

Un reto de los investigadores es capturar los hechos esenciales de los fenómenos en una formulación Matemática rigurosa que les permita apoyar la toma de decisiones con sólidos argumentos científicos. En la mentalidad de algunos “científicos y editores” persiste cierto nivel de subvaloración en cuanto a la utilización de la Matemática aplicada en las diferentes ciencias.

En este libro, un grupo de personas dedicadas a la investigación operacional (IO) realizan su aporte en el modelado matemático de fenómenos del medio ambiente y la salud, válidas para apoyar la toma de decisiones sustentada con informaciones validadas y analizadas oportunamente.

Un gran valor de este libro es el hecho de estar escrito en idioma Español, lo que demuestra que está dirigido fundamentalmente a Latinoamérica con intereses de divulgación científica y colaboración internacional. Existe gran interés de estos grupos de viabilizar investigaciones conjuntas para resolver problemas del área.



Universidad Autónoma de Guerrero



División Académica de Ciencias Económico
Administrativas de la Universidad Juárez
Autónoma de Tabasco de México .



Facultad de Matemática
Universidad de la Habana

MODELACIÓN MATEMÁTICA DE FENÓMENOS DEL MEDIO AMBIENTE Y LA SALUD

Editores

José Felix García Rodríguez

Carlos N. Bouza Herrera

Dante Covarrubias Melgar



ÍNDICE

MEDIO AMBIENTE

Uso de redes neuronales para caracterizar las estaciones meteorológicas del centro de cuba

GLADYS CASAS-CARDOSO Y CARLOS. PÉREZ RISQUET / 11

Estrategia de muestreo implementable para hacer un inventario de maguey papalote (agave cupreata trel et berger) en el estado de guerrero.

LORENA ALONSO, CARLOS N. BOUZA, DANTECOVARRUBIAS Y LUCIO DÍAZ. / 26

Modelos multivariados de regresión lineal jerárquica: una aplicación en biometría

MARIO MIGUEL OJEDA Y FERNANDO VELASCO / 40

Estudio del método de verosimilitud empírica y su aplicación a datos medioambientales

DIEGO IGLESIAS, MARÍA DEL MAR RUEDA / 54

Estimación y comparación de las razones de verosimilitud de tests diagnósticos binarios, con aplicación al diagnóstico de enfermedades medioambientales

JOSÉ ANTONIO ROLDÁN NOFUENTES, JUAN DE DIOS LUNA DEL CASTILLO, MIGUEL ÁNGEL MONTERO ALONSO / 65

Observabilidad de un sistema electroenergetico: heurística para la ubicación de dispositivos de medición

LISSETTE VALDES VALDES Y SIRA MA. ALLENDE ALONSO / 82

La sostenibilidad de las actividades turísticas, un aporte desde la investigación de operaciones.

VÍCTOR ERNESTO PÉREZ LEÓN, RAFAEL CABALLERO FERNÁNDEZ, MARÍA AMPARO / 101

El modelo delta para evaluar el grado de acuerdo entre dos observadores.

PEDRO FEMIA MARZO Y ANTONIO MARTÍN ANDRÉS, MARÍA ÁLVARIZ HERNÁNDEZ / 124

Valoración económica de bienes y servicios ambientales a partir de técnicas de decisión multicriterio. Estudio de caso: parque nacional viñales.

ALAIN HERNÁNDEZ SANTOYO, MARÍA AMPARO LEÓN SÁNCHEZ, RAFAEL CABALLERO FERNÁNDEZ, VÍCTOR ERNESTO PÉREZ LEÓN Y MAYRA CASAS VILARDELL. / 146

SALUD

Evaluación económica e investigación operacional en salud pública como respaldo a la toma de decisiones gerenciales

JOSÉ FÉLIX GARCÍA RODRÍGUEZ, ANAÍ GARCÍA FARIÑA, ANA MARÍA GÁLVEZ GONZÁLEZ / 160

Evaluación de la clasificación de los pacientes sospechosos de dengue, en el estado de Guerrero, México.

AGUSTÍN SANTIAGO JOSÉ M. SAUTTO MELINA FERNANDO / 160

Evaluación mediante curvas roc de las potencialidades de marcadores precoces de lesión cerebral en pacientes hipertensos neurológicamente asintomáticos.

CARLOS BOUZA HERRERA, TERESA LA ROSA CHIBÁS, PEDRO ANTONIO VALDÉS SOSA, GERTRUDIS DE LOS ÁNGELES HERNÁNDEZ GONZÁLEZ / 185

Modelo para la transmisión y control de la influenza ah1n1

BEATRIZ E. VILLANUEVA M., ANIBAL MUÑOZ L., IRMA PÉREZ C., JOSÉ BETANCOURT B., JOSÉ F. GARCÍA R., OMAR A. ARCE S., LEONARDO D. RESTREPO A. / 214

Análisis de riesgos proporcionales aplicado a la supervivencia de los casos de Sida notificados al registro andaluz de Sida

ANA MARÍA LARA-PORRAS, YOLANDA ROMÁN-MONTOYA Y AURORA BUENO-CAVANILLAS / 233

MODELOS MULTIVARIADOS DE REGRESIÓN LINEAL JERÁRQUICA: UNA APLICACIÓN EN BIOMETRÍA

Mario Miguel Ojeda

Fernando Velasco,

Facultad de Estadística e Informática, Universidad Veracruzana.

ABSTRACT

In this work it is presented the multivariate version of the hierarchical linear regression models with a brief description of the theoretical results and methodological strategies, including comments on computational aspects and on the steps for analyzing specific data; it is included an illustrative biometric application, which examines the trend and variability in height and diameter growth among families of conifers.

RESUMEN

En este trabajo se presenta la versión multivariada de los modelos de regresión lineal jerárquica con una descripción sucinta de los resultados teóricos y las estrategias metodológicas, incluyendo comentarios sobre aspectos computacionales y lo relativo a los pasos para analizar datos concretos; se incluye una aplicación ilustrativa en el área de la biometría, en la cual se estudia la tendencia y la variabilidad del crecimiento en altura y diámetro entre familias de plantas coníferas.

1 INTRODUCCIÓN

En la década de los 80's surgieron los modelos de regresión jerárquica como propuesta para modelar datos que presentan estructura de anidamiento o estructura jerárquica. Por ejemplo, los estudiantes se encuentran agrupados en salones, los salones están anidados en escuelas, las cuales a su vez están dentro de distritos escolares. Los habitantes de una ciudad están agrupados en colonias, las colonias en comunidades, las comunidades en municipios y así sucesivamente. En general se tienen J grupos con n_j unidades en el j -ésimo grupo, $j = 1, 2, \dots, J$. A cada grupo se le denomina unidad de nivel 2; así se tienen J unidades de nivel 2, y a cada una de las n_j unidades en cada grupo se les denomina unidad de nivel 1; con lo que se tienen n_j unidades de nivel 1 en la j -ésima unidad de nivel 2. El número n_j de unidades de nivel 1 no tienen que ser necesariamente igual en cada unidad de nivel 2. En general en un sistema con estructura jerárquica se pueden presentar varios niveles. Los datos obtenidos de poblaciones con estructuras jerárquicas no se pueden considerar como una muestra aleatoria simple. Existe una correlación entre las unidades dentro de grupo que no se debe ignorar. Los modelos de regresión

jerárquica consideran esta correlación en el proceso de modelaje estadístico y son por lo tanto, la herramienta de análisis estadístico cuando el investigador desea modelar con regresión alguna relación en datos agrupados jerárquicamente.

Los modelos de regresión lineal jerárquica han tenido un notable desarrollo en disciplinas como la sociología, la investigación educativa, la investigación social, la psicología, la economía, la epidemiología, la biología, y la ecología por mencionar algunas, por lo que podemos decir que forman parte nodal del desarrollo de las aplicaciones biométricas. Los modelos lineales jerárquicos tienen una gran historia, pero han recibido especial atención en las últimas décadas y sus áreas de aplicación se han multiplicado considerablemente (Goldstein, 1997; Raudenbush y Bryk, 2002; Arnau y Balluerka, 2004). Recientes desarrollos en cómputo han hecho que se incremente la atención en el uso de estos modelos en el análisis de datos con estructura jerárquica. En la actualidad existe software estadístico el cual permite analizar datos con estructura jerárquica: MLwiN (Rasbash *et al.*, 2009); SAS (Cody, 2001; Little, 2002); y S-PLUS (Pinheiro y Bates, 2000). Los modelos lineales jerárquicos también se denominan modelos multinivel (Snijders y Bosher, 1999; Gelman y Hill, 2007), modelos de componentes de varianza (Searle *et al.*, 2006), modelos de coeficientes aleatorios (Longford, 1993), y modelos de efectos mixtos (Laird y Ware, 1982; Mcculloch y Searle, 2001; West *et al.*, 2007).

La idea fundamental de estos modelos es considerar la estructura jerárquica de la muestra y/o la población de referencia en la postulación del modelo. En este sentido, los modelos lineales jerárquicos son más realistas que los modelos lineales convencionales, pero además los incluyen como casos particulares. Es así que los modelos jerárquicos permiten explicar la variabilidad a nivel de grupos y nivel de unidades anidadas dentro de grupos; y si la variabilidad entre grupos no es significativa, entonces el modelo a usar para explicar el fenómeno sería un modelo de regresión lineal. Cuando se tiene datos con estructura jerárquica de dos niveles lo más razonable es razonable postular un modelo que considere una posible diferencia entre las unidades de nivel 2; es decir, plantear un modelo tal que, para cada unidad de nivel 2, se tengan diferentes coeficientes; de allí el nombre de modelos lineales de coeficientes aleatorios. En los modelos lineales jerárquicos cada uno de los niveles de la estructura jerárquica es representado formalmente con su propio submodelo, los cuales se anidan en un modelo general; de allí lo de modelos jerárquicos (ver Figura 1).

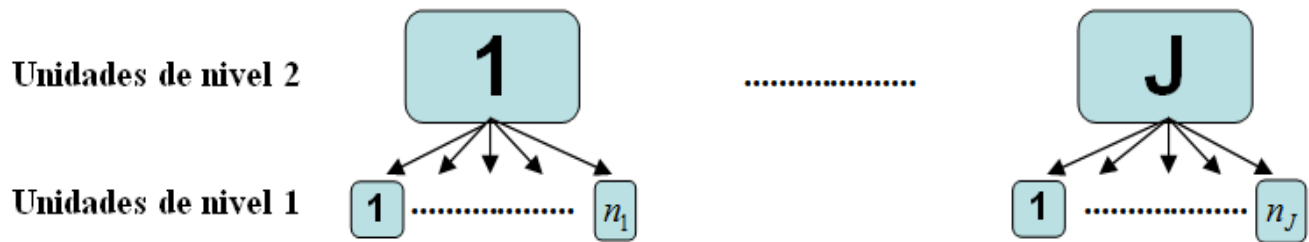


Figura 1. Diagrama de estructura jerárquica de dos niveles.

Por otro lado, las técnicas de análisis multivariado nos permiten lograr una comprensión del fenómeno bajo estudio ya que éste se estudia considerando a las variables respuesta que intervienen de manera simultánea. En la modelación estadística multivariada, la regresión lineal multivariada es la extensión natural de los modelos de regresión lineal. Sin embargo, los modelos de regresión jerárquica multivariada han tenido poca atención en su generalización al caso multivariado. Las contribuciones han sido relativamente pocas, pero el campo de aplicación es grande. Podemos decir que a pesar del desarrollo teórico y metodológico de la modelación jerárquica, estos modelos aún son pocos conocidos, incluso por especialistas en estadística. En este sentido, este trabajo se motiva por dar a conocer estos modelos y mostrar su aplicación en el área de la biometría. El artículo presenta de manera breve y concisa al modelo multivariado de regresión jerárquica junto con una aplicación en datos reales de curvas de crecimiento.

2. METODOLOGÍA

El modelo de regresión lineal jerárquico para una variable respuesta $Y^{(1)}$ para el nivel 1, con una variable explicatoria a nivel 1 X , tiene la forma

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \quad i = 1, \dots, n_j; \quad j = 1, \dots, J \quad (1)$$

donde β_{0j} y β_{1j} tienen la interpretación usual de un modelo de regresión simple. Estos coeficientes varían de unidad a unidad de nivel 2, y e_{ij} es el efecto aleatorio de nivel 1. La diferencia con el modelo de regresión usual es que cada una de las unidades de nivel 2 está caracterizada por un coeficiente β_{0j}

diferente, y similarmente para β_{1j} . El modelo para el nivel 2, con una variable explicatoria nivel 2, tiene la forma

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + u_{0j} \\ & \text{y} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_j + u_{1j}\end{aligned}\quad (2)$$

donde γ_{00} es el término independiente, γ_{01} es el coeficiente que indica la contribución de la variable explicatoria a nivel de grupo, γ_{10} es el coeficiente que indica la contribución de la variable explicatoria a nivel de individuo, y γ_{11} es el coeficiente que indica la contribución correspondiente a ambas variables. Los errores u_{0j} y u_{1j} son los efectos aleatorios de nivel 2. Los supuestos de los errores a nivel 1 y a nivel 2 son $e_{ij} \sim N(0, \sigma_e^2)$, $u_{0j} \sim N(0, \sigma_{u0}^2)$ y $u_{1j} \sim N(0, \sigma_{u1}^2)$, con los errores a nivel 2 u_{0j} y u_{1j} independientes entre ellos y con el error a nivel 1, e_{ij} .

En el caso de dos variables respuesta $Y^{(1)}$ y $Y^{(2)}$ el modelo de regresión lineal jerárquico con una variable explicatoria a nivel 1 X , tiene la forma

$$\begin{bmatrix} Y_{ij}^{(1)} \\ Y_{ij}^{(2)} \end{bmatrix} = \begin{bmatrix} \beta_{0j}^{(1)} + \beta_{1j}^{(1)} X_{ij} \\ \beta_{0j}^{(2)} + \beta_{1j}^{(2)} X_{ij} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(1)} \\ e_{ij}^{(2)} \end{bmatrix}; \quad i=1, \dots, n_j; \quad j=1, \dots, J \quad (3)$$

donde β_{0j} y β_{1j} denotan a los coeficientes de regresión en el nivel 1, $e_{ij}^{(1)}$ y $e_{ij}^{(2)}$ denotan los errores aleatorios de nivel 1 con los supuestos $E[e_{ij}^{(1)} e_{ij}^{(2)}] = 0$ y $\text{Var}[e_{ij}^{(1)} e_{ij}^{(2)}] = \sigma_e^2 \mathbf{I}_2 = \mathbf{\Omega}$. El modelo para el nivel 2, con una variable explicatoria nivel 2, tiene la forma

$$\begin{bmatrix} \beta_{0j}^{(1)} & \beta_{0j}^{(2)} \\ \beta_{1j}^{(1)} & \beta_{1j}^{(2)} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}W_j + u_{0j} & \gamma_{00} + \gamma_{01}W_j + u_{0j} \\ \gamma_{10} + \gamma_{11}W_j + u_{1j} & \gamma_{10} + \gamma_{11}W_j + u_{1j} \end{bmatrix}; \quad j=1, \dots, J \quad (4)$$

donde γ_{00} , γ_{01} , γ_{10} , y γ_{11} denotan los coeficientes de regresión en el nivel 2, W_j a la variable explicatoria de nivel 2 y los u_{0j} , u_{1j} , u_{0j} , y u_{1j} a los errores aleatorios de nivel 2. Los supuestos de los errores son que estos se distribuyen normalmente con:

$$E \begin{bmatrix} u_{0j} & u_{0j} \\ u_{1j} & u_{1j} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{y} \quad \text{Var} \begin{bmatrix} u_{0j} & u_{0j} \\ u_{1j} & u_{1j} \end{bmatrix} = \begin{bmatrix} \Psi^{\circ} & \\ \Psi^{\circ} & \Psi^{\circ} \end{bmatrix} = \Psi; \quad j=1, \dots, J \quad (5)$$

donde Ψ denota a la matriz de varianzas y covarianzas del vector de errores bivariado de nivel 2, formada por Ψ° que denota a la matriz de varianzas y covarianzas de los errores correspondientes a la variable respuesta (1) y Ψ° que denota a la matriz de varianzas y covarianzas de los errores correspondientes a la variable respuesta (2), y Ψ° que denota a la matriz de varianzas y covarianzas combinada.

El modelo de regresión lineal jerárquico para k variables respuesta $(Y^{(1)} \dots Y^{(k)})$ en su representación matricial en el primer nivel está dado por

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{e}_j, \quad j=1, \dots, J, \quad (6)$$

donde \mathbf{Y}_j denota la matriz de respuestas o realizaciones de k variables dependientes; \mathbf{X}_j la matriz con los valores de las variables explicatorias de primer nivel, $\boldsymbol{\beta}_j$ la matriz de coeficientes de regresión de primer nivel y \mathbf{e}_j la matriz de errores aleatorios de primer nivel con el supuesto de que: $E(\mathbf{e}_j) = \mathbf{0}$ y $\text{Var}(\mathbf{e}_j) = \mathbf{I}_{n_j} \otimes \boldsymbol{\Omega} = \boldsymbol{\Theta}$. Para un estudio detallado ver Muller y Stewart (2006). Aunque los errores de primer nivel se distribuyen independientemente e idénticamente, la forma de la matriz de varianzas y covarianzas $\boldsymbol{\Theta}$ se muestra de manera general, porque en realidad los errores asociados con respuestas multivariadas no son probablemente ni iguales ni independientes. Algunas de las estructuras que puede tener la matriz $\boldsymbol{\Theta}$ y que pueden llegar a modelarse

- (i) $\theta^2 \mathbf{I}_k$
- (ii) $\text{diag}(\theta_1^2, \theta_2^2, \dots, \theta_k^2)$
- (iii) $\boldsymbol{\Theta}$
- (iv) $\theta^2 \left\{ \begin{matrix} |^{l-1} \\ \vdots \\ |^{l-1} \end{matrix} \right\}, l=1, 2, \dots, k$
- (v) Restricciones arbitrarias de igualdad.

La estructura que plantea (i) para la matriz Θ implica que las k respuestas observadas se distribuyan idénticamente iguales, lo cual no difiere de la estructura de errores que se supone en las respuestas univariadas. La estructura que presenta (ii) relaja este supuesto proponiendo que los errores pueden ser heterocedásticos. La estructura simétrica de la matriz Θ en (iii) sugiere además que los errores pueden estar correlacionados. Por otro lado, la estructura en (iv) provee una correlación de primer orden ρ entre los errores adyacentes en datos con pequeñas series de tiempo. Finalmente, la estructura en (v) sirve para necesidades teóricas y prácticas muy específicas.

En el segundo nivel el modelo general está dado por

$$\beta_j = \mathbf{w}_j' \Gamma + \mathbf{u}_j, \quad j = 1, \dots, J, \quad (7)$$

donde β_j es la matriz de coeficientes de regresión de primer nivel; respuestas o realizaciones de las k variables dependientes, \mathbf{w}_j es la matriz con los valores de las variables explicatorias de segundo nivel; Γ es la matriz de coeficientes fijos de segundo nivel y \mathbf{u}_j es la matriz de errores aleatorios de segundo nivel, con el supuesto de que:

$$E(\mathbf{u}_j) = \mathbf{0} \quad \text{y} \quad \text{Var}(\mathbf{u}_j) = \mathbf{I}_k \otimes \Psi_p = \Psi,$$

donde Ψ es la matriz de varianzas y covarianzas de las matrices de errores k -variados de nivel 2.

3. APLICACIÓN: CURVAS DE CRECIMIENTO DE CONIFERAS

Interesa estudiar la tendencia y la variabilidad del crecimiento en altura y diámetro entre familias de plantas coníferas. Para esto se empleó una muestra de 28 perfiles de crecimiento (datos proporcionados por el Instituto de Genética Forestal de la Universidad Veracruzana), provenientes de igual número de familias de coníferas. Los perfiles de crecimiento fueron calculados a partir de mediciones mensuales de la altura $Y^{(1)}$ y diámetro $Y^{(2)}$ de 32 plantas (en promedio) por familia, en un lapso de 6 meses. Algunas de las familias no presentaban datos (mediciones de altura y diámetro) en ciertos meses; sin embargo, se

asignó un valor razonable en los casos en los que sucedió esto. El modelo se ajustó con el paquete MLwiN (Rasbash, J. *et al.*, 2009).

El modelo propuesto para los perfiles de crecimiento fue:

$$y = at^b \exp^{ct},$$

donde y denota la variable respuesta, t denota el tiempo, $a = \exp^{\beta_0}$, $b = \beta_1$, y $c = \beta_2$; donde aplicando la transformación logaritmo natural (\ln) a la ecuación anterior se obtiene el modelo lineal:

$$\ln y_{ij} = \beta_{0j} + \beta_{1j} t_{ij} + \beta_{2j} \ln t_{ij} + e_{ij}, \quad (8)$$

que contextualizado al estudio en particular se entiende a y_{ij} como la observación en el tiempo i ($i = 1, 2, \dots, 6$) de la familia j ($j = 1, 2, \dots, 28$). De esta forma, las ecuaciones que se propusieron al segundo nivel fueron:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}W_1 + \gamma_{02}W_2 + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_1 + \gamma_{12}W_2 + u_{1j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}W_1 + \gamma_{22}W_2 + u_{2j} \end{aligned} \quad j = 1, 2, \dots, 28 \quad (9)$$

De lo anterior el modelo jerárquico multivariado completo que se propuso finalmente para explicar las variaciones ocurridas en ambos niveles fue el siguiente:

Nivel 1:

$$\begin{bmatrix} \ln \text{Altura} \\ \ln \text{Diametro} \end{bmatrix} = \begin{bmatrix} \beta_{0j}^{(1)} + \beta_{1j}^{(1)} t_{ij} + \beta_{2j}^{(1)} \ln t_{ij} \\ \beta_{0j}^{(2)} + \beta_{1j}^{(2)} t_{ij} + \beta_{2j}^{(2)} \ln t_{ij} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(1)} \\ e_{ij}^{(2)} \end{bmatrix}$$

Nivel 2:

$$\beta_{0j}^{(1)} = \gamma_{00} + \gamma_{01} \text{ASNM} + \gamma_{02} \text{OIAMAD} + u_{0j}$$

$$\beta_{1j}^{(1)} = \gamma_{10} + \gamma_{21} \text{ASNM} + \gamma_{12} \text{OIAMAD} + u_{1j}$$

$$\beta_{2j}^{(1)} = \gamma_{20} + \gamma_{21} \text{ASNM} + \gamma_{22} \text{OIAMAD} + u_{2j}$$

$$\beta_{0j}^{(2)} = \gamma_{00} + \gamma_{01} \text{ASNM} + \gamma_{02} \text{OIAMAD} + u_{0j}$$

$$\beta_{1j}^{(2)} = \gamma_{10} + \gamma_{21} \text{ASNM} + \gamma_{12} \text{OIAMAD} + u_{1j}$$

$$\beta_{2j}^{(2)} = \gamma_{20} + \gamma_{21} \text{ASNM} + \gamma_{22} \text{OIAMAD} + u_{2j}$$

En las figuras 2 y 3 se presentan las distribuciones de crecimiento para las variables altura y diámetro respectivamente.

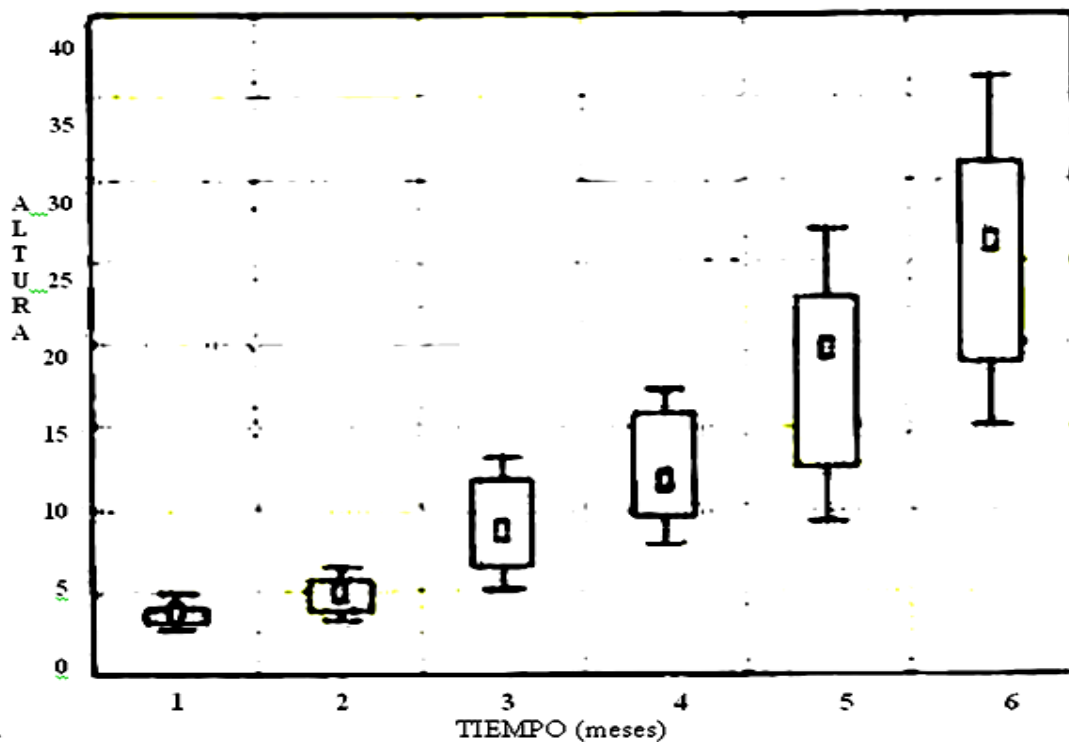


Figura 2. Distribución de los perfiles de crecimiento de las 28 familias para la variable altura. Cada caja representa la distribución de los promedios de 32 plantas.

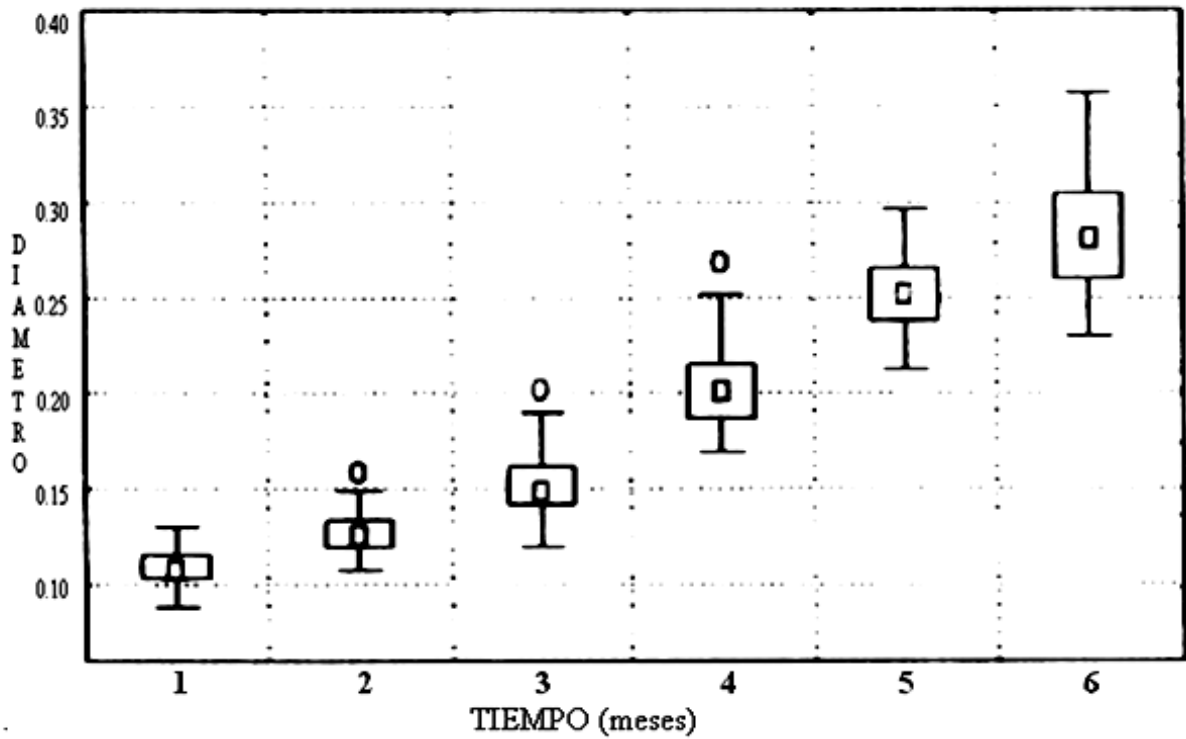


Figura 3. Distribución de los perfiles de crecimiento de las 28 familias para la variable diámetro. Cada caja representa la distribución de los promedios de 32 plantas, indicándose con círculos los valores atípicos.

En las figuras 4 y 5 se muestran los perfiles de crecimiento de las 28 familias mencionadas anteriormente (también para altura y diámetro respectivamente)

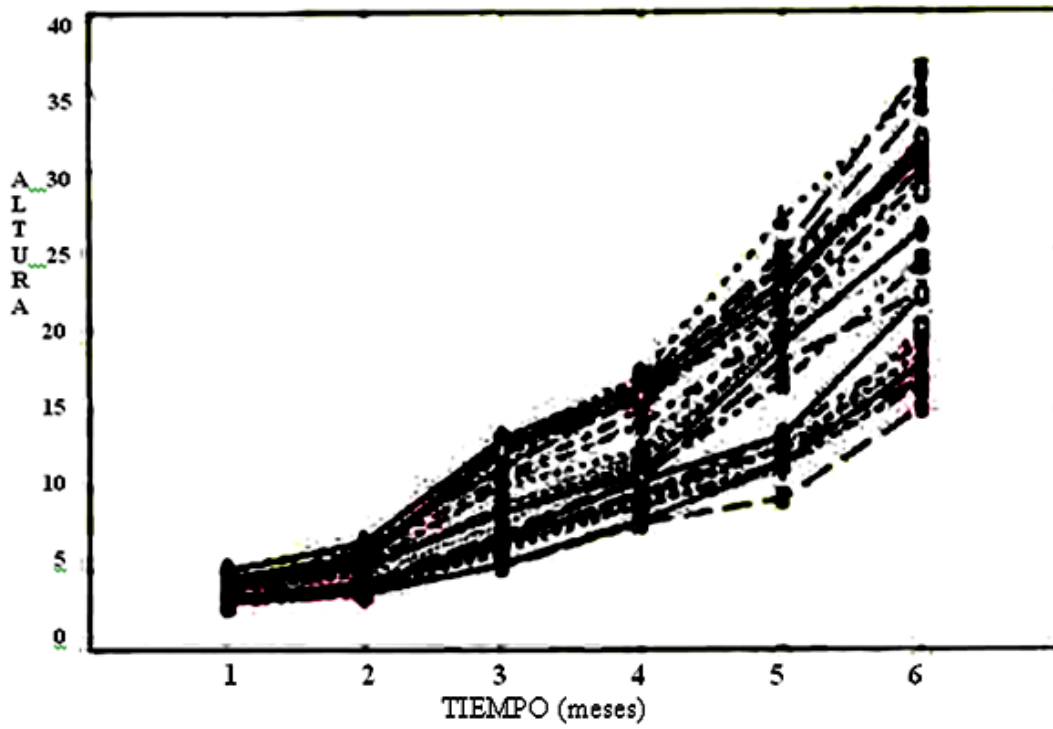


Figura 4. Perfiles de crecimiento de las 28 familias para la variable altura.

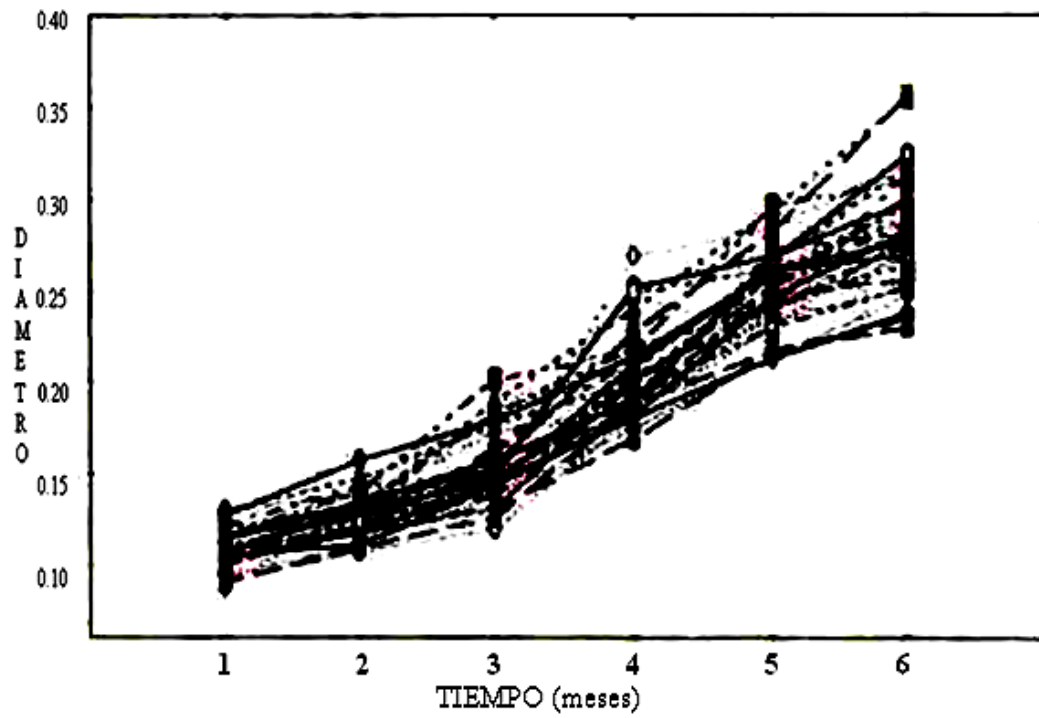


Figura 5. Perfiles de crecimiento de las 28 familias para la variable diámetro.

De acuerdo con el ajuste del modelo jerárquico bivariado, se observa que los coeficientes que resultaron significativos ($p < 0.05$) fueron los asociados con la constante y el tiempo en ambas variables respuesta conjuntamente. También el coeficiente asociado a ASNМ para la variable diámetro resultó ligeramente significativo ($p < 0.089$).

Tabla 1. Ajuste del modelo con ln(Altura) y ln(Diámetro) como variable respuesta.

Respuesta	Variable asociada	Estimación	Error estándar	p -value
ln(Altura)	β_{00} Constante	0.984346	0.281321	<0.01***
ln(Diámetro)	β_{00} Constante	-1.733342	0.184483	<0.01***
ln(Altura)	β_{01} ASNМ	-0.000205	0.000076	<0.01***
ln(Diámetro)	β_{01} ASNМ	0.000085	0.000050	0.089*
ln(Altura)	β_{02} DIAMAD	0.008011	0.005048	0.114
ln(Diámetro)	β_{02} DIAMAD	0.003505	0.003311	0.293
ln(Altura)	β_{10} T	0.510623	0.156151	<0.01***
ln(Diámetro)	β_{10} T	0.289564	0.130108	0.032**
ln(Altura)	β_{11} ASNМ x T	0.000005	0.000042	0.904
ln(Diámetro)	β_{11} ASNМ x T	-0.000013	0.000035	0.711
ln(Altura)	β_{12} DIAMAD x T	-0.004044	0.002802	0.149
ln(Diámetro)	β_{12} DIAMAD x T	-0.001387	0.002335	0.555
ln(Altura)	β_{20} ln T	-0.063474	0.440872	0.888
ln(Diámetro)	β_{20} ln T	-0.059033	0.367355	0.872
ln(Altura)	β_{21} ASNМ x ln T	-0.000153	0.000119	0.200
ln(Diámetro)	β_{21} ASNМ x ln T	-0.000033	0.000099	0.741
ln(Altura)	β_{22} DIAMAD x ln T	0.010975	0.007910	0.167
ln(Diámetro)	β_{22} DIAMAD x ln T	0.002299	0.006592	0.726

La matriz de componentes de varianza y covarianza que resultó de este ajuste se presenta a continuación (los errores estándar de los coeficientes se encuentran abajo entre paréntesis):

$$\hat{\Omega}_u = \begin{bmatrix} 0.019 & & \\ (0.005) & & \\ 0.008 & 0.006 & \\ (0.003) & (0.002) & \\ \mathbf{0}_{4 \times 1} & \mathbf{0}_{4 \times 1} & \mathbf{0}_{4 \times 4} \end{bmatrix}$$

Donde $\mathbf{0}$ denota matrices de 0's con errores estándar de 0's. En esta matriz vemos que todos los coeficientes aleatorios tienen una varianza de cero, excepto los coeficientes asociados con los interceptos y la covarianza entre ambos. Lo anterior indica en primera instancia que los perfiles de crecimiento varían en los interceptos (tanto en altura como en diámetro), lo cual se puede ver en las Figuras 4 y 5. Dicho de otra manera, solo se encontraron diferencias significativas ($p < 0.01$) entre los interceptos de los perfiles de crecimiento para ambas variables respuesta. También resultó significativa la covarianza entre los coeficientes asociados a las constantes, lo que quiere decir que entre más diferencia exista entre los interceptos del perfil de crecimiento de una de las variables respuesta, más diferencia habrá en los perfiles de crecimiento de la otra.

El hecho de no encontrar variación significativa en muchos de los coeficientes aleatorios nos está indicando, en principio, que las variables explicatorias de segundo nivel ($W's$) han explicado satisfactoriamente la variación existente entre los perfiles de crecimiento, con excepción de los parámetros aleatorios asociados con los interceptos de ambas variables respuesta. En este caso, las variables explicatorias de segundo nivel (ASNMM y DIAMAD) no explicaron completamente la variación en los crecimientos de las 28 familias al inicio de las mediciones, lo cual apunta a que existen otra u otras variables que no se midieron y que están provocando esa variación en el crecimiento inicial.

En estudios de este tipo, la variabilidad entre las mediciones efectuadas mensualmente (tanto en altura como en diámetro) resultó significativa ($p < 0.01$), como se aprecia en el estimador de la matriz de varianzas y covarianzas de errores al primer nivel:

$$\hat{\Omega}_e = \begin{bmatrix} 0.009 & & \\ (0.001) & & \\ 0.001 & 0.007 & \\ (0.001) & (0.001) & \end{bmatrix}$$

De acuerdo con los resultados obtenidos del ajuste del modelo de regresión lineal jerárquico multivariado para explicar la variación entre los perfiles de crecimiento en altura y diámetro de 28 familias de plantas coníferas, podemos escribir finalmente el modelo de la siguiente manera

$$\begin{bmatrix} \ln(\text{Altura}) \\ \ln(\text{Diámetro}) \end{bmatrix} = \begin{bmatrix} -0.9843 + 0.5106 \mu_j \\ -1.7333 + 0.2895 \mu_j \end{bmatrix} - 0.0002 \text{ASN}M_j$$

En el cual observamos que el tiempo es un variable explicatoria de primer nivel que está asociada con ambas variables respuesta, es decir, con la altura y el diámetro, sin embargo, ASN_M es una variable explicatoria de segundo nivel que sólo se encuentra asociada con altura.

4. COMENTARIO FINAL

La modelación jerárquica es una técnica estadística que permite el análisis de datos que presentan estructura jerárquica, lo cual es muy común en varios campos de aplicación en biometría y ciencias ambientales. Esta metodología permite estudiar la variabilidad tanto a nivel individuo como entre los grupos, ya que incorpora en el análisis modelos a nivel individuo y a nivel grupo, y por lo tanto errores en cada uno de los niveles de la estructura jerárquica. Por otra parte el investigador en la mayoría de las ocasiones tiene la necesidad de estudiar medidas múltiples para poder dar solución a su investigación. Por lo cual es necesario el análisis multivariado que tiene aplicaciones en diversas disciplinas, tales como: Biología, Agricultura, Medicina, entre otras. Así la mayoría de las investigaciones se dan en el contexto del análisis multivariado, en particular para la modelación tendríamos el caso de la regresión lineal multivariada. La modelación jerárquica multivariada es una técnica que permite analizar datos con estructura jerárquica cuando se tienen dos o más variables respuesta. En la biometría hay infinidad de estudios en los cuales se presenta esta situación: análisis de datos longitudinales, estudios de medidas repetidas y curvas de crecimiento multivariadas.

Agradecimientos

Los autores agradecen a Sergio Juárez y a Judith Montero por su apoyo en la escritura y edición de versiones preliminares de este trabajo.

REFERENCIAS

- [1] ARNAU, J. y BALLUERKA, N. (2004): Análisis de Datos Longitudinales y de Curvas de Crecimiento. Enfoque Clásico y Propuestas Actuales; **Psicothema**, 16. 156-162.
- [2] CODY, R. (2001) : **Longitudinal data and SAS. A Programmer`s Guide**. SAS Publishing. Cary, NC: SAS Institute, Inc.
- [3] GELMAN, A. and HILL, J. (2007) : **Data Analysis Using Regression and Multilevel/Hierarchical Models**. Cambridge University Press.
- [4] GOLDSTEIN, H. (1997) : Methods in School Effectiveness Research; **School Effectiveness and School Improvement**, 8. 369-390.
- [5] LAIRD, N. M. and WARE, J. H. (1982) : Random Effects Models for Data Longitudinal; **Biometrics**, 38. 963-974.
- [6] LITTLE, R. C. STROUP, W.W. and FREUND, R. J. (2002) : **SAS for Linear Models**. Cary, NC: SAS Institute, Inc.
- [7] LONGFORD, N. T. (1993) : **Random Coefficient Models**. Oxford: University Press, New York.
- [8] McCULLOCH, C. E. and SEARLE, S. R. (2001) : **Generalized, Linear and Mixed Models**. John Wiley. New York. U.S.A.
- [9] MULLER, K. E. and STEWART, P.W. (2006) : **Linear Model Theory: Univariate, Multivariate , and Mixed Models**. John Wiley. New Jersey. U.S.A.
- [10] PINHEIRO J. C. and BATES, D. M. (2000) : **Mixed Effects Models in S and S-PLUS**. Springer Verlag, New York.
- [11] RAUDENBUSH, S. W. and BRYK, A. S. (2002) : **Hierarchical Linear Models: Applications and Data Analysis Methods**. NewburyPark. Sage.
- [12] RASBASH, J., STEELE, F., BROWNE, W. J. and GOLDSTEIN, H. (2009) : **A User`s Guide to MLwiN. Version 2.10**. Centre for Multilevel Modelling. University of Bristol.
- [13] SEARLE, S. R., CASELLA, G. and McCULLOCH, C. E. (2006) : **Variance Components**. John Wiley, New Jersey, U.S.A.
- [14] SNIJDERS, T. A. B. and BOSHER, R. (1999) : **Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling**. Sage Publications, London.
- [15] WEST, B. T., WELCH, K. B. and GALECKI, A. T. (2007) : **Linear Mixed Models: A Practical Guide Using Statistical Software**. Chapman & Hall/CRC, Boca Raton, Florida. U.S.A.