

AUTOCORRELACIÓN EN SERIES TEMPORALES

Los diseños de series temporales interrumpidas (DSTI) permiten evaluar el impacto que solo una serie temporal puede ejercer en un tratamiento.

Este tipo de diseños presenta problemas al inferir hipótesis experimentales.

DEPENDENCIA SERIAL

Se define como el hecho de que las respuestas emitidas por un sujeto en un determinado momento están estrechamente relacionadas con las emitidas por el mismo sujeto en un tiempo pasado de la serie, es decir, no son puntuaciones independientes. Jones, Weinrott y Vaught afirman que la dependencia serial significa que puntuaciones temporalmente adyacentes tienden a estar relacionadas una con otra.

Es la correlación existente entre datos, por lo que se denomina correlación serial o autocorrelación.

Para una serie temporal dada, hay diversos coeficientes de autocorrelación. En concreto, se puede fijar el grado en que un valor en el tiempo t se ve influido por valores de los tiempos $t-1$, $t-2$, $t-3$, etc. El grado en que un valor del tiempo t se ve afectado por el tiempo $t-1$ se denomina autocorrelación de un retardo. Dentro del ámbito del análisis conductual, esta autocorrelación de retardo uno es la más frecuente. Con ella es suficiente para descubrir la dependencia serial de los datos.

La dependencia serial entre observaciones sucesivas se calcula mediante la autocorrelación o correlación entre pares de datos con un intervalo temporal fijo.

Si, además de las autocorrelaciones con un intervalo temporal igual a uno, se calculan también las correlaciones con otros intervalos temporales, se puede construir una representación gráfica de correlaciones denominada correlograma, cuyo examen ayuda a detectar la presencia de dependencia.

Las causas principales de autocorrelación son las tendencias o ciclos. Las tendencias se definen como una dirección natural en el nivel de la conducta observada que aumenta o disminuye progresivamente a lo largo del tiempo.

El coeficiente de autocorrelación es un indicador de la posible existencia de tendencias o ciclos.

EFFECTO DE LA DEPENDENCIA SERIAL EN EL ANÁLISIS DE DATOS CONDUCTUALES

En la investigación de caso único los datos pueden ser evaluados a través de la inspección visual, análisis estadísticos o ambos. La presencia de autocorrelación, característica propia de datos conductuales, constituye un problema al hacer inferencias tanto a partir del análisis visual como del basado en pruebas estadísticas convencionales.

✓ Análisis visual

Es un método comúnmente usado para evaluar el efecto de la intervención y consiste en examinar las representaciones gráficas de los datos.

Aquellas variables que muestren un efecto menos fuerte deben ser ignoradas, para no incurrir en errores de:

Tipo I : rechazar la hipótesis nula (H_0) cuando es verdadera.

Tipo II : aceptar la hipótesis nula (H_0) cuando es falsa.

- Ésta es una de las principales debilidades del uso exclusivo del análisis visual, puesto que puede ocurrir que no se tenga en cuenta el efecto de variables de gran importancia en el ámbito clínico y aplicado. Por otro lado, la simple inspección visual es un tanto problemática, sobre todo cuando existe una tendencia en la línea base o cuando los datos tienen mucha variabilidad.

Bajo tales circunstancias, los investigadores tienen dificultades en determinar si una intervención ha sido o no eficaz. Matyas y Greenwood observan que los analistas visuales detectan, en muchas ocasiones, intervenciones significativas cuando de hecho no las hay; por el contrario, muy raramente detectan efectos verdaderos cuando no existen. Esta afirmación debilita la utilización de la técnica visual con objeto de evaluar el impacto de una intervención.

- Otra dificultad radica en la elección de la escala de tiempo y, particularmente, de la variable que se registra: lo que en un gráfico aparece como una variación importante, puede verse reducido, en otro, a una variación insignificante debido a la modificación de la escala.

Otro problema asociado al análisis visual de datos autocorrelacionados hace referencia al bajo nivel de acuerdo entre el análisis visual y estadístico de los mismos.

El *análisis estadístico* produce siempre el mismo resultado.

El *análisis visual* puede llevar a respuestas distintas cuando se explica dos veces sobre los mismos datos.

Del estudio de *Jones* se desprenden tres conclusiones:

- 1- No se obtiene el mismo resultado a partir de inferencias visuales que a partir del AST (análisis de series temporales).
- 2- El acuerdo entre inferencia visual y AST varía en función de los niveles de dependencia serial (cuando la autocorrelación es alta se produce un mayor desacuerdo, y a medida que disminuye la dependencia serial, el acuerdo incrementa).
- 3- Los AST presentan cambios estadísticamente significativos, el acuerdo es relativamente bajo.

Es de esperar un desacuerdo entre el análisis visual y el AST. Por lo que los dos tipos de análisis deben ser complementarios, con lo que los investigadores, probablemente, inferirán efectos significativos de forma más frecuente que si se usaran solamente análisis visuales.

✓ **Análisis estadístico**

Desde este punto de vista los registros generados por un sujeto único en una serie de observaciones sucesivas son los más difíciles de analizar, debido a la presencia de autocorrelación.

Por lo general, en ciencias conductuales, el número de observaciones es pequeño y, por lo tanto, es difícil estimar una autocorrelación estadísticamente significativa. En este sentido, Sharpley afirma que las pruebas utilizadas con objeto de hallar la significación estadística de una autocorrelación son irrelevantes. Los procedimientos estadísticos aplicados al análisis de datos conductuales tienen por objetivo evaluar los cambios de nivel, tendencia o variabilidad de forma cuantitativa y controlada. Las pruebas estadísticas más conocidas son la *t* de Student- Fisher y la *F* de Snedecor. La autocorrelación sesga las estimaciones de la varianza del error y, por consiguiente, viola el supuesto de independencia de los residuales.

Han habido diferentes intentos a fin de resolver el análisis estadístico de datos serialmente dependientes. En efecto, cuando la aplicación de las pruebas *t* y *F* no resulta adecuada, se han propuesto una serie de procedimientos modificados cuyo objetivo es solventar el problema de dependencia serial. La propuesta más simple consiste en adaptar los procedimientos paramétricos convencionales (prueba *t*, AVAR y regresión múltiple). En esta línea se encuentran los trabajos de:

- Shine y Bower quienes resuelven el problema de la inferencia estadística en DSTI transformando la estructura del diseño clásico de medidas repetidas (sujetos x tratamiento) en un diseño de ensayos x tratamientos. La dificultad principal de este nuevo procedimiento es la dependencia existente entre las diferentes observaciones. Shine y Bower salvan esta dificultad, en el ámbito teórico, dando por supuesto que dichas respuestas de un sujeto ante un estímulo concreto varían aleatoriamente en torno a un valor central μ , en consecuencia, las respuestas pueden considerarse estadísticamente independientes y normalmente distribuidas.
- Gentile, Roden y Klein apuntan por una estructura bifactorial del diseño de medidas repetidas. Sugieren que una forma de reducir la dependencia serial es combinar las fases bajo el mismo tratamiento. La alternativa planteada por Gentile consiste en convertir el diseño de caso único en un diseño de AVAR unidireccional donde la acción de los tratamientos es el componente "entre" y el número de observaciones por tratamiento el componente "intra".

Las críticas a estos intentos de adaptar los estadísticos convencionales a los datos de diseños únicos no se hicieron esperar. Estas críticas se basan en el hecho de que tanto Shine y Bower como Gentile parten del supuesto teórico de que las respuestas de un sujeto se distribuyen de forma natural e independientemente. Se constató que los patrones de independencia variaban de un experimento a otro y, por lo tanto, es impropio el supuesto de que las conductas generadas por un mismo sujeto son independientes y que su distribución sea normal.

- Según *Hartmann*, la violación del supuesto de la independencia de los componentes del error (esto es que la correlación entre cualquier par de observaciones debe ser igual a cero) invalida la aplicación del modelo de AVAR unidireccional. La dependencia serial implica una pérdida importante de grados de libertad.
- Thoresen y Elashoff plantean una crítica principal al planteo de AVAR propuesto por Gentile, es que normalmente las pruebas t y F detectan si existen cambios significativos en las medias entre-fases μ , por tanto, se puede obtener un efecto no significativo debido a las tendencias. Cualquier tendencia presente dentro de las fases no es tenida en cuenta por las

técnicas estadísticas clásicas, lo que lleva a una pérdida de información.

- Busk y Marascuilo, Levin... han demostrado la dificultad principal radica en el hecho de que, al aplicar las técnicas de análisis tradicionales, las estimaciones de la tasa de error Tipo I aumentan notablemente. En consecuencia, no hay que olvidar que las pruebas estadísticas clásicas fueron planteadas bajo el supuesto de independencia de las observaciones.

El incumplimiento de la condición de independenciagenera una serie de efectos no deseables, al aplicar los estadísticos convencionales. Siguiendo la lógica de la estadística clásica, se compara la variabilidad entre-fases con la variabilidad intrafases que corresponde al error experimental. La autocorrelación sesga las estimaciones de la varianza del error, en el sentido de la existencia de una subvaloración de la probabilidad de cometer un error Tipo I cuando la autocorrelación es positiva y una sobrevaloración, cuando es negativa.

Posteriormente, diversos autores, (Padia, Greenwood y Matyas) al retomar esta cuestión, han llegado a las mismas conclusiones. En resumen, ante la presencia de autocorrelación positiva, la prueba estadística tiende a ser "liberal", y en caso de que la autocorrelación sea negativa, la prueba es demasiado "conservadora", es decir, está negativamente sesgada.

- Hibbs desarrolló una fórmula que permite calcular el grado en que una prueba t o F es hinchada debido a la presencia de autocorrelaciones positivas en los datos. Al igual como se ha señalado en otros trabajos se observa que, incluso con niveles bajos de autocorrelación, los valores t o F son distorsionados. En efecto, con una autocorrelación tan baja como 0.1, se espera que los estadísticos t o F sean hinchados en un 110%.

De todo lo anterior, se concluye que no es adecuado aplicar en DSTI aquellas pruebas estadísticas que requieren el supuesto de independencia, puesto que la presencia de autocorrelación sesga sustancialmente los resultados de dichas pruebas. Con objeto de solventar el problema de dependencia serial existen algunos procedimientos estadísticos que tienen en cuenta este aspecto y permiten estimar los efectos del tratamiento de un modo más preciso. Como por ejemplo el AST.

El AST desarrollado por Box y Jenkins, con este procedimiento permiten eliminar, a nivel estadístico, el efecto de la dependencia serial inherente a los datos. El AST evalúa los cambios de tendencia y nivel que presenta el curso de un fenómeno en el tiempo teniendo en cuenta la dependencia serial. El inconveniente

principal, que se plantea a partir de la utilización del AST en datos conductuales, es la gran cantidad de observaciones requeridas

para una correcta identificación del modelo autorregresivo integrado de medias móviles.

En los últimos años, se han propuesto una serie de procedimientos que no están afectados por la presencia de datos autocorrelacionados como, por ejemplo, la mayoría de las pruebas estadísticas no paramétricas basadas en los principios de la aleatoriedad y las técnicas de análisis por mínimos cuadrados generalizados.

AUTOCORRELACIÓN EN DATOS CONDUCTUALES: MITO O REALIDAD

La mayoría de los investigadores siguen la línea planteada por Box y Tiao(1965) y Glass et al. (1975) de que el registro de observaciones sucesivas conlleva normalmente dependencia serial. Sin embargo, los estudios que tratan de cuantificar el impacto de la autocorrelación sobre los resultados de las investigaciones llegan a ser contradictorios. Mientras Jones et al. (1977) tras el análisis de 24 series temporales de la revista *Journal of Applied Behavior Analysis (JABA)* encuentran autocorrelaciones, que oscilan entre el 0.40 y 0.93, en el 83% de éstas; un estudio no publicado Kennedy (1976) sólo encuentra autocorrelaciones de retardo uno significativas en el 29% de los casos. Además, la presencia de autocorrelación mostró falta de acuerdo entre evaluadores. DeProspero y Cohen (1979) obtuvieron un acuerdo entre jueces de 0.61, mientras que Jones, Weinrott y Vaught (1978) de sólo 0.39. También existía un bajo acuerdo entre los resultados obtenidos mediante el análisis visual y el análisis estadístico aplicado a los mismos datos(DeProspero y Cohen, 1979; Gottman y Glass, 1978; Jones et al. , 1978). A pesar de los resultados discrepantes de estos estudios para la mayoría de los investigadores parecía quedar claro que la autocorrelación era algo intrínseco a los estudios de series temporales.

Huitema en su trabajo de 1985 se convierte en la principal nota discordante en este tema. Para este autor la mayoría de las veces la presencia de autocorrelaciones significativas es totalmente ficticia y se debe a la utilización de técnicas estadísticas tradicionales. Huitema (1985) ha criticado la manera de calcular la autocorrelación en el estudio de Jones et al.(1977). Para él cuando las observaciones de línea base y tratamiento se toman en conjunto para calcular la autocorrelación de retardo uno se obtienen coeficientes de correlación altos, que se deben en parte a la presencia de diferentes niveles en la conducta. Como solución propone calcular la autocorrelación a partir de los residuales de las series.

De este modo en su artículo de 1985 se llevaron a cabo dos estudios de gran repercusión. En el primero de ellos se reanalizaron las 24 series temporales muestreadas por Jones et al. (1977). Comparando la distribución de frecuencias de los coeficientes de autocorrelación de retardo uno y la correspondiente distribución de los coeficientes estandarizados (mediante la Z de Fisher), llegó a la conclusión de que las dos distribuciones son bastante coincidentes salvo en la cola izquierda. Dicha cola es más gruesa en la distribución de los coeficientes directos que en la distribución de los coeficientes estandarizados. Para Huitema (1985) la causa era que las autocorrelaciones negativas altas se obtienen de cantidades extremadamente pequeñas de observaciones. En el segundo de estos estudios se calculó las autocorrelaciones de uno a cuatro retardos en 441 series temporales publicadas en JABA entre 1968 y 1977. Las autocorrelaciones de retardo uno, correspondientes a las fases de línea base, mostraban una distribución casi normal con un ligero sesgo y media de -0.01 . Con el objeto de evitar el problema de la longitud de fase estandarizó las autocorrelaciones y atribuyó la desviación de la distribución obtenida respecto a la normal a que diez conjuntos de datos tenían un coeficiente de autocorrelación estandarizado superior a 2.7.

Han sido estos resultados los que han llevado a Huitema a defender que la autocorrelación en el análisis conductual aplicado es un mito y en consecuencia los registros conductuales pueden ser analizados mediante las técnicas estadísticas convencionales.

Estos resultados han creado la mayor controversia de los últimos años en el análisis conductual aplicado. En un lado están los autores que opinan que los procedimientos estadísticos paramétricos no pueden ser empleados porque los errores entre observaciones no son estadísticamente independientes (Jones et al., 1977; Kratochwill, 1978; Levin et al., 1978; Phillips, 1983; Toothaker et al. 1983). En el otro lado, están quienes defienden procedimientos estadísticos clásicos, argumentando que las autocorrelaciones de los registros son cero o no difieren estadísticamente de cero (Huitema, 1985; Kazdin, 1980).

La postura defendida por Huitema (1985) ha sido rebatida por numerosos investigadores. Suen (1987) argumenta que la ausencia de autocorrelación no es concluyente por una serie de aspectos metodológicos. Entre estos destaca las limitaciones inherentes a la actual epistemología estadística, la cantidad reducida de observaciones en el análisis conductual aplicado, el uso de valores α

convencionales para probar la significación de la autocorrelación y, como más importante, los efectos del método de adquisición de datos. Todas estas cuestiones deberían ser resueltas antes de concluir la no-existencia de autocorrelación.

En cuanto al método de adquisición de datos, se utilizan diversos métodos. El método de observación directa del comportamiento es el más utilizado. La forma más común de llevar a cabo dicha observación es a través de sesiones discretas a lo largo del tiempo. No obstante, en un mismo intervalo de tiempo de cada sesión, así como la distancia entre dos sesiones consecutivas puede que no sea uniforme. En la práctica, cuando las observaciones se realizan en intervalos de distinta longitud, o con tiempos distintos entre intervalos, la muestra del patrón de conducta subyacente suele estar sesgada, con lo que el coeficiente de autocorrelación de retardo uno no representa en su totalidad la posible tendencia, periodicidad o cambios.

Para llegar a la conclusión de no-autocorrelación, habría que justificar que las explicaciones metodológicas explicadas por Suen (1987) son improbables. No obstante esto es difícil de conseguir.

Una solución para poder determinar la presencia de autocorrelación sería realizar un análisis de potencia estadística. De este modo podríamos estimar la probabilidad de aceptar una autocorrelación significativa cuando es verdadera, es decir la probabilidad de rechazar la H_0 correctamente.

Sharpley y Alavosius (1988) critican la propuesta de Huitema (1985), argumentando que incluso niveles bajos de autocorrelación pueden sesgar los valores de t y F . Estos autores critican a Huitema (1985) en dos puntos:

- a) Estandarizó la autocorrelación antes de comprobar si la distribución obtenida se alejaba de la normalidad, con lo que la fórmula empleada en el proceso de estandarización influye en la forma de la distribución.
- b) El método utilizado para estandarizar las distribuciones fue la Z de Fisher (el menos sensible de los posibles), y para probar la significación de las autocorrelaciones se utilizó el método de Barlett (1946) que es insensible a tamaños muestrales pequeños.
- c) Además concluyó que los datos conductuales no están autocorrelacionados debido a que la tendencia central de la distribución de r_1 es cero, cuando de hecho el valor esperado de r_1 es $-(n-1)^{-1}$. así la tendencia central estaría por debajo de 0.

Busk y Marascuilo (1988), también criticaron las conclusiones de Huitema (1985) por el hecho de que las autocorrelaciones son medidas en muestras pequeñas y las por ello las pruebas para identificar la autocorrelación tienen baja potencia.

Para Huitema (1988), todas estas críticas se basan en malas interpretaciones. En éste artículo expone las tres conclusiones principales de su trabajo de 1985:

- 1) Los residuales de los modelos convencionales tienen una autocorrelación escasa o nula, y por lo tanto deben considerarse los métodos estadísticos tradicionales.
- 2) El método más apropiado para cada estudio depende de los datos y de lo que el investigador esté buscando. El AVAR de un sólo factor o los modelos de discontinuidad en la regresión paso a paso serán eficaces en muchos casos.
- 3) En caso de que exista un número suficiente de datos debe considerarse el uso de modelos de series temporales.

Huitema (1988) nunca dijo que el AVAR fuera el modelo preferente, ello dependerá de si los datos se ajustan al modelo.

En cuanto a la potencia estadística, el argumento común entre los críticos es que las conclusiones de Huitema (1985) son inapropiadas porque la potencia del procedimiento de prueba es muy baja, debido a que el número de observaciones en cada una de las 441 series de datos es muy pequeño. Huitema (1988) alega que esto es irrelevante para el promedio de los coeficientes, puesto que la base de su estudio es la distribución de las 441 autocorrelaciones, más que el resultado de pruebas de significación.

Otro tema relevante es la naturaleza de la variación de r_1 en la distribución. Para Huitema (1988) es un error pensar que la variación de los valores de θ_1 es la responsable de la variación observada en r_1 . Esto es debido a que los críticos no distinguen entre parámetros (θ_1) y estadísticos (r_1), y se tiende a interpretar los valores muestrales como si se tratara de parámetros poblacionales. Que r_1 sea un estimador preciso dependerá del número de estimaciones en las que se base. Con un número de estimaciones grande r_1 no diferirá de θ_1 . Pero si se basa en el número de observaciones que suele ser habitual en los estudios de autocorrelación, la igualdad entre r_1 y θ_1 será muy baja. Además, la discrepancia entre r_1 y θ_1 es mayor con valores positivos altos de θ_1 que con valores negativos o positivos cercanos a cero, independientemente del tamaño de la muestra (Huitema y McKean, 1991; Kendall, 1954). Y puede haber

divergencias en los resultados dependiendo de que estimador del parámetro de la autocorrelación se halla utilizado (Huitema y McKean, 1991).

Huitema (1988) responde a la cuestión precedente, argumentando que la mayor parte de la variancia de r_1 se debe al error muestral, más que a la variancia entre los parámetros de θ_1 . Por este motivo, es razonable utilizar la media de la distribución de r_1 para estimar los parámetros θ_1 .

ESTIMACIÓN DE LA AUTOCORRELACIÓN

El estimador convencional de la autocorrelación de retardo k viene dado por la siguiente expresión:

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

La función de autocorrelación de retardo k es una medida de la autocorrelación entre Y_t y Y_{t+k} , para una determinada distancia o retardo k . El numerador de la ecuación estima la covarianza entre dos puntos de la serie separados por k retardos, y el denominador representa la variancia de las observaciones Y_t . Esta fórmula sólo es útil para muestras grandes. Para tamaños muestrales pequeños sería necesario aplicar algún factor de corrección. De ello hablaremos más tarde.

La autocorrelación de retardo uno (r_1) constituye la correlación entre la serie temporal original y la serie desplazada un lugar. Así r_1 se calcula a partir del emparejamiento de la primera observación con la segunda, de la segunda con la tercera, de la tercera con la cuarta... E igualmente en el resto de retardos, se tiene que r_k es el coeficiente de correlación estimado entre la serie temporal original y la serie en el retardo k .

Si r_1 es estadísticamente significativa, quiere decir que la respuesta en un determinado punto del tiempo es predicha por la respuesta en la ocasión anterior. La dirección de dicha predicción viene determinada por el signo de la autocorrelación.

Las propiedades fundamentales del coeficiente de autocorrelación son:

- 1) El coeficiente r_0 es igual a la unidad.
- 2) La función de autocorrelación es simétrica en relación con el retardo 0. En consecuencia, se cumple que el coeficiente r_k es el mismo tanto si se retarda hacia adelante como hacia atrás.
- 3) Cada vez que se retarda la serie temporal, se pierden un par de observaciones en la estimación de la correspondiente r_k .

Existen algunas pruebas para verificar la significación estadística de la autocorrelación, como la prueba de Barlett (Barlett, 1946), la prueba de Geary (Habibagahi y Pratschke, 1972) o el estadístico d de Durbin-Watson (Kmenta, 1971). Pero el mayor inconveniente de su uso es que sólo ponen a prueba que la H_0 no es cierta. Así cuando la autocorrelación de un conjunto de datos no es estadísticamente significativa no puede concluirse que no exista autocorrelación.

En un diseño de series temporales interrumpidas de dos fases existen tres formas aceptables de calcular la autocorrelación. La primera de ellas consiste en calcular los residuales de toda la serie.

La segunda implica promediarla función de autocorrelación de los residuales de la línea base y de los residuales de la intervención. La tercera consiste sólo en calcular los datos de la función de autocorrelación solamente de los datos de pre-intervención.

COEFICIENTE DE AUTOCORRELACIÓN Y TAMAÑOS MUESTRALES PEQUEÑOS

Los indicadores que suelen aplicarse en la evaluación de un estimador son el sesgo y el CME. El sesgo es la desviación o diferencia entre la r_1 y el valor del parámetro asociado. Cuando la diferencia se calcula entre el valor esperado de r_1 y el valor asociado al parámetro se obtiene el sesgo teórico, mientras que el sesgo empírico es la diferencia entre el valor medio calculado de r_1 y el valor del parámetro. Huitema y McKean (1991) comprobaron que ambos sesgos son similares en valores de \otimes_1 cercanos a cero, mientras que al alejarse se incrementa la diferencia. También existe discrepancia entre ambos indicadores dependiendo del tamaño de la muestra. Las muestras pequeñas (por debajo de 20) generan una discrepancia mayor.

El CME es definido por el valor esperado de la desviación cuadrada entre la estimación muestral y el parámetro de la población. El CME es función tanto de la variancia muestral como del sesgo. Debido al carácter asimétrico de las distribuciones muestrales Huitema y McKean (1991) recomiendan utilizar este indicador.

Prácticamente todos los estudios realizados en series temporales cuentan con tamaños muestrales muy pequeños. Ello implica que el coeficiente r_1 difícilmente muestre la existencia de dependencia serial, pues existe una alta probabilidad de cometer un error de tipo II.

La potencia de los estadísticos del coeficiente de autocorrelación es baja cuando el número de observaciones es pequeño. Busk y Marascuilo (1988) demostraron que para identificar un coeficiente de autocorrelación poblacional de 0.50 como estadísticamente significativo, con un riesgo de error tipo I de 0.05 y un riesgo de error tipo II de 0.10 es necesario un tamaño de muestra de 65 observaciones. Para un coeficiente de autocorrelación de 0.30, un error de tipo I de 0.05 y un error de tipo II de 0.20, se requieren 117 observaciones con objeto de hallar la significación estadística. Obviamente, estos valores muestrales superan en mucho las muestras habituales en los estudios de autocorrelación. Estos autores concluyen, que si bien la mayoría de las investigaciones obtienen valores de autocorrelación mayores de cero, éstas se miden en muestras donde las pruebas estadísticas tienen baja potencia para identificar la autocorrelación. Además existe el problema del intervalo de tiempo entre los registros: las conductas lejanas en el tiempo tienen una probabilidad menor de reflejar la existencia de autocorrelación. Así, cuando las series temporales son cortas y el intervalo entre observaciones es grande, las autocorrelaciones estimadas son bajas.

Huitema y McKean (1991) realizaron un estudio de simulación de Monte Carlo las propiedades estadísticas de cinco estimadores de autocorrelación, entre ellos r_1 . En este estudio generaron muestras de un modelo variando el tamaño de la muestra ($n = 6, 10, 20, 50, 100, 500$) y los valores del parámetro ρ_1 dentro del intervalo $[-0.90, 0.90]$, con un total de 114 combinaciones. En este estudio hallaron que los estimadores convencionales de autocorrelación no son válidos con tamaños de muestras pequeños (menos de 50 observaciones).

El estimador más utilizado ha sido el coeficiente de autocorrelación de retardo k . Huitema y McKean (1991) demostraron que r_k está sesgado porque el numerador se basa en menos términos que el denominador. Al aumentar el retardo, la discrepancia entre el número de términos del numerador y el denominador se incrementa: r_k se basa en un numerador con $n - k$ términos y un denominador

con n términos. Esta discrepancia lleva a estimaciones más cercanas a cero de lo esperado. De ahí que frecuentemente se multiplique por el término $n/(n-k)$, en un intento de eliminar la discrepancia en el número de términos. A este estimador se le denomina \hat{r}_k^* y se representa:

$$r_{1}^* = \left(\frac{n}{n-1} \right) r_1$$

Existen otros estimadores como la *fórmula exacta* de Kendall (1976), que tiene un parecido directo con la correlación de Pearson; o la estimación cíclica (Moran, 1948). A través de la simulación de Monte Carlo, Huitema y McKean (1991) demostraron que ninguno de estos estimadores es adecuado para tamaños de muestras pequeños. En el caso del estimador r_1 hallaron que generaba más sesgo, en especial en los valores positivos del parámetro. En vista de este resultado propusieron un estimador modificado de r_1 :

$$r_{1+} = r_1 + \frac{1}{n}$$

En esta ecuación, al añadir el término $1/n$ se corrige hasta cierto punto el sesgo negativo, es decir la subestimación de la autocorrelación negativa. Además r_{1+} converge con r_1 al aumentar el tamaño de la muestra, puesto que $1/n$ se aproxima a cero. Este estimador generó un CME más pequeño que los estimadores clásicos.

Una vez se ha estimado el coeficiente de autocorrelación, se lleva a cabo una comparación de los estadísticos con respecto a las pruebas de la hipótesis. La H_0 de autocorrelación se prueba comúnmente dividiendo r_1 por $1/\sqrt{n}$. El estadístico de prueba es comparado con los valores críticos de la distribución normal. Huitema y McKean (1991) encontraron que las tasas del error Tipo I empírico se desvían considerablemente del valor nominal. Por este motivo sugieren usar el estimador de error estándar de r_1 propuesto por Moran (1948):

$$e.e.(r_1) = \frac{n-2}{n\sqrt{n-1}}$$

Esta prueba mostró una mayor potencia que la convencional, aunque era baja para tamaños de muestra excesivamente pequeños.

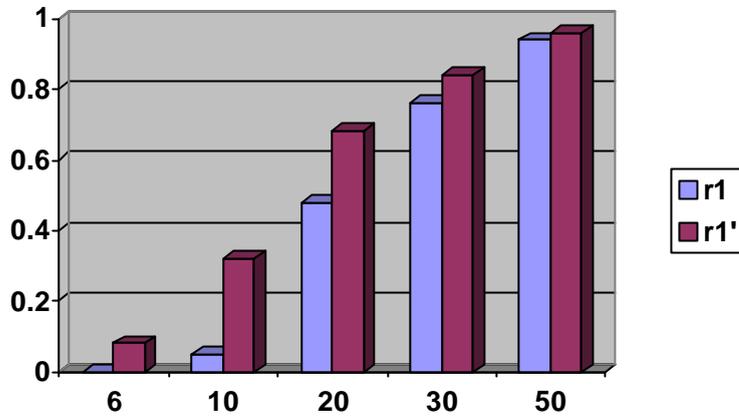
Existe un estimador de la autocorrelación propuesto para el análisis de series temporales cortas: el estadístico C. Un estudio de simulación de Monte Carlo demostró que este estadístico funciona tan bien o mejor que el estimador modificado r_{1+} (DeCarlo y Tyron, 1993).

$$C = 1 - \frac{\sum_{t=1}^{n-1} (Y_t - Y_{t+1})^2}{2 \sum_{t=1}^n (Y_t - \bar{Y})^2}$$

Aunque se han propuesto otros estimadores ninguno corrige el sesgo totalmente. Arnau y Bono (1998) y Arnau propone un estimador alternativo de autocorrelación de primer orden (r_1') que consiste en la corrección del estimador convencional r_1 por el valor absoluto de un modelo de ajuste. Este modelo se obtiene de la función polinómica del sesgo para cada tamaño de la muestra.

$$r_1' = r_1 + \text{modelo}$$

Este estimador reduce el sesgo tanto en valores positivos como negativos del parámetro. El CME presenta además valores inferiores a los obtenidos por el estimador convencional. Este estimador además es más preciso en tamaños de muestras pequeños.



Potencia de los estimadores r_1 y r_1' para $\rho_1 = 0.6$.

El sesgo de r_1' disminuye a partir de todos los valores de n y ρ_1 . Por ello, el estimador r_1' constituye la mejor alternativa para estimar la autocorrelación de primer orden con muestras inferiores a 50 observaciones.

BIBLIOGRAFÍA

- **CAPÍTULO DE LIBROS**

HUITEMA.B.E. (1986). Autocorrelation in behavioural research. Where art thou?. En A. Poling y R.W. Fuquea (Eds). **Research Methods in Applied Behavioral Análisis**. New York: Plenum Press. 187 – 208.

MATYAS.T.A. y GREENWOOD.K.M. (1997). Serial Dependency in Single- Case Time Series. En R.D. Franklin. D.B.Allison y B.S. Gorman (Eds). **Design and Análisis of Single-Case Research**. Mahwah.N.J. L.E.A. Publ. 215–243.

- **ARTÍCULOS**

ARNAU.J. (1999 a). Reducción del sesgo en la estimación de la autocorrelación en series temporales cortas. **Metodología de las Ciencias del Comportamiento. 1 (1)**. 25-37.

ARNAU.J. (1999 b). Series temporales cortas y mínimos cuadrados generalizados: análisis de la intervención. **Metodología de las Ciencias del Comportamiento. 1 (2)**. 119-135.

ARNAU.J. (1999 c). Series temporales cortas y mínimos cuadrados generalizados: respuestas y comentarios. **Metodología de las Ciencias del Comportamiento. 1 (2)**. 201-206.

BAER.D.M. (1998). An autocorrelated commentary on the need for a different debate. **Behavioral Assessment. 10**. 295-297.

BLANCA.M.J y ATO.M. (1999). Una aproximación empírica al problema de la autocorrelación. **Metodología de las Ciencias del Comportamiento. 1 (2)**. 185-200.

VALLEJO.G. (1986). Aplicación de análisis de series temporales en diseños con $n=1$. **Revista Española de Terapia del Comportamiento. 4. (1)** 1-29.