

Análisis Estadístico de Diseños AB Replicados Mediante Series Temporales Agrupadas

G. Vallejo*¹, P. Livacic-Rojas y N. Conejo***

**Universidad de Oviedo, **Universidad Central de Chile*

RESUMEN

Las pruebas basadas en la aleatorización de estímulos se consideran actualmente procedimientos estadísticos muy adecuados para evaluar los datos obtenidos con diseños de línea de base múltiple. Especialmente, cuando los investigadores, además de determinar al azar los puntos de intervención, también permutan el orden en que los sujetos, conductas o ambientes reciben el programa de tratamiento. Sin embargo, para que una prueba de aleatorización se considere válida se requiere seleccionar al azar algún aspecto del diseño antes de registrar los datos, lo cual desafortunadamente no suele ser frecuente en la investigación clínica aplicada. En estos casos las pruebas de la aleatorización resultan inadecuadas y es preferible utilizar algún procedimiento alternativo en orden a garantizar la validez de las inferencias relativas a efectividad de los tratamientos. En el presente artículo se desarrolla un método de análisis de estos diseños basado en las series temporales interrumpidas agrupadas y se presenta un ejemplo para ilustrar la técnica propuesta. *Palabras clave:* Diseños AB, Series temporales agrupadas.

ABSTRACT

At present the randomization tests are considered as suitable statistical methods for evaluating the data from multiple baseline AB designs. Particularly, when the researchers, in addition to determining the intervention points at random, also randomly selects the order in which the person, behaviors, or situations are allocated to the treatment programs. However, in order to conduct a valid randomization test it is required to randomize some aspect of the design, which unfortunately usually is not frequent in the applied clinical investigation. This is a situation in which a randomized test is inadequate and some alternative procedure may be preferable in order to draw valid statistical inferences about treatment effects. In this article a pooled interrupted time series method is presented for the analysis of data based on these designs, and an example is given to illustrate the corresponding technique. *Key words:* AB designs, Pooled time series.

¹La correspondencia concierne con este artículo puede enviarse a Guillermo Vallejo Seco, Departamento de Psicología, Universidad de Oviedo, Plaza de Feijóo s/n, 33003 Oviedo. E-mail: gvallejo@correo.uniovi.es. 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).

Por razones diversas, los psicólogos que trabajan en los ámbitos aplicados saben bien que cada día están más obligados a evaluar la efectividad de los tratamientos que aplican a los pacientes que demandan sus servicios. Al igual que Harris y Jenson (1984), no albergamos ninguna duda que los profesionales que desarrollan su práctica clínica con rigor metodológico gozan de un mayor éxito social y contribuyen más al desarrollo de los conocimientos psicológicos, dado que los resultados obtenidos son más fáciles de describir, interpretar e inclusive comunicar. Por esta razón, siempre que se aplique algún programa de intervención o tratamiento debería interesar descubrir, no sólo si existe relación entre el tratamiento que se aplica y la conducta que se registra, sino también, precisar cuán probable resulta dicha relación y qué importancia práctica tiene. Para garantizar la calidad de las inferencias causales, conviene abordar el éxito o fracaso de los programas de intervención complementando los criterios terapéuticos con potentes diseños de investigación conductual y adecuados métodos de evaluación de datos.

Durante las últimas décadas una de las estrategias que más ha granjeado el interés de los investigadores que trabajan en el contexto de la modificación de conducta ha sido la basada en los diseños de línea base múltiple (Barlow y Hersen, 1984). Estos diseños representan una alternativa inequívoca para aquellas situaciones en las cuales el investigador es incapaz de revertir las fases o eliminar el tratamiento dentro de su estudio, bien sea por limitaciones éticas o por limitaciones prácticas. Esta estrategia conlleva aplicar el mismo tratamiento de forma escalonada o secuencial a diferentes condiciones de línea base (sujetos, conductas o ambientes), coleccionar datos simultáneamente y determinar si una condición que no está bajo el control del tratamiento cambia a lo largo de otra que sí lo está. El razonamiento lógico que subyace a los diseños de línea base múltiple consiste en registrar los datos concurrentemente para poder desestimar posibles hipótesis rivales, en el caso que el cambio conductual acontezca a raíz de presentar el tratamiento en el punto de interés. De acuerdo con la estrategia propia de estos diseños, la relación entre el tratamiento y la conducta únicamente se inferirá si se produzcan cambios escalonados en función de la aplicación secuencial del tratamiento a los distintos sujetos, conductas o ambientes. Por ejemplo, si tras implementar un tratamiento en una de las diversas líneas de base del diseño se produce un cambio, mientras que las restantes líneas de base se mantienen estables, o a lo sumo tan sólo se observan pequeñas oscilaciones, el investigador cuenta con suficientes garantías para establecer relaciones causa-efecto. Con este formato se ejerce un alto grado de control del factor historia, lo que en palabras de Shadish, Cook y Campbell (2002), sería una de las principales amenazas contra la inferencia causal, al involucrar tanto en las comparaciones entre las fases como dentro de las fases.

Para evaluar los datos obtenidos en diseño de línea base múltiple en cualquiera de sus modalidades básicas existentes (entre sujetos, entre respuestas y entre ambientes), se han propuesto diversos métodos de análisis estadísticos no paramétricos (prueba de rangos, pruebas basadas en la aleatorización de estímulos y pruebas basadas en la aleatorización de respuestas) para complementar al tradicional análisis gráfico de los datos. Dentro de los procedimientos estadísticos, que dicho sea de paso va a ser donde

se va centrar nuestro trabajo, los métodos desarrollados por Wampold y Worsham (1986) y por Marascuilo y Busk (1988) son los que se han configurado como la alternativa de análisis más recomendable dada su validez y extremada versatilidad.

Wampold y Worsham (1986) desarrollaron una prueba estadística no paramétrica basada en el muestreo aleatorio de estímulos que sirve para analizar datos recogidos simultáneamente en varios sujetos, conductas o ambientes tras la aplicación secuencial del tratamiento. El diseño de línea base múltiple aleatorizado de Wampold y Worsham sirve a los mismos propósitos que el diseño tradicional, aunque a diferencia de éste el punto o intervalo de tiempo en que se aplica el programa de intervención en cada una de las diferentes líneas de base (secuencias de fases AB que componen el diseño) no se establece sistemáticamente a partir del patrón de datos que emerge del experimento. Por el contrario, en el modelo de estos autores, el investigador decide por adelantado las sesiones en las que será introducido el tratamiento y, después, emplea la asignación aleatoria para determinar el orden de aparición del tratamiento en las diversas series. En otras palabras, bajo el procedimiento de Wampold y Worsham se fijan los puntos de intervención para las diferentes líneas de base sin garantizar que los datos permanecerán estables entre los tiempos de asignación e intervención. De esta forma, lo más que se hace es seleccionar puntos de intervención (intencionalmente en su caso) que no coincidan con las puntuaciones iniciales ni con las puntuaciones finales de las diferentes series. Posteriormente, se selecciona al azar el orden en el cual los sujetos, conductas o ambientes recibirán el tratamiento en las sesiones prefijadas.

Para evaluar la efectividad de la intervención con el modelo de Wampold y Worsham, se compara el estadístico de prueba (p.e., la diferencia entre las medias de las fases A y B sumada a través de diferentes líneas de base del diseño) con la distribución de aleatorización obtenida tras calcular el valor del estadístico para todos los posibles órdenes en que los sujetos, conductas o ambientes podrían haber sido asignados al azar a los puntos de intervención predeterminados. De acuerdo con este formato, si se cuenta con tres sujetos a los que se planifica observar a lo largo de veinte sesiones, el tratamiento se puede asignar al azar de 6 maneras diferentes; mientras que si se dispone de cinco, el número de secuencias de tratamiento que podrían ocurrir conforme el principio del azar sería de 120. Por tanto, para una prueba unidireccional donde bajo hipótesis alternativa se predice que el promedio de respuestas obtenidas en la condición de tratamiento (fase B) será mayor que el promedio de respuestas obtenidas bajo la condición de control (fase A), la probabilidad que el estadístico de prueba sea el más largo de los seis o de los ciento veinte valores de los que consta la distribución de aleatorización del estadístico será $1/6$ ó $1/120$, respectivamente. Obviamente, una vez determinada la distribución de aleatorización es relativamente sencillo calcular el valor de probabilidad (valor p) o proporción de valores de la distribución de tamaño igual o superior al del valor muestral observado. Si el valor p es menor o igual que el nivel de significación especificado antes de recolectar los datos, se rechaza la hipótesis nula y se concluye a favor del efecto predicho en la hipótesis alternativa. Repárese que con esta prueba el investigador no siempre puede concluir si la intervención ha resultado significativa desde el punto de vista estadístico.

Como se desprende de lo expuesto en el apartado anterior, resulta evidente que

la potencia de la distribución de aleatorización obtenida con el modelo propuesto por Wampold y Worsham (1986) será pequeña cuando el número de sujetos, conductas o ambientes sea inferior a cinco. Para solucionar este problema, Marascuilo y Busk (1988) sugieren seleccionar al azar los puntos de intervención, en vez de prefijarlos intencionalmente de antemano.

De acuerdo con el modelo propuesto por Marascuilo y Busk (1988), la distribución de aleatorización no sólo se obtiene permutando el orden en que los sujetos, conductas o ambientes reciben el tratamiento, sino que, además se requiere localizar para cada sujeto (conducta o ambiente) todos los posibles puntos de intervención que pueden ser elegidos al azar. Por lo tanto, en una serie que constase de diez sesiones de observación, bajo el supuesto de establecer cuatro medidas en la condición de control y otras cuatro en la condición experimental, se podrían seleccionar al azar con idéntica probabilidad tres puntos de intervención diferentes. En concreto, los correspondientes a las sesiones quinta, sexta y séptima. De este modo, un diseño con tres líneas de base que permita seleccionar uno de los tres puntos de intervención independientemente para cada sujeto, conducta o ambiente, generará 3^3 asignaciones aleatorias posibles y, por consiguiente, la región crítica quedará definida por 1.36 sucesos ($.05 \times 27$). En cambio, si se hubiese dispuesto de diez puntos de intervención se tendrían 10^3 asignaciones aleatorias y la región crítica quedaría definida por 50 sucesos ($.05 \times 1000$). Obviamente, cuando se cuente con un número relativamente elevado de sujetos y/o de puntos de intervención, comparar el estadístico de prueba con la distribución de aleatorización construida a partir de todas las posibles disposiciones de los datos puede resultar poco menos que prohibitivo. Afortunadamente, además de poder aproximar la distribución muestral del estadístico de prueba mediante alguna de las distribuciones muestrales estándar (p.e., la normal), también cabe la posibilidad de utilizar permutaciones para una muestra aleatoria de datos, en lugar de realizar todas las permutaciones posibles de los datos y, por ende, utilizar pruebas de aleatorización aproximadas en vez de pruebas de aleatorización exactas. El programa SCRT desarrollado por Van Damme y Onghena (1993) permite obtener el valor p correspondiente a la permutación aleatoria y sistemática de los datos. Además, el programa no sólo está concebido para aplicar las pruebas basadas en la aleatorización de estímulos en los diseños de línea base múltiple, sino también, en la mayoría de los diseños de caso único existentes.

A pesar de que las pruebas de la aleatoriedad constituyen estrategias de análisis eficientes, versátiles y fácilmente aplicables a los datos de caso único replicados, ello no implica que el tratamiento estadístico de los datos recogidos con esta clase de diseños quede resuelto de forma adecuada en todas las situaciones. A nuestro juicio existen una serie de problemas que limitan el uso de estas pruebas con carácter general. En primer lugar, tal y como ha sido puesto de manifiesto en los apartados anteriores, el estadístico de prueba en los diseños de línea base múltiple permutados se obtiene sumando las diferencias entre las medias de las condiciones de tratamiento y control a lo largo de las diferentes líneas de base de las que conste el diseño. Por lo tanto, en el caso que se rechace la hipótesis que afirma que todas las diferencias de medias fuesen igualmente probables, al investigador tan sólo le estaría permitido conocer que algo ha ocurrido teniendo pocas probabilidades de acontecer por azar. Dicho de otra

forma, el rechazo de la hipótesis nula global únicamente le indicará al investigador que al menos una de las líneas de base del diseño se ha visto afectada por el tratamiento; no obstante, carecerá de información para saber cuál ha sido y si se ha visto afectada más de una (Edgington, 1996). En segundo lugar, aunque estas pruebas se hallan libres de los supuestos de la distribución (la población esta formada por el conjunto de todas las permutaciones posibles y la muestra es cada una de estas permutaciones), conviene no olvidar que la validez de las mismas descansa en aleatorizar algún aspecto del diseño antes de registrar los datos. En nuestro caso concreto, en permutar el orden en que se introduce el tratamiento en las diferentes líneas de base y en seleccionar al azar los puntos de intervención. Se excusa decir que lo expuesto choca frontalmente con la forma de proceder de muchos terapeutas de conducta, debido a que la mayoría de veces los puntos de intervención no se fijan intencionalmente ni tampoco se determinan al azar, sino que se establecen sistemáticamente a partir del patrón de datos que emerge del experimento. Si la manipulación de las condiciones tan sólo fuese parcialmente dependiente de los datos (p.e., contar en la fase A con al menos tres puntos antes de aplicar el programa de intervención que no configuren una tendencia secular), se podría seguir utilizando pruebas basadas en la aleatorización (ver Ferron y Ware, 1994). No obstante, coincidimos con Onghena (1992) en que la potencia del diseño podría llegar a verse sustancialmente afectada.

En tercer lugar, resulta obvio que los diseños de línea base múltiple ejercen un exquisito control de la variable tratamiento y con ello de la mayor parte de las amenazas que atentan contra la validez interna, en especial, de la maduración, regresión e historia. Sin embargo, lo cierto es que su implementación en ambientes aplicados puede ser una tarea difícil y, en ocasiones, poco práctica. Convendría resaltar que, rara vez los pacientes que padecen una determinada disfunción conductual demandarán asistencia al mismo tiempo y, no parece que sea recomendable, como acertadamente ha puesto de relieve Hayes (1981), retrasar la intervención hasta que se disponga de un número suficiente de participantes para satisfacer los requisitos del diseño. En estos casos, una solución que los investigadores tienen a su alcance consiste en utilizar los diseños de línea base múltiple no concurrente. Esta variedad de diseños AB replicados es más fácil de implementar que los diseños estándar, al tiempo que la calidad de la inferencia causal que se obtiene con esta estrategia se puede tildar de aceptable. Por ejemplo, es poco probable que tres o más pacientes que no comparten el mismo ambiente temporal y/o espacial experimenten idéntico evento histórico coincidiendo con la presentación del tratamiento. Por consiguiente, aunque resulta evidente que la validez interna del diseño se incrementa sustancialmente cuando el tratamiento se presenta de forma escalonada y los datos de todos los sujetos se registran simultáneamente, lo cierto es que por razones de diversa índole se dan casos en los que no es posible obtener observaciones concurrentes para todos los sujetos.

A raíz de lo dicho, el propósito del presente trabajo consiste en adaptar el modelo de los mínimos cuadrados generalizados al análisis de un diseño transversal de series temporales interrumpidas cortas. En nuestra opinión, la solución analítica que vamos a desarrollar sirve para que los terapeutas de conducta se enfrenten de una manera natural a los problemas reseñados en los apartados anteriores. Para facilitar la

exposición, en lo que resta del trabajo, haremos dos cosas muy concretas: Por un lado, describir de forma lisa y resumida la técnica de las series temporales interrumpidas cortas con una sola unidad de análisis y su generalización al análisis de series temporales con múltiples unidades. Y, por lado, ilustrar la técnica descrita adaptando una parte de los datos obtenidos por McClannahan, McGee, MacDuff y Krantz (1990) en una investigación encaminada a mejorar los hábitos de apariencia e higiene personal de doce niños autistas alojados en tres centros terapéuticos distintos.

ANÁLISIS DE SERIES TEMPORALES INTERRUMPIDAS ESTÁNDAR

Un examen de las publicaciones metodológicas en la investigación conductual pone de relieve que el análisis estadístico paramétrico de los diseños de series temporales interrumpidas está siendo objeto de un amplio debate entre los que defienden la vigencia de enfoques basados en alguna versión del modelo lineal general (tales como el análisis de la varianza y el análisis de la regresión múltiple) y los que defienden enfoques basados en modelos no lineales (tales como los procesos estocásticos formados por los denominados auto-regresivos, integrados y de medias móviles; procesos que reciben usualmente el nombre genérico de modelos ARIMA).

A pesar de que los modelos paramétricos tradicionales juegan un importante papel a la hora de estimar y probar cambios entre las medias de diferentes fases del diseño (ver Blanca y Ato, 1999), estas pruebas sólo son válidas si los errores asociados con las observaciones registradas con anterioridad y con posterioridad al evento de interés varían en torno a las medias de las respectivas fases normal e independientemente y con variancia constante (Tiao, 1985; Box y Tiao 1975). Cuando los datos registrados sucesivamente a lo largo del tiempo no están adecuadamente modelados, carecen de la gracia que habitualmente confiere el azar y, por ende, pueden ser dependientes y frecuentemente no estacionarios. Consecuentemente, todos aquellos procedimientos estadísticos que requieran para su correcta aplicación el supuesto de independencia tan sólo deberían emplearse cuando, tras una minuciosa evaluación se haya descartado la presencia de correlación serial, dado que la presencia de este fenómeno distorsiona seriamente las propiedades de las estimaciones obtenidas mediante pruebas que no le tienen en cuenta.

Cuando la dependencia serial está presente en los errores de un modelo aparentemente bien formulado para detectar cambios en la variable dependiente tras la introducción del impacto, no resulta aconsejable utilizar pruebas estadísticas basadas en el criterio de los mínimos cuadrados ordinarios. Para abordar el problema reseñado se hallan disponibles diversos métodos estadísticos. Dentro de este conjunto, la solución más popular, prometedor y también más practicada en el campo de las ciencias sociales es la adaptación efectuada por Glass, Willson y Gottman (1975) de la técnica de análisis basada en los modelos ARIMA, desarrollada inicialmente por Box y Tiao (1965) y Box y Jenkins (1970). Si se modela adecuadamente la secuencia temporal de observaciones en términos del proceso que la ha generado, el enfoque ARIMA permite eliminar la correlación serial de los residuales. No obstante, identificar la estructura de dependencia del proceso a fin de transformar el modelo en una secuencia de datos

independientes exige disponer de un número de observaciones considerable. Si bien apenas existen trabajos empíricos que se hayan centrado en investigar de manera sistemática la fiabilidad y exactitud correspondientes a la identificación de los modelos ARIMA, entre cincuenta y cien observaciones por fase suele ser lo recomendable. Por consiguiente, en la investigación conductual aplicada puede resultar problemático identificar correctamente la estructura serial del error con el propósito de removerla de las observaciones, puesto que muestras inferiores a cincuenta datos son la regla más que la excepción (Arnau, 1999).

Para solventar el espinoso problema de contar con series cronológicas cuya longitud sea inferior a 50 datos, sugerimos adoptar la estrategia que sigue: En primer lugar, incluir los componentes de variación sistemática relevantes dentro del modelo, incluyendo el programa de intervención, y efectuar el análisis utilizando el criterio de los mínimos cuadrados. Esto es, modelar las tendencias seculares, el impacto y posibles interacciones de interés. En segundo lugar, verificar si la relación funcional entre las variables ha sido bien establecida mediante el modelo especificado para explicar los cambios acaecidos en la conducta. Para realizar esta función diagnóstica conviene combinar alguna de las pruebas estadísticas disponibles (p.e., la prueba de Durbin-Watson de correlación entre los errores o la prueba de los signos de Geary) con el examen visual del comportamiento de los residuales. Por último, si tanto el análisis gráfico como el estadístico sugieren la presencia de dependencia serial, asumir que las series son representadas adecuadamente por un proceso autorregresivo de primer orden y utilizar el criterio de los mínimos cuadrados generalizados. A continuación se describe el procedimiento para un solo participante o unidad organizacional y después se procede a su generalización.

Modelo de series temporales

Supongamos que las respuestas emitidas a lo largo del tiempo por una simple unidad de análisis sean adecuadamente explicadas mediante la relación lineal de k variables predictoras, más un término de error que recoja el efecto conjunto de otras variables no directamente implicadas en el modelo y cuyo efecto no resulte relevante. Formalmente,

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t$$

donde Y_t es la puntuación dada a la respuesta medida en el tiempo t , β_0 es el interceptor del proceso, β_1, \dots, β_k son los coeficientes de regresión parcial del modelo, X_{1t}, \dots, X_{kt} son los valores de las variables X en el tiempo t y $\varepsilon_t = (Y_t - \beta_0 - \beta_1 X_{1t} - \beta_2 X_{2t} - \dots - \beta_k X_{kt})$ es el error del proceso en el tiempo t .

Para proceder a la estimación de los parámetros de la ecuación anterior, hemos de establecer algunos supuestos sobre el modo en que se han obtenido las observaciones. Estos supuestos son decisivos en el proceso inferencial, y en lo referente a los errores del modelo, se asume que se distribuyen normal e independientemente con media cero y varianza constante. En la práctica, alguna de las hipótesis nulas referidas

a este conjunto de supuestos puede ser rechazada, bien sea porque la serie resulte no estacionaria, o bien sea porque la serie presente dependencia serial. Si bien el interés de nuestro trabajo tiene que ver principalmente con la presencia de correlación serial, para su exacta identificación no sólo requiere que la secuencia de observaciones tenga un tamaño considerable, sino también se exige que la serie sea estacionaria.

Modelación de la tendencia para lograr la estacionariedad

En la práctica cabe esperar que la serie muestre alguna tendencia o inclinación a lo largo del tiempo. Esto plantea dos problemas al investigador: por un lado, determinar si está presente en la serie alguna forma de variación (secular o estocástica) y, por otro lado, proceder a su modelación o extracción. Para averiguar si el proceso muestra una tendencia creciente o decreciente suele ser suficiente con la mera inspección visual de la serie. No obstante, esta operación puede complementarse con la división de la serie temporal en dos segmentos de igual longitud, las cuales deben tener el mismo nivel, en caso contrario, la serie se desplaza y, por lo tanto, no se puede mantener que la media sea constante para todo t . En presencia de tendencia de carácter secular, antes que diferenciar la serie para remover las variaciones, recomendamos incorporar a la ecuación de regresión los componentes de tendencia como variables explicativas adicionales. La incorporación de las tendencias dentro de la ecuación de regresión, además de subsanar posibles errores de especificación del modelo, permite que sus efectos sean medidos, probados e interpretados. La omisión de tendencias seculares relevantes a la hora de describir los datos determina la presencia de correlación serial en los residuales, o al menos, contribuye a incrementarla de forma sustancial.

A continuación se describe cómo proceder para modelar tendencias temporales en diseños conductuales. Para ello nos vamos a centrar en las tres situaciones más factibles de acontecer cuando se utiliza un diseño AB:

- a. Cambio de nivel entre las dos fases de la serie y presencia de tendencia.
- b. Cambio de tendencia entre las fases de la serie y ausencia de cambio de nivel.
- c. Cambio de nivel y de cambio de tendencia entre la primera fase y la segunda.

Comprobar el efecto de la intervención en una situación como la especificada en el caso a, requiere llevar a cabo una corrección de la tendencia a fin de que los cambios operados en la variable dependiente sean atribuidos a la aplicación del tratamiento. Dicho de otra manera, en presencia de una tendencia, generalmente lineal o a lo sumo cuadrática (polinomios de orden más elevado raramente suelen ser requeridos), se necesita verificar si los datos sufren incrementos (o decrementos) como consecuencia de la introducción de la intervención, antes que lo esperado por la evolución natural de la tendencia. En definitiva, la cuestión a resolver es la siguiente: ¿La aplicación de un tratamiento produce algún cambio en las respuestas de los sujetos al margen de la curva evolutiva que siguen los datos? Para dar cuenta de esta cuestión se puede llevar a cabo un análisis de regresión en el cual, además de introducir el efecto de la intervención por

medio de una variable ficticia o *dummy* que asigna el valor 0 a las n_1 ocasiones de medida incluidas en la fase de línea de base y el valor 1 a las n_2 ocasiones de medida incluidas en la fase experimental, también se incorpore la tendencia como una variable explicativa adicional. Para representar dicho evento se puede hacer uso de una variable subsidiaria o *proxy*, de forma tal que el tiempo sea medido puntuando con 1 la primera observación, con 2 la segunda y, así hasta la última puntuación a la que se asignará el valor N, siendo $N = n_1 + n_2$. Si el ajuste a una tendencia lineal no fuese suficiente, se eleva la variable *proxy* tiempo al cuadrado con el fin de modelar la tendencia cuadrática a la que probablemente se ajusten mejor los datos. De este modo, el cambio de nivel en presencia de tendencia puede ser estimado mediante la ecuación de regresión múltiple que sigue:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \varepsilon_t$$

o, bien ajustando el modelo cuadrático que sigue:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 T_t^2 + \beta_3 X_t + \varepsilon_t$$

donde Y_t representa el valor de la variable dependiente en el tiempo t , β_0 es una constante que indica el nivel promedio de la serie temporal, β_1 es una constante que indica si acontece una tasa de cambio constante a lo largo del período de tiempo que duran las observaciones, β_2 es una constante que indica si la tasa de cambio describe una tendencia cuadrática, β_3 es una constante que indica si se ha producido un cambio de nivel como consecuencia de la introducción del impacto y ε_t son los residuos aleatorios que suponemos normales, con media cero y varianza constante a lo largo de todos los puntos de la serie temporal.

Continuando con la modelación de la tendencia. En el caso b, a diferencia de lo que ocurría en el caso a, existe cambio de tendencia entre las dos fases de la serie y ausencia de interrupción entre las mismas. En principio, cabe suponer, que la mayor o menor inflexión en el componente de tendencia guarde relación con el tamaño del efecto de la intervención. El cambio de dirección que experimenta la tendencia a raíz de la presentación del impacto se puede manejar incluyendo dos covariantes temporales distintas en el análisis, una para la fase pretratamiento y otra para la fase postratamiento. La primera covariante recoge los valores asociados con la variable de conteo de las observaciones, de esta forma, tanto los datos registrados con anterioridad a la introducción del programa, como los registrados con posterioridad al mismo se les asigna el valor temporal correspondiente a la ocasión de medida. La segunda covariante también recoge los valores de la variable *proxy* tiempo, pero a diferencia de lo dicho para la primera covariante, los n_1 registros previos a la intervención se codifican como 0, mientras que los valores correspondientes al cambio de pendiente se codifican como $T_t - (n_1 + 1)$.

Cuando exista un cambio de tendencia entre fases una ecuación que incluya las

dos covariantes citadas reproducirá los datos de una manera más exacta, que otra ecuación que tan sólo incorpore el tiempo como variable explicativa. De este modo, la ecuación de regresión que resulta más apropiada para detectar el cambio de tendencia queda convenientemente expresada como:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 Z_t + \varepsilon_t$$

donde Z_t es una variable *proxy* que toma el valor 0 para las $(n_1 + 1)$ primeras observaciones, el valor 1 para la segunda observación posterior a la intervención, el valor 2 para la tercera y el valor $N - (n_1 + 1)$ la última. El resto de los términos que aparecen en la ecuación ya han sido definidos.

Por último, vamos a proceder a modelar la tendencia correspondiente al caso *c*. Este caso representa una situación más general, pues ofrece la posibilidad de abordar simultáneamente ambos patrones de cambio entre fases. Esto es, tanto el cambio de nivel como el cambio de tendencia. Bajo esta situación el investigador debe verificar, no sólo si la presentación del tratamiento produce una discontinuidad en el punto de intervención de la serie temporal, sino también comprobar si a raíz de la introducción del impacto se produce interacción entre el tratamiento y la tendencia, o si se prefiere algún cambio en el sentido u orientación de la(s) tendencia(s) a través de las fases. Para ello es necesario incluir tres variables explicativas en el modelo, exceptuando la constante. La primera variable se refiere a la tendencia, la cual es convenientemente manejada puntuando con 1 el primer registro efectuado y con N el último. La segunda se refiere al tratamiento experimental, el cual es definido puntuando con 0 los registros previos a la intervención y con 1 los registros posteriores a la misma. La tercera variable recoge los cambios de tendencia posteriores a la intervención, al igual que en caso *b*, las primeras $(n_1 + 1)$ observaciones se codifican con el valor 0, la segunda observación posterior a la intervención con el valor 1 y, así sucesivamente, hasta la última que se codifica con el valor $N - (n_1 + 1)$. Bajo esta situación, la ecuación modelo que resulta más adecuada para detectar los cambios aludidos puede expresarse como sigue:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 Z_t + \varepsilon_t$$

Una vez seleccionada la ecuación que al menos en teoría ofrece una mejor explicación de los cambios observados en la conducta, se procede a comprobar si los residuales del modelo ajustado se comportan de acuerdo a un proceso de ruido blanco, o si por el contrario, presentan dependencia serial. La prueba *d* y *d'* generalizada de Durbin-Watson constituyen procedimientos ampliamente utilizados para verificar la significación estadística de la correlación de retardo uno o superior. Ambas se encuentran implementados en el módulo AUTOREG del programa SAS (SAS Institute, 2001).

Existen dos razones fundamentales para desear conocer si los errores satisfacen el supuesto de independencia. En primer lugar, si los errores no son estadísticamente independientes unos de otros, la varianza correspondiente a los estimadores obtenidos

mediante el procedimiento de los mínimos cuadrados ordinarios resulta sesgada. En concreto, si los errores están correlacionados positivamente los modelos estadísticos que no contemplan la dependencia serial producirán errores estándar negativamente sesgados, contrastes estadísticos positivamente sesgados y tasas de error de Tipo I infladas (Carvajal, Baumler, Harrist y Parcel, 2001; Krull y MacKinnon, 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, inducimos a concluir que un programa de intervención es efectivo cuando lo cierto es que resulta completamente ineficaz. En definitiva, la liberalidad de estas pruebas conduce a que el investigador capitalice sobre el azar con mayor frecuencia de la debida por obtener estimaciones ineficientes. Justamente lo contrario ocurriría en el caso que los errores del modelo estuviesen correlacionados negativamente. En segundo lugar, la presencia de dependencia serial en los errores denota una especificación equivocada de la expresión matemática utilizada para caracterizar la relación entre un conjunto de variables independientes y una o más variables dependientes; bien sea por la omisión en el modelo seleccionado de una o más variables explicativas importantes, o bien sea por la inclusión de una o más variables explicativas irrelevantes (Huitema y McKean, 2000).

Si los errores están libres de correlación serial, el analista está en condiciones de proceder a evaluar la significación de los coeficientes de regresión obtenidos tras aplicar alguna estrategia de modelado estadístico. La vía más sencilla de proceder consiste en computar el estadístico

$$t(\hat{\beta}_j) = \frac{\hat{\beta}_j}{\sigma_e \sqrt{c^{jj}(\hat{\beta}_j)}}$$

donde σ_e es la raíz cuadrada de la varianza residual y c^{jj} es el j -ésimo elemento de la diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$, siendo \mathbf{X} la matriz de diseño. A su vez, cada una de las hipótesis referidas a los $\hat{\beta}_j$ coeficientes de regresión es rechazada al nivel α si $t > t_{(1-\alpha);v}$, donde $t_{(1-\alpha);v}$ es 100 $(1-\alpha)$ ésimo percentil de la distribución t con $n=N-h-1$ grados de libertad, con h denotando el número variables independientes incluidas en el modelo.

De estar interesados no sólo en la estimación puntual, sino también en evaluar la exactitud con que la estimación de $\hat{\beta}_j$ ha sido realizada, se puede hacer uso de la estimación por intervalo que sigue:

$$Pr \{ \hat{\beta}_j - t_{1-\alpha, N-h-1} [SE(\hat{\beta}_j)] \leq \beta_j \leq \hat{\beta}_j + t_{1-\alpha, N-h-1} [SE(\hat{\beta}_j)] \}$$

donde $SE(\hat{\beta}_j)$ se refiere al error estándar del estimador $\hat{\beta}_j [\hat{\sigma}_e \sqrt{c^{jj}(\hat{\beta}_j)}]$.

Si tras analizar el modelo que mejor se ajusta a los datos los residuales presentan correlación serial, el analista debe dirigir sus esfuerzos a estimar el valor del parámetro autorregresivo (recuérdese que hemos asumido que la dependencia serial se acomoda

a un proceso autorregresivo de primer orden). Aunque para realizar este cometido existen diversas versiones del criterio de los mínimos cuadrados generalizados (método bietápico de Durbin, método iterativo de Cochrane-Orcutt, método de búsqueda de Hildreth-Lu, método de Yule-Walker), una buena aproximación al verdadero valor del parámetro se puede obtener fácilmente a partir de la relación $d \cong 2(1 - \phi_1)$. Donde ϕ_1 se refiere al coeficiente de correlación serial de primer orden y d al estadístico de Durbin-Watson. A partir de aquí se procede a remover la dependencia serial transformando las variables del modelo mediante la expresión

$$Z_t^* = Z_t - \phi_1 Z_{t-1}$$

donde Z_t^* denota cualquier variable que forme parte del modelo especificado.

Una vez que el vector de respuestas y la matriz de diseño han sido transformadas, se estima y se prueba la significación del efecto del programa de tratamiento, tendencias seculares, variaciones estacionales, y demás covariadas incluidas en la matriz del diseño utilizando el criterio de los mínimos cuadrados ordinarios. Por último, se comprueba si los residuales del modelo transformado o filtrado se comportan de acuerdo a un proceso de ruido blanco, de lo contrario se vuelve a revisar todo el proceso, procurando que no se convierta en bucle sin fin.

Así pues, si logramos que los residuales del modelo ajustado describan un proceso de ruido blanco, esto es, que estén distribuidos normal e independientemente con media cero y varianza constante, el procedimiento de estimación mínimo cuadrático ordinario, además de proporcionarnos estimadores de los parámetros que carecen de sesgos, también nos proporcionará estimadores que posean las propiedades de consistencia y eficiencia; ya que los estimadores muestrales tendrán entre los estimadores insesgados la varianza más pequeña. Así pues, podremos efectuar predicciones y utilizar los habituales contrastes de t y F , pues las distribuciones que se asumen sí que cumplen en este caso con las propiedades básicas.

En resumen, concluimos esta breve descripción de la técnica señalando los pasos más importantes que debe seguir el analista conductual a la hora de estimar el efecto del tratamiento mediante un diseño de series temporales interrumpidas simple:

1. Se selecciona un modelo en el que aparezcan adecuadamente incorporados los efectos relativos al programa de tratamiento, tendencias seculares, tendencias estacionales y demás variables de interés. Para especificar el modelo que mejor se ajusta a los datos conviene utilizar alguna estrategia de modelado estadístico.
2. Se estima la ecuación de series temporales interrumpidas por mínimos cuadrados ordinarios y se calculan los residuales del modelo.
3. Se utilizan los residuales obtenidos en el paso anterior para verificar si existe dependencia serial. En caso negativo, se concluye el análisis evaluando la significación de los coeficientes de regresión. En caso afirmativo, se continúa con el análisis estimando el coeficiente de correlación serial de los residuales.
- 4.- Se utiliza el coeficiente de correlación serial para transformar todas las

variables incluidas en el modelo formulado inicialmente, incluida la constante.

5. Se utilizan los valores de las variables transformadas para estimar por mínimos cuadrados ordinarios el efecto del tratamiento, por lo general, cambio de nivel y/o cambio de tendencia.
6. Por último, se utilizan los residuales del modelo transformado para comprobar que la dependencia serial ya no permanece en el modelo.

ANÁLISIS DE DISEÑOS DE SERIES TEMPORALES INTERRUMPIDAS REPLICADOS

Cuando se dispone de dos o más secuencias temporales el análisis se puede abordar con diferentes enfoques en función del número de unidades que se incluyan, la cantidad de observaciones que se realicen, la recepción de uno más tratamientos en todas las unidades o tan sólo una parte de las mismas y, del tipo de asunciones que se efectúen. Por ejemplo, Vallejo, Fernández y Velarde (2001) recomiendan analizar los datos de dos o más grupos de participantes medidos repetidamente a lo largo del tiempo empleando el enfoque del modelo mixto lineal general. Por su parte, Algina y Olejnik (1982) recomiendan analizar los datos de dos grupos de participantes medidos repetidamente antes y después de introducir el impacto utilizando la técnica de los perfiles multivariado estándar, mientras que Vallejo y Escudero (1999) efectúan el mismo cometido usando el análisis de tendencias basado en el procedimiento multivariado de Brown-Forsythe mejorado. Las técnicas de análisis citadas asumen que todas las unidades reciben el programa de tratamiento en las mismas ocasiones de medida. Sin embargo, la mayor parte de las veces las unidades que configuran el plan de investigación reciben el programa de tratamiento en diferentes momentos temporales, como ocurre en los diseños de línea de base múltiple. En estos casos, una solución enormemente atractiva y socorrida se obtiene generalizando el análisis de series temporales interrumpidas simple que acabamos de exponer en el apartado anterior. Un procedimiento similar ha sido sugerido por Vallejo y Escudero (1998) y Velicer y Fava (2003) dentro del contexto de los modelos ARIMA y del enfoque de la transformación general, respectivamente.

La extensión del análisis de series temporales con una sola línea de base a situaciones en las que se disponga de múltiples líneas de base es relativamente sencilla, para ello tan sólo se requiere descubrir la transformación algebraica adecuada, construir un supervector de observaciones con las puntuaciones de las diferentes unidades experimentales y, utilizar una matriz de diseño parametrizada correctamente. El último objetivo es relativamente sencillo de conseguir si se procede de una manera similar a lo dicho para modelar las tendencias, esto es, complementando el conocimiento disponible a partir del examen gráfico de las diferentes series AB con la aplicación de alguna estrategia de modelado estadístico.

Para que el modelo de series temporales interrumpidas AB replicado constituya una extensión directa del modelo de series temporales interrumpidas AB simple necesitamos examinar dos condiciones. Por un lado, que las diferentes secuencias de fases AB que componen el diseño constituyan realizaciones aleatorias del mismo proceso (en

nuestro caso, autorregresivo de primer orden) y, por otro, que las diferentes secuencias temporales sean independientes. La primera condición se puede tildar de excesivamente restrictiva, sin embargo, conviene tener presente dos cosas. Por un parte, que el proceso autorregresivo de primer orden es el modelo con mayor plausibilidad *a priori* en las ciencias sociales y comportamentales (Simonton, 1977) y, por otra parte, el contexto de investigación donde se propone implementar la técnica. En el ámbito de los diseños de línea base múltiple encontramos razonable asumir que las series están determinadas por un proceso al que subyace el mismo constructo de investigación. Obviamente, cuando se asume que los residuales del modelo siguen el mismo proceso estocástico en todas las líneas de base del diseño el primer supuesto se cumple por definición.

La segunda condición se puede comprobar fácilmente verificando si existe relación entre las matrices de correlación de los residuales de las diferentes series individuales $[R_{\epsilon}(k)]$. Ello exige ajustar un modelo para cada una de las secuencias y calcular los errores del modelo. Si existe correlación serial se recomienda ajustar los modelos asumiendo un proceso autorregresivo de primer orden, dado que como reseñó anteriormente, en la mayoría de los diseños utilizados en la investigación conductual aplicada no se dispone de la cantidad mínima de observaciones aconsejada. Al verificar esta segunda condición, pueden acontecer tres situaciones diferentes: a) que los errores de las diferentes secuencias sean mutuamente independientes entre sí (la *i*-ésima serie está afectada por su propio pasado), b) que estén relacionados contemporáneamente (la *i*-ésima serie está afectada por su propio pasado y por la historia actual de otras series implicadas en el diseño) y c) que los errores estén correlacionados a lo largo de diversos retardos (la *i*-ésima serie, además de estar afectada por su propio pasado, también lo está por la historia pasada de otras series implicadas en el diseño). De acontecer lo especificado bajo las condiciones b y c, resulta más conveniente efectuar un análisis de series temporales multivariado, en vez del análisis de series temporales agrupado que estamos describiendo. La significación de la matriz de correlaciones para *k* retardos que se estimen oportunos se puede evaluar mediante el criterio de *pormanteau* descrito por Anderson (1984). Bajo hipótesis nula el estadístico $\chi^2 = -[n - 1 - (2m + 5) / 6] R_{\epsilon}(k) |$ sigue una distribución ji-cuadrado con $m(m-1)/2$ grados de libertad, siendo *m* el número de líneas de base del diseño. La comprobación del supuesto especificado, al igual que el análisis de otros menesteres más complejos, pueden ser analizados mediante el programa WMTS-1 desarrollado por Tiao *et al.* (1979).

APLICACIÓN DE LA TÉCNICA AL ANÁLISIS DE UN EJEMPLO

Un estudio de McClannahan *et al.* (1990) nos servirá como ejemplo para ilustrar la técnica. Estos autores utilizaron la estrategia de línea de múltiple a través de participantes para evaluar la eficacia de un programa de intervención destinado a mejorar los hábitos de limpieza e higiene personal de 12 niños autistas institucionalizados y realojados para la ocasión en tres centros terapéuticos diferentes (con 5, 3 y 4 niños respectivamente). En el primer centro el tratamiento se implementó bajo condiciones de demora temporal de 6 días, en el segundo de 10 y el tercero de 16. La intervención experimental consistió en proporcionar refuerzo social por parte de sus cuidadores; en

concreto, éstos realizaban comentarios positivos de la apariencia de los participantes y elogiaban la importancia de una buena higiene y de un esmerado cuidado personal. La variable dependiente consistió en una medida de los hábitos higiénicos obtenida a partir del cuestionario PAI (personal appearance index). En la figura 1 aparecen recogidas las puntuaciones promediadas de aspecto personal obtenidas para los niños residentes en los diferentes centros terapéuticos.

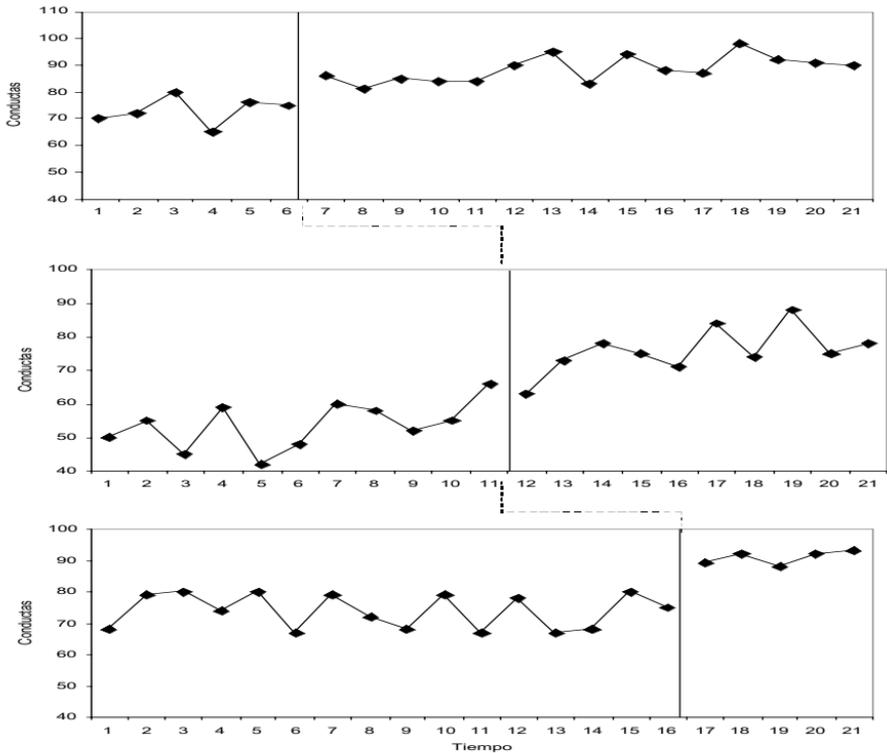


Figura 1. Porcentaje promedio de conductas de apariencia personal obtenidas por niños autistas residentes en tres centros terapéuticos diferentes. Datos tomado con ligeras modificaciones desde la figura 3 de McClannahan, L. E., McGee, G. G., McDuff, G. S., and Krantz, P. J. (1990). Assessing and improving child care: A personal appearance index for children with autism. *Journal of Applied Behavior Analysis*, 23, p. 478.

El primer paso para implementar correctamente la técnica descrita en el apartado anterior, implica identificar el modelo de dependencia serial para cada una de las series temporales de las que consta la investigación. Para el caso que nos ocupa la especificación de cada uno de los diferentes modelos aparece recogida en la tabla 1. Los tres

modelos son parsimoniosos y conducen a un buen ajuste de los datos, lo cual se traduce en que los residuales recalculados desde las variables transformadas constituyen ahora una secuencia de variables aleatorias normal e independientemente distribuidas con media cero y varianza constante. Para comprobar que los residuales de los modelos especificados siguen un proceso que no difiere de un ruido blanco, a partir de las seis primeras autocorrelaciones de los residuales se ha calculado el criterio de *portmanteau* Q de Ljung y Box (1976). Como queda puesto de relieve en la tabla 1 en ninguna de las tres series el valor empírico del coeficiente Q supera su correspondiente valor teórico.

Tabla 1. Modelos ajustados y correlaciones entre los errores para varios retardos

Modelos	DW	pDW	Test de Ljung-Box	R^2
LB 1: $(1 + .455B)\hat{\varepsilon}_t = v_t$	2.849	0.041	$Q = 5.43(\chi_{0.05}^2; 5 = 11.07)$.77
LB 2: $(1 + .438B)\hat{\varepsilon}_t = v_t$	2.843	0.048	$Q = 2.84(\chi_{0.05}^2; 5 = 11.07)$.81
LB 3: $(1 + .444B)\hat{\varepsilon}_t = v_t$	2.771	0.076	$Q = 7.99(\chi_{0.05}^2; 5 = 11.07)$.70
$[\mathbf{R}_{\hat{\varepsilon}}(0)]$	$[\mathbf{R}_{\hat{\varepsilon}}(1)]$	$[\mathbf{R}_{\hat{\varepsilon}}(2)]$	$[\mathbf{R}_{\hat{\varepsilon}}(3)]$	
$\begin{bmatrix} 1.000 & & & \\ -0.401 & 1 & & \\ & 0.291 & -0.317 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.000 & & & \\ 0.277 & 1 & & \\ & -0.138 & 0.041 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.000 & & & \\ -0.025 & 1 & & \\ & -0.252 & -0.106 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.000 & & & \\ -0.193 & 1 & & \\ & 0.395 & -0.178 & 1 \end{bmatrix}$	
$\chi^2 = 5.717 < \chi_{3,0.05}^2$	$\chi^2 = 1.934 < \chi_{3,0.05}^2$	$\chi^2 = 0.182 < \chi_{3,0.05}^2$	$\chi^2 = 4.006 < \chi_{3,0.05}^2$	

Tras ajustar un modelo individual para cada una de las líneas de base, se procedió a calcular las matrices de correlaciones muestrales $[\mathbf{R}_{\hat{\varepsilon}}(k)]$ entre los residuales de los modelos. En la parte inferior de Tabla 1 aparecen las matrices de correlaciones correspondientes a $[\mathbf{R}_{\hat{\varepsilon}}(0)]$, $[\mathbf{R}_{\hat{\varepsilon}}(1)]$, $[\mathbf{R}_{\hat{\varepsilon}}(2)]$, $[\mathbf{R}_{\hat{\varepsilon}}(3)]$. En ninguno de los cuatro retardos analizados se detectó presencia de correlaciones significativas. De acuerdo con la prueba ji-cuadrado global, cada una de las matrices anteriores no difiere de una matriz de identidad $[\mathbf{P}(0,1,2 \text{ y } 3) = \mathbf{I}]$.

Una vez verificado que las series, además de estar afectadas tan sólo por su propio pasado, constituyen la realización de procesos autorregresivos de primer orden, se estima la ecuación de series temporales interrumpidas por mínimos cuadrados ordinarios y se calculan los residuales del modelo. Para ello se utiliza como variable dependiente las puntuaciones resultantes de superponer las observaciones referidas a las tres secuencias temporales AB y como matriz de diseño los parámetros correspondientes al nivel general de las series (N), al cambio de nivel dentro de cada serie (C_N), a la diferencia de nivel entre la primera serie y la segunda (N_{E12}), a la diferencia de

nivel entre la primera serie y la tercera (N_{E13}), a la diferencia de cambio de cambio de nivel entre la primera serie y la segunda (N_{E12}), a la diferencia de cambio de nivel entre la primera serie y la tercera (N_{E13}) y al valor de la variable de conteo temporal o tendencia (T). La primera, la segunda y la última fueron utilizadas para ajustar los modelos individuales.

Tras el pertinente análisis de la regresión mínimo cuadrática ordinaria, se obtuvo la ecuación predicha que sigue:

$$\hat{Y}_t = 71.3 + 10.83X_{1t} - 20.48X_{2t} - 1.42X_{3t} + 6.73X_{4t} + 1.85X_{5t} + 0.44X_{6t}$$

(2.34) (3.33) (2.84) (2.81) (3.55) (3.88) (0.19)

$$R^2 = 0.844 \quad N = 63 \quad DW = 2.60 \quad \text{probDW} = 0.0416 \quad \phi_1 = -0.3016$$

Los valores entre paréntesis denotan los errores estándar asociados con los estimadores correspondientes al nivel, cambio de nivel dentro de cada serie, diferencia de nivel entre las series uno y dos, diferencia de nivel entre las series uno y tres, diferencia de cambio de nivel entre las series uno y dos, diferencia de cambio de nivel entre las series uno y tres y tendencia secular. R^2 es una medida de la bondad de ajuste del modelo e indica el porcentaje de varianza de la variable dependiente explicada por la combinación de las variables independientes, N indica el número de registros efectuados, DW se refiere al estadístico Durbin-Watson, probDW indica el valor de probabilidad asociado con estadístico Durbin-Watson y ϕ_1 representa la correlación serial de primer orden.

Antes de abordar consideraciones relativas a la significación de los coeficientes de regresión encontrados, se requiere transformar las variables por el modelo de dependencia serial ajustado, dado que el método de estimación utilizado (el convencional criterio de los mínimos cuadrados) no constituye una solución adecuada. Repárese que, de acuerdo con el estadístico Durbin-Watson, los residuales no conforman el supuesto de independencia a pesar de que el modelo parece contener todas las variables de interés para dar cuenta de los cambios conductuales operados a raíz de la intervención.

Para descubrir el modelo de dependencia serial que siguen los residuales de la ecuación de series temporales interrumpidas anterior, se puede continuar usando el módulo AUTOREG, o la rutina ARIMA disponible en la mayoría de los programas estadísticos existentes (p.e., SAS, SPLUS y SPSS). A partir de aquí se identifica el modelo utilizando alguno de los criterios de identificación semiautomática existentes (p.e., el AIC o el BIC), o utilizando las tradicionales claves estadísticas (p. e., FAC, FACI y FACP). Cuando no se disponga del mínimo de observaciones recomendado, lo sensato será asumir un modelo autorregresivo de primer orden y proceder a estimar el valor del parámetro conforme a lo expuesto más arriba. En nuestro caso el modelo de dependencia serial especificado fue un AR(1) $[(1 + 0.3026)\hat{\epsilon}_t = v_t]$. Dicho modelo proporciona un buen ajuste de los datos, lo cual se traduce en que los residuales recalculados desde las variables transformadas constituyen ahora una secuencia de variables aleatorias normal e independientemente distribuidas con media cero y varianza constan-

te.

Diagnosticada la estructura de correlación de los residuales, se transforman todas las variables del modelo, incluido el interceptor, y se efectúa el análisis de la regresión mínimo cuadrática ordinaria. Bajo esta nueva situación obtuvimos la ecuación predicha que sigue:

$$\hat{Y}_t^* = 71.7 + 11.44X_{1t}^* - 20.75X_{2t}^* - 1.12X_{3t}^* + 7.50X_{4t}^* + 1.35X_{5t}^* + 0.38X_{6t}^*$$

(1.77) (2.57) (2.40) (2.13) (2.69) (2.95) (0.15)

$$R^2 = 0.909 \quad N = 63 \quad DW = 2.020 \quad \text{probDW} = 0.752 \quad \phi_1 = -0.0149$$

El asterisco se utiliza para denotar que la regresión múltiple se realizó utilizando las variables transformadas a fin de remover la dependencia serial del modelo, la cual como se puede apreciar es mínima. Los términos restantes ya fueron definidos anteriormente.

De acuerdo con la prueba $|t|$, los coeficientes estimados correspondientes a N^* , $C_{N^*}^*$, N_{E12}^* y C_{N12}^* resultaron significativamente distintos de 0 al 1%, mientras que el coeficiente referido a T^* lo fue al 5%. Por su parte, los coeficientes estimados asociados con N_{E13}^* y C_{N12}^* no resultaron estadísticamente significativos. Centrándonos en los coeficientes de regresión más relevantes podemos destacar los siguientes hallazgos: En primer lugar, por lo que respecta a la eficacia del programa de intervención, se aprecian diferencias estadísticamente significativas al 0.1% entre los valores predichos en el tiempo $n_1 + 1$ (primer punto después de la intervención) en ausencia de tratamiento y los valores predichos en el tiempo $n_1 + 1$ en presencia del tratamiento. En segundo lugar, existen diferencias significativas desde el punto de vista estadístico entre los valores predichos correspondientes al nivel preintervención de la primera línea de base y los correspondientes al nivel preintervención de la segunda. Sin embargo, los niveles predichos de la primera y tercera líneas de base no difieren entre sí. En tercer lugar, los valores predichos correspondientes al cambio de nivel en las series registradas en el primer y segundo centro difieren entre sí más allá de lo que sería esperable que aconteciese por azar bajo hipótesis nula. Entre las series uno y tres tampoco observó ninguna diferencia en lo referido a los cambios de nivel.

Para concluir resaltar que la eficacia del programa de tratamiento también se hubiese puesto de relieve aplicando el modelo de Marascuilo y Busk (1988), obviamente asumiendo que los datos cumplen con los supuestos requeridos por implementar la prueba de la aleatoriedad. En concreto, el estadístico de prueba resultó ser el quinto valor más largo de las 1728 asignaciones aleatorias posibles de las que consta la distribución de aleatorización del estadístico; por ende, el valor $p = 0.0028$. Sin embargo, esta prueba únicamente nos informa que algo ha ocurrido cuya probabilidad de acontecer por azar es escasa. Como acabamos de ver, el procedimiento las series temporales agrupado no sólo nos proporciona esta información, sino que, además nos pone sobre la pista que las puntuaciones de apariencia personal registradas en el segundo centro son más bajas y también variables que las obtenidas en el primero y en el tercero, tanto

en ausencia del programa de refuerzo social como en presencia del mismo.

REFERENCIAS

- Algina, J. y Olejnik, S.F. (1982). Multiple group time-series design: An analysis of data. *Evaluation Review*, 6, 203-232.
- Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd Ed). New York: John Wiley & Sons.
- Arnau, J. (1999). Series temporales cortas y mínimos cuadrados generalizados: Análisis de la intervención. *Metodología de las Ciencias del Comportamiento*, 1, 119-135.
- Barlow, D.H. y Hersen, M. (1984). *Single Case Experimental Designs: Strategies for Studying Behavior Change* (2nd Ed). New York: Pergamon Press.
- Blanca, M.J. y Ato, M. (1999). Una aproximación empírica al problema de la autocorrelación. *Metodología de las Ciencias del Comportamiento*, 1, 185-200.
- Box, G.E.P. y Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.
- Box, G.E.P. y Tiao, G.C. (1965). A change in level of a nonstationary time series. *Biometrika*, 52, 181-192.
- Box, G.E.P. y Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70, 70-79.
- 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, 185-205.
- Edgington, E.S. (1996). Randomized single-subject experimental designs (invited essay). *Behavior Research Therapy*, 34, 565-574.
- Ferron, J. y Ware, W. (1994). Using randomization test with responsive single-case designs. *Behavior Research Therapy*, 32, 787-791.
- Glass, G.V., Willson, V.L. y Gottman, J.M. (1975). *Design and Analysis of Time-Series Experiments*. Boulder Co: Colorado Associated University Press.
- Harris, F.N. y Jenson, W.R. (1984). Comparisons of multiple-baseline across persons designs and AB designs with replication: Issues and confusions. *Behavioral Assessment*, 7, 121-127.
- Hayes, S.C. (1981). Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology*, 49, 193-211.
- Huitema, B.E. y McKean, J.W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Methods*, 3, 104-116.
- Huitema, B.E. y McKean, J.W. (2000). A simple and powerful test for autocorrelated errors in OLS intervention models. *Psychological Methods*, 87, 3-20.
- Kull, L.J. y MacKinnon, D.P. (2001). Multinivel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36, 249-277.
- Ljung, G.M. y Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.
- Marascuilo, L.A. y Busk, P.L. (1988). Combining statistics multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment*, 13, 137-157.
- McClannahan, L.E., McGee, G.G., McDuff, G.S. y Krantz, P.J. (1990). Assessing and improving child

- care: A personal appearance index for children with autism. *Journal of Applied Behavior Analysis*, 23, 469-482.
- SAS Institute (2001). *SAS/ETS Software: Version 8.2 (TS M0)*. Cary, NC: SAS Institute Inc.
- Shadish, W.R., Cook, T.D. y Campbell, D.J. (2002). *Experimental and Quasi Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Simonton, D.K. (1977). Cross-sectional time-series experiments: Some suggested statistical analysis. *Psychological Bulletin*, 84, 489-502.
- Tiao, G.C. (1985). ARMA models, intervention problems and outlier detection in time series. In E. J. Hannan, P. K. Krishnaiah and M. M. Rao (Eds.): *Handbook of Statistics, Vol. 5 Time Series in the Time Domain*, pp. 85-118. Amsterdam: North-Holland.
- Tiao, G.C., Box, G.E.P., Grupe, M.R., Hudak, G.B., Bell, W.R. y Chang, I. (1979). *The Wisconsin Multiple Time Series (WMTS-1) Program: A Preliminary Guide*. Department of Statistics, University of Wisconsin-Madison.
- Van Damme, G. y Onghena, P. (1993). *Single-Case Randomization Tests*. (Version 1.1) [Computer program]. Department of Psychology, Katholieke Universiteit Leuven (Belgium).
- Vallejo, G. y Escudero, J.R. (1998). Análisis de diseños de series temporales múltiples. *Análisis y Modificación de Conducta*, 24, 79-103.
- Vallejo, G. y Escudero, J.R. (1999). Evaluación de la intervención en diseños de series temporales cortas mediante el análisis polinómico de tendencias. *Metodología de las Ciencias del Comportamiento*, 1, 167-183.
- Vallejo, G., Fernández, P. y Velarde, H. (2001). Un estudio comparativo de pruebas robustas para el análisis de datos longitudinales. *Metodología de las Ciencias del Comportamiento*, 3, 35-52.
- Velicer, W.F. y Fava, J.F. (2003). Time series analysis for psychological research. In J. Schinka and F. Velicer (Eds), *Research Methods in Psychology* (pp. 581-606). *Handbook of Psychology* (I. B. Weiner, Editor-in-Chief). New York: John Wiley & Sons.
- Wampold, B.E. y Worsham, N.L. (1986). Randomization tests for multiple-baseline designs. *Behavioral Assessment*, 8, 135-143.

Recibido, 27 Febrero 2003

Aceptado, 17 Mayo 2003