

ANÁLISIS CON MODELOS MULTINIVEL

Lic. Flores López Juan Carlos

✉ caarloslopez1@gmail.com

RESUMEN

Esta investigación desarrolla una aplicación del modelo multinivel, a los datos simulados dado que en el medio no se pudo conseguir información. Por lo tanto, se toma en cuenta la información bajo las siguientes características, como variable respuesta se tomó en cuenta a las calificaciones obtenidas por los alumnos en Matemáticas y Lenguaje. Se asumió niños de 4 a 8 años, los cursos desde pre kínder, kínder, primero y segundo de primaria. Se consideró también 10 tipos de colegio y una muestra de 670 estudiantes

En este estudio, los estudiantes (nivel 1) se encuentran anidados tanto en grados escolares como en escuelas (colegios), por lo que en los modelos que se llevaron a cabo, se puso a prueba si debían considerarse los grados o escuelas como un segundo nivel o incluso a los grados como un segundo nivel y las escuelas como tercer nivel.

Para el análisis se obtuvo la gráfica de dispersión de la variable nota de matemáticas la puntuación centrada donde se observa una relación con tendencia lineal positiva de magnitud media. Se realizó el estudio de las características residuales donde se observa la aproximación al supuesto de normalidad.

En la aplicación del modelo mismo se evidencia que existe una relación importante entre las diferentes dimensiones medidas por el instrumento el rendimiento académico particularmente en el caso de las Matemáticas. Los factores más relevantes son los de “Atención y persistencia” y “Desempeño académico”. Y para la comprobación de los resultados puede observar los resultados obtenidos en esta investigación.

PALABRAS CLAVE

Modelo multinivel, grados o escuelas, puntuación centrada, rendimiento académico.

1. INTRODUCCIÓN

En la Universidad Mayor de San Andrés (U.M.S.A), la actividad de un docente tiempo completo está inmersa la investigación y como es de conocimiento de toda la comunidad universitaria de nuestro medio, esta actividad permite el avance y la actualización en el conocimiento científico, que exige nuestra realidad. A raíz de este hecho, se realiza la investigación para la gestión 2015 sobre la aplicación de los modelos multinivel. Este es con el propósito de contribuir con la comunidad estudiantil y todos lo que tengan interés en conocer estos métodos que es de gran importancia en las aplicaciones en el

análisis de datos en las distintas disciplinas. Estos métodos pueden ser aplicados en todas las ciencias, como ser ingeniería, psicología, salud, sociología, económico, la industria, administración y centros de investigación de ámbito universitario, etc.

Estamos conscientes de que toda investigación Estadística, se basa en información recogida, estos pueden ser a través de encuestas, censos, simulación, etc. Una de las etapas que sigue después de la recolección de datos es la obtención de resultados de la información. Desde este punto de vista el proyecto tuvo el propósito de desarrollar el soporte teórico y la aplicación de los modelos multinivel.

Dentro los aspectos metodológicos, podemos mencionar que el tipo de estudio es exploratorio y descriptivo. Y desde el punto de vista de los métodos de investigación que se aplica en la presente investigación es Estadístico y de análisis.

2. LOS DATOS JERÁRQUICOS Y EL PROBLEMA MULTINIVEL

En la vida real existe una infinidad de clases o tipos de datos, incluyendo datos recolectados, en las ciencias pedagógica, exactas, económicas, ingeniería, etc. tienen una estructura jerárquica o en forma de clúster¹.

Por ejemplo, estudios de herencia realizados en seres humanos y en animales se tratan con niveles o jerarquías naturales, donde los hijos (descendientes) están agrupados dentro de las familias. Estos, a su vez, tienden a parecerse más a sus padres en sus características físicas y mentales que individuos escogidos al azar de la población. En segundo lugar, muchos experimentos también crean datos jerárquicos, por ejemplo, los diseños experimentales en las clínicas son llevados a cabo en centros y grupos de individuos elegidos aleatoriamente (Goldstein; 1995).

El termino jerarquía define unidades agrupadas en diferentes niveles. Así, los hijos serían las unidades del nivel 1 en una estructura de datos de 2 niveles, donde el nivel 2 está dado por las familias; los estudiantes (las unidades del nivel 1) están agrupados dentro de las escuelas (las unidades del nivel 2).

¹ Se entiende por “clúster” un agrupamiento que contiene elementos de un menor nivel. Por ejemplo, en una muestra, el conjunto de familias en un vecindario. El “nivel”, por otra parte, es un componente de los datos jerárquicos. El nivel 1 es el menor nivel; por ejemplo, estudiantes dentro de las escuelas o medidas repetidas para un mismo individuo.

Los modelos multinivel o jerárquicos se presentan en distintos tipos de aplicaciones, por ejemplo: La aplicación en el área de la salud y la educación, en las que los datos se presentan de una forma anidada o jerárquica; por ejemplo, pacientes dentro de hospitales, o estudiantes dentro de escuelas. Las aplicaciones de los modelos jerárquicos también se enmarcan perfectamente en una amplia gama de aplicaciones del gobierno y los negocios, etc. Donde muestran clústers que son efectuadas en una o varias etapas, y ofrecen una aproximación unificadora por medio del análisis de los modelos de efectos aleatorios (random-effects), de varianza-covarianza (variance-components o ANOVA) y de los modelos mixtos (mixedmodels). Así sucesivamente.

Respecto a los métodos de estimación los modelos multinivel, se dividen en:

- 1) Aquellos que utilizan métodos de máxima verosimilitud.
- 2) Los que se basan en la Estadística Bayesiana.

A su vez, dentro de cada uno de estos grupos, se encuentran distintas formas de estimación. El cuadro 1 sintetiza las formas de estimación existentes para estructuras de datos multinivel.

Cuadro 1
Métodos de estimación para los modelos multinivel

	<i>Métodos</i>	<i>Abreviatura</i>
Máxima verosimilitud	Mínimos Cuadrados Generalizados	IGLS
	Mínimos Cuadrados Generalizados Iterativos	RIGLS
	Cuasi-verosimilitud Marginal	MQL
	Cuasi-verosimilitud Penalizada	PQL
	Full Bayes estimación	FB
Estadística Bayesiana	Empirical Bayes estimation	EM
	Cadenas de Markov-Monte Carlo	MCMC

A continuación, mencionamos un modelo jerárquico lineal de dos niveles (HLM)¹. Este modelo, está dado por:

$$\underline{y}_j = X_j \beta + Z_j \underline{\delta}_j + \underline{\epsilon}_j$$

$$j = 1, \dots, m,$$

Con
$$\begin{pmatrix} \underline{\epsilon}_j \\ \underline{\delta}_j \end{pmatrix} \sim N \left(\begin{pmatrix} \phi \\ \phi \end{pmatrix}, \begin{pmatrix} \sum_j \theta & \phi \\ \phi & \Omega(\mathcal{E}) \end{pmatrix} \right)$$

Y
$$(\underline{\epsilon}_j, \underline{\delta}_j) \perp (\underline{\epsilon}_\ell, \underline{\delta}_\ell) \quad \forall j \neq \ell$$

Las dimensiones de los vectores, $\underline{y}_j, \beta$ y $\underline{\delta}_j$ respectivamente, son n_j, r y s . Como en todos los modelos de tipo regresión, las variables explicativas X y Z son consideradas como variables fijas, que también pueden ser expresadas en las distribuciones de las variables aleatorias ϵ y δ son condicionales en X y Z . Las variables aleatorias $\underline{\epsilon}$ y $\underline{\delta}$ también se llaman vectores residuales en los niveles 1 y 2, respectivamente. Las variables δ también se denominan pendientes aleatorias. Las unidades de nivel dos también son llamadas clusters.

La especificación estándar más

¹ Tom A. B. Snijders and Johannes Berkhof

frecuentemente usada de las matrices de covarianza son los residuales de nivel uno, que son independientes e idénticamente distribuidos (i.i.d), por ejemplo:

$$\sum_j (\theta) = \sigma^2 I_{n_j}$$

Donde I_{n_j} es la matriz identidad n_j -dimensional; y que todos los elementos de la matriz de covarianza de nivel dos Ω son parámetros libres (lo que se podría identificar Ω con ξ), o algunos de ellos están limitados a 0 y los otros son parámetros libres.

El cuestionamiento de la especificación de este modelo puede ser dirigido a diversos aspectos: la elección de variables incluidas en X , la elección de variables para Z , los residuos después de haber esperado el valor 0, la homogeneidad de las matrices de covarianza a través de los clústers, la especificación de las matrices de covarianza, y las distribuciones normales multivariable. Hay que tener en cuenta que las variables explicativas X y Z son consideradas como determinísticas; la suposición de que los valores esperados de los residuos (para las variables explicativas fijas) son cero, es análoga a la suposición de que los residuos no están correlacionados con las variables explicativas en un modelo

con variables explicativas aleatorias.

3. GENERALIDADES Y CARACTERÍSTICAS

Después de una ardua búsqueda de información, no se pudo conseguir información para la aplicación de los modelos multinivel, toda información era incompleta y muy difícil de aplicar con este tipo de información. Dado el escenario incierto de la disponibilidad de una base de datos para su aplicación de este tipo de modelos, se tomó la decisión de utilizar información simulada.

4. RESULTADOS

En este estudio se utilizaron como *variables respuesta* las calificaciones obtenidas por los alumnos en Matemáticas y Lenguaje. Se asumió niños de 4 a 8 años, los cursos de pre kínder, kínder, primero y segundo de primaria. Se consideró también 10 tipos de colegios y una muestra de 670 estudiantes.

Se realizó el modelamiento como variable dependiente nota de la materia Matemáticas como variable respuesta, y variables independientes: grado y colegios, cuyos resultados fueron los siguientes: el intercepto cambió de 7.282 a 20.595. La variable “grado” demuestra efectos significativos, y el ajuste del modelo también mejora en forma significativa, disminuyendo la medida de desviación de 194.518 a 33.429.

Lo anterior señala que al haber controlado las diferencias individuales, se manifiestan las diferencias entre los grados. Señala también que la diferencia en el rendimiento entre alumnos no es muy amplia, y que varía significativamente, hay características homogéneas en las escuelas. El modelo

indica que existe una estructura que justifique un modelamiento multinivel.

Bajo este criterio también es necesario realizar el análisis de residuos, para la comprobación de la normalidad de la información.

Lo que se observa es la información de los residuos que se comporta aproximadamente normal, cumpliendo los requerimientos del modelo. En este estudio, los estudiantes (nivel 1) se encuentran anidados tanto en grados escolares como en escuelas (colegios), por lo que en los modelos que se llevaron a cabo, se puso a prueba si debían considerarse los grados o escuelas como un segundo nivel, o incluso a los grados como un segundo nivel y las escuelas como un tercer nivel.

Los siguientes apartados se organizan entorno a las dos **variables de respuesta** analizadas: Calificaciones de Matemáticas y Calificaciones de Lenguaje.

En cada apartado se ponen a prueba modelos con base en las siguientes estructuras de datos:

Nivel 1:	Grados
Nivel 2:	Colegios

Y posteriormente:

Nivel 1:	Alumnos.
Nivel 2:	Grado escolar.

Por último, se pone a prueba:

Nivel 1:	Alumnos
Nivel 2:	Grados escolar
Nivel 3:	Colegios.

Después de realizar un estudio de los modelos multinivel, tratando de responder el rendimiento en la asignatura de matemáticas y lenguaje considerando como variable de pendiente y el grado y colegio como niveles de primer y segundo orden, podemos decir lo siguiente:

- Se estableció que existe una relación importante entre las diferentes dimensiones medidas por el instrumento el rendimiento académico particularmente en el caso de las Matemáticas. Los factores más relevantes son los de “Atención y persistencia” y “Desempeño académico”.
- A pesar de que los resultados obtenidos en secciones anteriores señalaban que la relación entre las variables explicativas y las variables respuesta es distinta en las diferentes escuelas, los resultados del análisis multinivel no apoyan esto. Sin embargo, debe interpretarse este resultado con cuidado, ya que como se mencionó al inicio, pueden no haberse manifestado estas diferencias debido a que el número de colegios incluidos en la muestra es reducido.
- El grado escolar que cursa el niño cuando se lleve a cabo el seguimiento es un aspecto que debe ser incluido tanto en los modelos de regresión jerárquicos como en los tradicionales, ya que incrementa la precisión de las estimaciones. En los grados avanzados existe mayor exactitud en las predicciones.

En conclusión, los análisis presentados en esta sección señalan que no es importante la inclusión de un segundo nivel correspondiente al *grado* que cursaban los alumnos al momento de la evaluación. Esto es verdad

para Matemáticas, y aunque en Lenguaje no fue necesario *utilizar* un modelo multinivel, se confirma a través del uso de modelos de regresión lineal tradicional la necesidad de tomar en cuenta el *grado* que cursan. Entre más avanzado sea éste, se *alcanzará* mayor exactitud en la predicción.

De *igual* manera, aquellos modelos que utilizan las dimensiones medidas por el instrumento en lugar de la puntuación global, logran un mejor *ajuste*. Se determinó que los factores más significativos son el “Desempeño académico” y la “Atención”.

Los resultados de este estudio indican que el género no tiene efectos sobre el rendimiento.

Asimismo, el uso del modelamiento multinivel ha servido para comprobar que las diferencias que existen entre los colegios, no requieren una interpretación diferente para cada escuela de las puntuaciones obtenidas con el instrumento. Puede utilizarse la misma transformación de puntajes en todos los colegios incluidos en el estudio.

Tanto los análisis de regresión OLS como los modelos de regresión jerárquica, han permitido conocer la magnitud de la relación entre las puntuaciones obtenidas en la prueba de seguimiento y las calificaciones escolares posteriores. Han permitido, asimismo, establecer cuál de los factores medidos por la prueba tiene mayor efecto sobre el rendimiento. También han logrado establecer la relación que existe entre el momento en que se realiza aplicación del instrumento (grado escolar en que se aplica) y su capacidad predictiva, así como la importancia de los factores varia en los diferentes grados, y no en las diferentes escuelas (en forma significativa). Sin embargo, aún no se ha realizado la estimación de la exactitud de

la prueba para detectar niños en riesgo de presentar problemas de aprendizaje. Es decir, no se tuvo información acerca de los porcentajes de éxito que se tienen en la identificación y clasificación de los niños.

BIBLIOGRAFIA

- Gelman Andrew “*Data Analysis Using Regression and Multilevel/Hierarchical Models*”.
- Goldstein Harvey “*Handbook of Multilevel Analysis*”
- Hox J.J. “*Applied Multinivel Analysis*”
- Martínez-Garrido Cynthia y F. Javier Murillo “*Programas para la realización de Modelos Multinivel. Un análisis comparativo entre MLwiN, HLM*”.
- Min Yang, JonRasbash, Harvey Goldstein, Maria Barbosa “*MLwiN Macros for advanced Multilevel modeling*”.
- Snijders Tom A. B. “*Multilevel Analysis*”