación permitió que surgiera un proceso de investigación nuevo, la **Minería de Datos**. La cual se basa en el uso de algoritmos computacionales que permiten extraer nuevos conocimientos, de grandes bases de datos que surgen de la acumulación de información que se generan de las actividades cotidianas de las organizaciones. Este conocimiento permite entre otros, conocer anomalías no esperadas y tomar decisiones sobre nuevas situaciones generadas (Martínez Luna, 2011).

La minería o exploración de datos, conocida también como “**data mining**”, se puede definir como: un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al analizar grandes volúmenes de datos. Este proceso se utiliza hoy en diferentes campos de la ciencia, incluidos aplicaciones financieras, análisis de mercados y comercio, seguros, educación, etc. (Perez Lopez y Santín González, 2007). Cuando se analizan bases de datos académicas, se pueden utilizar herramientas estadísticas propias de la minería de datos. La minería de datos educativos, es una disciplina emergente que se centra en la aplicación de datos, herramientas y técnicas de minería a datos relacionados con la educación. Esto permite desarrollar modelos que mejoren las experiencias de aprendizaje y la eficacia institucional (Huebner, 2013).

La Sociedad Internacional de Minería de Datos Educativos, define la minería de datos educativos como "una disciplina emergente, relacionada con el desarrollo de métodos para explorar los tipos únicos de datos que vienen de los entornos educativos, y el uso de esos métodos para comprender mejor a los estudiantes y los entornos en los que aprenden". Dentro de las categorías que se usa en la minería de datos están, **Predicción:** por Clasificación, Regresión, Estimación de la densidad y Clustering o Agrupamientos. **Minería de Relaciones:** por Reglas de Asociación, Correlación, Extracción secuencial de patrones y Extracción de Datos Causales; **Destilación de Datos:** por el juicio humano; y **Descubrimientos con Modelos** (Baker y Yacef, 2009).

Cómo antecedente del uso de la minería de datos para obtener información de conductas ocurridas con los estudiantes de la UNI, Universidad Nacional de Ingeniería, Sede Regional Norte, se pueden citar la experiencia de Montenegro *et al* (2011), cuando utilizaron el análisis estadístico Determinante para clasificar potenciales estudiantes exitosos y desertores, así como la publicación de Dicovskiy (2013), quien determinó conductas diferenciadas según el tipo de formación de los docentes.

La forma más sencilla de análisis, es con el uso de la Estadística descriptiva a partir de datos ya obtenidos. Esto permite resumir y sintetizar información y volverla manejable para que se pueda interpretar y extraer conclusiones que, de otra manera, serían ininteligibles. Y una vez que se dispone de una síntesis de la información que hemos recogido de un conjunto de datos, también puede interesar

el preguntarnos si eso que se observa, vale también para otros grupos de datos, a los que se han observado, entonces se hace uso de herramientas de estadística inferencial (Bologna, 2013).

En un análisis estadístico inferencial, los modelos mixtos se pueden usar cuando los datos tienen algún tipo de estructura jerárquica o de agrupación, estos modelos mixtos permiten tener factores fijos, aquellos cuyos niveles son de interés para el experimentador y factores aleatorios, aquellos cuyos niveles son solo una realización de todos los posibles niveles procedentes de una población, y varios términos de error. En el lenguaje de estadística R hay dos paquetes que permiten ajustar modelos mixtos. El primero que apareció fue el paquete nlme, escrito inicialmente por José Pinheiro y Douglas Bates y al que luego se han sumado otros autores (Cayuela, 2010, pág. 110). El software InfoStat implementa una interfase amigable de la plataforma R para la estimación de modelos lineales generales y mixtos a través de los procedimientos gls y lme de la librería nlme (Di Rienzo *et al.*, 2011)

Los Criterios de Información de Akaike (AIC) y los Criterios Bayesianos de Información (BIC) proporcionan un juicio para la selección de modelos ajustados a un conjunto de datos (Di Rienzo, Macchiavelli, & Casanoves, 2011, pág. 3). Los Criterios de Información se utilizan cada vez más para abordar la selección de modelos. Sin embargo, con muestras pequeñas, y con factores aleatorios, en general el AIC supera al BIC, al seleccionarse el mejor modelo (Graft Acquah, 2010, pág. 4). Para interpretar estos indicadores, un valor de AIC y BIC más bajo, implica que el modelo se ajusta mejor (Gómez *et al.*, 2012).

De la reflexión sobre el rendimiento académico en las carreras de ingenierías y al disponer de una colección de 11 años de notas de estudiantes de la carrera de Ingeniería Agroindustrial en la UNI Norte, se planteó como objetivo general en este artículo *determinar el modelo que mejor explique la respuesta al efecto producido en las calificaciones por: el año académico, de I a V, el Género del estudiante, varón mujer y la interacción de ambos factores, en la carrera de ingeniería agroindustrial de la UNI Norte, período 2005-2015.*

2. METODOLOGÍA

Con enfoque mixto de investigación y aplicando estrategias de Minería de datos, se hizo aplicación de técnicas estadísticas descriptivas y modelación estadística avanzada para responder a las preguntas de investigación: *Efecto del año Académico, del Género y la interacción de ambos factores sobre las notas académicas, mediante la aplicación de modelos generales y mixtos.* El Sistema de Registro Académico de la Sede UNI Norte de Estelí facilitó de forma anónima, sin identificar nombres, parte de la base de datos histórica, 2005-2015 de registros académicos de la carrera de Ingeniería Agroindustrial, en formato de tabla en el software Excel, de 556 estudiantes. A partir de esta tabla se diseñó una base de datos estadística en el Software InfoStat, la cual se recodificó y se generaron en las nuevas variables de análisis.

De acuerdo a sus características particulares, como en esta investigación, los sujetos de estudio, los estudiantes, no fueron asignados al azar a los grupos de manera predefinida, sino que dichos grupos y sus repeticiones ya estaban prefijados antes del experimento, estos ya están formados de forma independiente, los datos se organizaron de acuerdo a la lógica de un Diseño Cuasiexperimental (Hernandez *et al.*, 2014), con datos desbalanceados.

En este trabajo, primero se caracterizó de forma gráfica y estadística descriptivas las variables numéricas que explican el rendimiento académico. Esto se hizo partiendo de la descomposición de la variable *Calificaciones de asignatura*, nota 0-100, según se utiliza en la UNI, generando tres nuevas variables: *“Deserción en Aula”*, nota final de 0”, *“Calificación sin deserción en aula”* y *“Calificación de los estudiantes Aprobados”*. La variable *“Deserción”* dio información sobre los alumnos que inscribieron las clases, pero luego desertaron”. La variable *“Calificación sin deserción en aula”* explicó el comportamiento de los estudiantes que participaron con trabajos en aula y evaluaciones en exámenes hasta finalizar la asignatura. La variable *“Calificación de los estudiantes Aprobados”* explicó el comportamiento de los estudiantes que aprobaron la asignatura, nota final mayor a 60.

Para la variable “Calificación de 0 a 100”, se analizaron 17,734 datos; con la variable “Calificación sin deserción en aula” se tuvieron 15,645 datos y con la variable “Calificación de Aprobados” se tuvo 14,747 datos. Para caracterizar estas variables se construyeron Histogramas, con su curva normal teórica generada a partir del promedio y de la variancia de los datos empíricos, también se hicieron gráficos de cuantiles tipo qq-plot, que se utilizaron para evaluar el grado de ajuste de las observaciones con una distribución teórica del tipo Normal. Sobre estos gráficos se puede consultar en el Manual del Usuario de InfoStat (Di Rienzo, *et al.*, 2008). Los gráficos Box plot sirvieron para verificar visualmente si las varianzas de los diferentes grupos eran constantes (Correa y Salazar, 2016).

Para la estimación de modelos lineales generales y mixtos a través de los procedimientos `gls` y `lme` de la librería `nlme` escritos en R, se utilizó como interfase el software InfoStat. (Di Rienzo *et al.*, 2011). Se evaluó el efecto de los **Factores Fijos: “Año Académico” y “Género” y sus interacciones, con la variable “Calificación de Aprobados”, nota final ≥ 60 puntos**. Se modeló este factor de forma secuencial con la inclusión de los **Factores aleatorios: Estudiante y Docente**, para reducir su efecto directo y el de pseudo réplica, que se pueden generar sobre los Factores Fijos de estudio. Se consideró el mejor modelo, a aquel que tuviera los valores de AIC y BIC más bajos.

Si “ Ω ” fue el efecto del año académico en la calificación, con $j=1\dots 5$ años académicos y “ β ” es el efecto del género con $k = 1, 2$ valores. Las tres hipótesis nulas a verificar fueron:

- $H_0: \mu_I = \mu_{II} = \mu_{III} = \mu_{IV} = \mu_V$, igualdad de notas entre los años académicos;
- $H_0: \mu_{varón} = \mu_{Mujer}$, igualdad de notas entre varón y mujer;
- $H_0: (\Omega \beta) = 0$, no existe la interacción entre Año Académico y Género.

El software InfoStat 2016 versión profesional, fue utilizado para hacer el análisis estadístico de los datos. Fueron utilizadas sus rutinas: Medidas Resumen, gráficos qq-plot, box-plot, Modelos generales lineales y mixtos (Di Rienzo *et al.*, 2015).

3. RESULTADOS Y DISCUSIÓN

Caracterización de las variables: Calificación de 0 a 100”, “Calificación sin deserción en aula” y “Calificación de Aprobados”.

De las 17,734 calificaciones totales, $0 \leq \text{nota} \leq 100$, el 12% correspondieron a desertores en aula, nota final igual a 0. Por otro lado, los no desertores y reprobados, $0 < \text{nota} < 60$, fueron apenas el 5% de todos los datos, quedando que el 83% de las calificaciones correspondían a notas de aprobado, $\text{nota} \geq 60$ puntos. En la tabla 1, se presentan las medidas resumen.

Tabla 1. Medidas Resumen de las variables en estudio

Variable	n	Media	D.E.	Mín.	Máx.
Calificación de 0 a 100	17,734	64.59	27.88	0	100
Calificación sin deserción en aula	15,645	73.22	15.80	1	100
Calificación de Aprobados	14,747	75.88	11.27	60	100

Como era de esperar, en las tres variables estudiadas, los promedios fueron mayores y los desvíos estándar, D.E., menores, a medida que se fue reduciendo la amplitud de la variable.

Los histogramas, los gráficos QQ Plot con la correlación, valor “r”, entre los valores observados y el ajusta a una distribución normal estándar, de las variables: “Calificación de 0 a 100”, “Calificación sin deserción en aula” y “Calificación de Aprobados”, se pueden observar en las figuras 1, 2 y 3. En estas figuras se resalta que los valores de “r”, mejoraron de la figura 1 a la figura 2, de 0.89 a 0.95, y de la figura 2 a la figura 3, de 0.95 a 0.98. Lo que mostró que, en la medida que se redujo el espacio de la variable Calificación, de 0-100 a 60-100, los datos se aproximaron mejor a una distribución normal.

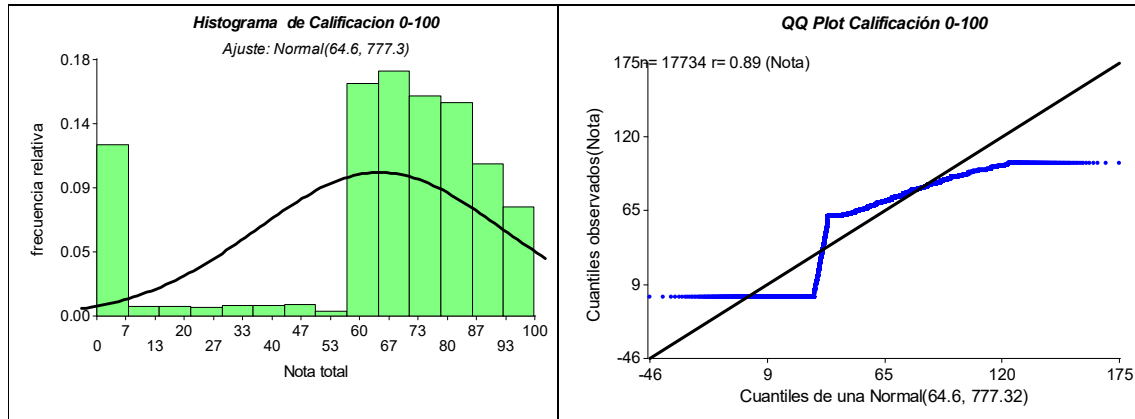


Figura 1. Histograma y QQ plot de la “Variable Calificación de 0 a 100” o Nota completa.

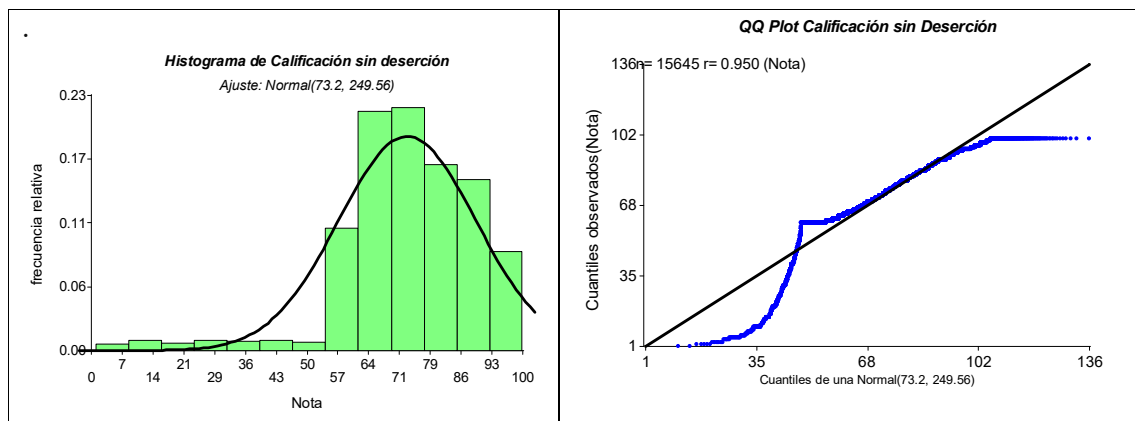


Figura 2. Histograma y QQ plot de la Variable “Calificación sin deserción en aula”

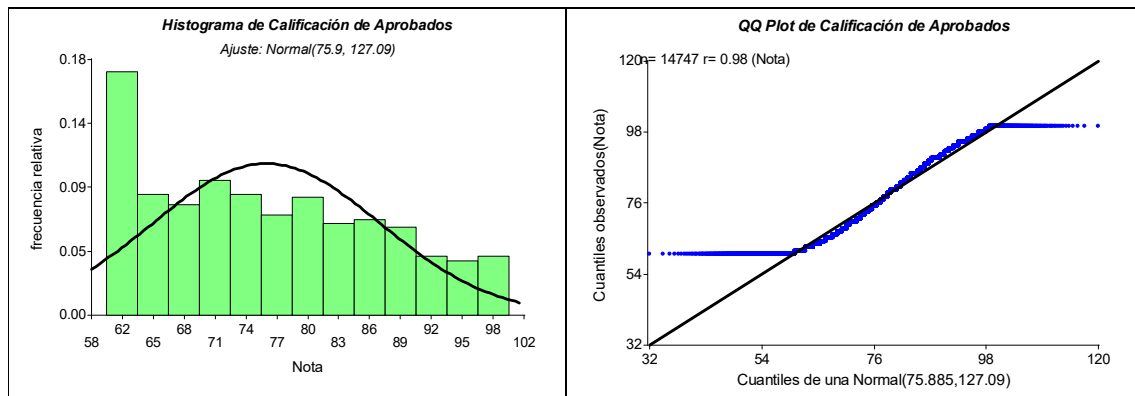


Figura 3. Histograma y QQ plot de la Variable “Calificación de los estudiantes Aprobados”, nota \geq a 60 puntos

Modelación de los efectos fijos “Año Académico” y “Género” con la variable “Calificación de los estudiantes Aprobados”, nota final \geq a 60 puntos

Se modeló para responder a la pregunta ¿el Año Académico, y/o el Género y/o la interacción de los dos factores, influyen en la nota de los estudiantes aprobados? En el Modelo 1: no se evaluó el efecto de Factores Aleatorios, lo cual fue equivalente a un análisis de variancia tradicional, con dos factores fijos y su interacción. En el Modelo 2: se incorporó como factor aleatorio a los Estudiantes. En el Modelo 3: se incorporó a los Estudiantes y Docentes como factores aleatorios.

A continuación, se muestran la estructura de los tres modelos evaluados:

Modelo 1: $Y_{ijkl} = \mu + \Omega_i + \beta_j + (\Omega * \beta)_{ij} + \varepsilon_{ijkl}$;

Modelo 2: $Y_{ijkl} = \mu + \Omega_i + \beta_j + (\Omega * \beta)_{ij} + \lambda_k + \varepsilon_{ijkl}$;

Modelo 3: $Y_{ijkl} = \mu + \Omega_i + \beta_j + (\Omega * \beta)_{ij} + \lambda_k + \gamma_l + \varepsilon_{ijkl}$; donde:

Y_{ijkl} : la calificación de los estudiantes, con $j=1\dots 5$ años académicos.

Ω_i : El efecto del año académico en la calificación de los estudiantes.

β_j : El efecto del género con $k = 1, 2$ valores.

$(\Omega * \beta)_{ij}$: Efecto de Interacción año académico*género.

λ_k : Es el efecto Aleatorio del estudiante ó sujeto o que se evalúa.

γ_l : Es el efecto Aleatorio del docente.

ε_{ijkl} : Efecto aleatorio de variación.

Se debe considerar que en general el factor sujeto o cualquier otro factor que represente una unidad experimental, en los modelos mixtos se debe analizar como una muestra aleatoria de la población de interés (Bates, 2005), que fue lo que se hizo en los Modelos 2 y 3 (Tabla 2).

Tabla 2. Comparación de modelos de análisis con diferentes factores aleatorios

Modelo	Factores Fijos	Factores Aleatorios	AIC	BIC	p-valor
1	Año Académico, Género y su Interacción		112,942.31	113,025.88	
2	Año Académico, Género y su Interacción	Estudiante	108,992.40	109,083.57	<0.0001
3	Año Académico, Género y su Interacción	Estudiante, Docente	108,463.41	108,562.19	<0.0001

Al comparar los tres modelos, en la tabla 2, se observó que los valores menores del AIC, del BIC y de los *p valor* indicaron que: el Modelo 2 explicó mejor los resultados obtenidos que el Modelo 1 y que el **Modelo 3**, explicó mejor los resultados obtenidos que el Modelo 2. La inclusión de ambos factores aleatorios mejoró la modelación, por tanto el mejor modelo es el **Modelo 3**, siendo en esencia un *Análisis de Variancia en DCA, Bifactorial con efecto de Interacción, incluyendo los efectos aleatorios de Estudiante y Docente*. Los resultados concuerdan con lo expresado por Gómez *et al.*, (2012), quienes afirman que la aplicación de los modelos mixtos, en general dan criterios de información necesarios para la selección del mejor modelo.

En las tablas 3 y 4, se observa que cuando se compararon las pruebas de hipótesis del Modelo 1, el referente, con el mejor Modelo, el Modelo 3, la modelación mixta disminuyó de manera notable los grados de libertad, GL, del denominador de la prueba. Quitándose así del modelo el efecto de pseudo réplica por estudiante y docente. En ambas tablas, se observó diferencias significativas entre los factores de efectos fijos “Año académico” y “Género”. Sin embargo, no se demostró interacción entre ambos factores. Estos resultados confirman que, los modelos mixtos son una propuesta de modelación estadística avanzada, que permiten mejorar la calidad del análisis de los factores fijos y factores aleatorios, modelando la variabilidad aleatoria y la correlación de los errores, siendo muy útiles en el análisis de datos desbalanceados, datos con pseudoréplicas, o datos con algún tipo de estructura jerárquica o de agrupación.

Tabla 3. Pruebas de hipótesis marginales del Modelo 1.

Factor	Num GL	Den GL	Valor F	p-valor
Año Académico	4	14737	5.67	0.0001
Género	1	14737	279.81	<0.0001
Interacción	4	14737	0.68	0.6044

Tabla 4. Pruebas de hipótesis marginales del Modelo 3.

Factor	Num GL	Den GL	Valor F	p-valor
Año Académico	4	5980	27.39	<0.0001
Género	1	495	38.42	<0.0001
Interacción	4	5980	0.7	0.5924

En lo que se refiere a la normalidad de los residuos del modelo, en los gráficos QQ Plot la Prueba de Shapiro Francia, aportó las evidencias de una alta correlación positiva de los residuos con respecto a una distribución Normal evidenciados por valores de $r = 0.98$ y $r = 0.99$ del Modelo 1, referente, al Modelo 3, con dos factores mixtos. En ambos modelos se considera que los residuos tuvieron una distribución muy cercana a la distribución normal, demostrándose la normalidad esperada.

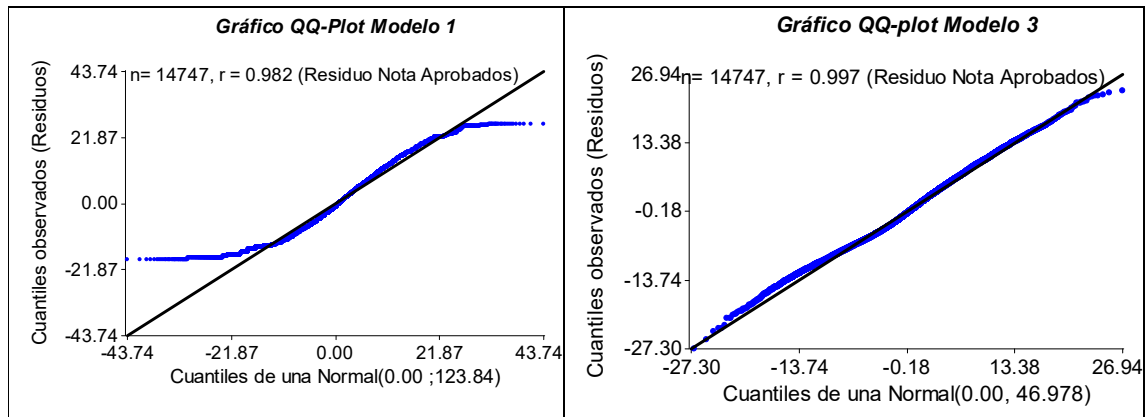


Figura 4. Gráfico QQ Plot de los residuos del Modelo 1 y del Modelo 3

La homocedasticidad de los residuos estandarizados para los dos factores fijos, se observa en los gráficos de Box Plot, presentados en las figuras 5 y 6, en donde las cajas de los diferentes grupos se observan de tamaño semejante, por lo que se evidencia que los tratamientos son comparable entre sí, es decir son homocedásticos. Sin embargo, se resalta que el Modelo 3, en general, tuvo cajas de menor tamaño, lo que fue una muestra de la reducción de las variancias de los errores para los diferentes grupos.

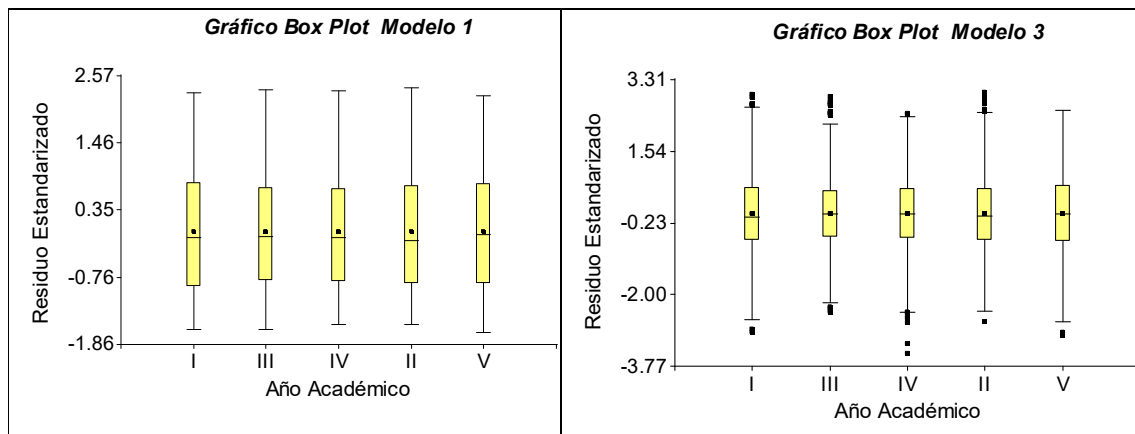


Figura 5. Gráficos Box Plot del Factor “Año Académico”, de los Modelos 1 y 3

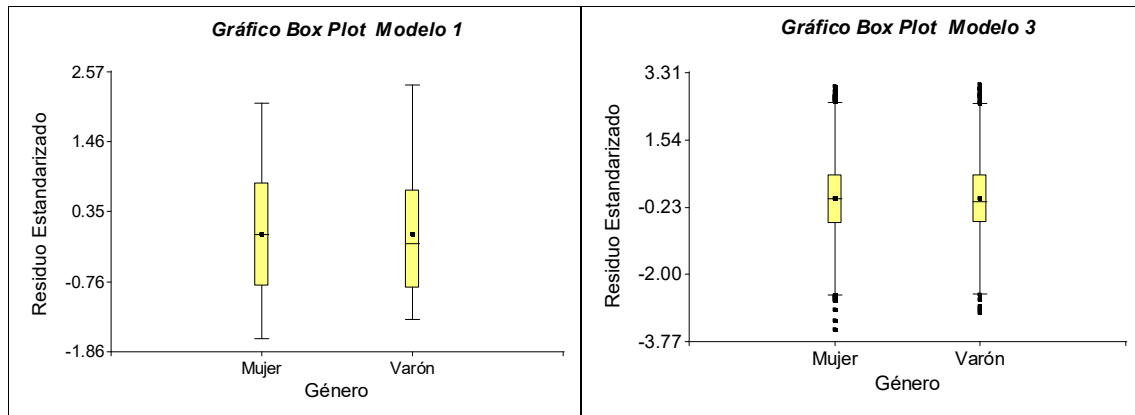


Figura 6. Gráficos Box Plot del Factor “Género”, de los Modelos 1 y 3

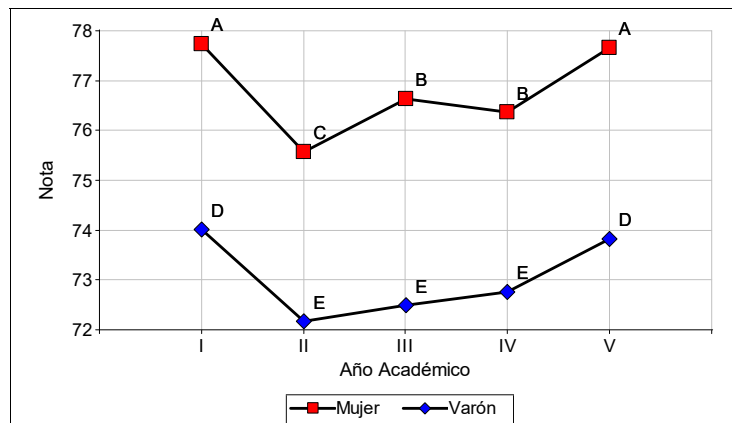
Resultados de las pruebas de hipótesis con el mejor modelo: el Modelo 3, con dos factores aleatorios.

Se realizó la separación de promedios fue aplicada mediante la prueba LSD de Fisher a los resultados obtenidos del mejor modelo, observándose en la tabla 5 que Varones y Mujeres tienen diferencias significativas entre los promedios de notas de aprobados, nota >60, las mujeres tuvieron mejores promedios. Por otra parte, fue aplicada la prueba LSD de Fisher a los resultados obtenidos para comparar el efecto de *Año Académico*, lo que permitió evidenciar que primer y quinto año tiene los mejores promedios y son estadísticamente iguales entre sí. En la figura 7, se observan las notas por año académico y género. Se destaca al observar la nota de los varones y mujeres que, ambos niveles del factor Género (mujer y varón) muestran paralelismo, no se cruzan, lo que explica gráficamente la falta de interacción Género y Año Académico.

Tabla 5. Comparación de promedios de notas para el Año Académico y Género, con el Modelo 3

Comparación de Promedios: factor Género con factores aleatorios: Estudiante y Docente			Comparación de Promedios: factor Año Académico con factores aleatorios: Estudiantes y Docentes		
Género	Nota	Significancia estadística	Año Académico	Nota	Significancia estadística
Mujer	76.8	A	I	75.9	A
Varón	73.0	B	V	75.7	A
			III	74.6	B
			IV	74.6	B
			II	73.9	C

Medias con una letra común no son significativamente diferentes ($p > 0.05$), prueba LSD de Fisher



Medias con una letra común no son significativamente diferentes ($p > 0.05$), prueba LSD de Fisher

Figura 7. Gráfico de Interacciones de Género y Año Académico, del Modelo 3

4. CONCLUSIONES

La variable “Calificación de 0 a 100”, $\text{nota} \geq 0$, se puede estudiar mejor al descomponerla en dos variables “Desertores”, $\text{nota} = 0$ y “Calificación de Aprobados”, $\text{nota} \geq 60$. Esta descomposición permitió crear una variable del tipo binaria: Desertor o no Desertor, la cual puede ser tratada para pruebas de hipótesis como una variable continua en el intervalo 0-1, bajo un Modelo Lineal Generalizado y Mixto. La variable “Calificación de Aprobados”, también se puede analizar como un caso de los Modelos Lineales Generales y Mixtos. La mayoría de los datos analizados correspondieron a estudiantes aprobados, $\text{nota} \geq 60$ puntos. En la medida que se redujo el espacio de la variable Calificación, de 0-100 a 60-100, los datos se aproximaron mejor a una distribución normal, obtenido para los Modelos 2 y 3.

Al modelar el Año Académico, el Género y la interacción como factores fijos, en relación con la nota de los estudiantes aprobados, se determinó qué, la inclusión de los factores aleatorios: Estudiante y Docente, mejoraron los criterios de AIC, BIC del modelo, así como la normalidad y homocedasticidad de residuos de los datos, mejorándose así la calidad del análisis estadístico realizado. Al disponer de una base de datos desbalanceada y con pseudo réplicas, la modelación por modelos mixtos, permitió mejorar el proceso de modelación, que finalmente fue el de un *Análisis de Variancia en DCA, Bifactorial con efecto de Interacción, incluyendo los efectos aleatorios de Estudiante y Docente*.

El mejor modelo, con menores AIC y BIC, fue el Modelo 3, que incluyó los dos factores aleatorios, (estudiantes y docentes) permitió determinar diferencias significativas en las notas de los estudiantes aprobados por los factores: “Año Académico” y “Género”, con $p \text{ valor} = <0.0001$ y $p \text{ valor} = <0.0001$ respectivamente. Sin embargo, no se determinó interacción significativa entre ambos factores con $p \text{ valor} = 0.5924$.

5. REFERENCIAS

- Baker, R. y Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, 1(1), 3-17.
- Bates, D. (2005). Fitting linear mixed models in R. *R News*, 5(1), 27-30.
- Bologna, E. (2013). *Estadística para Psicología y Educación* (Tercera ed.). Córdoba, Argentina: Brujas.
- Casanoves, F. (2016). *Curso de Posgrado de Control Estadístico de la calidad*. Estelí: UNAN-FAREM.
- Cayuela, L. (2010). *Modelos lineales mixtos en R (versión 1.0)*. Granada: EcoLab, Centro Andaluz de Medio Ambiente, Universidad de Granada - Junta de Andalucía.
- Correa, J. y Salazar, J. (2016). *Inroducción a los modelos mixtos*. Medellín: Universidad Nacional de Colombia, Facultad de Ciencias. Escuela de Estadística.
- Di Rienzo, J. A., Macchiavelli, R. E. y Casanoves, F. (2011). *Modelos lineales mixtos : aplicaciones en InfoStat*. Córdoba, Argentina: Grupo Infostat.
- Di Rienzo, J., Casanoves, F., Balzarini, M., Gonzalez, L., Tablada, M. y Robledo, C. (2008). *Manual del Usuario InfoStat*. Córdoba, Argentina: Grupo InfoStat, FCA, Universidad Nacional de Córdoba.
- Di Rienzo, J., Casanoves, F., Balzarini, M., Gonzalez, L., Tablada, M. y Robledo, C. (2015). *InfoStat versión 2015*, Grupo InfoStat, FCA, Universidad Nacional de Córdoba. Córdoba, Argentina.

- Dicoyskiy, L. M. (2013). Evolución Académica de la carrera de Ingeniería Agroindustrial, UNI Sede Regional Norte, en el periodo 2005-2012. *El Higo*, 3(1), 18-20.
- Garbanzo Vargas, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la Educación Superior Pública. *Revista Educación*, 3(1), 43-63.
- Gómez, S., Torres, V., García, Y. y Navarro, J. (2012). Procedimientos estadísticos más utilizados en el análisis de medidas repetidas en el tiempo en el sector agropecuario. *Revista Cubana de Ciencia Agrícola*, 46(1), 1-7.
- Graft Acquah, H. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, 2(1), 1-6. Recuperado el 15 de julio de 2017, de <http://www.academicjournals.org/journal/JDAE/article-full-text-pdf/650D3294276>
- Hernandez, R., Fernández, C. y Baptista, P. (2014). *Metodología de la Investigación* (Sexta ed.). México: MacGraw-Hill.
- Huebner, R. (Abril de 2013). A survey of educational data-mining research. *Research in Higher Education Journal*, 19, 1-13. Recuperado el 5 de Abril de 2017, de <https://www.aabri.com/manuscripts/121328.pdf>
- Martínez Luna, G. L. (Octubre-Diciembre de 2011). La Minería de Datos. Cómo hallar una aguja en un pajar. *Ingenierías*, XIV(53), 53-63.
- Montenegro, J., López, A. y Fonseca, E. (2011). *Automatización de un Sistema Predictivo de Deserción Estudiantil en la UNI Norte, integrado al SIRA*. Estelí: UNI Norte.
- Morales Vallejos, P. (15 de Enero de 2002). La evaluación académica: conceptos y planteamientos básicos. *Cuaderno Monográficos del ICE. Serie Didáctica Número 2*, 52.
- Navarro, R. (2003). El rendimiento académico: concepto, investigación y desarrollo. *REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2), 1-15.
- Pedroza, M. (2015). *El Modelo de gestión de la investigación, modelo I+D+i de la UNAN-Managua*. Managua: UNAN-Managua.
- Perez Lopez, C. y Santín González, D. (2007). *Minería de datos: técnicas y herramientas*. Madrid: Paraninfo.
- Ruiz, G., Ruiz, J. y Ruiz, E. (25 de Abril de 2010). Indicador global de rendimiento. (OEI, Ed.) *Revista Iberoamericana de Educación*(52), 1-11.
- Santos Guerra, M. Á. (2014). *La Evaluación como Aprendizaje. Cuando la flecha impacta la diana* (Segunda ed.). Madrid, España: Narcea, S. A.
- Souza Silva, J. (1999). "El Cambio de Época, el Modo Emergente de Producción de Conocimiento y los Papeles Cambiantes de la Investigación y Extensión en la academia del Siglo XXI". Panamá: IICA. Recuperado el 11 de 09 de 2017, de Grupo Chorlavi: <http://www.grupochorlavi.org/php/doc/documentos/docabr1.html>

UNI. (2006). *Reglamento del Régimen Académico de la Universidad Nacional de Ingeniería*. Managua: UNI.

SEMBLANZA DE LOS AUTORES

Luis María Dicovskiy Riobóo: es graduado de Ingeniero Agrónomo por la Universidad Nacional de Rosario, Argentina, recibió el título de maestro en Estadística e Investigación de Operaciones por la Universidad de Valencia, España, el título de maestro en Investigación Social Cualitativa por la UPOLI, Nicaragua, actualmente es doctorante de la UNAN Managua en “Gestión y Calidad de la Investigación Científica” y docente de la UNI Sede Regional Norte.

Manuel Enrique Pedroza Pacheco: es graduado de Ingeniero Agrónomo Fitotecnista, en la Facultad de Ciencias Agropecuarias de la UNAN, hoy UNA y Doctor en Ciencias Agrícolas, en la Universidad Agraria de Plovdiv, Bulgaria. Es miembro fundador de la Academia de Ciencias de Nicaragua (ACN). Es el creador del Programa de Maestría en “Métodos de Investigación Científica” y el Programa de Doctorado en “Gestión y Calidad de la Investigación, ambos de la UNAN Managua. Actualmente es profesor titular e investigador de UNAN-Managua.