

AMENAZAS A LA VALIDEZ ANALÍTICA DE LAS TÉCNICAS USADAS HABITUALMENTE EN LA EVALUACIÓN DE PROGRAMAS

**Guillermo Vallejo
J. Ramón Fernández
Roberto Secades**

Universidad de Oviedo

RESUMEN

Técnicas basadas en el modelo lineal general son frecuentemente recomendadas para el análisis de datos obtenidos desde diseños cuya particularidad más relevante es la asignación de grupos intactos a las condiciones investigación. A pesar del apoyo que tradicionalmente tan recibido estos procedimientos, con en el presente trabajo se pretende, por un lado, alertar a los investigadores aplicados de los errores de inferencia que pueden cometer cuando en el proceso de evaluación científica utilizan los modelos estadísticos convencionales para establecer la efectividad de un tratamiento administrado a grupos de sujetos, más que sujetos individuales. Y, por otro lado, sugerirles procedimientos analíticos válidos y fáciles de implementar mediante alguno de los diferentes programas genéricos existentes (p.e., mediante el módulo PROC MIXED del programa SAS®).

Los autores agradecen los comentarios realizados por Dr. Manuel Ato acerca del trabajo. Sus sugerencias nos han permitido mejorar sustancialmente la exposición y la calidad del mismo. Este trabajo ha sido financiado mediante el proyecto de investigación concedido por el MEC (Ref.: SEJ2005-01883).

Correspondencia: Guillermo Vallejo Seco. Universidad de Oviedo. Departamento de Psicología. Plaza de Benito Feijóo, s/n, 33003 Oviedo (España. E-mail: gvallejo@uniovi.es.

Palabras clave: *EVALUACIÓN DE PROGRAMAS, UNIDAD DE ANÁLISIS, MODELO MIXTO LINEAL GENERAL, SESGO DE AGREGACIÓN, DATOS ANIDADOS.*

SUMMARY

Techniques based on the general linear model are frequently recommended for the analysis of data obtained from designs whose more notable particularity is that intact groups are assigned to specific treatment conditions. Despite the advocacy received by these procedures, with this paper is sought, on one hand, to alert the applied investigators of the inferential errors that can make when in the process of scientific evaluation the usual statistical models use to found the effectiveness of an intervention administered to groups of subjects rather than to individual subjects. And, on the other hand, to suggest them valid and easy analytic procedures of implementing by means of some of the different existent generic programs (e.g., trough the MIXED procedure of SAS® program).

Key words: *EVALUATION OF PROGRAMS, UNIT OF ANALYSIS, GENERAL LINEAR MIXED MODEL, AGGREGATION BIAS, NESTED DATA.*

INTRODUCCIÓN

En las ciencias del comportamiento resulta bastante habitual que los datos obtenidos en contextos clínicos, educativos y organizacionales presenten una estructura jerárquica que permite distinguir diferentes niveles de agregación. Por ejemplo, los empleados de una empresa están agrupados en departamentos, los pacientes de un hospital en plantas y los estudiantes de una escuela en clases. En la última situación, los estudiantes constituyen el nivel de agregación inferior, las clases el nivel intermedio y las escuelas el nivel superior. Tanto en psicología como en otras disciplinas (p.e., epidemiología, producción animal, salud pública o sociología), los investigadores cada día están más interesados en evaluar los cambios acaecidos en

sistemas donde unidades de observación de un determinado nivel de agregación se hallan anidadas dentro de un nivel de agregación superior.

Hasta bien entrada la década de los ochenta los investigadores aplicados apenas prestaron atención a las agrupaciones naturales presentes en los datos. La mayor parte de los mismos, fueron registrados en el nivel más bajo de la jerarquía y analizados mediante técnicas basadas en el modelo lineal general. Este modelo constituye el pilar fundamental en el que se apoya el análisis estadístico en la investigación social, e integra procedimientos tan familiares como el análisis de la regresión, el análisis de la varianza, o el análisis de la covarianza (ANCOVA), y por extensión la mayor parte de los procedimientos multivariados utilizados actualmente (Ato, 2002). Por ejemplo, una técnica que se emplea con mucha reiteración para evaluar la efectividad de los programas de prevención e intervención social es el ANCOVA. Si se ignora el problema de la unidad de análisis, al que nos referiremos posteriormente, y sus posibles efectos sobre la precisión de las estimaciones y la amplitud de las inferencias, el ANCOVA constituye un buen método para determinar si la variable independiente afecta a la dependiente. Con todo, conviene tener presente que si el programa resulta efectivo, la técnica en cuestión, no indica ni dónde ni cuándo tiene su efecto óptimo. Mientras que si el programa no surte los efectos deseados, la técnica tampoco permite explorar qué otras influencias están actuando sobre la respuesta, ni establecer si la relación entre el programa y la respuesta está afectada por la acción moderadora de otras variables presentes en el estudio, ni tan siquiera conocer si el comportamiento de estas variables es similar de registrarse en más de un nivel de la jerarquía. Si bien las amenazas a la validez de los análisis realizados con el ANCOVA u otros enfoques similares son diversas, a nuestro juicio, las dos más importantes son las que se derivan de la falta de independencia de las unidades y de la agregación de las mismas.

Mediante el presente trabajo nos proponemos, por un lado, alertar a los investigadores aplicados de los errores inferenciales que pueden cometer cuando en el proceso de evaluación científica utilizan los modelos estadísticos convencionales para establecer la efectividad de un tratamiento administrado a grupos de sujetos, por lo general,

intactos. Y, por otro lado, sugerirles procedimientos analíticos válidos, y también fáciles de implementar mediante alguno de los programas informáticos globales o generales disponibles en la actualidad (p.e., SAS, S-PLUS o SPSS). Para alcanzar los objetivos propuestos, en las secciones 2 y 3 se revisan detalladamente los problemas que surgen cuando se emplean técnicas basadas en el modelo lineal general o en el modelo lineal generalizado para analizar datos provenientes de unidades que no son independientes entre sí; así como la forma de abordarlos mediante modelos estadísticos que tanto la investigación teórica, como la experimentación realizada mediante simulación Monte Carlo consideran más apropiados. En la sección 4 se presentan conclusiones de carácter general.

EL PROBLEMA DE LA UNIDAD DE ANÁLISIS

En muchos diseños de investigación desarrollados para evaluar programas de prevención e intervención social, unidades de análisis colectivas, más que unidades de análisis individuales, constituyen el referente observacional al que va dirigido el programa o tratamiento. Cuando un investigador utiliza unidades de investigación colectivas por necesidades logísticas, de viabilidad política, de validez ecológica o por cualquier otra razón, lo que hace usualmente es asignar agrupaciones existentes de manera natural, tales como aulas escolares, departamentos empresariales o plantas hospitalarias, a las condiciones de estudio. Antes de continuar queremos dejar bien sentado que, con independencia de la regla de asignación empleada (aleatoria o intencional), desde el punto de vista analítico no es lo mismo que las unidades de asignación utilizadas para configurar las condiciones de tratamiento y control sean colectivas, en lugar de individuales. Tal confusión no sólo limita la habilidad del investigador para comprender el diseño de investigación utilizado, sino que también puede invalidar las inferencias obtenidas tras la aplicación de técnicas de análisis basadas en el modelo lineal general o en el modelo lineal generalizado.

Para aclarar lo dicho sin refugiarnos en el paraguas de la terminología utilizaremos una serie de ejemplos concatenados. Supongamos, en primer lugar, que un investigador está interesado en

estudiar cómo afectan determinadas variables explicativas de componente social, psicológico y fisiológico al número de cigarrillos que diariamente consumen los sujetos de una muestra de fumadores seleccionada al azar a partir de una lista censal expuesta en un colegio electoral de la ciudad de Oviedo. En esta situación, una parte de la variación observada en la conducta de fumar de los sujetos será adecuadamente explicada por la acción de las variables independientes reseñadas, mientras que otra parte quedará sin explicar y estará alojada en la única fuente de variación aleatoria existente. Así pues, los datos empíricos que se desean explicar son función de un componente sistemático representado por una combinación de variables explicativas y sus respectivos coeficientes (vector del predictor lineal) y un componente aleatorio que recoge la discrepancia o error entre las respuestas observadas y las pronosticadas por el modelo. Por consiguiente, en este caso el modelo lineal general resulta apropiado, pues al ser la respuesta de naturaleza métrica, se asume que los errores asociados con las unidades de observación se distribuyen normalmente y que el vector de valores esperados de la respuesta es una función lineal de un vector de coeficientes o parámetros desconocidos.

Sigamos con nuestro ejemplo ficticio referido a la conducta de fumar, pero admitamos ahora que el investigador estuviese interesado en estudiar cómo afectan las variables explicativas de componente social, psicológico y fisiológico al grado de adicción a la nicotina. Para alcanzar dicho objetivo, los sujetos que fumen 20 cigarrillos o más al día serán clasificados como altamente dependientes, los que fumen entre 10 y 20 cigarrillos como moderadamente dependientes y los que fumen menos de 10 como levemente dependientes. En esta nueva situación las variables explicativas son las mismas que en el ejemplo anterior, sin embargo, la variable dependiente es categórica. Salvo que el teorema central del límite diga otra cosa, cuando la variable de respuesta carece de naturaleza métrica no parece apropiado asumir que los errores sigan una distribución normal o que el valor esperado de la respuesta sea una función lineal de un vector de parámetros desconocidos. La solución natural para analizar datos que presenten esta característica la ofrece el modelo lineal generalizado. En concreto, este modelo resuelve los problemas reseñados transformando el vector

de valores esperados a la misma escala de medida que se utilice en el proceso de estimación de la función de regresión. Obviamente, la transformación operada en las observaciones será una u otra dependiendo de la distribución de probabilidad que se seleccione para la variable de respuesta. Por ejemplo, si la distribución de probabilidad que se elige para la variable de respuesta es la distribución gamma la función que sirve para relacionar el vector de valores esperados y el vector del predictor lineal será la recíproca de la media, si se selecciona la distribución binomial la función de enlace será la transformación logit, si se selecciona la distribución de poisson la función de enlace será la logarítmica, si se selecciona la inversa gaussiana la función de enlace será la recíproca de la media al cuadrado, mientras que si se selecciona la distribución multinomial la función de enlace será el logit generalizado.

El modelo lineal generalizado al incluir cualquier miembro de la familia exponencial de distribuciones para describir la probabilidad del término de error, constituye una versión generalizada del modelo lineal clásico; sin embargo, nos seguimos encontrando con un modelo que para explicar el comportamiento observado en la respuesta incluye un único término aleatorio distribuido independientemente con arreglo a la distribución de error seleccionada para los datos.

Para comprender mejor las limitaciones inherentes a ambos modelos, imagine el lector, por último, que a nuestro investigador conductual no sólo le hubiese interesado estudiar qué variables sociales, psicológicas y fisiológicas tenían mayor relevancia en la conducta de fumar, sino que también hubiese tenido curiosidad por verificar la eficacia de un programa de autoayuda para dejar de fumar, basado esencialmente en información suministrada por correo y en llamadas telefónicas. Para alcanzar dicho objetivo el investigador estimó conveniente utilizar un proceso de muestreo en dos etapas. En la primera eligió una muestra aleatoria de los colegios electorales de la ciudad reseñada con anterioridad, mientras que en la segunda seleccionó una muestra aleatoria de sujetos dentro de cada colegio. Posteriormente, la mitad de los grupos resultantes del proceso de muestreo fueron asignados al azar a la condición de tratamiento y la otra mitad a la condición de comparación o control. En una situación como la descrita resultaría inapropiado utilizar un modelo que incluyera

una única fuente de variación aleatoria. De hecho, si un investigador procediese de este modo ignoraría tanto la variabilidad existente entre los diferentes grupos (habitualmente los grupos no están formados al azar, y por ende, cabe esperar que difieran en múltiples y variados aspectos), como la variabilidad existente dentro de los grupos, y con ello, la variabilidad individual. Sea como fuere, mal en cualquier caso.

Como señala Murray (1998), cuando la variable de respuesta tiene una estructura métrica, pero existe más de una fuente de variación aleatoria, ninguno de los procedimientos anteriores resulta apropiado. En este caso, la solución natural la proporciona el modelo mixto lineal general o, alternativamente, el modelo mixto lineal generalizado si la naturaleza de la variable de respuesta origina datos categóricos. Además, ambos modelos se pueden aplicar cuando existe desgaste de muestra (observaciones perdidas) y los datos se desvían de los supuestos de homogeneidad e independencia. Por lo tanto, para aplicar correctamente el modelo clásico o el modelo generalizado a este tipo de datos, se requiere satisfacer ciertos supuestos, en especial, el que alude a la ausencia de relación entre las observaciones. Sin embargo, la independencia estocástica queda puesta en entredicho cuando lo que se asigna al azar a las condiciones de control y tratamiento son unidades de análisis colectivas, en lugar de unidades de análisis individuales. Cuando los grupos constituyen la unidad de análisis, lo sensato es pensar que las características específicas de los grupos queden reflejadas en los datos, pues las observaciones que están próximas en el tiempo, en el espacio o en ambas dimensiones a la vez, tienden a ser más homogéneas que las observaciones que se encuentran alejadas. Es muy probable que los datos que se extraen desde agrupaciones naturales tales como municipios, comunidades sanitarias o escuelas mantengan un cierto grado de parecido entre sí, dado que se hallan expuestas a influencias comunes. Por ejemplo, los estudiantes de una clase conversan unos con otros rutinariamente, comparten el mismo tipo de experiencias y están expuestos a las mismas contingencias educativas. Por consiguiente, como señalan Shadish, Cook y Campbell (2002), las observaciones registradas desde cada unidad reflejarán tanto los efectos que la propia individualidad tiene sobre la conducta, como los efectos que las variables colectivas ejercen sobre los individuos. Los primeros efectos variarán dentro de

las unidades colectivas y a través de las mismas, mientras que los segundos tan sólo variarán entre las distintas unidades colectivas.

Debido a que el agrupamiento induce correlación (la homogeneidad dentro del grupo implica que los errores entre los miembros del mismo estén relacionados, por lo general, positivamente) los modelos estadísticos que no contemplen la dependencia entre las unidades producirán errores estándar de los parámetros de interés negativamente sesgados, contrastes estadísticos positivamente sesgados y tasas de error infladas (Carvajal, Baumler, Harrist y Parcel, 2001; Kish, 1965; Krull y MacKinnon, 2001). En la práctica esto supone que cuando los datos estén positivamente correlacionados, tanto las pruebas basadas en el modelo lineal clásico como las basadas en el modelo lineal generalizado, incrementarán sustancialmente la probabilidad de rechazar la hipótesis nula y, por consiguiente, inducirán al investigador a concluir que su tratamiento es efectivo cuando lo cierto es que puede haber resultado completamente ineficaz. En consecuencia, la liberalidad de estas pruebas facilita que se capitalice sobre el azar con mayor frecuencia de la debida por obtener estimaciones ineficientes.

El precio a pagar cuando se utilizan pruebas estadísticas que ignoran el agrupamiento de los datos puede ser mostrado analíticamente con relativa facilidad efectuando dos pequeñas operaciones: Por un lado, elaborando algún índice que nos informe de la relación existente entre las respuestas de los sujetos que forman parte del mismo grupo. Como es bien sabido la correlación dentro del grupo y debida, entre otras posibles causas, al proceso de selección, exposición, interacción mutua o alguna combinación de éstas, queda convenientemente reflejada mediante la razón de componentes de varianzas o coeficiente de correlación dentro del grupo (r). Y, por otro lado, descubriendo la varianza correspondiente al efecto de la intervención. Para los datos del ejemplo anterior, la varianza correspondiente a la diferencia entre las dos condiciones de medias, asumiendo grupos (q) de igual tamaño (n), resultó ser (ver Anexo 1):

$$\sigma_{\Delta}^2 = 2 \left[\frac{\sigma_y^2}{nq} (1 + (n-1)\rho) \right] = \frac{2[\sigma_e^2 + n\sigma_{\beta(A)}^2]}{nq} \quad (1)$$

Dos implicaciones sumamente importantes se desprenden de la ecuación (1). En primer lugar, si los miembros que configuran los grupos son muy heterogéneos entre sí, la varianza de los tratamientos será sustancialmente más pequeña que si los miembros son muy homogéneos. De ahí que las tasas de error de Tipo I sean muy elevadas cuando se obvia la dependencia existente entre las puntuaciones. De hecho, cuando la correlación es mayor que cero, cualquier modelo matemático que ignore la varianza debida a la unidad de asignación (grupos de sujetos intactos en nuestro caso) generará tasas de error por encima del nivel de significación estipulado. Por consiguiente, cualquier prueba que asuma correlación nula, falsamente por supuesto, tendrá un comportamiento liberal y proporcionará estimaciones ineficientes. Lo afirmado puede ser comprobado fácilmente examinando los resultados analíticos recogidos en la Tabla 1. Dicha tabla ofrece la relación existente entre las tasas de error de Tipo I y las variables número de grupos por condición (2, 4, 8, 16 y 32), tamaño de los grupos (5, 10, 20 y 40) y grado de correlación existente dentro de los grupos (0.0, 0.1, 0.2, y 0.3). En segundo lugar, en presencia de correlación positiva, cualquier prueba estadística que reconozca la varianza debida a la unidad de asignación reducirá la potencia, comparado con un estudio en el cual el grado de correlación entre las observaciones dentro del grupo tienda a cero. En la Tabla 1 también se puede verificar la relación que existe entre el tamaño del efecto requerido para satisfacer una potencia de 0.80 con un nivel de significación del 5% y las variables número de grupos por condición, tamaño de los grupos y grado de correlación existente dentro de los grupos. Dichos valores han sido obtenidos utilizando el programa SAS (2001; ver códigos Anexo 2).

Además de lo dicho más arriba, tres patrones emergen de la tabla anterior: Uno, las tasas de error de Tipo I mantienen una relación monótonica creciente con las variables número de grupos, tamaño de los mismos y grado de correlación existente entre los miembros que configuran los grupos. Dos, el tamaño del efecto requerido para satisfacer una determinada potencia disminuye conforme se incrementa el número de grupos, el tamaño de los mismos y el grado de dependencia existente entre sus miembros. Tres, la precisión del diseño depende más del número de grupos que del número de sujetos dentro de los grupos. Por consiguiente, para alcanzar la potencia deseada resulta más satisfactorio

Tabla 1. Tasas de error y tamaño del efecto en unidades estándar requerido para satisfacer una potencia de 0.80

rho = 0.00			rho = 0.10			rho = 0.20			rho = 0.30						
q	n	tipol	te	q	n	tipol	te	q	n	tipol	te	q	n	tipol	te
2	5	.050	2.399	2	5	.075	2.992	2	5	.103	3.598	2	5	.136	4.252
2	10	.050	1.696	2	10	.098	2.464	2	10	.148	3.173	2	10	.202	3.898
2	20	.050	1.199	2	20	.139	2.153	2	20	.221	2.938	2	20	.299	3.710
2	40	.050	0.848	2	40	.206	1.979	2	40	.324	2.813	2	40	.419	3.612
4	5	.050	1.060	4	5	.097	1.332	4	5	.154	1.590	4	5	.217	1.880
4	10	.050	0.750	4	10	.143	1.089	4	10	.239	1.402	4	10	.328	1.724
4	20	.050	0.530	4	20	.222	0.952	4	20	.356	1.298	4	20	.459	1.640
4	40	.050	0.375	4	40	.335	0.875	4	40	.488	1.243	4	40	.587	1.597
8	5	.050	0.674	8	5	.108	0.840	8	5	.175	1.011	8	5	.246	1.194
8	10	.050	0.476	8	10	.162	0.692	8	10	.271	0.891	8	10	.367	1.095
8	20	.050	0.337	8	20	.252	0.605	8	20	.396	0.825	8	20	.499	1.042
8	40	.050	0.238	8	40	.374	0.556	8	40	.528	0.790	8	40	.622	1.015
16	5	.050	0.458	16	5	.112	0.571	16	5	.183	0.687	16	5	.258	0.812
16	10	.050	0.324	16	10	.170	0.470	16	10	.284	0.606	16	10	.381	0.744
16	20	.050	0.229	16	20	.264	0.411	16	20	.411	0.561	16	20	.514	0.708
16	40	.050	0.162	16	40	.388	0.378	16	40	.543	0.537	16	40	.635	0.690
32	5	.050	0.318	32	5	.114	0.397	32	5	.188	0.477	32	5	.264	0.564
32	10	.050	0.225	32	10	.174	0.327	32	10	.289	0.421	32	10	.388	0.517
32	20	.050	0.159	32	20	.270	0.286	32	20	.428	0.390	32	20	.521	0.492
32	40	.050	0.113	32	40	.395	0.263	32	40	.549	0.373	32	40	.645	0.479

Nota : $\sigma_e^2 = 1$; rho = correlación dentro de los grupos; q = número de grupos por condición de tratamiento; n = número de sujetos dentro de cada grupo; tipo1 = tasas de error; te = tamaño del efecto estandarizado.

incrementar el número de unidades colectivas que el tamaño de las mismas, aunque esta no deja de ser una cuestión delicada si nos atenemos a los costos del diseño. Por regla general, el muestreo de los grupos resultará más costoso que el muestreo de sus miembros.

EL PROBLEMA DEL SESGO DE AGREGACIÓN

La utilización de unidades de investigación colectivas conduce de una manera casi inevitable al denominado sesgo de agregación. Dicho sesgo se presenta cuando se obtienen resultados diferentes para una misma variable, dependiendo de cual sea nivel de análisis adoptado. Por ejemplo, Raudenbush y Bryk (2002) señalan que en la investigación educacional el promedio de la variable clase social medido en el nivel de la escuela puede tener un efecto superior, e inclusive más duradero, que el efecto de la clase social medido en el nivel del estudiante. Así pues, la elección de distintos niveles de análisis nos puede llevar a obtener conclusiones diferentes acerca de los efectos de una variable particular, a no ser que la relación entre la variable dependiente y la variable independiente fuese la misma dentro y a través de las unidades de análisis. En el contexto longitudinal dicha confusión no se presentaría cuando las observaciones dentro y a través de los sujetos fuesen intercambiables. Esto es, cuando la relación entre la variable dependiente y la variable independiente para un determinado sujeto en dos puntos cualesquiera del tiempo, no difiriese de la relación obtenida por dos sujetos diferentes en los mismos puntos.

Los supuestos que acabamos de exponer, aunque no resulten imposibles de cumplir, sí son poco probables de satisfacer en la mayor parte de los ámbitos de investigación. Por lo tanto, de utilizarse un análisis de la regresión de un solo nivel, surgirían problemas interpretativos, pues los coeficientes de regresión obtenidos serían diferentes dependiendo de que el análisis se efectuase en el nivel superior utilizando medidas de respuesta del grupo, o en el nivel inferior utilizando medidas de respuesta individual. En concreto, limitándonos al ejemplo de Raudenbush y Brik, las inferencias que se efectúen de los estudiantes a partir del análisis realizado en el nivel escuela diferirán de las inferencias que se realicen de los mismos a partir del análisis realizado en el nivel estudiante y, viceversa. Dado que los resultados

obtenidos variarán dependiendo de cual sea la unidad de análisis utilizada, la pregunta que surge de inmediato es la siguiente: ¿resulta apropiado ignorar la variabilidad individual agregando las medidas dentro de los grupos para garantizar la independencia de las observaciones? A pesar que Murray (1998) considera al análisis de medias como una estrategia adecuada para este tipo de datos, pensamos que esta práctica constituye más un problema que una solución; pues, además de reducir la potencia de las pruebas estadísticas (sensibilidad para captar las diferencias existentes entre los grupos que conforman las condiciones de investigación), olvida que la interdependencia entre los individuos tiene interés por derecho propio.

Los modelos mixtos con términos aleatorios anidados, además de permitirnos manejar diseños con desgaste de muestra, también resuelven los problemas expuestos en los apartados anteriores. Mediante estos modelos el problema de la unidad de análisis se maneja computando los componentes de varianza correspondientes a las unidades de asignación y de observación. A su vez, la confusión derivada del sesgo de agregación se maneja descomponiendo la relación observada entre variables, tales como el logro académico y la clase social, dentro de componentes separados de primer y segundo nivel. Mediante esta técnica ambos problemas son tratados formulando modelos e hipótesis para procesos y relaciones dentro y a través de los diferentes niveles. Apuntar, finalmente, que este enfoque capacita al investigador para formalizar la estructura anidada de sus datos de una manera compacta. El modelo resultante de combinar los diferentes modelos expresa las relaciones entre las variables propias de ese nivel y específica de qué modo las variables de un nivel se relacionan con las relaciones que acontecen en otro nivel diferente. Al combinar la información obtenida a través de los niveles, la técnica permite examinar simultáneamente los efectos de las variables en cada uno de los niveles implicados, así como la posible interacción entre las variables pertenecientes a niveles que difieren entre sí.

A MODO DE CONCLUSIÓN

Cuando se trabaja con datos organizados jerárquicamente, lo usual es que se incumpla el supuesto de independencia, que existan múltiples

fuentes de variación aleatoria y que acontezcan problemas interpretativos derivados de la frecuente agregación de los datos. En estos casos, ni las técnicas basadas en el modelo lineal general, ni en el modelo lineal generalizado resultan apropiadas. Pues, en todas ellas tan sólo se permite captar la variación de un único componente. Si los datos siguen una distribución normal, la solución natural la suministra el modelo mixto lineal general (ver, Harville, 1977; Henderson, 1975); mientras que si los datos siguen cualquier otro miembro de la familia exponencial, la solución la natural la proporciona modelo mixto lineal generalizado (ver, Breslow y Clayton, 1993; Wolfinger y O'Connell, 1993). El uso de estos modelos, además de proporcionar una solución adecuada a los problemas planteados, también reduce los efectos del azar asociados con la realización de pruebas múltiples (Rindskopf y Saxe, 1998). Cuando se emplean las técnicas analíticas convencionales se intenta examinar cada nivel separadamente, lo que ocasiona que se efectúe un número considerable de pruebas. Mediante el uso métodos basados en el enfoque del modelo mixto, el número de pruebas estadísticas para cada variable dependiente se reduce básicamente a dos pruebas globales, a saber: una para los efectos fijos del modelo y otra para los efectos aleatorios del mismo.

En definitiva, cuando las intervenciones se aplican a través de múltiples unidades de agrupamiento se requiere un enfoque analítico distinto del empleado tradicionalmente para contrastar las hipótesis de interés. Como se ha puesto de relieve en un trabajo previo (Vallejo, Fernández y Secades, 2003), si se cuenta con tamaños de muestra moderados de cada uno de los niveles implicados, los modelos multinivel, un caso especial del modelo mixto lineal general (Mass y Snijders, 2003; Cnaan, Laird y Slasor, 1997), desarrollados gracias al esfuerzo de numerosos investigadores por preservar la estructura jerárquica de los datos, proporcionan un conjunto de herramientas potentes y flexibles. Estos modelos no sólo permiten efectuar estimaciones más precisas que las obtenidas con las técnicas tradicionales, en especial cuando existe desgaste de muestra, sino también desarrollar modelos conceptualmente más adecuados para dar cuenta de las interrelaciones que se dan entre los individuos y entre éstos y el medio ambiente, tanto en contextos transversales como longitudinales.

REFERENCIAS BIBLIOGRÁFICAS

- Ato, M. (2002). *El modelo Lineal General y sus Aplicaciones*. Documento mimeografiado. Universidad de Murcia.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9-25.
- Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, *16*, 2349-2380.
- Carvajal, S. C., Baumler, E., Harrist, R. B., & Parcel, G. S. (2001). Multinivel models and unbiased tests for group based interventions: Examples from the safer choices study. *Multivariate Behavioral Research*, *36*(2), 185-205.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*, 320-338.
- Henderson, C. R. (1975). The best linear unbiased estimation and prediction under a selection model. *Biometrics*, *31*, 423-447.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.
- Kull, L. J., & MacKinnon, D. P. (2001). Multinivel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*, 249-277.
- Maas, C. J. M., & Snijders, T. A. B. (2003). The multinivel approach to repeated measures for complete and incomplete data. *Quality and Quantity*, *37*, 71-89.
- Murray, D. M. (1998). *Design and Analysis of Group-Randomized Trials*. New York: Oxford University Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and Data*. Thousand Oaks, CA: SAGE Publications, Inc.
- Rindskopf, D., & Saxe, L., (1998). Zero effects in substance abuse programs: Avoiding false positives and false negatives in the evaluation of community-based programs. *Evaluation Review*, *22*, 78-94.
- SAS Institute Inc. (2001). *SAS/STAT Software: Version 8.2 (TS M0)*. Cary, NC: SAS Institute Inc. Shadish, W. R., Cook, T. D., & Campbell,

- D. J. (2002). *Experimental and QuasiExperimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Vallejo, G., Fernández, J. R. y Secades, R. (2003). Análisis estadístico y consideraciones de potencia en la evaluación de programas de muestreo de dos etapas. *Psicothema*, 15, 300-308.
- Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48, 233-243.

ANEXO 1

En un diseño entre grupos al azar la varianza de y_{ij} es igual a σ_ϵ^2 , mientras que en un diseño jerárquico de grupos al azar tenemos

$$\text{Var}(y_{ijk}) = \sigma_\epsilon^2 + \sigma_{B(A)}^2$$

$$\text{Cov}(y_{ijk}, Y'_{i'jk}) = \sigma_{B(A)}^2$$

donde $\sigma_{B(A)}^2$ es el componente de varianza atribuible a la unidad de asignación o de agrupamiento y $\sigma_\epsilon^2 = \sigma_y^2 - \sigma_{B(A)}^2$ (también $\sigma_\epsilon^2 = \sigma_y^2 (1-\rho)$) el componente de varianza atribuible a la unidad de observación. Siendo la correlación existente entre las respuestas de dos sujetos de un mismo grupo

$$\rho = \frac{\sigma_{B(A)}^2}{\sigma_\epsilon^2 + \sigma_{B(A)}^2}$$

la porción de varianza total atribuible a la unidad de asignación o de agregación. A su vez, la varianza de la media del grupo del diseño jerárquico de grupos alzar viene dada por:

$$\sigma_{\bar{y}_B}^2 = \frac{\sigma_\epsilon^2}{n} + \sigma_{B(A)}^2 = \frac{\sigma_\epsilon^2 + n\sigma_{B(A)}^2}{n}$$

$$\sigma_{\bar{y}_B}^2 = \frac{\sigma_\epsilon^2 + \sigma_{B(A)}^2 + n\sigma_{B(A)}^2 - \sigma_{B(A)}^2}{n}$$

$$\sigma_{\bar{y}_B}^2 = \frac{\sigma_\epsilon^2 + \sigma_{B(A)}^2 + (n-1)\sigma_{B(A)}^2}{n}$$

$$\sigma_{\bar{y}_B}^2 = \frac{(\sigma_\epsilon^2 + \sigma_{B(A)}^2) + \frac{(\sigma_\epsilon^2 + \sigma_{B(A)}^2)(n-1)\sigma_{B(A)}^2}{\sigma_\epsilon^2 + \sigma_{B(A)}^2}}{n}$$

$$\sigma_{\bar{y}_B}^2 = \frac{(\sigma_e^2 + \sigma_{\beta(A)}^2) \left(1 + (n-1) \frac{\sigma_{\beta(A)}^2}{\sigma_e^2 + \sigma_{\beta(A)}^2} \right)}{n}$$

$$\sigma_{\bar{y}_B}^2 = \frac{\sigma_y^2}{n} (1 + (n-1)\rho)$$

dividiendo la expresión anterior por el número de grupos asignados a la condición tratamiento se tiene que la varianza de la media de dicha condición es

$$\sigma_{\bar{y}_A}^2 = \frac{\sigma_y^2}{nq} (1 + (n-1)\rho) = \frac{\sigma_e^2 + n\sigma_{\beta(A)}^2}{nq}$$

y la varianza correspondiente al efecto de la intervención, en nuestro caso la diferencia entre dos condiciones de medias, vendrá dado por:

$$\sigma_{\Delta}^2 = 2 \left[\frac{\sigma_y^2}{nq} (1 + (n-1)\rho) \right] = \frac{2[\sigma_e^2 + n\sigma_{\beta(A)}^2]}{nq}$$

ANEXO 2

Códigos programa SAS;

data;

options nodate;

filename Tabla1 'a:\Tabla1.txt';

filename Tabla2 'a:\Tabla2.txt';

filename Tabla3 'a:\Tabla3.txt';

array grupos{4}(2,4,8,16);

array muestra{4}(5,10,20,40);

array rho{4}(0.0 0.1,0.2,0.3);

alpha=0.05; tipo2=.20; sigmae=1;

do g=1 to 4;

do i=1 to 4;

do j=1 to 4;

c1=(1+(muestra{i}-1)*rho{j});

c2=(1-rho{j});

t1=(1-probf(finv(1-alpha,1,2*(grupos{g}-1))*c2/c1,1,2*(grupos{g}-1),0));

var=(2*(((sigmae/(1-rho{j}))/muestra{i}*grupos{g}))*c1);

te=(sqrt(var*(tinv(1-alpha/2,2*(grupos{g}-1))+tinv(1-tipo2,2*(grupos{g}-1))))**2);

po=1-probt(tinv(1-alpha/2,2*(grupos{g}-1)),2*(grupos{g}-1),te/sqrt(var));

file Tabla1;put@1 grupos{g}2.0@8muestra{i}3.2@16rho{j}5.2@26 t1 5.3;

file Tabla2;put@1 grupos{g}2.0@8muestra{i}3.2@16rho{j}5.2@26 te 5.3;

file Tabla3;put@1 grupos{g}2.0@8muestra{i}3.2@16rho{j}5.2@26 po 5.3;

end;

end;

end;

run;

proc print;