Psicothema
2000. Vol. 12, Supl. nº 2, pp. 74-76
ISSN 0214 - 9915 CODEN PSOTEG

Test de la aleatorización en diseños de serie temporal interrumpida con grupo de control

María José Blanca
Universidad de Málaga

Se presenta un método alternativo de análisis estadístico para los datos provenientes de los diseños de serie temporal interrumpida basado en el test de la aleatorización. El procedimiento se basa en la asignación aleatoria de los sujetos a los grupos experimentales y determinación aleatoria del punto de intervención. Se discuten los estadísticos pertinentes para evaluar la hipótesis acerca del cambio conductual basado en el cambio de medias, o en su caso, de tendencias entre grupos y fases de tratamiento.

Randomization test for interrupted time-series design. The randomization test is discussed for the analysis of interrupted time-series design with a control group. Three random assignment procedures are suggested, which include assignment of subjects to treatments, point intervention, or both of them. Also different sensitive statistical tests are discussed for change of level or trends.

El diseño de serie temporal interrumpida con grupo de control implica la medida de la variable dependiente en varios puntos del tiempo antes y después de la intervención. Se han sugerido diferentes procedimientos estadísticos, tales como el análisis de regresión o el análisis de serie temporal. Sin embargo, en ocasiones, estos análisis no son correctos por el número reducido de observaciones o por el incumplimiento de alguno de los supuestos.

El objetivo de este trabajo es exponer un método alternativo basado en el test de la aleatorización, el cual ha sido desarrollado principalmente en el contexto de los diseños de caso único (Edgington, 1967, 1972; 1975, 1980a; Kratochwill y Levin, 1980; Levin, Marascuilo y Hubert, 1978; Wampold y Furlong, 1981; Wampold y Worsham, 1986; Marascuilo y Busk, 1988; Onghena, 1992), pero puede ser extendido a los diseños que incluyen grupos múltiples. El procedimiento a seguir se puede resumir de la siguiente manera:

a) Especificación del procedimiento de asignación aleatoria, el cual depende del diseño seguido y del número de unidades experimentales.

b) Especificación de la hipótesis nula y alternativa.

c) Elección del nivel de significación (usualmente establecido en .05) y del estadístico a utilizar. Éste debe estar basado en la hipótesis sobre el cambio conductual (cambio de nivel, cambio de tendencia, etc.)

d) Recogida de datos y cálculo del estadístico (estadístico observado).

e) Generación de la distribución del estadístico, la cual se usa para determinar la significación estadística asociada al resultado muestral. La distribución es generada a través del cálculo del estadístico en cada permutación de los datos muestrales.

f) Cálculo del nivel de significación asociado al estadístico, basado en la ratio entre el número de estadísticos igual o mayor (o menor) que el observado y el número total de permutaciones.

g) Decisión acerca de la aceptación o no de la hipótesis nula. Si el nivel de significación asociado al estadístico es menor o igual al especificado en el punto c, la hipótesis nula no será aceptada como probable.

El test de la aleatorización es aplicable a cualquier diseño siempre que se incluya algún procedimiento de asignación aleatoria. En el presente trabajo, se presentarán tres procedimientos de aleatorización para los diseños de serie temporal interrumpida con grupo de control, basados en la aleatorización de sujetos, aleatorización del punto de intervención o ambos conjuntamente.

Aleatorización de sujetos

El procedimiento más básico de aleatorización en el diseño de serie temporal interrumpida (DSTI) con grupo de control se puede centrar en la asignación aleatoria de los sujetos a los grupos experimentales, siguiendo el mismo procedimiento que para un diseño unifactorial al azar. El número de asignaciones posibles viene dado por:

$$P = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

donde N es el número total de sujetos, y n el número de sujetos en un grupo.

El estadístico seleccionado dependerá de los cambios esperados entre las fases y entre los dos grupos. Si se hipotetiza un cambio de nivel diferencial del pretest al postest entre el grupo control (GC) y grupo experimental (GE), el estadístico $D = (\bar{Y}_{22} - \bar{Y}_{21}) - (\bar{Y}_{12} - \bar{Y}_{11})$ puede ser útil, donde $\bar{Y}_{jk}$ representa la media de los sujetos en el j-ésimo grupo (j=1 para el GC y j=2 para el GE) y k la

Correspondencia: M. José Blanca
Facultad de Psicología
Universidad de Málaga
29071 Málaga (Spain)
E-mail: Blamen@uma.es

condición de la intervención (k=1 para el pretest y k=2 para el postest). Así, $\overline{Y}_{11}$ representa la media obtenida en el GC en el pretest y $\overline{Y}_{12}$ simboliza la media en el mismo grupo en el postest, y así sucesivamente. La lógica del análisis es demostrar que la diferencia observada es la mayor diferencia encontrada en los datos. Con ello, se podría afirmar que ha habido un efecto del tratamiento con una probabilidad igual o inferior al nivel de significación establecido. Por tanto, si la hipótesis predice que la media será superior en el postest, y este cambio será mayor en el grupo experimental que en el grupo control, la probabilidad de significación será:

$$p = \frac{\textit{Número de diferencias encontradas mayor o igual a } D}{\textit{Número total de permutaciones}}$$

Si la hipótesis del cambio conductual implica un cambio diferencial en la tendencia de los datos entre el pretest y postest con respecto al GC y GE, se deberían calcular otros estadísticos más sensibles. El procedimiento más simple, adaptando las recomendaciones de Marascuilo y Busk (1988) para el diseño AB, consiste en estimar cuatro líneas de regresión, una para cada fase y grupo, siguiendo el modelo

$$Y_t = \beta_0 + \beta_1 T_t + \varepsilon_y$$

donde Y_t es la media de los sujetos pertenecientes a cada grupo en cada ocasión medida y T_t es una variable *proxy* que representa la tendencia y es codificada como 1 para la primera observación, 2 para la segunda, 3 para la tercera y así sucesivamente. Una vez calculadas las pendientes de cada modelo de regresión, se sustituye el estadístico D por el estadístico $B = (\beta_{122} - \beta_{121}) - (\beta_{112} - \beta_{111})$, donde β_{111} es la pendiente del pretest en el grupo control, es la pendiente del pretest en el grupo experimental, y así sucesivamente.

Sin embargo, el análisis de regresión múltiple puede ser más adecuado y completo, tanto para el cambio de tendencia como de nivel, incluyendo todos los predictores pertinentes. El modelo general que introduce tanto diferencias en medias como en tendencia sería:

$$Y_t = \beta_0 + \beta_1 G_j + \beta_2 X_t + \beta_3 T_t + \beta_4 G_j X_t + \beta_5 G_j T_t + \beta_6 X_t T_t + \beta_7 G_j X_t T_t + \varepsilon_t$$

donde Y_t denota la observación de la variable dependiente en la serie temporal.

G_j es una variable *dummy* que representa al grupo de pertenencia (G_j =0 para el GC y G_j =1 para el GE).

X_t es una variable *dummy* que representa la intervención (X_t =0 para el pretest y X_t =1 para el postest).

T_t es la variable que representa la tendencia, como anteriormente se explicó.

β_0 es el término constante y el valor esperado de Y_t cuando todas las variables independientes son iguales a cero.

β_1 es interpretado como el cambio en Y_t entre el GC y el GE, manteniendo constante el resto de predictores.

β_2 es el cambio en Y_t entre el pretest y postest, manteniendo constante el resto de predictores.

β_3 representa el cambio en Y_t por cada unidad de cambio en T_t. Es decir, la pendiente de la serie temporal.

β_4 es el coeficiente de regresión de la interacción entre la pertenencia al grupo y la intervención y puede ser interpretado como el cambio diferencial de medias entre el pretest y postest en el GC y GE, manteniendo el resto de variables constantes.

β_5 es interpretado como el cambio en tendencia entre el GC y el GE, manteniendo el resto de variables constantes.

β_6 es el cambio de tendencia entre el pretest y postest, manteniendo el resto de variables constantes.

β_7 es el coeficiente de regresión de la interacción entre la pertenencia al grupo, intervención y tendencia. Se puede interpretar como el cambio diferencial en tendencia entre el pretest y postest en cada grupo, manteniendo el resto de variables constantes.

ε_t es el residual del modelo.

Si se esperan grupos equivalentes en el pretest, no cambio de tendencia ni nivel en el GC, pero un cambio de tendencia en el GE, el coeficiente de interés será β_7. Así, para calcular la significación de este coeficiente a partir del test de la aleatorización, en vez de utilizar la prueba convencional t, cuyos supuestos son más restrictivos, se ha de realizar un análisis de la regresión para cada permutación de datos, analizando la proporción de coeficientes con valores mayores o igual al observado en la combinación real de datos.

El procedimiento de aleatorización basado en la asignación aleatoria de los sujetos para los DSTI con grupo de control está limitado por su bajo poder estadístico si el número de sujetos es pequeño. Así, por ejemplo, si se incluyen 6 sujetos, entonces la región crítica para rechazar la hipótesis nula viene definida por un solo valor. Esto significa que si el estadístico observado en la combinación real de los datos no es el mayor (o menor) no hay posibilidad de rechazar la hipótesis nula. Si hay dos valores en la distribución del estadístico igual o mayor al observado, la probabilidad asociada será 2/20=.10, la cual no alcanza la significación estadística. Por tanto, el procedimiento será más útil cuanto mayor número de sujetos se incluya en el diseño.

Determinación aleatoria del punto de intervención

El procedimiento propuesto por Edgington (1980b) puede ser extendido a los diseños DSTI con grupo de control, donde se determina aleatoriamente el punto de intervención. Este procedimiento puede incluir limitaciones en la asignación, tal como restringir el segmento de observaciones donde se puede introducir el punto de intervención, para asegurar que las fases sean suficientemente largas como para obtener un patrón claro de datos.

El procedimiento general sigue los mismos pasos que en el caso previo, pero la distribución del estadístico es realizada tomando en cuenta la nueva asignación aleatoria. Así, no se permutan las observaciones pertenecientes a los sujetos, sino la longitud de las fases. Para una extensión de este punto, el lector puede consultar los trabajos de Edgington (1980b, 1995) y Wampold y Furlong, (1981).

Para alcanzar la significación estadística en este modelo, serían necesario al menos 20 puntos de observación posibles para introducir el tratamiento. Sin embargo, incluso en este caso, la prueba tiene baja potencia estadística ya que el estadístico observado debería ser el mayor (o menor) de la distribución para alcanzar la probabilidad de 0,05 (p=1/20=0,05). Por tanto, para aumentar la potencia, se debería incrementar el número de puntos para determinar la intervención. No obstante, este número de observaciones no es siempre posible en la investigación aplicada por lo que un procedimiento que combinara la asignación aleatoria de los sujetos así como el punto de intervención podría constituir una solución al problema.

Asignación aleatoria de los sujetos y determinación aleatoria del punto de intervención

La aplicación conjunta de los dos procedimientos de aleatorización en los DSTI con grupo de control aumentaría la potencia de la prueba. Así, el número de permutaciones posibles de los datos ascendería a $P = PI\binom{N}{n}$, siendo PI los puntos posibles de intervención. Supongamos que se desea formar aleatoriamente dos grupos con tres sujetos cada uno y determinar aleatoriamente la intervención entre la sesión 10 y 12 de la serie temporal (3 puntos posibles de intervención). El número de posibilidades vendrá dado por $P = 3\binom{6}{3} = (3)20 = 60$. La significación del estadístico pertinente, según se hipotetice un cambio diferencial de nivel o de tendencia o ambos, se realizaría calculando su valor para cada una de estas 60 permutaciones. Posteriormente, el nivel de significación vendrá dado por la proporción de estadísticos con valor igual o superior al observado. En este caso la potencia de la prueba se incrementa con respecto a los otros dos procedimientos, ya que el número de posibilidades es bastante superior.

Discusión

Se han presentado tres procedimentos de análisis estadístico basado en el test de la aleatorización, los cuales pueden constituir una solución en aquellos casos en los que los supuestos del análisis paramétrico no se satisfagan. El único requisito necesario radica en introducir algún tipo de aleatorización en el diseño: asignación aleatoria de los sujetos a los grupos, determinación aleatoria del punto de intervención, o ambos conjuntamente.

En general, el test de la aleatorización implica una planificación previa a la recogida de datos. El primer procedimiento tiene la ventaja de que permite introducir la intervención cuando el patrón de datos sea adecuado (e.g. estabilización de la línea de base en las unidades experimentales). En el contexto de los diseños de caso único, el hecho de que el patrón de datos no determine el punto de intervención ha sido considerado una limitación de los tests de la aleatorización (Ferron y Ware, 1994; Kazdin, 1980; Matyas y Greenwood, 1991; Onghena, 1992). Por tanto, este procedimiento soluciona el problema cuando la aplicación de la intervención dependa de los datos obtenidos en la línea de base.

Por otra parte, si la adjudicación aleatoria de los sujetos a los grupos no es posibles porque éstos están formados por grupos naturales, se podría seguir el segundo procedimiento expuesto consistente en la aleatorización del punto de intervención. Sin embargo, este procedimiento tiene la desventaja de que necesita un elevado número de observaciones que no siempre es posible en la investigación aplicada. Por otro lado, como ya se ha mencionado, no podrá ser seguido cuando la introducción del tratamiento dependa del patrón de datos encontrado en el pretest.

Finalmente, el empleo conjunto de los dos tipos de adjudicación aleatoria se convierte, desde un punto de vista metodológico, en el más adecuado ya que disminuye las amenazas contra la validez interna. Igualmente, esta clase de aleatorización incrementa el número de permutaciones y, consecuentemente, incrementa la potencia estadística.

En conclusión, cada procedimiento tiene sus propias ventajas y desventajas y es posible que en algunas situaciones ninguno de ellos pueda ser utilizado. En los casos donde se pueda introducir algún tipo de aleatorización, el investigador deberá seleccionar el procedimiento más adecuado y proceder a la recogida de datos de acuerdo con él. Igualmente, deberá seleccionar el estadístico más sensible que represente el cambio conductual esperado.

Referencias

Edgington, E.S. (1967). Statistical inference from N=1 experiments. *Journal of Psychology, 65*, 195-199.

Edgington, E.S. (1972). N=1 experiments: hypothesis testing. *The Canadian Psychologist*, *13*(2), 121-134.

Edgington, E.S. (1975). Randomization tests for one-subject operant experiments. *The Journal of Psychology*, *90*, 57-68.

Edgington, E.S. (1980a). *Randomization test*. New York: Marcel Dekker, Inc.

Edgington, E.S. (1980b). Random Assignment and statistical tests for one-subject experiments. *Behavioral Assessment*, *2*, 19-28.

Edgington, E.S. (1980c).Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics*, *5*(3), 235-251.

Edgington, E.S. (1995). *Randomization test (3ª ed. rev.).* New York: Marcel Dekker, Inc.

Ferron, J. y Ware, W. (1994). Using randomization tests with responsive single-case designs. *Behavior Research and Therapy*, *32*(7), 787-791.

Kazdin, A.E. (1980). Obstacles in using randomization tests in single-case experimentation. *Journal of Educational Statistics*, *5*(3), 253-260.

Kratochwill, T. R., y Levin, J. R. (1980). On the applicability of various data analysis procedures to the simultaneous and alternating treatment design in behavior therapy research. *Behavioral Assessment*, *2*, 353-360.

Levin, J.R., Marascuilo, L.A. y Hubert, L.J. (1978). N = nonparametric randomization test. En T.R. Kratochwill. *Single subject research: strategies for evaluating changes*. (pp.167-196) New York: Academic Press.

Marascuilo, L.A. y Busk, P.L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment, 10*, 1-28.

Matyas, T.A. y Greenwood, K.M. (1991). Problems in the estimation of autocorrelation in brief time series and some implications for behavioral data. *Behavioral Assessment*, *13*, 137-157.

Onghena, P. (1992). Randomization tests for extensions and variations of ABAB single-case experimental designs: A Rejoinder. *Behavioral Assessment, 14*, 153-171.

Onghena, P. y Edgington, E.S. (1994). Randomization tests for restricted alternating treatments designs. *Behavior Research and Therapy*, *32*(7), 783-786.

Wampold, B.E. y Furlong, M.J. (1981). Randomization tests in single-subject designs: illustrative examples. *Journal of Behavioral Assessment, 3*(4), 329-341.

Wampold, B.E. y Worsham, N.L. (1986). Randomization tests for multiple-baseline designs. *Behavioral Assessment*, *8*, 135-143.