

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

Análisis estadístico y consideraciones de potencia en la evaluación de programas mediante diseños de muestreo de dos etapas

Guillermo Vallejo Seco, José Ramón Fernández Hermida y Roberto Secades Villa
Universidad de Oviedo

La evaluación de programas de prevención acarrea errores de decisión derivados principalmente de la dificultad de asignar de manera aleatoria las unidades individuales a las condiciones de investigación. La elección de la unidad de análisis apropiada a la hora de evaluar la efectividad del impacto está determinada por la naturaleza de la intervención y por el diseño de investigación seleccionado. Cuando las unidades de asignación y de observación difieren entre sí, esto es, cuando entidades colectivas más que individuales son asignadas al azar a los tratamientos, los análisis realizados en los niveles más bajos de la jerarquía proporcionan estimaciones ineficientes de los parámetros y a menudo conducen a que las pruebas de significación sean inadecuadas. La meta de este trabajo es doble. Por un lado, presentar un método analítico que permite utilizar los datos de cualquier nivel del diseño sin inflar las tasas de error. Y, por otro lado, determinar el número de grupos y el tamaño de éstos en función de la variabilidad existente y de los costos.

Statistical analysis and considerations of power in the programs evaluation through two-stage sampling designs. Evaluation of prevention programs involves decision errors due to the difficulty of randomly assigning individuals to research conditions. The nature of the intervention and design of the study determine the choice of the appropriate unit of analysis in impact assessments. When units of assignment and units of observation differ, that is, when clusters of people rather individuals are assigned at random to treatments, the analyses conducted at lower levels of the study hierarchy provide inefficient parameter estimates, and often result in inappropriate significance tests. Therefore, the purpose of this paper is (a) to present an analytical method that permits the use of data at all levels of design without increasing Type I error rates, and (b) to determine the number of clusters and the sample size per group according to variability and cost.

Los datos provenientes de muchos diseños de investigación basados en la evaluación de programas, tienen una estructura que resulta muy similar a que se observa en los diseños de muestreo por conglomerados de dos o más etapas. En el caso más simple, el investigador selecciona inicialmente una muestra aleatoria de conglomerados o unidades primarias de muestreo y posteriormente elige al azar las unidades secundarias de muestreo dentro de cada uno de los conglomerados. Las unidades primarias pueden ser escuelas, clases, clínicas o cualquier otra entidad y las unidades secundarias maestros anidados dentro de escuelas, estudiantes anidados dentro de clases o pacientes anidados dentro de clínicas.

En todos los casos reseñados, unidades de análisis colectivas, más que unidades de análisis individuales, constituyen el referente observacional al que va dirigido el tratamiento o programa intervención social. 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 aleatoriamente unas unidades configuradas con antelación a la intervención a la condición de tratamiento y otras a la condición de control. Raramente un investigador que ha procedido conforme a lo expuesto, seleccionará aleatoriamente unidades individuales de las unidades primarias de muestreo para enviarlas después conforme al azar al programa. De todos modos, aunque fuese posible asignar personas dentro de los agrupamientos al programa no parece una opción muy deseable debido, entre otras cosas, a la probable difusión de los tratamientos. Más frecuente resulta que el investigador cree el problema de la unidad de análisis, administrando el tratamiento colectivamente a unidades asignadas en un principio a los grupos de forma individual. Por

tanto, uno debe de ser muy recatado con estas cuestiones. Obviamente, no es lo mismo asignar al azar unidades colectivas al programa de prevención que hacerlo individualmente. Tal confusión, no sólo limita la habilidad del investigador para comprender el diseño de investigación utilizado, sino que también puede invalidar la utilización de las técnicas de análisis basadas en el modelo lineal general o generalizado para evaluar las consecuencias derivadas de la aplicación del programa.

La aplicación de las técnicas basadas en los modelos reseñados requiere satisfacer ciertos supuestos, especialmente, el que alude a la independencia entre las observaciones. 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 altamente 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. Un indicador de la porción de variabilidad total atribuible a la unidad de asignación se obtiene por medio del coeficiente de correlación intraclase.

Debido a que los miembros de una unidad colectiva tienden a ser más similares que los miembros que no lo son, un conjunto de observaciones correlacionadas proporciona menos información que un número similar de observaciones independientes. Por esta razón, cuando los modelos estadísticos que asumen independencia entre las unidades son aplicados a datos que están correlacionados, se produce una infraestimación de los errores típicos de medida (Carvajal, Baumler, Harrist y Parcel, 2001). En la práctica, esto supone que tanto las pruebas basadas en el modelo lineal clásico, como las basadas en el modelo lineal generalizado incrementen sustancialmente la probabilidad de rechazar la hipótesis nula cuando de hecho es verdadera y, por ende, inducirnos a concluir que un programa es efectivo cuando resulta completamente ineficaz. En definitiva, conduce a que el investigador capitalice sobre el azar con mayor frecuencia de la debida por obtener estimaciones ineficientes (Rinndskopf y Saxe, 1998).

Además del incumplimiento del supuesto de independencia, cuando los datos están organizados jerárquicamente existe más de una fuente de variación aleatoria en los mismos. Por lo tanto, ni las técnicas basadas en el modelo lineal general, ni en el modelo lineal generalizado son 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 proporcionan los modelos lineales mixtos, también conocidos como modelos de efectos aleatorios, modelos de componentes aleatorios, modelos lineales multinivel, modelos jerárquicos o modelos de la regresión de efectos mixtos (Goldstein, 1995, Aitkin y Longford, 1986; Laird y Ware, 1982; Raudenbush y Bryk, 2002, Oliver, Rosel y Jara, 2000).

Mientras que si los datos siguen cualquier otro miembro de la familia exponencial la solución natural la proporcionan los modelos mixtos generalizados (Breslow y Clayton, 1993; Wolfinger y O'Connell, 1993). Todos estos modelos reconocen la estructura anidada de los datos y permiten estimar las variaciones que se producen en los distintos estratos producidos por el agrupamiento, tanto en estudios de carácter transversal como de carácter longitudinal.

En las secciones que siguen se proporciona una breve introducción al modelo mixto lineal general, se describen las técnicas de estimación y se especifican los procedimientos de inferencia para contrastar las hipótesis correspondientes a los efectos fijos del modelo, a los efectos aleatorios y a los componentes de varianza. Finalmente, utilizaremos un diseño jerárquico de grupos aleatorizados con pretest y postest para ilustrar como determinar tamaños de muestra óptimos para determinar los efectos del diseño. Pues como señala Raudenbush (1997), estos modelos tienen un enorme atractivo, pero suelen despertar algún que otro recelo entre los investigadores debido a su relativa complejidad analítica y a su posible falta de precisión estadística por una incorrecta elección del tamaño de las unidades de muestreo.

El modelo mixto lineal general

El modelo lineal estándar para explicar n observaciones tomadas para cada una de las p covariantes (predictores) y/o factores (variables independientes) puede ser escrito como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (1)$$

donde \mathbf{y} es un vector de dimensión $n \times 1$ conteniendo los valores de la variable de respuesta y_{ij} para la unidad i en el grupo j (o bien para el sujeto i en el tiempo j), \mathbf{X} es una matriz de diseño de dimensión $n \times p$ que especifica los valores de efectos fijos que corresponden a cada parámetro para cada una de las observaciones (vectores de ceros y unos denotan la ausencia y presencia de efectos categoriales para variables carentes de una estructura métrica, mientras que vectores numéricos denotan los efectos de las variables mediadas en una escala cuantitativa), $\boldsymbol{\beta}$ es un vector de parámetros no aleatorios estimado desde los datos que puede incluir variables de diversa naturaleza y \mathbf{e} es vector de errores desconocidos de dimensión $n \times 1$ distribuidos normal e independientemente con media cero y varianza constante. Los coeficientes del vector $\boldsymbol{\beta}$ son parámetros de efectos fijos que describen el comportamiento promedio de la población. No obstante, puede ocurrir que no todos los términos del modelo tomen valores constantes en las sucesivas repeticiones de la investigación, sino que algunos puedan ser vistos como el resultado de extraer muestras al azar de una distribución normal (modelos de efectos aleatorios). Además, si la matriz de diseño incluye covariantes, también es posible que los parámetros del vector $\boldsymbol{\beta}$ no representen correctamente la relación entre \mathbf{X} e \mathbf{y} para algunos sujetos o grupos. Por consiguiente, para tener en cuenta lo dicho se requiere disponer de algún enfoque que permita establecer una relación global entre las variables para todos los sujetos y modelar relaciones separadas que varíen aleatoriamente entre los sujetos (modelos de coeficientes aleatorios).

El modelo lineal mixto proporciona la solución adecuada para abordar la problemática reseñada, pues no requiere asumir que la selección de los niveles de las variables tenga que ser efectuada de una manera arbitraria, ni que todos los coeficientes del modelo sean constantes fijas. Además, el enfoque del modelo mixto, tam-

bién extiende el modelo lineal general al permitir una especificación más flexible de la matriz de la matriz de covarianza de \mathbf{e} . En concreto, relaja los supuestos de homogeneidad de las varianzas e independencia de los errores. Usando la notación matricial el modelo mixto es representado como sigue:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{2}$$

donde el componente de efectos fijos, $\mathbf{X}\boldsymbol{\beta}$, se define igual que en la ecuación anterior y el componente de efectos aleatorios, $\mathbf{Z}\mathbf{u}$, permite definir diferentes relaciones entre las unidades o entre los sujetos en el caso longitudinal. \mathbf{Z}_j es una matriz de diseño de dimensión $n_j \times k$ para una determinada unidad de segundo nivel o para un sujeto individual en el caso longitudinal (siendo n_j el número de unidades de primer nivel que se hallan anidadas dentro de cada unidad de segundo nivel, o el número de veces que un sujeto es observado y k el número de predictores incluidos). Para muchos modelos, los k predictores son un subconjunto de los p predictores incluidos en la matriz \mathbf{X} , y el subconjunto de predictores es el mismo o similar para cada unidad o para cada sujeto. \mathbf{Z} es una segunda matriz de diseño diagonal de bloques de orden $n_j \times Jk$ para el componente aleatorio (siendo J el número de unidades de nivel 2, o el número de sujetos). \mathbf{u}_j es un vector de parámetros de efectos aleatorios de dimensión $k \times 1$ asociado con \mathbf{Z}_j . Los vectores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_j$ se encuentran reunidos dentro de \mathbf{u} , un vector de dimensión $Jk \times 1$ que contiene los parámetros de efectos aleatorios específicos para todas las unidades, o para todos los sujetos en el caso longitudinal. Por último, \mathbf{e} es un vector de parámetros desconocidos de dimensión $n \times 1$, cuyos elementos, a diferencia del modelo clásico, no necesitan ser independientes ni homogéneos.

Los supuestos distribucionales del modelo implican que los coeficientes residuales \mathbf{u} se distribuyen normal e independientemente con media $\mathbf{0}$ y matriz de covarianza \mathbf{G} . Donde \mathbf{G} es una matriz diagonal de bloques de dimensión $Jk \times Jk$, con cada bloque \mathbf{G}_j de dimensión $k \times k$ conteniendo las varianzas y covarianzas de los efectos aleatorios para cada una de las unidades de segundo nivel, o para cada uno de los sujetos de contar con medidas repetidas. También se asume que el vector de errores \mathbf{e} , además de ser independiente del vector \mathbf{u} , se distribuye normalmente con media $\mathbf{0}$ y matriz de covarianza \mathbf{R} . Donde \mathbf{R} es una matriz diagonal a bloques de orden $n \times n$, con cada bloque \mathbf{R}_i conteniendo las varianzas y covarianzas de los errores dentro de los sujetos. Si se cumple que $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ y $cov(\mathbf{r}, \mathbf{u}) = \mathbf{0}$, entonces

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}(\theta)) \tag{3}$$

donde $\mathbf{V}(\theta) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ y θ se refiere a los componentes de varianza de la matriz \mathbf{V} . El enfoque del modelo clásico es un caso particular del enfoque del modelo mixto. Cuando $\mathbf{R} = \sigma^2 \mathbf{I}$ y $\mathbf{Z} = \mathbf{0}$ ambos enfoques son plenamente coincidentes.

Estimación de los parámetros $\boldsymbol{\beta}$, \mathbf{u} y $\mathbf{V}(\theta)$ del modelo mixto

El procedimiento estándar para obtener estimaciones de $\boldsymbol{\beta}$ y \mathbf{u} , asumiendo que las matrices \mathbf{G} y \mathbf{R} resultan conocidas, consiste en resolver las ecuaciones del conocido modelo mixto de Henderson

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix} \tag{4}$$

Aplicando las identidades

$$\begin{aligned} \hat{\mathbf{R}}^{-1} - \hat{\mathbf{R}}^{-1}\mathbf{Z}(\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{R}}^{-1} &= (\mathbf{Z}\hat{\mathbf{G}}^{-1}\mathbf{Z}' + \hat{\mathbf{R}})^{-1} \\ (\hat{\mathbf{G}}^{-1} + \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{R}}^{-1} &= \hat{\mathbf{G}}\mathbf{V}(\theta)^{-1} \end{aligned} \tag{5}$$

los estimadores de $\boldsymbol{\beta}$ y \mathbf{u} que resuelven las ecuaciones del modelo mixto de Henderson son

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= [\mathbf{X}'\mathbf{V}(\theta)^{-1}\mathbf{X}]^{-1}\mathbf{X}'\hat{\mathbf{V}}(\theta)^{-1}\mathbf{y} \\ \hat{\mathbf{u}} &= \mathbf{G}\mathbf{Z}'\hat{\mathbf{V}}(\theta)^{-1}[\mathbf{y} - \mathbf{X}'\hat{\boldsymbol{\beta}}(\theta)] \end{aligned} \tag{6}$$

Tanto $\hat{\boldsymbol{\beta}}$, como $\hat{\mathbf{u}}$ tienen la propiedad de ser el mejor estimador lineal insesgado y el mejor predictor lineal insesgado (también referido como estimador empírico de Bayes o estimador encogido), respectivamente, de $\boldsymbol{\beta}$ y \mathbf{u} (McCulloch y Searle, 2001). A su vez, las varianzas de $\hat{\boldsymbol{\beta}}$ y $\hat{\mathbf{u}}$ son obtenidas como

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\hat{\mathbf{V}}(\theta)^{-1}\mathbf{X})^{-1} \\ \text{var}(\hat{\mathbf{u}}) &= \mathbf{G} - \mathbf{G}\mathbf{Z}'\mathbf{P}\mathbf{Z}\mathbf{G} \end{aligned} \tag{7}$$

donde la matriz de proyección

$$\mathbf{P} = \hat{\mathbf{V}}(\theta)^{-1} - \hat{\mathbf{V}}(\theta)^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}(\theta)^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}(\theta)^{-1}$$

Si el tamaño de muestra es reducido y el diseño no está perfectamente equilibrado los errores estándar correspondientes a los efectos fijos y aleatorios pueden resultar negativamente sesgados. Para ajustar el sesgo varias soluciones han sido sugeridas. Por ejemplo, Kenward y Roger (1997) han propuesto calcular el factor específico de inflación y ajustar los grados de libertad. Alternativamente, Liang y Zeger (1986) sugieren computar los errores estándar para los parámetros de efectos fijos usando un estimador consistente asintóticamente, conocido como estimador de la varianza empírica o estimador sandwich. En esencia, el estimador sandwich implica utilizar el patrón de covarianza observado de los datos, en lugar de un patrón de covarianza seleccionado estadísticamente. Ambos procedimientos se encuentran disponibles en el módulo PROC MIXED del programa SAS (2001, SAS Institute, versión 8.2).

El símbolo $(\cdot)^{-}$ usado en la expresión $[\mathbf{X}'\hat{\mathbf{V}}(\theta)\mathbf{X}]^{-}$ de la ecuación referida a los efectos fijos indica que una inversa generalizada se requiere si \mathbf{X} no tiene rango completo. El vector θ contiene los elementos únicos de \mathbf{G} y los parámetros en \mathbf{R} . De lo expuesto se aprecia rápidamente que estimar un modelo multinivel equivale a estimar un modelo mixto o combinado. Pues, aunque se pueden formular modelos separados para cada nivel, dichos modelos están conectados estadísticamente. Observando las ecuaciones referidas a los estimadores $\boldsymbol{\beta}$ y \mathbf{u} , se aprecia que la estimación del vector de efectos fijos depende de la matriz de componentes de varianza, mientras que la estimación del vector de efectos aleatorios depende tanto de la matriz $\mathbf{V}(\theta)$, como del estimador de mínimos cuadrados generalizados $\hat{\boldsymbol{\beta}}$.

Salvo contadas excepciones, las matrices **G** y **R** son desconocidas, lo que nos obligará a determinar los componentes de varianza de **V(θ)** a partir de los datos mediante alguno de los diferentes procedimientos de estimación existentes. Si el diseño de investigación está equilibrado, se pueden utilizar procedimientos algebraicos basados en el método de los momentos (Searle, Casella y McCulloch, 1992). Sin embargo, el tradicional método de los momentos consistente en resolver sistemas de ecuaciones simultáneas relacionando los valores esperados con los observados, tiene difícil acomodo en el contexto de la evaluación de programas, debido a que en los ámbitos aplicados las unidades de muestreo suelen estar anidadas en grupos no equilibrados con matrices de dispersión parametrizadas arbitrariamente. Cuando el diseño de investigación está desequilibrado, los componentes de la matriz **V(θ)** se estiman iterativamente mediante procedimientos numéricos. Por regla general, estos procedimientos están basados en técnicas de estimación de máxima verosimilitud (MV), o de máxima verosimilitud restringida (MVR) para evitar obtener estimaciones sesgadas. Otro procedimiento disponible para estimar los elementos de la matriz **V(θ)** se basa en el enfoque Bayesiano. No obstante, van der Leeden (1998), pone de relieve que el esfuerzo computacional exigido por este procedimiento puede ser considerable cuando los modelos son complejos y los tamaños de muestra de los niveles elevados. Además de los procedimientos reseñados, PROC MIXED tiene incorporados otros métodos.

Los estimadores MV de θ son obtenidos maximizando el logaritmo natural de la función de verosimilitud correspondiente a la densidad del vector **y** para **β** y θ , donde

$$l_c(\beta, \theta) = \frac{1}{2} \left[(n \log(2\pi) + \log |V(\theta)| + e' V(\theta)^{-1} e) \right] \tag{8}$$

con $e = y - X\beta$. Si n es pequeño, más que estimar los componentes de la varianza desde la verosimilitud global, puede interesar maximizar la parte de la verosimilitud que es invariante de los efectos fijos del modelo mediante el método MVR. En concreto, de acuerdo con derivaciones efectuadas por Harville (1977), maximizando el logaritmo de la función de verosimilitud

$$l_r(\beta, \theta) = \frac{1}{2} \left[(n-h) \log(2\pi) + \log |V(\theta)| + \log |X' V(\theta)^{-1} X| + e' V(\theta)^{-1} e \right] \tag{9}$$

Bajo el modelo normal, los estimadores MV o MVR de **β** y θ son usualmente determinadas mediante el algoritmo de Newton-Raphson (NR) o el algoritmo Expectación-Maximización (EM) descrito por Dempster, Laird y Rubin (1977). No obstante, existen algunas razones para preferir el algoritmo NR al EM. De acuerdo con Lindstrom y Bates (1988), el algoritmo NR requerir un menor número de iteraciones para converger que el algoritmo EM. Otra ventaja del algoritmo NR sobre el EM, reside en que el primero permite computar los errores estándar de los elementos de θ desde la matriz de información de empírica (inversa de la matriz hessiana cambiada de signo). Cuando se utiliza el algoritmo EM no se calcula dicha matriz y, por ende, los errores estándar para los elementos de θ (Jennrich y Schluchter, 1986). Detalles de las derivadas matriciales del algoritmo NR implementado en el módulo PROC MIXED se encuentran disponibles en el trabajo de Wolfinger, Tobias y Sall (1994).

Contraste de hipótesis en el modelo mixto

Los procedimientos que se ofrecen en este apartado para probar los efectos fijos, aleatorios y los componentes de varianza son generales. Para contrastar los efectos fijos y los efectos aleatorios SAS utiliza pruebas estadísticas basadas en las distribuciones *F* o *t*, mientras que para probar los componentes de varianza SAS utiliza el estadístico *Z* de Wald. Una vez que la matriz de dispersión ha sido identificada y sus parámetros convenientemente estimados, se procede a estimar **β** y **u** mediante los procedimientos definidos anteriormente y a probar que las *h* funciones estimables tienen el valor especificado en las hipótesis nulas que siguen: $H_0: L'\beta = 0$ y $H_0: L'u = 0$. Por ejemplo, los diferentes coeficientes correspondientes a los efectos fijos y aleatorios del modelo son probados calculando la razón entre los estimadores MV o (MVR) y sus respectivos errores estándar como sigue:

$$t(\hat{\beta}) = \frac{L'\hat{\beta}}{\sqrt{L'[X'V(\hat{\theta})^{-1}X]^{-1}L}} \text{ y } t(\hat{u}) = \frac{L'\hat{u}}{\sqrt{L'(G-GZ'PZG)^{-1}L}} \tag{10}$$

Cada una de las hipótesis referidas a los efectos fijos y aleatorios del modelo son rechazada al nivel α si $t > t_{(1-\alpha/2),v}$, donde $t_{(1-\alpha/2),v}$ es el 100(1- α /2)-ésimo percentil de la distribución *t* con *n* grados de libertad. Cuando los datos no están equilibrados, como suele ser usualmente el caso, y el número de unidades de nivel 2 es pequeño, la anterior prueba puede ofrecer resultados liberales. Para tratar de solventar este problema SAS ofrece la posibilidad de hacer uso de las opciones DDFM= SATTERTH o DDFM= KENWARDROGER para ajustar *n*.

A su vez, las hipótesis nulas correspondientes a los componentes de varianza (**R** y **G**) son de la forma $H_0: \theta = 0$. Para probar este tipo de hipótesis SAS proporciona el estadístico *Z* de Wald. Dicho estadístico se obtiene dividiendo cada uno de los parámetros estimados vía MV (o MVR) por su correspondiente error estándar:

$$Z = \frac{\hat{\theta}}{\sqrt{[I(\theta)]^{-1}}}; Z = \frac{\hat{\theta}}{\sqrt{[I(\theta)]^{-1}}} \tag{11}$$

donde $[I(\theta)]^{-1}$ es la matriz de covarianzas asintótica (inversa de la matriz de información) correspondiente a la solución MV (o MVR). El estadístico de Wald, computado por defecto por SAS, únicamente es exacto asintóticamente (Wolfinger, 1996), esto es, para muestras relativamente grandes. En consecuencia, cuando el número de unidades de segundo nivel sea reducido, se debe tener cierto recato a la hora de interpretar los resultados.

Resaltar, finalmente, que cuando se analizan datos longitudinales registrados regular y consistentemente de un reducido número sujetos, resulta usual fijar la matriz $Z = 0$ y modelar el patrón de covarianza que sigue la matriz **R**. Obviamente, esto también dependerá de si el investigador está interesado en los interceptores y tendencias globales (efectos fijos) o en los individuales (efectos aleatorios). De estar interesado en los efectos fijos, más que probar un simple parámetro de covarianza, lo que procede es verificar si un determinado patrón de covarianza causa una significativa mejora en relación con otro patrón. Cuando se utiliza PROC MIXED para discriminar entre modelos anidados sin cambios en los efectos fijos, procede utilizar el contraste de razón de verosimili-

tudes residuales; mientras que si se desea comparar modelos anidados entre sí con diferentes efectos fijos, lo correcto es utilizar el contraste de razón de verosimilitudes completas (ver Singer, 2002, para una justificación detallada de este tópico). Para modelos que no están anidados, criterios de selección, tales como el criterio de información de Akaike (AIC) o el criterio de información bayesiano de Schwarz (BIC), han sido usualmente adoptados (Wallace y Green, 2002). No obstante, también se puede seguir utilizando la prueba de la razón de verosimilitud si se compara cada uno de los modelos de interés con uno más sencillo, pero que se encuentre anidado simultáneamente dentro de ambos y seleccionar el modelo que ofrezca la mejora más importante (Brown y Prescott, 1999).

Análisis de potencia para detectar los efectos de un diseño multinivel

En la actualidad cada día son más los investigadores que consideran de utilidad conocer la probabilidad que tiene un impacto de un determinado tamaño de resultar significativo estadísticamente. Por ende, los métodos destinados a detectar tamaños de muestra óptimos, diferencias entre los tratamientos de interés y potencia de prueba (en adelante, análisis de potencia) deberían ser parte esencial del proceso de planificación de una investigación. Sin embargo, la creencia relativamente extendida de que la información requerida para efectuar el análisis de potencia es difícil de obtener y que los trabajos de mera conjetura constituyen un ejercicio de irresponsabilidad, han contribuido a que el análisis de potencia sea frecuentemente olvidado.

Por este motivo, no resulta extraño toparse con estudios que incluyen inadecuados tamaños de muestra, por regla general, un número de participantes insuficiente. Como resulta conocido, la inclusión de escasos tamaños de muestra lleva implícito el riesgo de aceptar más veces de las debidas la hipótesis de nulidad, cuando de hecho es falsa en la población. En estos casos se dice que el diseño adolece de potencia probatoria, ya que existe una probabilidad elevada de aceptar el modelo de azar como explicación más plausible de las diferencias encontradas. Por lo tanto, una cuestión clave a la hora de diseñar una investigación es elegir correctamente el tamaño de muestra.

El análisis de potencia resulta útil tanto *a priori* como *a posteriori*. Sin embargo, a nuestro juicio, cuando verdaderamente resulta crucial es antes de llevar a cabo la investigación. Aunque defensores de la idea de que un trabajo de potencia inadecuado puede llegar a constituir un trabajo de autodecepción, no tenemos ningún reparo en admitir que un análisis detallado de los factores que determinan la potencia capacita a los investigadores para diseñar mejores estudios. Coincidimos con D'Amico, Neilands, y Zambano (2001), en que un riguroso análisis de potencia *a priori* permite verificar si el esfuerzo, el tiempo y los costos requeridos por un diseño de investigación están plenamente justificados.

Como ha sido ejemplificado con anterioridad, existen diversos ámbitos de investigación donde los tratamientos son administrados a grupos de personas, más que a personas individuales. Más aún, en la mayoría de las ocasiones los miembros que conforman los grupos no han sido asignados conforme a las reglas del azar a los mismos. En el mejor de los casos grupos intactos son enviados aleatoriamente a las condiciones de tratamiento. Por consiguiente, en estas circunstancias se requiere precisar el tamaño de las unidades de observación y de asignación para detectar diferencias entre tratamientos e interacciones, o los efectos moderadores que las

características de los sujetos y/o grupos ejercen sobre la variabilidad del impacto. Pues, las unidades de asignación utilizadas ya no son los sujetos, sino los grupos. Además, como probablemente los recursos que conlleven muestrear ambos tipos de unidades difieran sustancialmente entre sí, será conveniente determinar los tamaños de muestra, no sólo en función del tamaño del efecto de interés y de las variaciones dentro y a través de los grupos, sino también de los costos implicados en detectar dichos efectos.

En lo que resta de este trabajo ilustraremos como determinar tamaños de muestra óptimos para poner de manifiesto el efecto principal de la intervención y de su interacción con el tiempo en un diseño jerárquico de grupos aleatorizados con pretest y posttest. En principio, mediante el procedimiento que vamos a desarrollar es inmediato comprobar como este diseño es uniformemente más óptimo para detectar diferencias entre los tratamientos que el diseño jerárquico de grupos aleatorizados estándar. Para dar cuenta de los objetivos planteados seguiremos un procedimiento similar al expuesto por Cohen (1988) y Raudenbush y Liu (2000). En concreto, utilizaremos un modelo estandarizado en el se conjeturan medidas del tamaño del efecto pequeñas (0.3), medianas (0.5) y grandes (0.70), varianzas dentro de los grupos igual a la unidad y varianzas a través de los grupos iguales al cuadrado de los tamaños del efecto estandarizados que acabamos de especificar.

Ejemplificación de la forma de obtener los tamaños de muestra

Para el diseño referido en la sección anterior el modelo puede ser formulado en términos escalares como sigue:

$$y_{ijkl} = \mu + \alpha_j + \beta_{k(j)} + \gamma_l + (\alpha\gamma)_{jl} + (\beta\gamma)_{k(j)l} + \epsilon_{i(jkl)} \quad (12)$$

donde el valor observado del *i*-ésimo sujeto anidado dentro de la *j*-ésima condición y del *k*-ésimo grupo en el *l*-ésimo tiempo (y_{ijkl}) es expresado como una función de la media general (μ), del efecto de la *j*-ésima condición de tratamiento (α_j), del efecto aleatorio del *k*-ésimo grupo anidado dentro de la condición *j* ($\beta_{k(j)}$), del efecto del *l*-ésimo tiempo (γ_l), del efecto conjunto de la *j*-ésima condición y el *l*-ésimo tiempo ($(\alpha\gamma)_{jl}$), de la combinación aleatoria del *k*-ésimo grupo y el *l*-ésimo tiempo ($(\beta\gamma)_{k(j)l}$) y de la variación aleatoria entre los miembros del grupo ($\epsilon_{i(jkl)}$).

Alternativamente, la Ecuación 12 puede ser rescrita en términos de un modelo multinivel. Para ello comenzamos escribiendo en el primer nivel un modelo similar al de la regresión clásica incorporando el tiempo como variable explicativa

$$y_{ij} = b_{0j} + b_{1j} T_{ij} + e_{ij} \quad (13)$$

donde y_{ij} denota la puntuación del sujeto *i*-ésimo en el *j*-ésimo grupo, el interceptor, b_{0j} , es igual a la media del grupo *j*, la pendiente, b_{1j} , representa el cambio promedio en el posttest asociado con una unidad de cambio en el pretest y e_{ij} denota la diferencia entre la puntuación del sujeto *ij*-ésimo y la media del *j*-ésimo grupo. Por simplicidad asumimos que el error sigue una distribución normal con media cero y varianza constante a través de los grupos, esto es, $e_{ij} \sim N(0, \sigma^2)$.

A continuación, incorporamos la naturaleza jerárquica de los datos en el modelo. Para ello consideraremos a los coeficientes de regresión b_{0j} y b_{1j} como variables dependientes que fluctúan a través de los grupos como una función de una media más el tratamiento y el error. Específicamente, los coeficientes de regresión se relacionan con el tratamiento como sigue:

$$\begin{aligned} b_{0j} &= \beta_{00} + \beta_{01} \text{Trat}_j + u_{0j} \\ b_{1j} &= \beta_{10} + \beta_{11} \text{Trat}_j + u_{1j} \end{aligned} \tag{14}$$

En los modelos de nivel 2 se asume que en cada grupo los parámetros b_{0j} y b_{1j} se distribuyen normalmente con medias β_{00} y β_{10} , respectivamente, y matriz de varianzas-covarianzas

$$G = \begin{bmatrix} \text{var}(\beta_{00}) & \text{cov}(\beta_{00}, \beta_{10}) \\ \text{cov}(\beta_{00}, \beta_{10}) & \text{var}(\beta_{10}) \end{bmatrix} = \begin{bmatrix} \text{var}(u_{0j}) & \text{cov}(u_{0j}, u_{1j}) \\ \text{cov}(u_{0j}, u_{1j}) & \text{var}(u_{1j}) \end{bmatrix} = \begin{bmatrix} \omega_{00} & \omega_{01} \\ \omega_{10} & \omega_{11} \end{bmatrix} \tag{15}$$

También se asume que los errores correspondientes a los niveles 1 y 2 son independientes entre sí, esto es, $\text{cov}(u_{0j}, e_{ij}) = \text{cov}(u_{1j}, e_{ij}) = 0$. Sustituyendo las expresiones correspondientes a la Ecuación 15 dentro de la Ecuación 14 se obtiene el modelo mixto que sigue:

$$y_{ij} = \beta_{00} + \beta_{01} \text{Trat}_j + u_{0j} + \beta_{10} T_{ij} + \beta_{11} T_{ij} \text{Trat}_j + u_{1j} T_{ij} + e_{ij} \tag{16}$$

donde y_{ij} denota la puntuación del sujeto i -ésimo en el grupo j -ésimo, β_{00} representa el valor que resulta de promediar las medias de los grupos, β_{01} representa la diferencia de medias en la respuesta de interés entre los grupos que reciben el tratamiento y los que no lo reciben, u_{0j} indica si existen diferencias entre las medias de los grupos en la variable dependiente controlando el efecto del tratamiento, β_{10} representa la diferencia promedio entre el pretest y el postest, β_{11} representa la diferencia media en la relación pretest-postest entre los grupos que reciben el tratamiento y los grupos que sirven de control, u_{1j} indica si la relación entre el pretest y el postest dentro del grupo varía a través de los mismos manteniendo constante el efecto del tratamiento, e_{ij} denota la diferencia entre la puntuación del sujeto ij -ésimo y la media del j -ésimo grupo. El sistema de codificación asumido es el siguiente: 1 para la constante, |0.5| para el tratamiento y el tiempo y |0.25| para la interacción de ambos.

Determinación del tamaño de muestra sin contemplar los costos de la investigación

Además especificar la forma y magnitud del efecto del diseño, de interés clave a la hora de llevar a cabo el análisis de potencia resultan los tres aspectos que siguen (ver también Murray, 1998):

- a. Seleccionar una prueba estadística para evaluar los efectos del diseño.
- b. Determinar la distribución de la prueba estadística seleccionada.
- c. Desarrollar los parámetros de nocentralidad de los efectos de interés, así como sus correspondientes varianzas.

En el caso del efecto principal del tratamiento, una prueba válida de $H_0: \beta_{01} = 0$ nos la ofrece el contraste estadístico

$$F_{\beta_{01}} = \frac{E(\hat{\beta}_{01})}{E(\hat{u}_{0j})} \tag{17}$$

Cuando la hipótesis contrastada es verdadera la distribución del estadístico F es aproximado mediante una distribución F central con $\nu_1 = p - 1$ y $\nu_2 = Q - 2$ grados de libertad para el numerador y denominador, respectivamente. En el caso de la interacción la prueba F se construye de una manera similar, en concreto

$$F_{\beta_{11}} = \frac{E(\hat{\beta}_{11})}{E(\hat{u}_{1j})} \tag{18}$$

con $\hat{\beta}_{01}$ y $\hat{\beta}_{11}$ definidos como en (22) y (26).

Si $H_0: \beta_{11} = 0$ es verdadera, estadístico F sigue una distribución F central con $\nu_1 = (p - 1)(q - 1)$ y $\nu_2 = (r - 1)(Q - 2)$ grados de libertad para el numerador y denominador, respectivamente.

Sin embargo, bajo hipótesis alternativa $F_{\beta_{01}}$ y $F_{\beta_{11}}$ siguen una distribución F nocentral con los grados de libertad especificados y los parámetros de no centralidad que siguen:

$$\lambda_{\beta_{01}} = \nu_1 F_{\beta_{01}} = \frac{nQr\beta_{01}^2}{4(\sigma_\epsilon^2 + n\omega_{11} + n\omega_{00})} \tag{19}$$

y

$$\lambda_{\beta_{11}} = \nu_1 F_{\beta_{11}} = \frac{nQ\beta_{11}^2}{8(\sigma_\epsilon^2 + n\omega_{11})} \tag{20}$$

Especificados los parámetros de nocentralidad, seleccionada la prueba estadística y determinada su distribución, podemos obtener la potencia correspondiente a los efectos fijos del diseño calculando la probabilidad de que una F nocentral con grados de libertad ν_1 y ν_2 y parámetro de nocentralidad para λ exceda el correspondiente valor crítico (Muller, La Vange Ramey y Ramey, 1992). Formalmente

$$\text{Potencia} = 1 - \text{Prob}[F(\nu_1, \nu_2; \lambda) < \text{Finv}(1 - \alpha, \nu_1, \nu_2)] \tag{21}$$

donde $\text{Finv}(1 - \alpha, \nu_1, \nu_2)$ representa el valor crítico obtenido a partir de una función de distribución F central inversa. Los valores de potencia pueden ser descubiertos utilizando adecuadas rutinas computacionales. Por ejemplo, se pueden utilizar las siguientes expresiones del programa SAS para obtener la potencia correspondiente a los efectos del diseño jerárquico de grupos aleatorizados con pretest y postest:

$$\begin{aligned} \text{Potencia}_{\beta_{01}} &= 1 - \text{Prob f}(\text{Finv}(1 - \alpha, \nu_1, \nu_2), \nu_1, \nu_2, \lambda_{\beta_{01}}) \\ \text{Potencia}_{\beta_{11}} &= 1 - \text{Prob f}(\text{Finv}(1 - \alpha, \nu_1, \nu_2), \nu_1, \nu_2, \lambda_{\beta_{11}}) \end{aligned}$$

Con propósitos ilustrativos, en la Tabla 1 aparece recogida la potencia obtenida para diversos los valores de n , Q , tamaños del efecto y varianzas a través de los grupos.

Los resultados de la Tabla 1 favorecen planificar el diseño con un mayor número de grupos que de miembros dentro del grupo, especialmente, cuando el tamaño del impacto postulado es pequeño y la varianza de los tratamientos a través de los grupos grande. Sin embargo, desde un punto de vista económico la conclusión puede resultar poco realista, puesto que por regla general el muestreo de los grupos resulta más costoso que el muestreo de los miembros del grupo. Por consiguiente, resulta de interés efectuar el análisis de potencia considerando también los costos implicados en el proceso de muestreo.

Tabla 1
Potencia para el efecto principal del tratamiento sin tener en cuenta el costo

Q	N	ω_{ϵ}	δ	Potencia	Q	Potencia	Q	Potencia
20	20	0.60	0.3	0.127	30	0.171	40	0.216
20	20	0.60	0.5	0.268	30	0.387	40	0.496
20	20	0.60	0.7	0.466	30	0.648	40	0.779
20	20	0.35	0.3	0.180	30	0.254	40	0.327
20	20	0.35	0.5	0.409	30	0.579	40	0.711
20	20	0.35	0.7	0.677	30	0.856	40	0.941
20	20	0.20	0.3	0.268	30	0.387	40	0.496
20	20	0.20	0.5	0.607	30	0.796	40	0.901
20	20	0.20	0.7	0.877	30	0.974	40	0.995
20	30	0.60	0.3	0.128	30	0.173	40	0.218
20	30	0.60	0.5	0.271	30	0.391	40	0.501
20	30	0.60	0.7	0.471	30	0.654	40	0.784
20	30	0.35	0.3	0.183	30	0.259	40	0.333
20	30	0.35	0.5	0.416	30	0.588	40	0.721
20	30	0.35	0.7	0.686	30	0.863	40	0.945
20	30	0.20	0.3	0.276	30	0.399	40	0.511
20	30	0.20	0.5	0.623	30	0.811	40	0.911
20	30	0.20	0.7	0.889	30	0.978	40	0.996
20	40	0.60	0.3	0.128	30	0.174	40	0.219
20	40	0.60	0.5	0.272	30	0.394	40	0.504
20	40	0.60	0.7	0.473	30	0.657	40	0.787
20	40	0.35	0.3	0.184	30	0.261	40	0.336
20	40	0.35	0.5	0.420	30	0.593	40	0.726
20	40	0.35	0.7	0.691	30	0.867	40	0.948
20	40	0.20	0.3	0.281	30	0.406	40	0.518
20	40	0.20	0.5	0.631	30	0.818	40	0.917
20	40	0.20	0.7	0.894	30	0.980	40	0.997
20	50	0.60	0.3	0.129	30	0.174	40	0.221
20	50	0.60	0.5	0.273	30	0.395	40	0.505
20	50	0.60	0.7	0.475	30	0.659	40	0.789
20	50	0.35	0.3	0.185	30	0.262	40	0.338
20	50	0.35	0.5	0.422	30	0.596	40	0.728
20	50	0.35	0.7	0.694	30	0.869	40	0.949
20	50	0.20	0.3	0.283	30	0.410	40	0.523
20	50	0.20	0.5	0.636	30	0.822	40	0.920
20	50	0.20	0.7	0.898	30	0.981	40	0.997

Nota: Q= número de grupos; N = miembros dentro del grupo, $\alpha = .05$; ω_{ϵ} = Varianza debida a los grupos + varianza debida a la interacción. $\omega_{11} = 0.10$

Determinación del tamaño de muestra en función de los costos del muestreo

Para abordar este análisis conviene conocer la varianza de los efectos del diseño. Siguiendo un procedimiento similar al descrito por Murray (1998) y Raudenbush y Liu (2000) el error estándar del efecto principal

$$\hat{\beta}_{01} = \bar{Y}_E - \bar{Y}_C \tag{22}$$

se puede obtener fácilmente si expresamos la varianza de la media de un grupo basado en n observaciones dependientes y r medidas repetidas como

$$\sigma_y^2 = \frac{\sigma_{\epsilon}^2}{nr} + \frac{\omega_{11}}{r} + \omega_{00} \tag{23}$$

y la varianza de la condición de tratamiento j basada en q grupos del mismo tamaño

$$\sigma_{y_j}^2 = \left(\frac{\sigma_{\epsilon}^2}{nr} + \frac{\omega_{11}}{r} + \omega_{00} \right) / q \tag{24}$$

Por tanto, asumiendo que las varianzas son homogéneas a través de los grupos nos queda

$$\text{Var}(\hat{\beta}_{01}) = \frac{4(\sigma_{\epsilon}^2 + n\omega_{11} + nr\omega_{00})}{nQr} \tag{25}$$

Operando de la misma forma encontramos que la varianza correspondiente al efecto de la interacción

$$\hat{\beta}_{11} = (\bar{Y}_{ED} - \bar{Y}_{EA}) - (\bar{Y}_{CD} - \bar{Y}_{CA}) \tag{26}$$

resulta

$$\text{Var}(\hat{\beta}_{11}) = \frac{8(\sigma_{\epsilon}^2 + n\omega_{11})}{nQ} \tag{27}$$

Observando las Ecuaciones 25 y 27 se aprecia que tanto el número de grupos, como el número de miembros dentro de cada grupo afecta a la precisión de las estimaciones. Sin embargo, la falta de precisión estadística será mayor cuando Q sea reducida que cuando lo sea n . Por consiguiente, dado que las unidades de asignación afectan más a la sensibilidad del diseño que las unidades de observación, el investigador deberá negociar hábilmente en función del costo de las mismas, los tamaños de Q y n que incluye en el estudio para tener una potencia apropiada.

De acuerdo con Cochran (1977), en muchos estudios de muestreo de dos etapas el costo que conlleva recolectar los datos se puede aproximar por una relación de la forma

$$C = C_1nQ + C_2 Q \tag{28}$$

donde C se refiere al costo total del estudio, C_1 al costo que conlleva muestrear los miembros dentro de cada uno de los grupos y C_2 el costo asociado con cada uno de los grupos.

Una vez determinado el coste total del estudio el investigador está en condiciones de seleccionar el valor de n que minimiza la varianza de los efectos del diseño. Para ello tan sólo se requiere efectuar dos sencillas operaciones. Por un lado, expresar las varianzas de los efectos del diseño teniendo en cuenta los costos del estudio

$$\text{Var}(\hat{\beta}_{01}) = \frac{4(\sigma_{\epsilon}^2 + n\omega_{11} + nr\omega_{00})(nC_1 + C_2)}{nrC} \tag{29}$$

y

$$\text{Var}(\hat{\beta}_{11}) = \frac{8(\sigma_{\epsilon}^2 + n\omega_{11})(nC_1 + C_2)}{nC} \tag{30}$$

Y, por otro lado, descubrir el valor de n que minimiza las varianzas de las Ecuaciones 29 y 30. Derivando con respecto a n se encuentra

$$n_{\beta_{01}}(\text{óptimo}) = \left[\frac{(\sigma_{\epsilon}^2 + r\omega_{11})C_2}{r\omega_{00}C_1} \right]^{1/2} \tag{31}$$

y

$$n_{\beta_{11}}(\text{óptimo}) = \left[\frac{\sigma_{\epsilon}^2 C_2}{\omega_{11} C_1} \right]^{1/2} \tag{32}$$

Idénticos valores hubiésemos obtenido de haber maximizado los parámetros de nocentralidad de las Ecuaciones 19 y 20 con respecto *n*.

Asumiendo el costo relativo entre C_2/C_1 es estimado en 2, 6, 8 y 10, en la Tabla 2 aparecen recogidos los valores de *n*, *Q* y potencia, para diversos tamaños del efecto, varianzas a través de los grupos y razones de costo.

De los resultados que aparecen recogidos en la Tabla 2 cuatro aspectos son dignos de mención. En primer lugar, manteniendo constante el tamaño del efecto y la varianza a través de los grupos, la potencia se incrementa conforme la razón de costos decrece. En segundo lugar, magnitudes de efectos medianos y grandes producen potencias que se aproximan al valor considerado ideal. Sin embargo, también se puede observar que cuando los grupos de tratamiento están separados por 0.2, unidades estándar varianzas inferiores a 0.10 serán requeridas para proporcionar potencias que detecten el efecto del tratamiento, al menos en el 50% de los casos; dado que la potencia se incrementa conforme disminuye la varianza. En tercer lugar, cuanto más costoso resulta muestrear los grupos en relación con el número de miembros que componen el grupo, mayor es el tamaño de *n* y menor el de *Q*. Resaltar, por último, que detectar el efecto de la interacción exige mayores tamaños de muestra que detectar el efecto principal. No obstante, desde el punto de vista cualitativo se puede observar como las funciones de potencia del efecto principal y de la interacción son idénticas.

Conclusiones

Las derivaciones presentadas en este trabajo muestran que el modelo lineal general no puede ser utilizado para estimar los parámetros del modelo mixto de la Ecuación 2, pues el procedimiento de los mínimos cuadrados ordinarios asume que los errores son independientes con media cero y varianza constante. Sin embargo, en un modelo como el de la Ecuación 2 existen múltiples fuentes de variación aleatoria, los errores no tienen porque se independientemente y las varianzas pueden ser distintas entre sí. En estos casos ni el modelo lineal general ni el generalizado resultan apropiados, la solución natural a los problemas reseñados la ofrece el modelo mixto lineal general si la distribución de probabilidad de la variable respuesta no se desvía de la normalidad o el modelo mixto lineal generalizado si los datos siguen cualquier otro miembro de la familia exponencial de distribuciones. Por ejemplo, si nos ceñimos al modelo de la Ecuación 16 se aprecia que existen tres fuentes de variación aleatoria, las cuales son convenientemente estimadas e interpretadas mediante el enfoque del modelo mixto. Además, de existir dependencia entre las unidades de primer nivel anidadas dentro de las unidades de segundo nivel, está se obtiene independientemente del error estimando la variación en la unidades de segundo nivel inducida por el agrupamiento (Carvajal, Baumler, Harrist y Porcel, 2001).

En el trabajo también se ejemplifica para el diseño jerárquico de grupos al azar con pretest y postest, en palabras de Murray (1998) uno de los diseños utilizados con mayor frecuencia en la evaluación de programas de prevención basados en organizaciones, como maximizar la potencia para poner de relieve los efectos de los tratamientos seleccionando tamaños de muestra óptimos, tanto en lo referido al número de grupos como al tamaño de los mismos. Aunque para alcanzar esta meta hemos utilizado, al igual que Raudenbush y Liu (2000), un modelo estandarizado, no faltaran ocasiones en las que los investigadores sean capaces de anticipar el valor de los componentes de varianza y del tamaño del efecto utilizando datos de algún estudio previo o de algún estudio piloto. En cualquier caso, los resultados mostrados en las tablas ponen claramente de relieve algunas pautas de comportamiento que los investigadores que utilizan este tipo de diseños deben tener bien presentes a la hora de planificar su investigación para lograr

C_2/C_1	VAR	TE	N(T)	Q(T)	POTEN(T)	N(I)	Q(I)	POTEN(I)
2	0.50	0.3	2	125	0.438	2	125	0.227
2	0.50	0.5	2	125	0.853	2	125	0.521
2	0.50	0.7	2	125	0.988	2	125	0.804
2	0.25	0.3	3	100	0.505	3	100	0.280
2	0.25	0.5	3	100	0.908	3	100	0.630
2	0.25	0.7	3	100	0.996	3	100	0.894
2	0.10	0.3	4	83	0.597	4	83	0.365
2	0.10	0.5	4	83	0.957	4	83	0.767
2	0.10	0.7	4	83	0.999	4	83	0.965
6	0.50	0.3	3	56	0.242	4	50	0.142
6	0.50	0.5	3	56	0.555	4	50	0.310
6	0.50	0.7	3	56	0.836	4	50	0.536
6	0.25	0.3	4	50	0.300	5	45	0.179
6	0.25	0.5	4	50	0.667	5	45	0.409
6	0.25	0.7	4	50	0.917	5	45	0.677
6	0.10	0.3	7	38	0.355	8	36	0.256
6	0.10	0.5	7	38	0.753	8	36	0.584
6	0.10	0.7	7	38	0.959	8	36	0.860
8	0.50	0.3	4	42	0.200	4	42	0.126
8	0.50	0.5	4	42	0.459	4	42	0.267
8	0.50	0.7	4	42	0.739	4	42	0.465
8	0.25	0.3	5	38	0.247	6	36	0.160
8	0.25	0.5	5	38	0.564	6	36	0.358
8	0.25	0.7	5	38	0.843	6	36	0.608
8	0.10	0.3	8	31	0.304	9	29	0.224
8	0.10	0.5	8	31	0.673	9	29	0.515
8	0.10	0.7	8	31	0.920	9	29	0.798
10	0.50	0.3	4	36	0.177	5	33	0.113
10	0.50	0.5	4	36	0.402	5	33	0.231
10	0.50	0.7	4	36	0.669	5	33	0.401
10	0.25	0.3	6	31	0.213	6	31	0.143
10	0.25	0.5	6	31	0.490	6	31	0.314
10	0.25	0.7	6	31	0.773	6	31	0.541
10	0.10	0.3	9	26	0.264	10	25	0.206
10	0.10	0.5	9	26	0.602	10	25	0.474
10	0.10	0.7	9	26	0.872	10	25	0.755

Nota: C_2/C_1 = razón de costos; VAR= varianza del impacto a través de los grupos; TE= tamaño del efecto estandarizado; N(T)= valor requerido de *n* para el efecto principal; Q(T)= número de grupos requeridos para el efecto principal; POTEN(T)= potencia de prueba correspondiente al efecto principal; N(I)= valor requerido de *n* para la interacción; Q(I)= número de grupos requeridos para la interacción; POTEN(I)= potencia de prueba correspondiente a la interacción

suficiente potencia estadística. Probablemente manteniendo constante el diseño, el tipo de análisis y el número de réplicas la precisión estadística se vea mejorada incluyendo alguna variable auxiliar y utilizando más tandas de observaciones.

Por último, no queremos pasar por alto algunas limitaciones que tiene el trabajo. En concreto, todas las derivaciones han sido referidas al análisis de potencia para un diseño relativamente sencillo con dos condiciones experimentales (condición de tratamiento y de comparación), en que se ha sumido que la variable dependiente era continua, con datos registrados regularmente, grupos equilibrados y ausencia de observaciones perdidas. Aunque cae

fuera de las pretensiones del trabajo extender las derivaciones encontradas a situaciones más reales, como por ejemplo a diseños no equilibrados, no resultaría mucho más complicado.

Agradecimientos

Los autores agradecen los valiosos comentarios de diversos profesores 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 MCT (Ref.: BOS-2000-0410).

Referencias

- Aitkin, M.A. y Longford, N.T. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, 149, 1-43.
- Breslow, N.E. y Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.
- Brown, H. y Prescott, R. (1999). *Applied Mixed Models in Medicine*. New York: John Wiley & Sons.
- Carvajal, S.C., Baumler, E., Harrist, R.B. y 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.
- Cochran, W. (1977). *Sampling Techniques*, 3rd Ed. New York: John Wiley and Sons.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Ed. Hillsdale, NJ: Erlbaum
- D'Amico, E.J., Neilands, T.B. y Zambarano, R. (2001). Power analysis for multivariate and repeated measures designs: A flexible approach using SPSS MANOVA procedure. *Behavior Research Methods, Instruments, and Computers*, 33, 479-484.
- Dempster, A.P., Laird, N.M. y Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Ser. B*, 39, 1-38.
- Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd Ed. London: Arnold.
- Lair, N.M. y Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Lindstrom, M.J. y Bates, D. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83, 1.014-1.022.
- Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-338.
- Jennrich, R.I. y Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42, 805-820.
- Kenward, M.G. y Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983-997.
- Liang, K.Y. y Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- McCulloch, C.E. y Searle, R.S. (2001). *Generalized, Linear, and Mixed Models*. New York: John Wiley and Sons, Inc.
- Muller, K.E., La Vange, L.M., Ramey, S.L. y Ramey, C.T. (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association*, 87, 1.209-1.226.
- Murray, D.M. (1998). *Design and Analysis of Group-Randomized Trials*. New York: Oxford University Press.
- Oliver, J.C., Rosel, J. y Jara, P. (2000). Modelos de regresión multinivel: Aplicación en Psicología escolar. *Psicothema*, 12, 487-494.
- Raudenbush, S.W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, 2, 173-185.
- Raudenbush, S.W. y Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5, 199-213.
- Raudenbush, S.W. y Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and Data*. Thousand Oaks, CA: SAGE Publications, Inc.
- Rinndskopf, D. y 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
- Shadish, W.R., Cook, T.D. y Campbell, D.J. (2002). *Experimental and Quasi Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- SAS Institute (2001). *SAS/STAT Software: Version 8.2 (TS M0)*. Cary, NC: SAS Institute Inc.
- Searle, S.R., Casella, G. y McCulloch, C.E. (1992). *Variance Components*. New York: John Wiley and Sons, Inc.
- Singer, D.J. (2002). Fitting individual growth models using SAS PROC MIXED. In D.S. Moskowitz and S.L. Hershberger (Eds.), *Modeling Intraindividual variability with Repeated Measures Data: Methods and Applications* (pp. 135-170). Mahwah, NJ: Lawrence Erlbaum Associates.
- Van der Leeden, R. (1998). Multilevel analysis of longitudinal data. In C.H.J. Bijleveld and L.J. Th. van der Kamp (Eds.), *Longitudinal Data Analysis: Designs, Models and Methods* (pp. 269-317). Thousand Oaks, CA: SAGE Publications, Inc.
- Vallace, D. y Green, B.S. (2002). Analysis of repeated measures designs with linear mixed models. In D.S. Moskowitz and S.L. Hershberger (Eds.), *Modeling Intraindividual Variability with Repeated Measures Data: Methods and Applications* (pp.135-170). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wolfinger, R. y O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48, 233-243.
- Wolfinger, R.D. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 205-230.