

Power analysis to detect treatment effect in longitudinal studies with heterogeneous errors and incomplete data

Guillermo Vallejo¹, Manuel Ato², Paula Fernández García¹, Pablo E. Livacic Rojas³ and Ellián Tuero Herrero¹
¹ Universidad de Oviedo, ² Universidad de Murcia and ³ Universidad de Santiago de Chile

Abstract

Background: S. Usami (2014) describes a method to realistically determine sample size in longitudinal research using a multilevel model. The present research extends the aforementioned work to situations where it is likely that the assumption of homogeneity of the errors across groups is not met and the error term does not follow a scaled identity covariance structure. **Method:** For this purpose, we followed a procedure based on transforming the variance components of the linear growth model and the parameter related to the treatment effect into specific and easily understandable indices. At the same time, we provide the appropriate statistical machinery for researchers to use when data loss is unavoidable, and changes in the expected value of the observed responses are not linear. **Results:** the empirical powers based on unknown variance components were virtually the same as the theoretical powers derived from the use of statistically processed indexes. **Conclusions:** The main conclusion of the study is the accuracy of the proposed method to calculate sample size in the described situations with the stipulated power criteria.

Keywords: Sample size planning, two-level nested design, longitudinal data analysis, heterogeneity variance, missing data.

Resumen

Análisis de potencia para detectar el efecto del tratamiento en estudios longitudinales con errores heterogéneos y datos incompletos.

Antecedentes: S. Usami (2014) describe un método que permite determinar de forma realista el tamaño de muestra en la investigación longitudinal utilizando un modelo multinivel. En la presente investigación se extiende el trabajo aludido a situaciones donde es probable que se incumpla el supuesto de homogeneidad de los errores a través de los grupos y la estructura del término de error no sea de identidad escalada. **Método:** para ello, se ha seguido procedimiento basado en transformar los componentes de varianza del modelo de crecimiento lineal y el parámetro relacionado con el efecto del tratamiento en índices de fácil comprensión y especificación. También se proporciona la maquinaria estadística adecuada para que los investigadores puedan usarlo cuando la pérdida de información resulte inevitable y los cambios en el valor esperado de las respuestas observadas no sean lineales. **Resultados:** las potencias empíricas basadas en componentes de varianza desconocidos fueron virtualmente idénticas a las potencias teóricas derivadas a partir del uso de índices estadísticos transformados. **Conclusiones:** la principal conclusión del trabajo es la exactitud del enfoque propuesto para calcular el tamaño de muestra bajo las situaciones reseñadas con el criterio de potencia estipulado.

Palabras clave: planificación tamaño de la muestra, diseño de dos niveles anidado, análisis de datos longitudinales, heterogeneidad de la varianza, datos perdidos.

Los estudios longitudinales con datos continuos son cada vez más comunes en las investigaciones sociales y conductuales. En algunos casos, los sujetos se miden en repetidas ocasiones a lo largo del tiempo con el fin de examinar su crecimiento individual y las posibles diferencias existentes entre ellos. En otros casos, los sujetos asignados a las distintas condiciones de tratamiento se tratan durante un período específico de tiempo y a la conclusión del estudio se comparan con respecto a sus tasas de crecimiento promedio. Con independencia de cuál sea el objetivo del estudio,

lo habitual y razonable es modelar el cambio en la respuesta de interés asumiendo un crecimiento lineal (Willet, 1988) y expresar el efecto de la intervención en términos de la interacción grupo \times tendencia lineal (Fitzmaurice, Laird, & Ware, 2011).

Un problema que no debe ser ignorado en los estudios que implican mediciones repetidas de los sujetos en sucesivos intervalos temporales es la presencia de datos omitidos en la variable de respuesta, lo que añadido a la falta de independencia de los errores de medida puede dar lugar a conclusiones erróneas cuando los datos longitudinales se analizan empleando modelos lineales clásicos de análisis de varianza y regresión. Además de otras, por esta razón los modelos lineales jerárquicos se han convertido en el método de elección para modelar el cambio en la respuesta a través del tiempo y los factores que influyen en el cambio.

Cuando se analizan datos longitudinales mediante modelos jerárquicos, el estudio debe tener un tamaño suficiente a fin

de detectar efectos de importancia teórica o práctica al nivel de potencia estadística deseado (e.g., 80%). Aunque existe una gran cantidad de literatura disponible explicando la forma de calcular el tamaño de muestra óptimo en este tipo de estudios (e.g., Muthén & Curran, 1997; Heo, Xue, & Kim, 2013; Raudenbush & Liu, 2001; Usami, 2014; Wänström, 2009), pocos trabajos se han ocupado de informar a los investigadores sobre cómo determinar el tamaño muestral cuando se espera que algunos sujetos abandonen prematuramente el estudio (Hedeker, Gibbons, & Waternaux, 1999; Roy, Bhaumik, Aryal, & Gibbons, 2007; Heo, 2014) y, hasta donde nosotros conocemos, ningún trabajo se ha ocupado de informar acerca del número de sujetos y de medidas repetidas que se deben incluir en el estudio para detectar el efecto de la intervención con un mínimo de garantía en presencia de errores con varianzas heterogéneas a través de los grupos del tratamiento.

Por este motivo, en el presente trabajo nos proponemos un doble objetivo. En primer lugar, desarrollar fórmulas para evaluar la potencia y determinar el tamaño de muestra para detectar el efecto de interacción grupo por tiempo, en el marco de un modelo de regresión de dos niveles con varianzas heterogéneas y datos completos e incompletos. En segundo lugar, comprobar si existen diferencias entre la potencia estadística teórica basada en estimaciones por Mínimos Cuadrados Ordinarios (MCO) y la potencia estadística empírica basada estimaciones por máxima verosimilitud.

Formulación del modelo estadístico

Supóngase que estamos interesados en comparar las respuestas de un grupo con las respuestas de otro (e.g., experimental *versus* control), respecto a sus tasas de cambio promedio. Si se considera que las medidas tomadas a lo largo del tiempo se hallan anidadas en los sujetos, los datos de este tipo se pueden analizar usando un modelo de regresión jerárquico de dos niveles. En el primer nivel se representa el cambio que esperamos experimente cada sujeto de la población durante un período específico de tiempo, mientras que en el segundo nivel se describe la relación conjeturada entre los parámetros de crecimiento individual y las variables explicativas que se asumen estables o independientes del tiempo a lo largo de la duración del estudio.

Adoptando un modelo de crecimiento individual en el cual el cambio es una función lineal del tiempo, el modelo de nivel 1 se puede formular como sigue:

$$Y_{it} = b_{0i} + b_{1i}X_{it} + e_{it} \quad (1)$$

donde Y_{it} denota la respuesta del i -ésimo sujeto ($i = 1, \dots, N$) en la t -ésima ocasión ($t = 1, \dots, T$), la variable X_{it} define el tiempo específico (días, semanas, meses o años) en que es observado cada sujeto y los parámetros aleatorios b_{0i} (intersección en el origen), b_{1i} (pendiente o tasa de cambio) y e_{it} (término de error) representan, respectivamente, el verdadero valor de la respuesta del i -ésimo sujeto al inicio del estudio (i.e., cuando $X_{i1} = 0$), la tasa de cambio de la respuesta del i -ésimo sujeto durante el periodo de recogida de los datos y el error de medida ocasionado por la desviación de la linealidad. Aquí, asumiremos que las T medidas repetidas de la respuesta serán tomadas en los tiempos $X_1 = 0, X_2 = 1, X_3 = 2, \dots, X_T = T-1 = D$, donde D denota la duración del estudio en alguna métrica significativa (e.g., años).

En el segundo nivel, los parámetros resultantes de modelar las trayectorias de cambio individual en función del tiempo,

se relacionan con las variables explicativas que describen las diferencias entre los sujetos en las intersecciones y pendientes. Si solo se dispone de una variable explicativa (e.g., un programa de intervención conductual para mejorar el lenguaje de niños autistas), el modelo de nivel 2 se convierte en

$$b_{0i} = \beta_{00} + \beta_{01}W_i + \mu_{0i} \quad (2)$$

$$b_{1i} = \beta_{10} + \beta_{11}W_i + \mu_{1i} \quad (3)$$

donde la variable indicador del programa de intervención $W_i = 0$ si la i -ésima unidad de nivel 2 se asigna al grupo control y $W_i = 1$ si se asigna al grupo experimental. En este modelo, β_{00} denota la respuesta media del grupo control al inicio del estudio, β_{01} la diferencia entre las respuestas promedio de ambos grupos al inicio del estudio, β_{10} la tasa de cambio promedio del grupo control y β_{11} la diferencia entre las tasas de cambio promedio de los grupos objeto de estudio o interacción grupo \times tiempo. Como consecuencia, la tasa de cambio promedio del grupo experimental se corresponde con la suma de $\beta_{10} + \beta_{11}$. Las variables aleatorias μ_{0i} y μ_{1i} son independientes de e_{it} y se asume que siguen una distribución normal bivariada con media cero, varianzas condicionales τ_{00} y τ_{11} , respectivamente, y covarianza condicional τ_{01} . En este caso, los componentes de varianza residual, τ_{00} y τ_{11} , representan la variabilidad que permanece en los parámetros b_{0i} y b_{1i} tras controlar por el efecto atribuible al programa.

Sustituyendo las ecuaciones (2) y (3) dentro de la ecuación (1), el modelo mixto o combinado se puede expresar como sigue:

$$Y_{it} = \beta_{00} + \beta_{01}W_i + \beta_{10}X_{it} + \beta_{11}W_iX_{it} + (u_{1i}X_{it} + u_{0i} + e_{it}) \quad (4)$$

En el modelo dado por la ecuación (4), se asume que los errores e_{it} condicionales en u_{1i} y u_{0i} se distribuyen normal e independientemente con media cero y varianza σ^2 constante. No obstante, en el trabajo actual también se considera la presencia de errores con varianzas heterogéneas a través de los grupos del tratamiento.

Los cuatro primeros términos del lado derecho de la ecuación (4) constituyen la parte fija del modelo y los tres últimos términos conforman la parte aleatoria. En este modelo, el valor esperado y la varianza de un sujeto cualquiera en una determinada ocasión de medida, así como la covarianza entre dos ocasiones de medida, son dadas por (para más detalles, véase el Apéndice A)

$$E(Y_{it} | X_{it}) = \beta_{00} + \beta_{01}W_i + (\beta_{10} + \beta_{11}W_i)X_{it} \quad (5)$$

$$Var(Y_{it} | X_{it}) = \tau_{00} + 2X_{it}\tau_{01} + X_{it}^2\tau_{11} + \sigma^2 \quad (6)$$

$$Cov(Y_{it}, Y_{it'} | X_{it}) = \tau_{00} + (X_{it} + X_{it'})\tau_{01} + X_{it}X_{it'}\tau_{11} \quad (7)$$

En algunas situaciones podría ser demasiado simplista asumir que el cambio a través del tiempo es lineal. Cuando los cambios en el valor esperado de las respuestas observadas se desvíen de la linealidad, se pueden considerar ecuaciones polinómicas más complejas. Por ejemplo, si la respuesta media aumenta de manera monótona en función del tiempo hasta que se produzca una estabilización en la mejora, entonces se podría considerar el modelo de crecimiento curvilíneo que sigue:

$$Y_{it} = \beta_{00} + \beta_{10}W_i + (\beta_{10} + \beta_{11}W_i)X_{it} + (\beta_{20} + \beta_{21}W_i)X_{it}^2 + \quad (8)$$

$$(u_{0i} + u_{1i}X_{it} + u_{2i}X_{it}^2 + e_{it})$$

En este caso, el valor esperado, varianza y covarianza de la variable respuesta, dados los predictores de nivel-1 (i.e., término lineal y término cuadrático), se pueden escribir como

$$E(Y_{it} | X_{it}, X_{it}^2) = \beta_{00} + \beta_{10} W_i + (\beta_{10} + \beta_{11} W_i) X_{it} + (\beta_{20} + \beta_{21} W_i) X_{it}^2 \quad (9)$$

$$Var(Y_{it} | X_{it}, X_{it}^2) = \tau_{00} + 2 X_{it} \tau_{01} + X_{it}^2 \tau_{11} + 2 X_{it}^2 \tau_{02} + 2 X_{it}^3 \tau_{12} + X_{it}^4 \tau_{22} + \sigma^2 \quad (10)$$

$$Cov(Y_{it}, Y_{it'} | X_{it}, X_{it'}, X_{it}^2, X_{it'}^2) = \tau_{00} + (X_{it} + X_{it'}) \tau_{01} + X_{it} X_{it'} \tau_{11} + (X_{it}^2 X_{it'}^2) \tau_{02} + (X_{it}^2 X_{it'} + X_{it} X_{it'}^2) \tau_{01} + (X_{it}^3 X_{it'}^2) \tau_{22} \quad (11)$$

Las ecuaciones (5)-(7) y (9)-(11) son esenciales para planificar de forma adecuada un estudio longitudinal, pues, como veremos más adelante, proporcionan la maquinaria que permite calcular la potencia estadística y el tamaño muestral. Para determinar la potencia de una prueba estadística, es necesario especificar el valor de los parámetros incluidos en los modelos de las ecuaciones (1)-(3). Sin embargo, en muchas ocasiones puede resultar complicado y poco realista adelantar el valor de los parámetros sin haber realizado el experimento. De ahí que, en la práctica, el uso de los métodos existentes para derivar el tamaño muestral queda limitado a situaciones donde los investigadores son capaces de anticipar un rango de valores probables de los parámetros de interés, a partir de los datos de estudios previos.

En un intento por optimizar el enfoque para realizar un análisis de potencia en estudios longitudinales, Usami (2014) sugiere transformar los componentes de varianza y covarianza (i.e., σ^2 , τ_{00} , τ_{01} , τ_{11}) y el parámetro del modelo de la ecuación (4) relacionado con el efecto del tratamiento (i.e., β_{11}) en índices más fáciles de especificar e interpretar. Esencialmente, en los cuatro que siguen:

1. *Fiabilidad de la medida al comienzo del estudio.* Este índice puede expresarse como la proporción entre la varianza de la intersección en el origen y la varianza de la variable respuesta al inicio del estudio. Formalmente,

$$\rho_1 = \frac{Var(u_{0i})}{Var(u_{0i} + e_{it})} = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \quad (12)$$

Repárese que cuanto menor sea el error de medida mayor será la fiabilidad.

2. *Tamaño del efecto estandarizado.* No existe unanimidad de criterio respecto a cómo definir este índice. No obstante, varios autores (Muthén & Curran, 1997; Fraser *et al.*, 2005; Rye *et al.*, 2005; Wänström, 2009; Heo *et al.*, 2013; Usami, 2014) utilizan el cociente entre la diferencia de pendientes promedio en el último punto de tiempo y la desviación estándar de la respuesta al final del estudio. Formalmente,

$$d_T = \frac{E(Y_{iT} | W_i = 1) - E(Y_{iT} | W_i = 0)}{\sqrt{Var(Y_{iT})}} = \frac{D\beta_{11}}{\sqrt{\tau_{00} + 2D\tau_{01} + D^2\tau_{11} + \sigma^2}} \quad (13)$$

donde $D = T-1$. Se excusa decir que otras estandarizaciones son posibles. Por ejemplo, Raudenbush y Liu (2001) definen el tamaño del efecto dividiendo el valor de la diferencia entre las tasas de crecimiento promedio por la desviación estándar entre las pendiente de los sujetos (i.e., $d = \beta_{11} / \sqrt{\tau_{11}}$). Una revisión

detallada de las diferentes fórmulas usadas para calcular el tamaño del efecto puede verse en Feingold (2009).

3. *Correlación de la tasa de cambio con el nivel inicial.* En los modelos de crecimiento lineal tenemos que:

$$r = \frac{Cov(u_{0i}, u_{1i})}{\sqrt{Var(u_{0i}, u_{1i})}} = \frac{\tau_{01}}{\sqrt{\tau_{00}\tau_{11}}} \quad (14)$$

Una correlación elevada entre la intersección y la pendiente indica que la evolución de la respuesta depende de su valor inicial.

4. *Razón entre la varianza de la respuestas al final y al inicio del estudio.* Para una estructura de crecimiento lineal este índice se puede expresar como:

$$k = \frac{Var(Y_{iT})}{Var(Y_{i1})} = \frac{\tau_{00} + 2D\tau_{01} + 2D\tau_{11} + \sigma^2}{\tau_{00} + \sigma^2} \quad (15)$$

Resolviendo simultáneamente las ecuaciones (14) y (15) se obtienen la varianza y covarianza que siguen (para más detalles, véase el Apéndice B):

$$\tau_{11} = \frac{(\sqrt{r^2\tau_{00} + (k-1)(\tau_{00} + \sigma^2)} - r\sqrt{\tau_{00}})^2}{D^2} \quad (16)$$

$$\tau_{01} = \frac{r\sqrt{r^2\tau_{00}^2 + \tau_{00}(k-1)(\tau_{00} + \sigma^2)} - r^2\tau_{00}}{D} \quad (17)$$

Procediendo de forma similar con las ecuaciones (13) y (14) se obtiene una expresión para calcular el efecto del tratamiento sin estandarizar usado en este trabajo

$$\beta_{11} = \frac{d_T \sqrt{k(\tau_{00} + \sigma^2)}}{D} \quad (18)$$

Nótese que usando el modelo de intercepto aleatorio

$$\beta_{11} = d_T \sqrt{\tau_{00} + \sigma^2}, k = 1 \text{ y } r = 0.$$

Se puede suponer sin pérdida de generalidad que la varianza de la respuesta inicial sea igual a uno, i.e., $\tau_{00} + \sigma^2 = 1$. Entonces las ecuaciones (12)-(18) se reducen a las dadas por Usami (2014). La restricción anterior posibilita conocer los parámetros del modelo con solo especificar el valor de ρ_1 , d_T , r y k . Repárese, al respecto, que estos índices se pueden detallar de forma bastante intuitiva; lo cual evita, en buena medida, la dificultad que supone en los estudios de tipo exploratorio definir el valor de los parámetros antes haber realizado el experimento. Además, hay que añadir que Usami (2014) constata que los índices ρ_1 , r y k ejercen una menor influencia en el cálculo del tamaño muestral que d_T , especialmente cuando $d_T > 0.5$. Estos hechos han motivado que abordemos el análisis prospectivo de potencia para detectar efectos estadísticamente significativos del tratamiento transformando los parámetros del modelo.

Varianza del efecto estimado del tratamiento

El efecto de la intervención se puede definir de distintas maneras. No obstante, en los estudios longitudinales es fundamental comparar la diferencia existente entre los grupos de tratamiento con respecto a sus tasas de crecimiento promedio, $\hat{\beta}_{11} = (\hat{\delta}_1 - \hat{\delta}_0)$, donde $\hat{\delta}_j (j = 0, 1)$ es el estimador de la pendiente lineal en el j -ésimo grupo por el método MCO. Específicamente,

$$\hat{\beta}_{11} = \frac{\sum_{i=1}^{N_E} \sum_{t=1}^T (X_{it} - \bar{X}_i) Y_{it}}{\sum_{i=1}^{N_E} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2} - \frac{\sum_{i=1}^{N_C} \sum_{t=1}^T (X_{it} - \bar{X}_i) Y_{it}}{\sum_{i=1}^{N_C} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2} \quad (19)$$

donde N_E y N_C son los tamaños muestrales de los grupos experimental y control, respectivamente, y $\sum_{t=1}^T (X_{it} - \bar{X}_i)^2$ es la suma de las desviaciones al cuadrado de los tiempos de medición con respecto a la media.

Para probar la hipótesis nula de igualdad de pendientes entre los niveles de la variable independiente ($H_0 : \beta_{11} = 0$) se necesita obtener la varianza del estimador MCO de β_{11} . Haciendo uso de las ecuaciones (6) y (7) y considerando que la varianza de una diferencia se reduce a la suma de las varianzas cuando los grupos son independientes, álgebra ordinaria muestra que (para más detalles, véase el Apéndice C):

$$Var(\hat{\beta}_{11}) = \frac{4}{N} \left(\frac{\sigma^2}{\sum_{i=1}^T (X_{it} - \bar{X}_i)^2} + \tau_{11} \right) \quad (20)$$

donde $N (=N_E + N_C)$ denota el número total de unidades de segundo nivel incluidas en el estudio, con $N/2$ sujetos en cada grupo.

Si las T mediciones repetidas son tomadas a intervalos temporales regulares durante todo el estudio (i.e., desde $X_1 = 0$ hasta $X_T = D$), entonces la ecuación (20) se puede reformular como (véase Fitzmaurice *et al.*, 2011):

$$Var(\hat{\beta}_{11}) = \frac{4}{N} \left(\frac{12\sigma^2(T-1)}{D^2T(T+1)} + \tau_{11} \right) \quad (21)$$

donde $D=(T-1)/f$ y f es la frecuencia de observación por unidad de tiempo (e.g., número de observaciones por año). Alternativamente, en términos de f

$$Var(\hat{\beta}_{11}) = \frac{4}{N} \left(\frac{12f^2\sigma^2}{T^3 - T} + \tau_{11} \right) = \frac{4}{N} (V_1 + \tau_{11}) \quad (22)$$

donde V_1 es la varianza muestral de la pendiente de un sujeto específico alrededor de algún valor central, la cual está basada en el error de medida, el número de mediciones efectuadas y la frecuencia de las mismas y τ_{11} la varianza de la pendiente del sujeto que no se debe ni al error de medida ni a la falta de ajuste. La suma de ambos componentes, en adelante σ_{b1}^2 , representa la variabilidad correspondiente al estimador MCO de la tasa de cambio del modelo formulado en la ecuación (1).

Cuando el cambio se asume lineal y la frecuencia de observación por unidad de tiempo igual a la unidad (i.e., $X_j = 0, 1, 2, \dots, T-1; f=1; D$

$=T-1$), la varianza muestral de las pendientes intra-sujeto se simplifica a: $V_1 = 12\sigma^2 / (T^3 - T)$. Para funciones de crecimiento más complejas (e.g., cuadráticas) e intervalos de tiempo equidistantes a partir de la línea base (e.g., $X_j = 0, 2, 4, \dots, 2T-2; D = (T-1)/f; T = fD+1$), la varianza muestral de las pendientes intra-sujeto alrededor de algún valor central tendría la forma general (véase, Raudenbush & Liu, 2001):

$$V_p = \frac{\sigma^2 f^{2p} (T-p-1)!}{l_p (T+p)!} \quad (23)$$

donde p denota el orden polinomial del cambio de respuesta y l_p es una constante, donde $l_1 = 1/12, l_2 = 1/720$ y $l_3 = 1/100800$.

Hasta ahora hemos asumido una matriz de identidad escalada para caracterizar estructura de covarianza del error de medida (i.e., σ^2). Sin embargo, no es descartable que otros modelos menos parcos describan de una manera más adecuada los datos (e.g., simetría compuesta, autorregresivo de orden uno o Toeplitz). Para modelar estructuras de covarianza alternativas a la escalar, sugerimos calcular la variabilidad del estimador MCO de la pendiente para el i -ésimo sujeto, $Var(\hat{b}_{1i})$ como sigue:

$$\sigma_b^2 = \left[c' \begin{pmatrix} v_i \\ a_p^2 \end{pmatrix} c \right] f^{2p} - m_p \quad (24)$$

donde c es un vector contraste basado en polinomios ortogonales de grado p , los cuales pueden ser obtenidos adoptando el método empleado por Raudenbush y Liu (2001), $V_i (=Z_i G_i Z_i' + \Sigma_i)$ es la matriz de covarianza de las mediciones repetidas, Z_i es una matriz diseño que especifica cómo cambia la respuesta promedio de un sujeto concreto a lo largo del estudio (e.g., función constante, lineal, cuadrática o cúbica), G_i es la matriz de dispersión de los efectos aleatorios de nivel 2, Σ_i es la estructura de covarianza de los errores de nivel 1, m es una constante cuyo valor depende del orden del polinomio seleccionado y $a_p = \sum_{t=1}^T c_{pt}^2$, aquí c_{pt} denota el coeficiente de un polinomio ortogonal de grado p en la ocasión de medida t . Si, como ocurre en este estudio, se busca una aproximación polinomial de primer grado $m_1 = 1 \times \tau_{11}$, mientras que si se buscara una aproximación polinomial de segundo grado $m_2 = 3 \times \tau_{22}$ y si se buscara una de tercer grado $m_3 = 35 \times \tau_{33}$. Bajo el polinomio de grado cero, el valor de la constante m es nulo. Otra ventaja del enfoque matricial descrito será ejemplificada en una sección posterior al escribir las fórmulas para tratar de determinar el tamaño de muestra más adecuado cuando se prevea la posibilidad de que existan datos perdidos.

Como ha sido dicho, es bastante común asumir los cambios en la respuesta como un proceso lineal. En este caso, una forma rápida de proceder para especificar la matriz de diseño Z_i consiste en usar los coeficientes de tendencia lineal (i.e., $t_1 = 0, t_2 = 1, \dots, t_T = T-1$) divididos por la frecuencia de observación $f = (T-1)/D$. Ordinariamente $f = 1$, pero dependiendo del valor de D otros muchos resultados son posibles (e.g., $f = .5$ o $f = 2$). Si el efecto del tiempo se asume cuadrático el vector de coeficientes del polinomio de la matriz de diseño Z_i se divide por f^2 , mientras que si el cambio se asume cúbico el vector coeficientes se divide por f^3 .

Análisis de potencia de la prueba del efecto del tratamiento

En orden a probar la hipótesis nula de pendientes homogéneas ($H_0 : \beta_{11} = 0$) frente a la hipótesis alternativa de pendientes

heterogéneas ($H_1 : \beta_{11} \neq 0$) el siguiente estadístico, basado en las ecuaciones (19) y (22), se puede utilizar

$$V_p = \frac{\sigma^2 f^{2p} (T - p - 1)!}{l_p (T + p)!} \quad (25)$$

Cuando H_0 es verdadera, el estadístico F sigue una distribución F central con grados de libertad $df_1 = 1$ para el numerador y $df_2 = N \times T - (N + 2)$ para el denominador. Por el contrario, si se acepta que H_0 es falsa, entonces el estadístico F sigue una distribución F con df_1 y df_2 grados de libertad y parámetro de no-centralidad

$$\lambda = \frac{N k_1 k_2 \beta_{11}^2}{\sigma_{b1}^2} \quad (26)$$

donde $k_1 = N_C / N$, $k_2 = N_E / N$ y $N = N_C + N_E$.

A pesar del interés que despierta el uso de los estudios aleatorizados entre los investigadores, en determinadas circunstancias la eficacia de los tratamientos deberá basarse en estudios controlados. En estos casos no es descartable que se incumpla el supuesto de homogeneidad de los componentes de varianza de nivel-1, de nivel-2 o de ambos niveles a la vez a través de los grupos. Para más detalles, véase el ejemplo descrito en Vallejo *et al.* (2015). Bajo el escenario más general (i.e., heterogeneidad en los niveles 1 y 2), el parámetro de no-centralidad λ se puede construir como sigue:

$$\lambda^* = \frac{N k_1 k_2 \beta_{11}^2}{\sigma_{b1(C)}^2 + \sigma_{b1(E)}^2} \quad (27)$$

donde $\sigma_{b1(C)}^2 = k_2 [12\sigma_{(C)}^2 / (T^3 - T) + \tau_{11}^{(C)}]$ y $\sigma_{b1(E)}^2 = k_1 [12\sigma_{(E)}^2 / (T^3 - T) + \tau_{11}^{(E)}]$.

Cuando se desee comparar más de dos grupos, el cálculo del parámetro de no centralidad se puede simplificar usando un método similar al discutido por Shieh (2003) en el marco del modelo lineal general multivariante. Específicamente,

$$\lambda = tr [(A \mathbf{V} \mathbf{A}')^{-1} (\mathbf{C} \mathbf{B} \mathbf{A}')' (\mathbf{C} \mathbf{M}^{-1} \mathbf{C}')^{-1} (\mathbf{C} \mathbf{B} \mathbf{A}')] \quad (28)$$

donde tr denota la traza de la matriz $[\times]$, $\mathbf{A} = (\mathbf{I}_{NG} \mid -\mathbf{I}_{NG})$ y $\mathbf{C} = (\mathbf{I}_{NG-1} \mid -\mathbf{I}_{NG-1})$ son matrices de contrastes entre sujetos con rango fila completo, \mathbf{I}_{NG} es un vector columna de unos, \mathbf{I}_{NG} es una matriz de identidad y el símbolo \mid representa la matriz aumentada que resulta de anexas las columnas de las matrices \mathbf{A} y \mathbf{C} . La matriz de valores esperados a través de las T mediciones, $\mathbf{B} = [\mu_{(C)0} \dots \mu_{(C)T-1}; \mu_{(E)0} \dots \mu_{(E)T-1}]$ se puede obtener fácilmente a partir de la ecuación (5) fijando $\beta_{00} = \beta_{01} = \beta_{10} = 0$, \mathbf{M} es una matriz diagonal que tiene por entradas diagonales el número de sujetos de los grupos, en nuestro caso el experimental y control, respectivamente, y la matriz \mathbf{V} se construye utilizando las ecuaciones (6) y (7). Si los componentes de varianza a través de los grupos son heterogéneos, entonces $\mathbf{V} = k_2 \mathbf{V}_{(C)} + k_1 \mathbf{V}_{(E)}$.

Dicho lo anterior, el método usado en este trabajo para aproximar la potencia asociada con la prueba del efecto del tratamiento se esboza en cuatro pasos:

1. Definir el nivel de significación α y el tamaño de muestra de los grupos control y experimental, a saber: N_C y N_E . Sin

pérdida de generalidad, podemos establecer que $\beta_{00} = \beta_{01} = \beta_{10} = 0$

2. Fijar el valor de los índices ρ_1 , d_r , r y k , determinar el valor de los parámetros σ^2 , τ_{00} , τ_{01} , τ_{11} y β_{11} y calcular el parámetro de no centralidad definido en las ecuaciones (26)-(28).
3. Especificar el valor crítico de la inversa de la función de distribución F central, es decir: $F_0 = \text{FINV}(1-\alpha, df_1, df_2)$.
4. Calcular la probabilidad de que la razón F exceda el valor crítico F_0 cuando H_0 es falsa. Bajo H_1 , la función de potencia asociada con la prueba F es dada por

$$\psi = 1 - \beta = P [F'(df_1, df_2, \lambda) > F_0] = 1 - P [F'(df_1, df_2, \lambda) < F_0] \quad (29)$$

donde $P [F'(df_1, df_2, \lambda) < F_0]$ es la función de distribución acumulativa para la F no central.

Determinación del tamaño de muestra

En esta sección se proporciona una expresión que nos informa del tamaño de muestra (N) requerido para probar la hipótesis nula $H_0 : \beta_{11} = 0$ (i.e., la interacción del tratamiento con el tiempo es nula o la diferencia entre el promedio de las tasas de crecimiento de los grupos experimental y control no difiere de cero) frente a la hipótesis alternativa $H_1 : \beta_{11} \neq 0$ con un potencia $100\%(1-\beta)$ y un nivel de confianza $100\%(1-\alpha)$.

La hipótesis nula se rechaza si el estimador consistente de β_{11} excede el valor crítico ($\hat{\beta}_{11} > c$). De acuerdo con Amatya, Bhaumik y Gibbons (2013), este valor define el límite entre las regiones de aceptación y rechazo y es determinado bajo las dos condiciones siguientes:

$$P(\hat{\beta}_{11} > c = Z_{1-(\alpha/2)} \sqrt{(N k_1 k_2)^{-1} \sigma_{b1}^2} \mid H_0 \text{ verdadera}) = \alpha \quad (30)$$

$$P(\hat{\beta}_{11} > c = \beta_{11} - Z_{1-\beta} \sqrt{(N k_1 k_2)^{-1} \sigma_{b1}^2} \mid H_1 \text{ verdadera}) = 1 - \beta \quad (31)$$

Igualando las ecuaciones (30) y (31), dado que c se asume idéntico bajo ambas hipótesis estadísticas, y resolviendo para N obtenemos (para más detalles, véase Apéndice D)

$$N = \frac{[Z_{1-(\alpha/2)} \sqrt{\sigma_{b1}^2} + Z_{1-\beta} \sqrt{\sigma_{b1}^2}]^2}{\beta_{11}^2 k_1 k_2} \quad (32)$$

donde $Z_{1-(\alpha/2)}$ y $Z_{1-\beta}$ son los percentiles $100\% (1-\alpha/2)$ y $100\% (1-\beta)$ de la distribución normal estándar para una prueba bilateral, α denota el nivel de significación y β la probabilidad de error Tipo II, la cual es igual a 1 menos la potencia.

Reescribiendo la varianza del estimador MCO de la tasa de cambio en términos de sus dos fuentes de variabilidad, produce un sustituto útil para la ecuación (32)

$$N = \left(\frac{[Z_{1-(\alpha/2)} + Z_{1-\beta}]^2}{\beta_{11}^2 k_1 k_2} \right) \left(\frac{12 f^2 \sigma^2}{T^3 - T} + \tau_{11} \right) \quad (33)$$

En el caso de que los componentes de varianza no sean constantes, sino que varíen en función de la participación de los

sujetos en el programa (grupo experimental vs. grupo control), entonces la fórmula para aproximar el tamaño de muestra bajo el escenario más general (i.e., heterogeneidad en los niveles 1 y 2) es dada por

$$N^* = \left(\frac{[Z_{1-(\alpha/2)} + Z_{1-\beta}]^2}{\beta_{11}^2 k_1 k_2} \right) \left(\frac{12f^2 [k_2 \sigma_{(C)}^2 + k_1 \sigma_{(E)}^2]}{T^3 - T} + k_2 \tau_{11}^{(C)} + k_1 \tau_{11}^{(E)} \right) \quad (34)$$

Hasta ahora hemos asumido que no existían datos faltantes intermitentes en la variable de respuesta, ni tampoco participantes que abandonasen prematuramente la investigación. Sin embargo, la pérdida de información es un problema generalizado en estudios de carácter longitudinal. La presencia de valores perdidos puede sesgar las estimaciones de los parámetros, incrementar las tasas de error, reducir la potencia estadística y limitar la representatividad de los hallazgos. Cuando un investigador prevea la pérdida de datos, deberá aumentar el tamaño muestral requerido en el caso completo, con el fin de garantizar que la validez de las inferencias y el alcance de las conclusiones no sean cuestionables, o, cuando menos, lo sean en la menor medida posible.

Como señalan Fitzmaurice *et al.* (2011), el impacto de los datos faltantes es difícil de cuantificar con precisión, ya que depende del mecanismo generador del patrón de no respuestas, y en este caso las fórmulas empleadas anteriormente para computar la varianza en las pendientes de los participantes, $Var(\hat{b}_{1i})$, pueden dejar de ser aplicables o no ser realistas. Un método que puede resultar conservador, pero que permite modelar razonablemente los abandonos prematuros del estudio, al respecto, O’Kelly y Ratitch (2014) señalan que en los estudios relacionados con el sector salud se produce un porcentaje sustancialmente menor abandonos temporales que definitivos, consiste en dividir la matriz V_i de la ecuación (24) por la matriz postulada para las pérdidas \mathbf{P} . Por tanto, bajo esta situación la variabilidad del estimador de la tasa de cambio puede ser derivada desde la expresión siguiente:

$$\sigma_{b_i}^2 \approx \left[\mathbf{c}' \left(\frac{\mathbf{V}_i}{\mathbf{P} \mathbf{a}_p^2} \right) \mathbf{c} \right] f^{2p} - m_p \quad (35)$$

La elección de la matriz \mathbf{P} dependerá del modelo de pérdida que se quiera resaltar. Por ejemplo, si la tasa de abandonos se asume constante en cada ocasión de medida, pero teniendo en cuenta las deserciones acaecidas en las ocasiones previas, la matriz \mathbf{P} adoptaría la forma que sigue:

$$\mathbf{P} = \begin{bmatrix} \rho^0 & \sqrt{\rho} & \sqrt{\rho^2} & \sqrt{\rho^3} & \Lambda & \sqrt{\rho^{t-1}} \\ \sqrt{\rho} & \rho^1 & \sqrt{\rho^3} & \sqrt{\rho^4} & \Lambda & \sqrt{\rho^t} \\ \sqrt{\rho^2} & \sqrt{\rho^3} & \rho^2 & \sqrt{\rho^5} & \Lambda & \sqrt{\rho^{t+1}} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \Lambda & \mathbf{M} \\ \sqrt{\rho^{t-1}} & \sqrt{\rho^t} & \sqrt{\rho^{t+1}} & \sqrt{\rho^{t+2}} & \Lambda & \rho^{t-1} \end{bmatrix}$$

donde ρ es la tasa de permanencia en el estudio. Para el caso homogéneo, la fórmula (32) proporciona resultados similares a los obtenidos utilizando los métodos descritos en Hedeker, Gibbons y Waternaux (1999).

Ejemplos ilustrativos

Para facilitar el uso de las fórmulas correspondientes al análisis de potencia y a la determinación del tamaño de muestra, nos apoyaremos en los datos de un estudio longitudinal llevado a cabo por Núñez, Rosário, Vallejo y González-Pienda (2013). Entre otras cosas, estos autores examinaron si la aplicación de un programa de mentoría mejoraba la autoeficacia para el aprendizaje autorregulado. El desempeño de dicho programa fue probado en 94 estudiantes de sexto de Primaria asignados al azar a dos condiciones experimentales, los cuales fueron evaluados al inicio del estudio y al cabo de 3, 6 y 9 meses. Por tanto, si medimos el paso del tiempo trimestralmente, este estudio implica una observación por unidad de tiempo ($f = 1$), una duración de tres trimestres ($D = 3$) y cuatro de medidas repetidas ($T = fD + 1 = 4$).

Primeramente, comenzaremos calculando la potencia estadística de la prueba del efecto tratamiento del estudio descrito en el apartado anterior. Tras analizar los datos con el módulo PROC MIXED del programa SAS, se obtuvieron las estimaciones siguientes: $\hat{\sigma}^2 = .45$, $\hat{\tau}_{00} = .2$, $\hat{\tau}_{01} = .035$, $\hat{\tau}_{11} = .012$ y $\hat{\beta}_{11} = .165$. A partir de estos resultados, calculamos la varianza de la pendiente y el parámetro de no-centralidad definidos en las ecuaciones (24) y (26). Conocido el parámetro de no-centralidad λ , el valor de potencia se puede obtener consultando las tablas de la distribución F no central (p.e., Ato & Vallejo, 2015) o mediante algún programa estadístico. Asumiendo un error $\alpha = 0.05$, con 1 y 280 grados de libertad y $\hat{\lambda} = 6.2$, la potencia observada es $\hat{\varphi} \cong 0.70$; un valor ligeramente inferior al criterio recomendado de 0.80. Eliminando 56 datos conforme a un patrón aleatorio monótono, lo que supone un 15% de abandonos, $\hat{\lambda} = 4.8$ y $\hat{\varphi} \cong 0.59$. En presencia de errores con varianzas heterogéneas a través de los grupos del tratamiento (p.e., relación 1:3), $\hat{\lambda} = 3.1$ y $\hat{\varphi} \cong 0.42$. Así pues, cuanto mayor sea el desgaste de muestra y/o el grado de heterogeneidad entre los grupos, menor será potencia.

Seguidamente, determinamos el tamaño muestral que habría permitido obtener la potencia deseada ($\varphi = 1 - \beta = 0.80$). A partir de la ecuación (33), con $Z_{1-(\alpha/2)} = 1.96$, $Z_{1-\beta} = .84$, $\hat{\beta}_{11} = .165$, $k_1 = k_2 = .5$, $f = 1$, $\hat{\sigma}^2 = .45$, $\hat{\tau}_{11} = 0.12$ y $T = 4$, tenemos que el N requerido para rechazar H_0 en el 80% de los casos con un nivel de significación del 5% es 120. De haber sido los errores de nivel 1 y 2 heterogéneos (relación 1:3), el valor de N proporcionado por la ecuación (34) ascendería a 238. Es importante señalar que la potencia mejoraría notablemente aumentado f , e indirectamente, T . Por ejemplo, doblando el valor de la frecuencia ($f = 2$) pero manteniendo constante el tamaño de muestra ($N = 94$), la duración del estudio ($D = 3$) y el valor del efecto del tratamiento sin estandarizar ($\hat{\beta}_{11} = .165$), la potencia observada para el caso homogéneo completo hubiese sido $\hat{\varphi} = 0.87$, en vez del valor obtenido cuando $f = 1$ (i.e., $\hat{\varphi} = 0.70$). Un resultado similar se obtendría prolongando D y manteniendo constante f .

Hasta ahora se ha asumido que el investigador conoce de antemano el valor de los parámetros incluidos en los modelos de las ecuaciones (1)-(3), no obstante, en muchas ocasiones carece de dicha información. Para ilustrar cómo calcular el tamaño muestral de la manera más fácil y realista posible, comenzaremos especificando el valor de los índices estadísticos definidos en las ecuaciones (12)-(15); asumiendo, sin pérdida de generalidad, que la varianza de la respuesta inicial es igual a uno ($\tau_{00} + \sigma^2 = 1$). Los valores conjeturados son los siguientes: $\rho_1 = 0.4$ (se espera una fiabilidad relativamente baja, pues con frecuencia las respuestas

proviene de cuestionarios), $d_T = 0.50$ (se espera un tamaño del efecto estandarizado medio), $r = 0.5$ (se espera una correlación moderadamente alta entre los residuales de nivel 2) y $k = 2$ (se espera que la varianza de las respuestas aumente ligeramente con el paso del tiempo). Usando las ecuaciones (17) y (18) se obtiene el valor de los parámetros $\tau_{11} = 0.60$ y $\beta_{11} = .236$. Sustituyendo los valores de estos parámetros en la ecuación (33), se logra el tamaño de muestra necesario para rechazar H_0 , cuando es falsa, en el 80% de las ocasiones; en concreto, $N = 102$. Si el tamaño del efecto postulado fuese bajo, $d_T = 0.20$, N ascendería a 634, mientras que descendería hasta 40 de haberse postulado un tamaño alto, $d_T = 0.80$.

Finalmente, queremos resaltar la exactitud de las fórmulas propuestas, ya que las estimaciones empíricas obtenidas empleando 5.000 repeticiones de Monte Carlo son virtualmente idénticas a las teóricas. Por ejemplo, cuando el tamaño del efecto es alto ($d_T = 0.80$; $N = 40$), el valor de la potencia teórica es $\varphi = 0.803$ y el de la potencia empírica $\hat{\varphi} = 0.797$. La diferencia existente entre ambas potencias es igual a .006, la cual no excede los límites del intervalo confianza al 95% $\pm 1.96\sqrt{(0.8 \times 0.2) / 5000} = \pm 0.011$. Similares resultados fueron obtenidos para el tamaño del efecto medio ($d_T = 0.50$; $N = 102$; $\varphi = 0.802$; $\hat{\varphi} = 0.799$) y bajo ($d_T = 0.20$; $N = 634$; $\varphi = 0.800$; $\hat{\varphi} = 0.797$).

Discusión

En este trabajo hemos extendido el procedimiento desarrollado por Usami (2014) para evaluar la potencia estadística de la prueba del efecto del tratamiento a situaciones que no impliquen asumir datos completos, ni homogeneidad de los errores a través de los grupos del tratamiento. A diferencia de Usami (2014), no hemos seguido el procedimiento de aproximar potencia desarrollado por Satorra y Saris (1985) en el contexto de los modelos de ecuaciones estructurales sino que, al igual que en el trabajo de Heo *et al.* (2013), la función de potencia ha sido derivado dentro del contexto de los modelos de regresión multinivel por el método MCO.

Como demuestran los resultados de las simulaciones descritas en los ejemplos, la principal conclusión de este trabajo es que

las potencias empíricas basadas en componentes de varianza desconocidos fueron virtualmente idénticas a las potencias teóricas derivadas a partir del uso de índices estadísticos transformados. Este descubrimiento confirma la exactitud de la función de potencia derivada en la ecuación (29) usando el método MCO para detectar un efecto de intervención al final del estudio. De este resultado se desprende la exactitud del enfoque propuesto para calcular el tamaño de muestra que nos permite rechazar la hipótesis nula cuando, de hecho, es falsa con el criterio de potencia estipulado en las distintas situaciones investigadas.

Además de lo dicho en el apartado anterior, la evidencia obtenida a partir de las ecuaciones (33) y (34) respalda, en términos generales, otras tres conclusiones de interés. Una, la varianza de la pendiente ejerce un notable efecto en el tamaño de muestra, sobre todo cuando el número de medidas repetidas es reducido (i.e., $T \leq 4$); a partir de $T = 8$ el valor de T se deja sentir con menor intensidad. Dos, el tamaño de muestra requerido aumenta conforme lo hace el error de medida, si bien dicho aumento se desacelera a medida que se incrementa T . Tres, a medida que aumenta T disminuye el tamaño de muestra. Nótese que T aparece en el numerador y en el denominador de las fórmulas, no obstante, el efecto positivo es mayor que el efecto negativo.

Para finalizar, destacar que el trabajo proporciona la maquinaria para que los investigadores puedan generalizar el método descrito a situaciones donde los cambios no sean lineales. También sería fácil de extender método propuesto a un número arbitrario de grupos y variables explicativas. No obstante, somos conscientes de que presenta algunas limitaciones. La más importante tiene que ver con su posible generalización a modelos multinivel de tres niveles con varianzas heterogéneas y datos perdidos.

Agradecimientos

Agradecemos los comentarios del editor y revisores porque permitieron una mejora sustancial del presente manuscrito. Este documento ha sido financiado por el Ministerio de Ciencia e Innovación español (Ref: PSI-2011-23395/PSI2015-6730-P).

Apéndice A

$$\begin{aligned} \text{Var}(Y_{ij} | X_{1j}, \dots, X_{pij}) &= E[Y_{ij} - E(Y_{ij})]^2 = E(Y_{ij})^2 - E[(Y_{ij})]^2 \\ &= E\left(u_{0j} + \sum_{p=1}^P u_{pj} X_{pj} + e_{ij}\right)^2 \\ &= E\left(u_{0j}^2 + \sum_{p=1}^P u_{pj}^2 X_{pj}^2 + 2 \sum_{p=1}^P u_{0j} u_{pj} X_{pj} + e_{ij}^2\right) \\ &= E(u_{0j}^2) + \sum_{p=1}^P X_{pj}^2 E(u_{pj}^2) + 2 \sum_{p=1}^P X_{pij} E(u_{0j} u_{pj}) + E(e_{ij}^2) \\ &= \tau_{00} + 2 \sum_{p=1}^P \tau_{0p} X_{pij} + \sum_{p=1}^P \tau_{pp} X_{pij}^2 + \sigma^2 \\ \text{Cov}(Y_{ij}, Y_{i'j} | X_{1ij}, \dots, X_{pij}) &= E[(Y_{ij} - E(Y_{ij}))(Y_{i'j} - E(Y_{i'j}))] = E(Y_{ij} Y_{i'j}) - E(Y_{ij})E(Y_{i'j}) \\ &= E\left[\left(u_{0j} + \sum_{p=1}^P u_{pj} X_{pj} + e_{ij}\right)\left(u_{0j} + \sum_{p=1}^P u_{pj} X_{pi'j} + e_{i'j}\right)\right] \\ &= E(u_{0j}^2) + \sum_{p=1}^P X_{pij} E(u_{0j} u_{pj}) + \sum_{p=1}^P X_{pi'j} E(u_{0j} u_{pj}) + \sum_{p=1}^P (X_{pij} X_{pi'j}) E(u_{pj}^2) \\ &= \tau_{00} + \sum_{p=1}^P \tau_{0p} (X_{pij} + X_{pi'j}) + \sum_{p=1}^P \tau_{pp} (X_{pij} X_{pi'j}) \end{aligned}$$

Apéndice B

$$\begin{aligned}
 r &= \frac{\tau_{01}}{\sqrt{\tau_{00}}\sqrt{\tau_{11}}}, y \quad k = \frac{\tau_{00} + 2(T-1)\tau_{01} + (T-1)^2\tau_{11} + \sigma^2}{\tau_{00} + \sigma^2} \\
 &\Rightarrow \begin{cases} k(\tau_{00} + \sigma^2) = \tau_{00} + 2(T-1)\tau_{01} + (T-1)^2\tau_{11} + \sigma^2 \\ \tau_{11} = \frac{\tau_{01}^2}{r^2\tau_{00}} \end{cases} \\
 &\Rightarrow (k-1)(\tau_{00} + \sigma^2) = 2(T-1)\tau_{01} + (T-1)^2\tau_{11} = 2(T-1)\tau_{01} + \frac{(T-1)^2\tau_{01}^2}{r^2\tau_{00}} \\
 &\Rightarrow (T-1)^2\tau_{01}^2 + 2(T-1)r^2\tau_{00}\tau_{01} - r^2\tau_{00}(k-1)(\tau_{00} + \sigma^2) = 0 \\
 \tau_{01} &= \frac{-2(T-1)r^2\tau_{00} + \sqrt{4(T-1)^2r^4\tau_{00}^2 + 4(T-1)^2r^2\tau_{00}(k-1)(\tau_{00} + \sigma^2)}}{2(T-1)^2} \\
 &= \frac{-r^2\tau_{00} + r\sqrt{r^2\tau_{00}^2 + \tau_{00}(k-1)(\tau_{00} + \sigma^2)}}{T-1} \\
 \tau_{11} &= \frac{\tau_{01}^2}{r^2\tau_{00}} = \frac{r^4\tau_{00}^2 + r^2(r^2\tau_{00}^2 + \tau_{00}(k-1)(\tau_{00} + \sigma^2)) \pm 2r^3\tau_{00}\sqrt{r^2\tau_{00}^2 + \tau_{00}(k-1)(\tau_{00} + \sigma^2)}}{(T-1)^2r^2\tau_{00}} \\
 &= \frac{2r^2\tau_{00}(k-1)(\tau_{00} + \sigma^2) \pm 2r\sqrt{r^2\tau_{00}^2 + \tau_{00}(k-1)(\tau_{00} + \sigma^2)}}{(T-1)^2}
 \end{aligned}$$

Apéndice C

Si asumimos que las respuestas de los sujetos son independientes, la varianza del numerador de ecuación (19) se puede descomponer como sigue:

$$\text{Var}\left(\sum_{i=1}^{N_j} \sum_{t=1}^T (X_{it} - \bar{X}_i) Y_{it}\right) = A + B$$

Desarrollando los términos del segundo miembro de la expresión anterior encontramos,

$$\begin{aligned}
 A &= \sum_{i=1}^{N_j} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2 \text{Var}(Y_{it}) \\
 &= \sum_{i=1}^{N_j} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2 (\tau_{00} + 2\tau_{01}X_{it} + X_{it}^2\tau_{11} + \sigma^2) \\
 &= \sum_{i=1}^{N_j} \sum_{t=1}^T [(X_{it} - \bar{X}_i)^2\tau_{00} + (X_{it} - \bar{X}_i)^2 2X_{it}\tau_{01} + (X_{it} - \bar{X}_i)^2 X_{it}^2\tau_{11} + (X_{it} - \bar{X}_i)^2\sigma^2] \\
 &= (\tau_{00} + \sigma^2)N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2 + 2\tau_{01}N_j \sum_{i=1}^T X_{it} (X_{it} - \bar{X}_i)^2 + \tau_{11}N_j \sum_{i=1}^T X_{it}^2 (X_{it} - \bar{X}_i)^2 \\
 B &= \sum_{i=1}^{N_j} \sum_{t=1}^T \sum_{t' \neq t} (X_{it} - \bar{X}_i)(X_{it'} - \bar{X}_i) \text{Cov}(Y_{it}, Y_{it'}) \\
 &= \sum_{i=1}^{N_j} \sum_{t=1}^T \sum_{t' \neq t} Z_t Z_{t'} [\tau_{00} + (X_{it} + X_{it'})\tau_{01} + X_{it}X_{it'}\tau_{11}] \\
 &= \tau_{11}N_j \sum_{i=1}^T \sum_{t' \neq t} (X_{it} - \bar{X}_i)(X_{it'} - \bar{X}_i) X_{it} X_{it'} - \tau_{00}N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2 - 2\tau_{01}N_j \sum_{i=1}^T X_{it} (X_{it} - \bar{X}_i)^2
 \end{aligned}$$

De lo que sigue que:

$$\begin{aligned}
 \text{Var}\left(\sum_{i=1}^{N_j} \sum_{t=1}^T (X_{it} - \bar{X}_i) Y_{it}\right) &= (\tau_{00} + \sigma^2 - \tau_{00})N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2 + 2\tau_{01}N_j \sum_{i=1}^T X_{it} (X_{it} - \bar{X}_i)^2 - \\
 &2\tau_{01}N_j \sum_{i=1}^T X_{it} (X_{it} - \bar{X}_i)^2 + \tau_{11}N_j \left[\sum_{i=1}^T X_{it}^2 (X_{it} - \bar{X}_i)^2 + \sum_{i=1}^T \sum_{t' \neq t} (X_{it} - \bar{X}_i)(X_{it'} - \bar{X}_i) X_{it} X_{it'} \right] \\
 &= \sigma^2 N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2 + \tau_{11}N_j \left[\sum_{i=1}^T X_{it}^2 (X_{it} - \bar{X}_i)^2 + \sum_{i=1}^T \sum_{t' \neq t} (X_{it} - \bar{X}_i)(X_{it'} - \bar{X}_i) X_{it} X_{it'} \right]
 \end{aligned}$$

$$\begin{aligned}
 &= \sigma^2 N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2 + \tau_{11} N_j \sum_{i=1}^T \sum_{t=1}^T (X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i) X_{it} X_{it} \\
 &= \sigma^2 N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2 + \tau_{11} N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^4
 \end{aligned}$$

Por tanto, la varianza de $\hat{\delta}_j$ es

$$\begin{aligned}
 \text{Var}(\hat{\delta}_j) &= \frac{\text{Var}\left(\sum_{i=1}^{N_j} \sum_{t=1}^T (X_{it} - \bar{X}_i) Y_{it}\right)}{\text{Var}\left(\sum_{i=1}^{N_j} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2\right)} = \frac{\sigma^2 N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2}{N_j^2 \sum_{i=1}^T (X_{it} - \bar{X}_i)^4} + \frac{\tau_{11} N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^4}{N_j^2 \sum_{i=1}^T (X_{it} - \bar{X}_i)^4} \\
 &= \frac{\sigma^2}{N_j \sum_{i=1}^T (X_{it} - \bar{X}_i)^2} + \frac{\tau_{11}}{N_j} = \frac{1}{N_j} \left(\frac{\sigma^2}{\sum_{i=1}^T (X_{it} - \bar{X}_i)^2} + \tau_{11} \right)
 \end{aligned}$$

Finalmente, utilizando las propiedades de la varianza cuando las variables son independientes, se tiene que

$$\text{Var}(\hat{\beta}_{11}) = \text{Var}(\hat{\delta}_1) + \text{Var}(\hat{\delta}_0) = \frac{4}{N} \left(\frac{\sigma^2}{\sum_{i=1}^T (X_{it} - \bar{X}_i)^2} + \tau_{11} \right)$$

Recuérdese que aquí N denota el número total de unidades de segundo nivel incluidas en el estudio, con $N/2$ sujetos en cada grupo.

Apéndice D

Partiendo de la ecuación fundamental de las dos situaciones muestrales:

$$\begin{aligned}
 Z_{1-\alpha/2} \sqrt{\sigma_b^2 / N k_1 k_2} &= \beta_{11} - Z_{1-\beta} \sqrt{\sigma_b^2 / N k_1 k_2} \\
 \beta_{11} &= Z_{1-\alpha/2} \sqrt{\sigma_b^2 / N k_1 k_2} + Z_{1-\beta} \sqrt{\sigma_b^2 / N k_1 k_2} \\
 \beta_{11} &= (Z_{1-\alpha/2} + Z_{1-\beta}) \sqrt{\sigma_b^2 / N k_1 k_2} \\
 \beta_{11}^2 N k_1 k_2 &= (Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma_b^2 \\
 N &= \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma_b^2}{\beta_{11}^2 k_1 k_2}
 \end{aligned}$$

References

- Amatya, A., Bhaumik, D., & Gibbons, R.D. (2013). Sample size determination for clustered count data. *Statistics in Medicine*, 32, 4162-4179.
- Ato, M., & Vallejo, G. (2015). *Diseños de investigación en Psicología*. Madrid: Pirámide.
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, 14, 43-53.
- Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. R., et al. (2005). Social information-processing skills training to promote social competence and prevent aggressive behavior in the third grades. *Journal of Consulting and Clinical Psychology*, 73, 1045-1055.
- Hedeker, D., Gibbons, R.D., & Waternaux, C (1999). Sample size estimation for longitudinal designs with attrition. *Journal of Educational and Behavioral Statistics*, 24, 70-93.
- Heo, M. (2014). Impact of subject attrition on sample size determinations for longitudinal cluster randomized clinical trials. *Journal of Biopharmaceutical Statistics*, 24, 507-522.
- Heo, M., Xue, X., & Kim, M. Y. (2013). Sample size requirements to detect and intervention by time interaction in longitudinal cluster randomized clinical trials with random slopes. *Computational Statistics and Data Analysis*, 60, 169-178.
- Núñez, J. C., Rosário, P., Vallejo, G., & González-Pienda, J. A. (2013). A longitudinal assessment of the effectiveness of a school-based mentoring program in middle school. *Contemporary Educational Psychology*, 38, 11-21.
- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2, 371-402.

- O'Kelly, M., & Ratitch, B. (2014). Analysis under missing-not-at-random assumptions. In M. O'Kelly & B. Ratitch (Eds.), *Clinical Trials with Missing Data. A Guide for Practitioners* (pp. 257-368). New York, NY: Wiley.
- Raudenbush S. W., Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6, 387-401.
- Roy, A., Bhaumik, D. K., Aryal, S., & Gibbons, R. D. (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, 63, 699-707.
- Rye, M. S., Pargament, K. I., Pan, W., Yingling, D. W., Shogren, K. A., & Ito, M. (2005). Can group interventions facilitate forgiveness of an ex-spouse? A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 73, 880-892.
- Satorra, A., & Saris, W. E. (1985). The power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50, 83-90.
- Shieh, G. (2003). A comparative study of power and sample size calculations for multivariate general linear models. *Multivariate Behavioral Research*, 38, 285-307.
- Usami, S. (2014). A convenient method and numerical tables for sample size determination in longitudinal-experimental research using multilevel models. *Behavior Research Methods*, 46, 1207-1219.
- Vallejo, G., Fernández, M. P., Cuesta, M., & Livacic-Rojas, P. E. (2015). Effects of modeling the heterogeneity on inferences drawn from multilevel designs. *Multivariate Behavioral Research*, 50, 73-88.
- Wänström, L. (2009). Sample sizes for two-group second-order latent growth curve models. *Multivariate Behavioral Research*, 44, 588-619.