

TESIS DOCTORAL

**DISEÑOS DE SERIES TEMPORALES INTERRUMPIDAS:
TÉCNICAS ALTERNATIVAS DE ANÁLISIS**

ROSER BONO CABRÉ

7D 332

UNIVERSIDAD DE BARCELONA

**DEPARTAMENTO DE METODOLOGÍA DE LAS CIENCIAS DEL
COMPORTAMIENTO**

**PROGRAMA DE DOCTORADO: MODELOS DE INTERVENCIÓN
PSICOLÓGICA (BIENIO 1990-92)**

**DISEÑOS DE SERIES TEMPORALES INTERRUMPIDAS:
TÉCNICAS ALTERNATIVAS DE ANÁLISIS**

TESIS DOCTORAL

Presentada por:

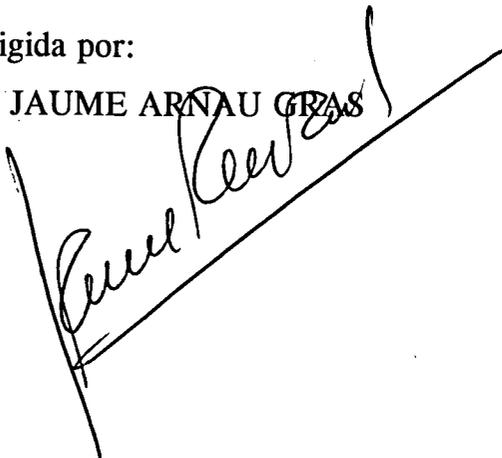
ROSER BONO CABRÉ



Para optar al título de Doctor en Psicología

Dirigida por:

Dr. JAUME ARNAU GRAS



Barcelona, agosto de 1994

Als meus pares

AGRADECIMIENTOS

Deseo expresar mi más sincero agradecimiento a todas aquellas personas que, de una u otra forma, han contribuido tan generosamente a que el presente trabajo de investigación pudiera realizarse. Es por ello que, antes de dar comienzo al mismo, quisiera dejar constancia de su ayuda.

En primer lugar, y de forma muy especial, reconozco mi deuda con el Dr. Jaume Arnau Gras, director de la presente investigación. Él ha sido quien, en mis tiempos de estudiante y durante el transcurso de sus clases, despertó mi interés por los *diseños experimentales* y quien, después de mi graduación, se ha preocupado intensamente por mi formación académica y científica. Su apoyo y disponibilidad constante ha posibilitado el llevar a su fin el presente trabajo. Le agradezco sus sugerencias acerca de los objetivos centrales de la investigación, sus valiosas ideas y aportaciones a la hora de planificar, organizar y estructurar cada una de las partes de este trabajo, y su rigurosa revisión de la obra completa. Debo también agradecerle el haber puesto a mi disposición material bibliográfico, así como instrumentos indispensables para la realización de la investigación. Por otra parte, me gustaría dejar constancia de mi gratitud por otras razones como, por ejemplo, las amenas conversaciones mantenidas sobre el tema y la mutua cooperación durante este período de tiempo. De igual forma, quiero agradecerle su paciencia por haber soportado mis múltiples consultas y angustias. Tarea difícil que no muchos hubiesen resistido. Sin embargo, con su optimismo ha sabido infundirme nuevas energías. El Dr. Arnau no sólo ha dirigido esta tesis sino que ha sido un gran amigo en los momentos difíciles, siempre dispuesto a escuchar y aconsejarme oportunamente. Por último, deseo pedirle disculpas por haber abusado, en infinidad de ocasiones, de su confianza. También quiero que sepa que de él he aprendido un ritmo intenso de trabajo y ésta es, sin lugar a dudas, la mejor influencia que he recibido, hasta el momento, en mi vida profesional.

También es mi deseo dejar constancia expresa de la gratitud que debo al Dr. Pedro Sánchez Algarra por el soporte matemático necesario para la comprensión de algunos de los artículos fundamentales en que se basa el presente trabajo. Igualmente merecen mi agradecimiento los Dres. Salvador Herrando Borge y Joaquim Veà Baró por enseñarme el manejo de algunos programas de ordenador y por la ayuda prestada en todas las ocasiones que la he solicitado. En la misma línea se dirige mi gratitud hacia el Dr. Vicenç Quera Jordana, a quien se debe la elaboración del *software* para el cálculo del estadístico C. De igual modo deseo manifestar mi agradecimiento, aún sin conocerlo personalmente, al Dr. Guillermo Vallejo Seco por su disposición en proporcionarme material bibliográfico, así como por sus comentarios telefónicos con respecto al tema. Al Dr. Francesc Salvador Bertran, a la Dra. Teresa Anguera Argilaga y a otros profesores, aún sin hacer mención explícita de ellos, he de agradecerles sinceramente la constante preocupación y el interés mostrado en el avance de mi trabajo. También agradezco la ayuda económica que ha recibido la presente investigación por parte de la División de Ciencias de la Salud de la Universidad de Barcelona.

Es importante destacar la amabilidad y eficiencia mostrada por el personal de la Biblioteca de Psicología de la Universidad de Barcelona. Gracias a ellos ha sido posible encontrar aquel material que parecía inaccesible.

A mi amiga, la Dra. Ana Adan Puig, por su ayuda y colaboración desinteresada en la tarea de corrección del texto y de los aspectos formales de presentación, y por sugerirme aclaraciones en algunas partes del trabajo. De la misma forma, he de recordar la corrección de estilo de la obra completa llevada a cabo por la filóloga Adela Mejías.

De manera muy especial quisiera manifestar un profundo agradecimiento a mi amiga y "compañera de fatigas" Loly Montilla Garzón, por su esfuerzo en animarme a la consecución de este trabajo, y por su inagotable paciencia y comprensión. No quisiera dejar de recordar a otros compañeros que, de un modo u otro, me han dado su apoyo, como la profesora del Departamento de Psicología Social, Antonia Ferrer, y la secretaria del Departamento de Metodología de las Ciencias del Comportamiento, Carina Sala.

Por último, quiero dedicar este trabajo a mis padres, por haber mostrado una comprensión sin límites en los momentos de preocupación, por sus consejos, atenciones y desvelos y, especialmente, por la confianza depositada en mí. A mis hermanos Carles y Montserrat, y -como no- a la pequeña Tànit que con sus siete años se ha preocupado mucho por mis altibajos. También a mis amigos y a otros familiares, quienes han sabido comprender las ausencias durante todo este tiempo. Espero no haber abusado en demasía.

A todos vosotros, mi cariño y mi recuerdo.

Barcelona, agosto de 1994

ÍNDICE

INTRODUCCIÓN	1
1. VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD	11
1.1. Variable aleatoria: conceptualización	12
1.2. Función de densidad de probabilidad y función de distribución de probabilidad	16
1.3. Distribuciones absolutamente continuas	23
1.3.1. Funciones Gamma y Beta de Euler	25
1.3.2. Relación entre las funciones Gamma y Beta	29
1.4. Momentos de una función de distribución	33
1.5. Función generatriz de momentos	38
1.6. Función característica	42
1.7. A modo de resumen	46
2. ESTADÍSTICO C: DESCRIPCIÓN	47
2.1. Media cuadrada de las diferencias sucesivas	47
2.2. Descripción del estadístico C	50
2.2.1. Características	53
2.2.2. Momentos de la distribución de C	54
2.2.3. Distribución de C	62
2.3. Tablas de la Función $I_x(p,q)$	66
2.4. Analogía entre la prueba de Durbin-Watson y el estadístico C	70
2.5. A modo de resumen	72

3. ESTADÍSTICO C: APLICACIÓN A DISEÑOS DE SERIES TEMPORALES INTERRUMPIDAS	75
3.1. Estrategias de uso	75
3.2. Estadístico C y diseños de series temporales interrumpidas	81
3.3. Ámbitos de utilización	85
3.4. Limitaciones de aplicación	86
3.5. Principales críticas	88
3.6. A modo de resumen	97
4. CÁLCULO DE LA POTENCIA DEL ESTADÍSTICO C: ESTUDIO DE SIMULACIÓN	99
4.1. Consideraciones generales entorno a los estudios de simulación	99
4.2. Hojas de cálculo: generación y análisis de datos	101
4.3. Potencia de una prueba estadística	102
4.3.1. Desarrollo histórico	103
4.3.2. Concepto de potencia: su relación con errores de Tipo I y Tipo II	107
4.3.3. Elementos de la potencia estadística	110
4.3.3.1. Tamaño muestral	110
4.3.3.2. Nivel de significación	111
4.3.3.3. Tamaño del efecto	112
4.3.4. Tipos de análisis de potencia	114
4.4. Objetivos	115
4.5. Estudio 1	118
4.5.1. Método	118
4.5.2. Resultados y discusión	120

4.6. Estudio 2	130
4.6.1. Método	130
4.6.2. Resultados y discusión	131
4.7. Estudio 3	141
4.7.1. Método	141
4.7.2. Resultados y discusión	144
4.8. Discusión general	163
5. ANÁLISIS DE SERIES TEMPORALES CORTAS:		
ESTADÍSTICO C VS MODELO DE EDGINGTON	169
5.1. <i>Software</i> para el cálculo del estadístico C	170
5.1.1. Características	171
5.1.2. Funcionamiento del programa	172
5.2. Modelo de Edgington	177
5.3. Estudio comparativo: estadístico C vs prueba de Edgington	180
5.4. A modo de resumen	223
6. CONCLUSIONES	225
REFERENCIAS	229

APÉNDICES	263
1. Ficha técnica de las distribuciones continuas más comunes	265
2. Desarrollo matemático del estadístico C para el cálculo de sus momentos	281
3. Diagrama de las áreas y curvas asociadas con las diferentes soluciones de la ecuación diferencial de Karl Pearson	285
4. Cálculo de los valores p y q para entrar en las tablas de la Función $I_x(p,q)$	289
5. Tablas de la Función $I_x(p,q)$ para valores de $p=0.5$ (Pearson, 1924)	293
6. Ejemplo propuesto por Tryon (1982) para el cálculo del estadístico C	309
7. Cuadro resumen de estudios donde se ha aplicado el estadístico C	323
8. Tablas de Crosbie (1989)	327
9. Análisis de la regresión <i>stepwise</i> a partir de las tablas 1-3 de Crosbie (1989)	331
10. Tablas de error Tipo I y potencia del estadístico C para la tercera estrategia de Tryon	337
11. <i>Software</i> para el cálculo del estadístico C	365

INTRODUCCIÓN

Uno de los diseños más comunes para evaluar el impacto que sobre una serie temporal puede ejercer un tratamiento, tanto en Ciencias Sociales como en Ciencias del Comportamiento, es el diseño de series temporales interrumpidas (DSTI), denominado así por Campbell y Stanley (1966). Se trata de una serie de registros, tomados a lo largo del tiempo, de una variable dependiente de una sola unidad observacional o sujeto, donde en algún punto de la secuencia de observaciones se produce una circunstancia o intervención -manipulación de la variable independiente- que puede afectar a las medidas de la variable dependiente (Arnau, en prensa). Este tipo de diseños presenta problemas al evaluar la inferencia del impacto o efecto de los tratamientos en datos conductuales, lo cual significa que, a pesar de la existencia de una gran cantidad de estudios que utilizan DSTI, nos hallamos todavía lejos de disponer de unos métodos apropiados de análisis.

Desde el punto de vista estadístico, los registros generados por un sujeto único en una serie de observaciones sucesivas son los más difíciles de analizar, debido a la presencia de dependencia serial (Glass, Willson y Gottman, 1975; Gottman, 1981; Hartmann, 1974; Hartmann *et al.*, 1980; Kratochwill *et al.*, 1974; Thoresen y Elashoff, 1974). En efecto, Jones, Vaught y Weinrott (1977) comprobaron, sobre la base de 24 series temporales seleccionadas del *Journal of Applied Behavior Analysis (JABA)*, que el 83% de las series presentaban una autocorrelación significativa en el retardo 1 que oscilaba entre 0.40 y 0.93. Este fenómeno, cuando no es tenido en cuenta, puede sesgar sustancialmente la prueba estadística (Velicer y McDonald, 1984). De entre todos los procedimientos estadísticos existentes, el análisis de series temporales interrumpidas (ASTI) es el más recomendado por la mayoría de investigadores, puesto que permite eliminar, a nivel estadístico, el efecto de la dependencia serial inherente a los datos (Arnau, en prensa; Barlow y Hersen, 1973; Glass *et al.* 1975; Gottman, 1973, 1981; Gottman y Glass, 1978; Gottman, McFall y Barnett, 1969; Hartmann *et al.*, 1980; Jones, Weinrott y Vaught, 1978). El ASTI fue propuesto, inicialmente, por Box y Jenkins (1970) y Box y Tiao (1965); se adaptó al ámbito social por Glass *et al.* (1975) y, posteriormente, se sistematizó

en los trabajos de Gottman (1981) y McCleary y Hay (1980). Este procedimiento estadístico consiste en la identificación de un modelo auto-regresivo integrado de medias móviles (ARIMA). De esta forma, se modela la dependencia serial de los datos teniendo en cuenta la evolución de la respuesta en el tiempo; es decir, se transforman las puntuaciones directas en puntuaciones serialmente independientes (no correlacionadas).

El inconveniente principal, que se plantea a raíz de la utilización de los análisis de series temporales para datos conductuales, es la gran cantidad de observaciones requeridas para una correcta identificación del modelo. No existen normas claras sobre cuál ha de ser el número mínimo de observaciones en cada período: dependerá en gran medida de otros factores, tales como la estabilidad o variabilidad de la respuesta. Sin embargo, Glass *et al.* (1975) y McCleary y Hay (1980) establecen un mínimo de 50 observaciones por fase, mínimo que Horne, Yang y Ware (1982) sitúan en 35-40 y Box y Jenkins (1970) en 100. Desafortunadamente, no suelen encontrarse investigaciones dentro del campo aplicado que satisfagan este mínimo. Esto, sin duda, puede constatarse al revisar los gráficos publicados en la mayoría de revistas que utilizan este tipo de diseños. Más concretamente, en *JABA*, el número de observaciones en la línea base oscila, frecuentemente, en torno a cuatro puntos (Huitema, 1985). Por otra parte, Velicer y Harrop (1983) han demostrado que, aun cuando se dispone del número de observaciones requeridas, los investigadores entrenados tienen dificultad en identificar adecuadamente el modelo.

Como se acaba de indicar, al utilizar el enfoque propuesto por Box y Jenkins (1970), el analista se encuentra ante la dificultad del reducido tamaño de muestra, es decir, cómo detectar un efecto de tratamiento con pocos datos por fase. Si a esto se añade la dificultad de identificar la estructura de la autocorrelación, puede concluirse que aún se está lejos de disponer de un análisis estadístico apropiado para datos conductuales.

En un intento de hallar modelos más simples y de identificación menos difícil, que salven tanto la dificultad de la dependencia serial como la gran cantidad de observaciones

requeridas para identificar un modelo ARIMA, surgen una serie de trabajos que plantean diversas alternativas. Tres son las propuestas que prescinden de la fase de identificación del modelo: la asunción de un modelo AR(1), cuando se trata de experimentos de series temporales transversales (Simonton, 1977); la estimación de la matriz de transformación desde los datos empíricos empleando un análisis de perfiles (Algina y Olejnik, 1982; Algina y Swaminathan, 1977, 1979; Swaminathan y Algina, 1977), y la utilización de un modelo auto-regresivo de quinto orden (Velicer y McDonald, 1984). Cada una de estas aproximaciones posee limitaciones propias (véase, por ejemplo, a Velicer y McDonald, 1984). Otra postura alternativa a los ASTI, y que guarda cierto parecido con la de Simonton (1977), es la defendida por Gottman (1981) y Williams y Gottman (1982), quienes abogan por el uso de modelos lineales auto-regresivos de orden p . Una descripción detallada de los problemas que presenta la utilización de este último enfoque viene dada por Greenwood y Matyas (1990).

Dejando de lado los análisis de series temporales, así como los procedimientos alternativos, y retomando la evolución histórica a partir de la década de los setenta, se observa el surgimiento de la polémica acerca de cuáles son los métodos de análisis más adecuados para datos conductuales. El análisis visual ha sido el enfoque dominante durante años (Kratochwill y Brody, 1978; Wampold y Furlong, 1981a) y, en la actualidad, sigue vigente. Entre los autores que lo han adoptado destacan Baer (1977), Ballard (1983), DeProspero y Cohen (1979), Kazdin (1982) y Parsonson y Baer (1978, 1986, 1992).

Los defensores de la interpretación visual consideran que aquellas variables que muestran un efecto menos fuerte deben ser ignoradas, a riesgo de incurrir en un error de Tipo I (Baer, 1977; Parsonson y Baer, 1978). No obstante, no tener en cuenta variables menos potentes puede generar errores de Tipo II (Arnau, 1994). Ésta es una de las principales debilidades del uso exclusivo del análisis visual, ya 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 (Barlow y Hersen, 1984; Morley y Adams, 1991). Bajo tales circunstancias, los investigadores tienen dificultades en determinar si una intervención ha sido o no eficaz. Matyas y Greenwood (1990) comprueban que los analistas visuales encuentran, en muchas ocasiones, intervenciones significativas cuando de hecho no las hay. Además, suelen darse discrepancias interjueces en la interpretación de los gráficos (DeProspero y Cohen, 1979; Furlong y Wampold, 1982; Gottman y Glass, 1978; Jones *et al.*, 1978; Parsonson y Baer, 1986; Sharpley, 1986; Wampold y Furlong, 1981a). Campbell (1974) y Sharpley (1986) también dan importancia a la dificultad que 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. Por otra parte, la orientación analítica o teórica del investigador puede sesgar el análisis gráfico, en el sentido de que algunos investigadores ponen, por ejemplo, mayor énfasis en la variabilidad, mientras que otros dan mayor importancia a la magnitud del cambio de nivel. Las principales limitaciones de la inspección visual de los datos, sobre todo cuando se intenta aplicar única y exclusivamente este procedimiento, están descritas con más detalle por Gottman (1981), Gottman y Glass (1978) y Sharpley (1981, 1986). A pesar de estas críticas, Morley y Adams (1991) han impulsado la actual expansión de las técnicas relativas a las representaciones gráficas (Cleveland, 1985) y el análisis exploratorio de datos (Tukey, 1977; Velleman y Hoaglin, 1981), argumentando que la exploración de los gráficos juega un papel central en la comprensión del conjunto de datos, y proporciona una ayuda a la aplicación de técnicas estadísticas. No se trata, por tanto, de técnicas inferenciales, sino simplemente descriptivas. Para una revisión más amplia del tema, puede consultarse a Arnau (1993, 1994).

Paralelamente a este desarrollo inicial, basado en los análisis visuales o gráficos, aparece la adaptación de las pruebas paramétricas convencionales a datos de replicación intrasujeto, tales como las pruebas *t* de Student-Fisher y *F* de Snedecor (Gentile, Roden y Klein, 1972; Shine y Bower, 1971), o el análisis de la regresión múltiple corregida por

la autocorrelación (Lewis-Beck, 1986; Simonton, 1977). La utilización de las pruebas *t* y *F* ha provocado una serie de controversias. Por un lado, están quienes consideran que estas pruebas son inadecuadas cuando los datos presentan dependencia serial (Greenwood y Matyas, 1990; Hartmann, 1974; Hartmann *et al.*, 1980; Jones *et al.*, 1977; Kratochwill, 1978; Kratochwill *et al.*, 1974; Levin, Marascuilo y Hubert, 1978; Phillips, 1983; Thoresen y Elashoff, 1974; Toothaker *et al.*, 1983), ya que la autocorrelación sesga las estimaciones de la variancia del error, en el sentido de la existencia de una sub-valoración de la probabilidad de cometer un error Tipo I cuando la autocorrelación es positiva y una sobre-valoración, cuando es negativa (Scheffé, 1959). En contrapartida, hay quienes defienden estos procedimientos clásicos, argumentando que las autocorrelaciones entre medidas son cero o no difieren significativamente de cero (Center, Skiba y Casey, 1985-1986; Huitema, 1985, 1986, 1988). Estos argumentos tienen su punto de inicio en los resultados de la investigación de Huitema (1985) quien, a partir de un reanálisis de los datos del trabajo de Jones *et al.* (1977), así como de los publicados en la revista *JABA*, concluye que la presencia de autocorrelación en los residuales de datos conductuales es baja y que, por lo tanto, es adecuado utilizar los métodos estadísticos tradicionales. Esta conclusión ha suscitado una serie de críticas que continúan siendo vigentes (Busk y Marascuilo, 1988; Greenwood y Matyas, 1990; Matyas y Greenwood, 1991; Sharpley y Alavosius, 1988; Suen, 1987; Suen y Ary, 1987). Uno de los argumentos esenciales de estas críticas radica en el tamaño de la muestra (número de observaciones) que suele ser muy pequeño y, por lo tanto, los estimadores de la autocorrelación son insensibles.

Debido a las críticas que ha suscitado el intento de aplicar los estadísticos convencionales se han desarrollado una serie de técnicas estadísticas que, con mayor o menor fortuna, han sido aplicadas a los datos de diseños de sujeto único o $N=1$. Éstas van desde pruebas paramétricas sofisticadas, como el análisis de series temporales, a procedimientos no paramétricos relativamente sencillos. Como se ha visto, el análisis de series temporales requiere una gran cantidad de observaciones con objeto de identificar el proceso. Dada, pues, la escasa viabilidad de los modelos ARIMA para los diseños de caso

único, las técnicas no paramétricas parecen ser, en la actualidad, la alternativa de análisis estadístico más deseable para diseños de series temporales cortas. Algunos de los procedimientos no paramétricos se basan en los principios de aleatorización, tales como las pruebas estadísticas para la aleatoriedad (Edgington, 1975, 1980a, 1980b, 1980c, 1984, 1992; Levin *et al.*, 1978) que no requieren asunciones sobre las poblaciones ni sobre el muestreo aleatorio de las unidades o sujetos (Edgington, 1992; Wampold y Furlong, 1981b). Una aproximación a la prueba para la aleatoriedad es la *U* de Mann-Whitney que requiere la transformación de la variable dependiente a rangos.

Hay una serie de situaciones donde el empleo de las pruebas para la aleatoriedad es muy limitado, por no decir imposible (Arnau, 1986; Barlow y Hersen, 1984). Se trata de aquellas situaciones en que: a) no aparecen cambios muy rápidos o bien no existe ningún cambio en la respuesta al pasar de la condición A a la B; b) los efectos de la variable de tratamiento son irreversibles y su retirada no permite esperar que las respuestas de los individuos regresen a los niveles anteriores, y c) la inversión de la conducta no es aconsejable por razones éticas o médicas. Además, las pruebas estadísticas para la aleatoriedad se adaptan con mucha dificultad a los diseños simples intra-sujeto, con una sola unidad y una sola intervención (Arnau, 1986, 1993), pero a medida que aumenta el número de fases o sujetos, estos análisis son bastante eficaces (Busk y Marascuilo, 1992; Edgington, 1992). En términos generales, los problemas asociados al uso de las pruebas para la aleatoriedad están relacionados con la necesidad de asignar aleatoriamente las condiciones a las ocasiones.

Con base también a los principios del azar, se ha propuesto un procedimiento que permite probar la variación sistemática de una secuencia de medidas ordenadas, es decir, la presencia de tendencias y cambios de tendencia de una fase a la siguiente. Tryon (1982) ha sido quien ha impulsado la utilización de este enfoque. Se trata del estadístico C de Young (1941), para el análisis de series temporales cortas. Esta prueba es considerada como no paramétrica, aunque no lo es estrictamente hablando. Sin embargo, como la

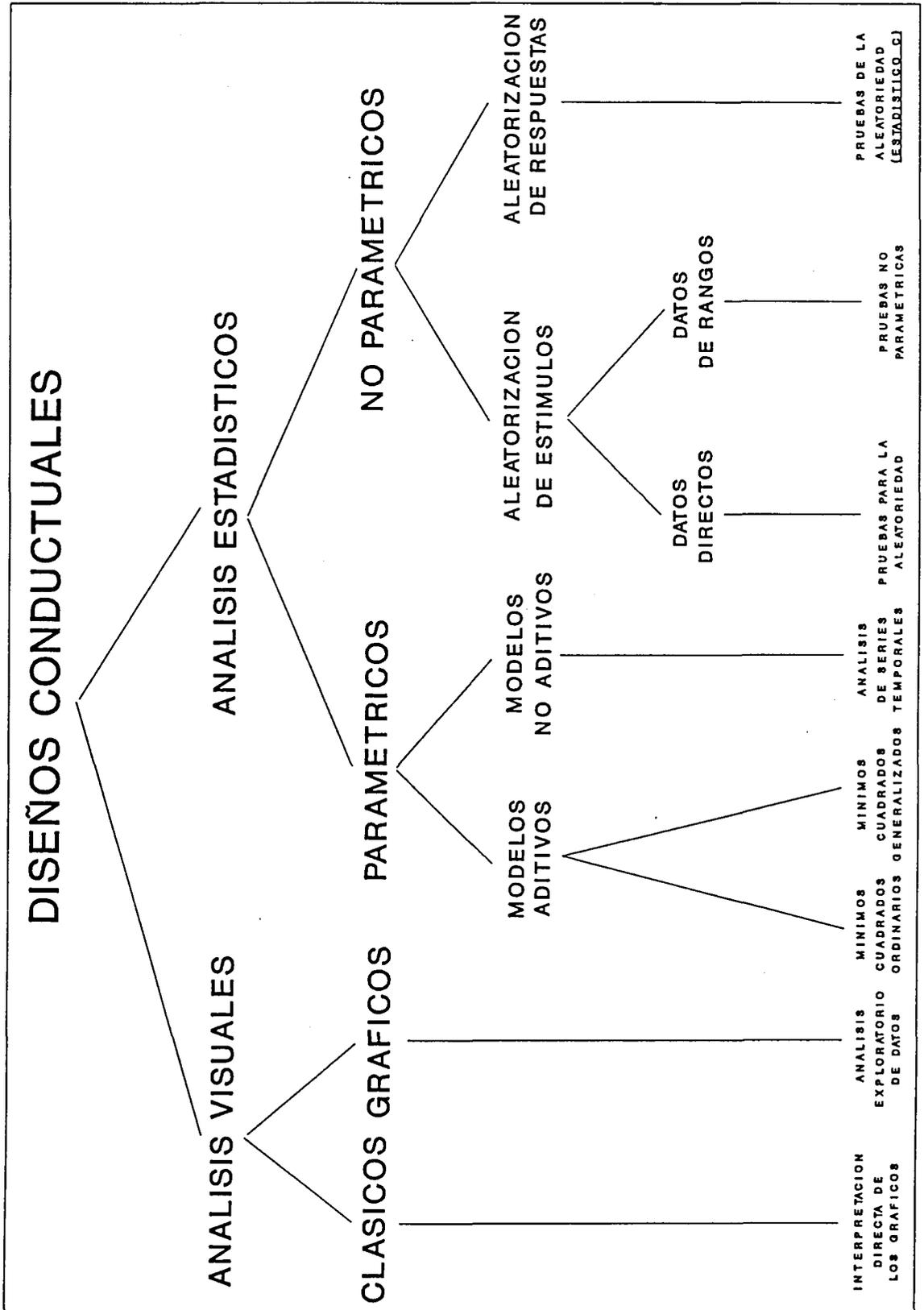
mayoría de las pruebas basadas en la aleatoriedad son no paramétricas y como que ésta es una prueba de la aleatoriedad, generalmente se llama -aunque de forma impropia- no paramétrica (Wonnacott y Wonnacott, 1990).

Hecho este breve repaso de las técnicas más utilizadas y de los autores más relevantes dentro del ámbito de los diseños de $N=1$, presentamos al final de esta introducción un cuadro resumen de una posible clasificación de los análisis de datos conductuales (Bono, 1994). En primer lugar, se hace una distinción entre análisis visuales y análisis estadísticos de los datos. De ahí surge un amplio debate entre los que defienden los análisis visuales y los que consideran mucho más ventajosas las técnicas estadísticas. Dentro de los procedimientos estadísticos, Arnau (1994) plantea un criterio de clasificación más amplio que el propuesto por Robinson, Morran y Hulse-Killackey (1989) entre series temporales y pruebas no paramétricas. En efecto, los análisis estadísticos quedan divididos en paramétricos y no paramétricos. A su vez, se subdivide la primera de estas dos categorías en dos subgrupos: a) modelos aditivos (Lewis-Beck, 1986; Simonton, 1977) y b) modelos no aditivos (Glass *et al.*, 1975; Gottman y Glass, 1978). En cuanto a los análisis estadísticos no paramétricos, se desglosan en: a) modelos basados en la aleatorización de estímulos (Edgington, 1967; Kazdin, 1976; Levin *et al.*, 1978) y b) modelos basados en la aleatorización de respuestas (Tryon, 1982). Finalmente, dentro de la aleatorización de estímulos se establece una distinción entre datos directos (pruebas para la aleatoriedad de Edgington) y datos de rangos (U de Mann-Whitney). Una descripción detallada de cada uno de los subapartados de la clasificación que se presenta se encuentra en Arnau (1994).

A partir de toda la amplia variedad de métodos de análisis aplicables a los datos de diseños de caso único, la presente investigación se centrará exclusivamente en el estadístico C . Para ello, hemos dividido el estudio en cinco capítulos. En el primero, se presentan una serie de conceptos fundamentales de la teoría de probabilidades, necesarios para la posterior descripción, en el segundo capítulo, del estadístico que proponemos. En el tercer

capítulo se ofrecen todos aquellos aspectos referentes a su aplicación. La parte principal del trabajo corresponde al capítulo cuarto, donde se lleva a cabo un estudio de simulación de series temporales cortas, con la finalidad de hallar la potencia estadística de la prueba C. Por último, en el capítulo quinto, se realiza una aplicación del estadístico C a datos reales extraídos de gráficos de artículos de revistas conductuales. Finalmente, también en el último capítulo, se aplica la prueba de Edgington para diseños experimentales A-B, con el propósito de establecer una comparación entre los resultados obtenidos por esta prueba y por el estadístico C.

CLASIFICACION GENERAL DE LOS ANALISIS DE DATOS CONDUCTUALES
(Bono, 1994)



1. VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD

Como se ha indicado en la introducción, el propósito del presente estudio es describir y evaluar el estadístico C de Young como alternativa a los análisis de series temporales, únicamente cuando se dispone de un número insuficiente de observaciones para poder identificar el modelo ARIMA tentativo. No obstante, antes de pasar a su descripción matemática, es necesario, para una mayor comprensión del tema, hacer una referencia a las distribuciones de probabilidad continuas, puesto que, como se verá en el siguiente capítulo, la distribución de C está directamente relacionada con la Función $I_x(p,q)$, e indirectamente con la Función Beta Incompleta.

En primer lugar, se introducirá el concepto de variable aleatoria que es la base para el estudio de la teoría de probabilidades y, posteriormente, los conceptos de función de densidad de probabilidad y función de distribución de probabilidad, así como las distribuciones continuas más comunes, principalmente las funciones Gamma y Beta (relacionadas con la Función $I_x(p,q)$ que se describirá en el apartado 2.3). Por otra parte, se tratarán los momentos de una función de distribución, la función generatriz de momentos y, aunque más brevemente, la función característica. Ello permitirá una mejor comprensión del apartado 2.2.2 dedicado a los momentos de la distribución de C.

Todas estas cuestiones serán expuestas de forma clara y simple, con una terminología acorde con la usada por la mayoría de textos de probabilidad. Las matemáticas necesarias para un tratamiento riguroso de esta temática están más allá de nuestro interés. Sin embargo, se intentará dar una presentación completa e inteligible de esta materia sin tener que recurrir a un cálculo excesivo.

Insistimos, una vez más, en que únicamente se hará referencia al caso continuo debido a que, como se ha indicado anteriormente, la distribución de C está relacionada con distribuciones absolutamente continuas. Sin embargo, cualquier discusión o demostración acerca de variables aleatorias continuas es válida, en casi todos los aspectos, para el caso

discreto, cambiando la notación. Esto es, en las funciones de probabilidad continuas se utiliza el signo \int que es reemplazado por el signo \sum en el caso discreto. No obstante, es importante aclarar que, en el caso continuo, la función de densidad de probabilidad, designada por $f(x)$, no es por sí misma una probabilidad, y sólo se la considerará como tal cuando esté multiplicada por una diferencial dx .

1.1. Variable aleatoria: conceptualización

En general, para describir los fenómenos aleatorios únicamente se requiere una información cuantitativa, es decir, tan sólo son indispensables datos numéricos (Ostle, 1965; Polo, 1985). Por ejemplo, el número de cigarrillos fumados por un individuo en un período de tiempo, las calificaciones de los estudiantes en un examen, etc. Incluso en los casos donde los resultados no son numéricos, como clasificar un artículo manufacturado según las categorías "defectuoso"/"no defectuoso", se asigna un número a cada uno de los resultados del experimento. Esto significa que en muchas situaciones experimentales se asigna un número real x a cada uno de los elementos s del espacio muestral S , y son estas magnitudes las que reciben el nombre de variables aleatorias (Feller, 1989; Fisz, 1963; Loève, 1963; McPherson, 1990; Papoulis, 1980; Quirin, 1978). Este concepto está directamente relacionado con el de función (Feller, 1989; Glass y Stanley, 1970). Por esta razón, es necesario definir, en primer lugar, el concepto general de función:

Un objeto X o $X(\cdot)$ es una función definida en un espacio S , si para todo elemento s de S hay un número real, denotado por $X(s)$, al cual denominamos el valor de la función X en s (Parzen, 1960; p. 269).

Esta función se llama variable aleatoria si cumple las condiciones generales especificadas en la siguiente definición:

Un objeto X es una variable aleatoria a) si es una función de valores reales definida en un espacio de descripciones muestrales, sobre una familia de cuyos subconjuntos se ha definido una función de probabilidades $P[\cdot]$, y b) si para todo conjunto boreliano B de números reales, el conjunto $\{s: X(s) \in B\}$ pertenece al dominio de $P[\cdot]$ (Parzen, 1960; p. 269).

Una variable aleatoria X es una función definida con base al resultado de un fenómeno aleatorio (Bendat y Pierson, 1986). Esto es, el valor de una variable aleatoria es un fenómeno aleatorio con resultados numéricos cuyo espacio de descripciones muestrales S es el conjunto R (que contiene todos los números reales desde $-\infty$ a $+\infty$) en cuyos subconjuntos está definida una función $P[\cdot]$, que asigna a cada conjunto boreliano de números reales E (también llamado evento) un número real no negativo, representado por $P[E]$, de acuerdo con los siguientes axiomas (Barr y Zehna, 1983; Parzen, 1960; Thomas, 1986):

AXIOMA 1. $P[E] \geq 0$ para todo evento E .

AXIOMA 2. $P[R] = 1$.

AXIOMA 3. Para toda sucesión de eventos $E_1, E_2, \dots, E_n, \dots$, que sea mutuamente exclusiva, se tiene que

$$P\left[\bigcup_{n=1}^{\infty} E_n\right] = \sum_{n=1}^{\infty} P[E_n]$$

Así pues, no todo conjunto de números reales puede considerarse como un fenómeno aleatorio. Existen ciertos conjuntos definidos por operaciones bastante complicadas de límites que no son probabilizables (Feller, 1978; Majó, 1964; Papoulis, 1980; Polo, 1985), en el sentido de que no cumplen los axiomas anteriores. Por esta razón, los conjuntos reales se restringen a los conjuntos borelianos; es decir, a una familia más pequeña que posee las mismas propiedades que la familia de conjuntos probabilizables \mathcal{S} , y puesto

que B está contenido en \mathcal{S} , todo conjunto boreliano es probabilizable. Por lo tanto, se entiende por fenómeno aleatorio no sólo un conjunto de números reales sino, también, un conjunto probabilizable de números reales.

La función de probabilidad $P[E]$ representa la probabilidad de que el resultado del fenómeno aleatorio esté en E o simplemente la probabilidad de E , y debe tener siempre las propiedades indicadas anteriormente. Estas propiedades pueden interpretarse aproximadamente como sigue: la probabilidad de que ocurra cualquier suceso es no negativa, la certeza se representa por la probabilidad 1 y las probabilidades se suman. Estos tres axiomas se hallan planteados, con más detalle, en Johnson y Kotz (1969) y Lippman (1971).

En la mayor parte de las aplicaciones de la teoría de probabilidades, las variables aleatorias suelen ser de dos tipos: discretas o continuas (Beaumont, 1986; Clarke y Disney, 1985; Glass y Stanley, 1970; Lippman, 1971; Taylor, 1974). Aunque a lo largo de este estudio sólo se hará referencia a las continuas, es conveniente establecer una distinción entre estos dos tipos de variables:

Sea X una variable aleatoria. Si el número de valores posibles de X es finito o infinito numerable, se denomina X a una *variable aleatoria discreta*. Esto es, se pueden anotar los valores posibles de X como $x_1, x_2, \dots, x_n, \dots$. En el caso finito la lista termina y en el caso infinito numerable la lista continúa indefinidamente (Meyer, 1986; p. 60).

Se define X como una *variable aleatoria continua* si existe una función $f(x)$, llamada función de densidad de probabilidad (f.d.p.) de X , que satisface las siguientes condiciones:

a) $f(x) \geq 0$ para todo x .

b)
$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

c) Para cualquier a, b , tal que $-\infty < a < b < +\infty$, se tiene

$$P[a \leq X \leq b] = \int_a^b f(x) dx .$$

(Meyer, 1986; p. 68.)

En conclusión, se afirma que X es una variable aleatoria continua si puede tomar todos los valores en algún intervalo (a, b) . Es decir, si su función de distribución no tiene discontinuidades. Un ejemplo sería el punto en que se produce un accidente en una carretera de 100 km. En este caso, el espacio muestral es un intervalo continuo: $S=(0,100)$. Por el contrario, un ejemplo de variable aleatoria discreta sería observar la cara superior de un dado lanzado sobre una mesa. En este último caso, el espacio muestral contiene seis resultados posibles, luego es discreto.

En la sección 1.3.1, se presentarán dos variables aleatorias continuas. Se trata de las funciones Gamma y Beta, particularmente importantes para nuestros propósitos. Sin embargo, antes de pasar a este punto, es conveniente establecer una distinción entre las funciones de densidad de probabilidad y las funciones de distribución de probabilidad, puesto que son dos instrumentos valiosísimos que definen la distribución de una variable aleatoria continua, tanto desde el punto de vista teórico como práctico.

1.2. Función de densidad de probabilidad y función de distribución de probabilidad

La función de probabilidad $P[E]$ es la probabilidad del suceso representado por el conjunto de Borel. Para muchas funciones de probabilidad existe una función $f(x)$, definida para todos los números reales x , a partir de la cual, por integración, puede obtenerse la probabilidad de E para cualquier evento E (Loéve, 1963; Rozanov, 1973; Turner, 1974):

$$P[E] = \int_E f(x) dx \quad (1.1)$$

Cuando una función de probabilidad $P[E]$ pueda representarse mediante la fórmula 1.1 en términos de alguna función $f(x)$, se afirma que $f(x)$ es la función de densidad de probabilidad (f.d.p.) de la función de probabilidad $P[E]$, o que la función de probabilidad $P[E]$ está especificada mediante la f.d.p. $f(x)$ (Fisz, 1963).

De acuerdo con la definición de variable aleatoria continua, propuesta por Meyer (1986), una función $f(x)$ debe poseer ciertas propiedades para ser una f.d.p. de una variable aleatoria continua X (Barr y Zehna, 1983; Cuadras, 1985; Feller, 1989; McPherson, 1990; Parzen, 1960; Turner, 1974; Uspensky, 1947; Wadsworth y Bryan, 1960).

La primera propiedad a satisfacer es que $f(x)$ sea mayor o igual que cero, para toda x contenida en R :

$$f(x) \geq 0 \quad -\infty < x < +\infty \quad (1.2)$$

Esta propiedad expresa que la curva representativa de la función de densidad $f(x)$ se encuentra por encima del eje de abscisas.

La segunda exige que la integral de la expresión 1.1 esté bien definida. Si se sustituye E por R en la ecuación 1.1, se tiene:

$$P[R] = \int_R f(x) dx = \int_{-\infty}^{+\infty} f(x) dx = 1 \quad (1.3)$$

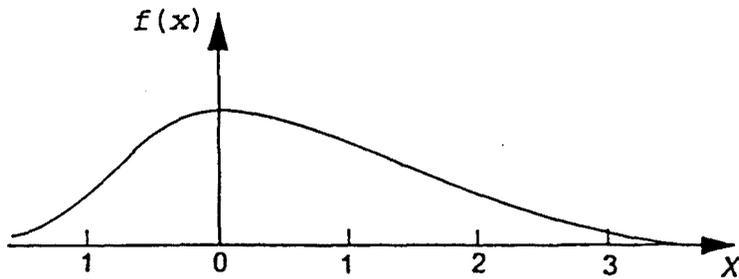
El área encerrada por la curva $f(x)$ contiene todos los valores posibles de la variable aleatoria y, puesto que a cada uno de estos valores le corresponde una probabilidad, la suma de las probabilidades de todos los resultados posibles es igual a la unidad.

La tercera propiedad a cumplir es que la probabilidad de un suceso E está situada en el intervalo designado por dx , es decir, el área bajo la curva de la f.d.p. $f(x)$ entre $x=a$ y $x=b$ queda representada por $P[a \leq X \leq b]$:

$$P[E] = \int_a^b f(x) dx \quad (1.4)$$

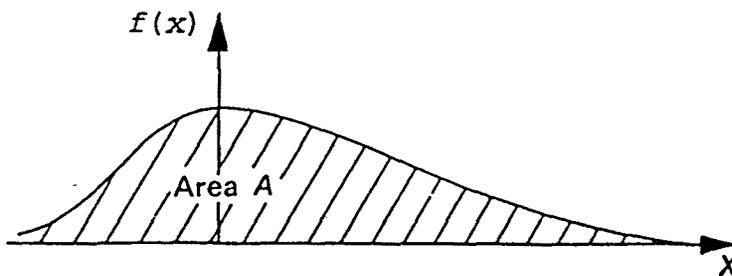
En los diagramas de la figura 1.1 se ilustran estas propiedades esenciales de las f.d.p.

a) $f(x) \geq 0$, siempre



Espacio de muestras: conjunto de los números reales

b) Área total = $\int_{-\infty}^{+\infty} f(x) dx = 1$



c) Suceso $E_1 = x \leq 0$: $P[E_1] = \int_{-\infty}^0 f(x) dx$

Suceso $E_2 = a \leq x \leq b$: $P[E_2] = \int_a^b f(x) dx$

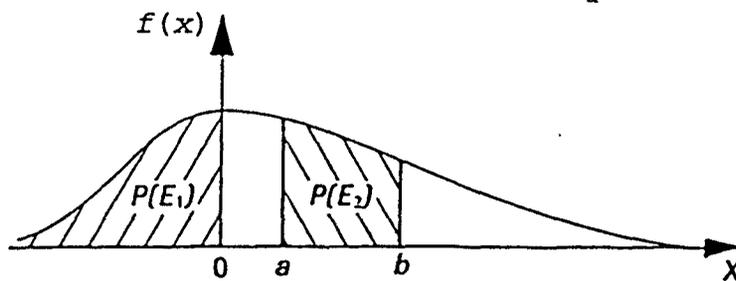


Figura 1.1. Ilustración de las propiedades de las f.d.p. para distribuciones continuas. (Turner, 1974; p. 215.)

Cabe destacar que las variables aleatorias no son de interés por sí mismas. A ellas están asociadas las funciones de distribución. Esto significa que las probabilidades de las variables aleatorias se expresan por medio de sus correspondientes funciones de distribución (Fisz, 1963; Gnedenko, 1976; Gnedenko, Belyayav y Solovyev, 1969; Loève, 1963; Quesada y Pardo, 1987). En efecto, la función de distribución contiene toda la información necesaria para describir las propiedades probabilísticas de un fenómeno aleatorio.

La función de distribución $F(\cdot)$ de un fenómeno aleatorio con resultados numéricos es más manejable que la de probabilidad y se define, para cualquier número real x , como la probabilidad de que un valor observado del fenómeno aleatorio sea menor o igual que el número x (Clarke, 1975; Papoulis, 1980; Parzen, 1960). Así, si X es una variable aleatoria, se denominará función de distribución de X a la función $F(x)$, definida por

$$F(x) = P[X \leq x] \quad (1.5)$$

para todo número real x (Cramer, 1946; Cuadras 1985; Larson, 1982; Thomas, 1986).

Si se especifica la función de probabilidad por una f.d.p. $f(\cdot)$, entonces la función de distribución $F(\cdot)$ correspondiente, para cualquier número real x , viene dada por la siguiente ecuación (Parzen, 1960; Tucker, 1967):

$$F(x) = \int_{-\infty}^x f(x') dx' \quad (1.6)$$

Las propiedades que cumple la función de distribución (Larson, 1982; Ostle, 1965; Quirin, 1978; Rozanov, 1973; Thomas, 1986; Tucker, 1967) son las siguientes:

- 1) $F(x)$ toma valores en el intervalo $[0,1]$:

$$0 \leq F(x) \leq 1 \quad \forall x \in R \quad (1.7)$$

Propiedad lógica, puesto que $F(x)$ es una probabilidad y, por definición, las probabilidades están comprendidas entre 0 y 1.

- 2) $F(x)$ comprende los extremos 0 y 1. El primer caso no tiene acumulación posible, y el segundo representa toda la masa de probabilidad correspondiente al intervalo total,

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P[X \leq x] = P[\emptyset] = 0 \quad (1.8)$$

$$\lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} P[X \leq x] = P[R] = 1 \quad (1.9)$$

- 3) $F(x)$ nunca es decreciente, puesto que el incremento del límite x implica una mayor acumulación del área, o sea, una mayor $F(x)$

$$\text{si } x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2) \quad \forall x_1, x_2 \in R \quad (1.10)$$

- 4) Conocida la función de distribución $F(x)$ de una variable aleatoria X , se puede calcular fácilmente la probabilidad de que la variable tome valores comprendidos entre a y b sin necesidad de integrar:

$$P[a < X \leq b] = F(b) - F(a) \quad (1.11)$$

- 5) $F(x)$ es continua por la derecha, es decir,

$$\lim_{\substack{x \rightarrow a \\ x > a}} F(x) = F(a), \quad \forall a \in R \quad (1.12)$$

- 6) Una última propiedad interesante de la función de distribución es que, por su carácter de función integral de la función de densidad, su derivada será igual a la función de densidad, esto es:

$$F'(x) = \frac{dF(x)}{dx} = f(x) \quad (1.13)$$

En resumen, toda función de distribución $F(x)$ es una función monótona, no decreciente, no negativa, continua a la derecha en cada punto y de valores comprendidos entre 0 y 1. Su forma general queda representada por la figura 1.2.

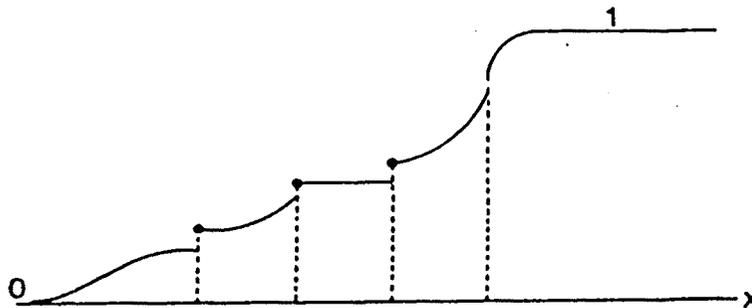


Figura 1.2. Forma general de la función de distribución. (Sánchez, 1988a; p. 147.)

Para un estudio más detallado de estas propiedades, puede consultarse a Fisz (1963), Gnedenko (1976), Polo (1985), Sánchez (1988a), Thomas (1986), Tucker (1966) y Wadsworth y Bryan (1960), entre otros.

Recíprocamente, toda función $F(x)$ define la función de probabilidad de una cierta variable aleatoria. En consecuencia, la ley de probabilidad de una variable aleatoria queda definida, indistintamente, por su función de probabilidad o por su función de distribución.

Si se considera un fenómeno aleatorio cuya función de probabilidad está especificada por una f.d.p. $f(\cdot)$, de manera que su función de distribución $F(\cdot)$ está dada por 1.6, entonces la gráfica de $y=F(x)$ aparece como una curva sin cortes (figura 1.3), es decir la función F es continua. Por lo tanto, se puede afirmar que una variable aleatoria es continua si su función de distribución no presenta discontinuidades.

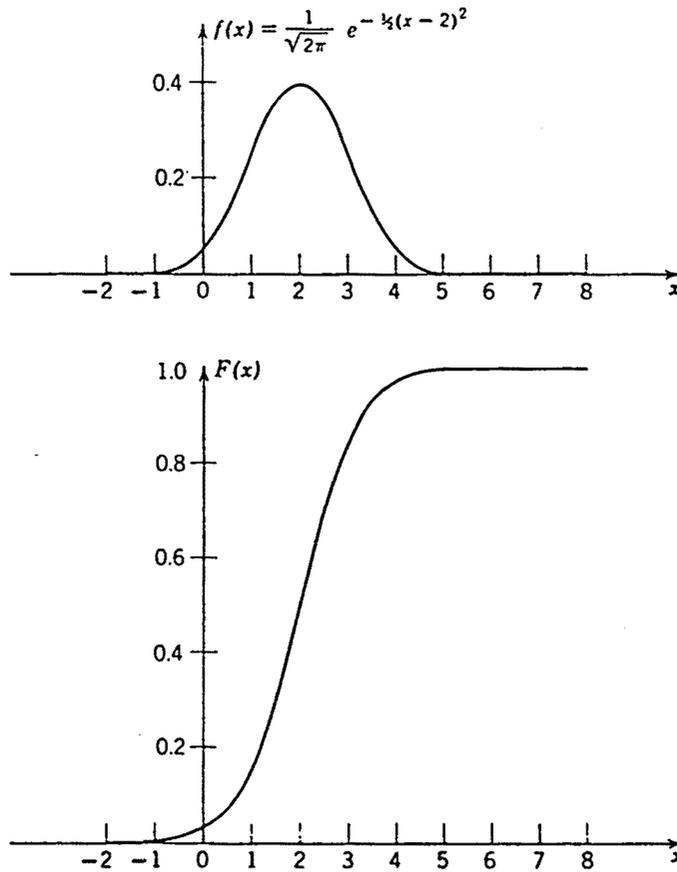


Figura 1.3. Gráfica de una función de distribución continua $F(\cdot)$ y de la f.d.p. $f(\cdot)$ en cuyos términos está dada $F(\cdot)$ por la ecuación 1.6. (Parzen, 1960; p. 168.)

Al utilizar una función de probabilidad continua como modelo para un experimento, hay que asumir dos proposiciones acerca de las probabilidades (Meyer, 1986; Turner, 1974; Uspensky, 1947; Wadsworth y Bryan, 1960). En primer lugar, cuando X es una

variable aleatoria continua toma una infinidad no numerable de valores. Por lo tanto, no resulta apropiado hallar la probabilidad de cada valor de la variable, ya que es de esperar que cada uno de ellos tenga probabilidad nula. Matemáticamente se expresaría como

$$P[X=x]=0 \quad \text{para cada suceso puntual } \{x\}.$$

Según esta primera proposición, la función de probabilidad para una variable aleatoria continua es igual a cero en todos los puntos. En consecuencia, debe cumplirse la segunda proposición, en la que se igualan cuatro probabilidades,

$$P[a < x < b] = P[a \leq x < b] = P[a < x \leq b] = P[a \leq x \leq b]$$

o sea, no importa si se incluyen los puntos extremos al calcular la probabilidad de que x tome algún valor en el intervalo de a a b .

En estos dos primeros apartados han quedado delimitados, de forma sencilla, los fundamentos de una teoría matemática de las variables aleatorias y de las distribuciones de probabilidad. El próximo objetivo será el estudio de dos distribuciones continuas: la Gamma y la Beta. El interés de ambas radica en su conexión con la función $I_x(p,q)$, muy relacionada con el estadístico C .

1.3. Distribuciones absolutamente continuas

Según Tucker (1966), muchos investigadores que se inician en la teoría de probabilidades infieren, a partir de las lecturas de manuales de estadística, la existencia de dos distribuciones de probabilidad: las discretas y las continuas. Sin embargo, estas distribuciones tan sólo son dos clases de todo el conjunto de distribuciones de probabilidad, con la particularidad de ser las más fáciles de manejar para quien no esté muy

familiarizado con la teoría de la medida y la integral de Lebesgue. Puede consultarse, para un mayor conocimiento de las medidas Lebesgue y de Lebesgue-Stieltjes, a Adams y Guillemin (1986), Bass (1974), Cramer (1946), Fabian y Hannan (1985), Feller (1978), Milton y Tsokos (1976), Pesin (1970), Pitt (1963) y Quesada y Pardo (1987).

Antes de continuar, y para que no se infiera la idea errónea de que las distribuciones de probabilidad discretas y continuas dicotomizan el conjunto de todas las distribuciones de probabilidad, es importante resaltar que hay muchas funciones de distribución que no son discretas ni continuas. Por ejemplo, una función de distribución puede ser continua y, sin embargo, puede no obtenerse por la fórmula 1.6 mediante una función de densidad (Quesada y Pardo, 1987). A ese tipo pertenecen las distribuciones singulares. Se tienen, por tanto, tres tipos "puros" de distribuciones en R : las distribuciones de tipo discreto, las de tipo continuo y las singulares. Por último, también existen las variables aleatorias de tipo mixto (Meyer, 1986; Papoulis, 1980).

Otro aspecto a tener en cuenta es que, muchas veces, se utiliza el término continuo aunque, en realidad, el nombre matemático es absolutamente continuo (Gnedenko, 1976; Gnedenko y Kolmogorov, 1968; Sánchez, 1988c). Ejemplos de variables aleatorias absolutamente continuas son el peso de los individuos de una determinada población, la estatura de una persona, las calificaciones de un examen, etc.

En el apéndice 1 se muestra la ficha técnica de las distribuciones continuas más comunes, extraídas de Chambers y Fowlkes (1966), Christensen (1989), Fowlkes (1987) y Rothschild y Logothetis (1986), así como su representación gráfica.

Sin duda, la distribución normal es la más importante, y más ampliamente utilizada, puesto que está relacionada con diversas distribuciones. Sin embargo, no se llevará a cabo un estudio de la misma, ya que el interés del presente trabajo va en otra dirección. Por lo tanto, nos centraremos tan sólo en las funciones Gamma y Beta.

1.3.1. Funciones Gamma y Beta de Euler

La función Gamma, denominada también integral de Euler de segunda especie, se denota por Γ y queda definida en el intervalo $(0, \infty)$ por la siguiente integral:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \quad (1.14)$$

La integral impropia anterior existe sólo si $p > 0$.

Una variable aleatoria continua X , que toma sólo valores positivos, tiene una distribución de probabilidad Gamma si su f.d.p. viene dada por

$$f(x) = \frac{\lambda^p}{\Gamma(p)} e^{-\lambda x} x^{p-1} \quad \text{si } x > 0$$

(1.15)

$$f(x) = 0 \quad \text{si } x \leq 0$$

Esta distribución depende de dos parámetros positivos, λ y p , y se simboliza por $G(\lambda, p)$. En el apéndice 1 (p. 276) se muestra gráficamente la f.d.p. de la ecuación 1.15 para $\lambda=1$ y diversos valores de p , pudiéndose observar la variedad de situaciones susceptibles de representarse mediante esta distribución.

Para ciertos valores de λ y p se pueden obtener otras distribuciones de la misma familia que la Gamma (Kalbfleisch, 1979; Meyer, 1986; Sánchez, 1988c; Wadsworth y Bryan, 1960). Por ejemplo, cuando $p=1$, la distribución Gamma coincide con la distribución exponencial y, cuando $\lambda=1/2$ y $p=n/2$ (donde n es un entero positivo), con la distribución ji-cuadrado con ν grados de libertad. Así pues, la distribución exponencial y la ji-cuadrado son casos especiales de la distribución Gamma. También existen relaciones con otras distribuciones, como la de Poisson. Sin embargo, no se abordará aquí esta

temática, tan sólo queremos dejar bien claro el papel central que juega la distribución Gamma en la estadística, pues muchas distribuciones son casos particulares de ésta. Para un mayor desarrollo de este aspecto puede consultarse a Meyer (1986).

Las principales propiedades de la función Gamma (Caridad, 1985; Cramer, 1946; Cuadras, 1985; Feller, 1978; Fisz, 1963; Guillén, 1991; Jambunathan, 1954; Johnson y Kotz, 1969; Meyer, 1986; Parzen, 1960; Sánchez, 1988c) son las que siguen:

- 1) La integral $\Gamma(p)$ existe si y sólo si $0 < p < \infty$.
- 2) La función $\Gamma(p)$ es absolutamente continua en $]0, \infty[$.
- 3) La función $\Gamma(p)$ es infinitamente derivable y para todo entero k , la derivada k -ésima de Γ es la integral

$$D^k \Gamma(p) = \int_0^{\infty} x^{p-1} \ln^k x e^{-x} dx, \quad 0 < p < \infty \quad (1.16)$$

- 4) $\Gamma(1) = 1$.
- 5) Integrando por partes, se obtiene de la ecuación 1.14, para todo $0 < p < \infty$, que $\Gamma(p+1) = p\Gamma(p)$ y, en particular, si n es un entero positivo, $\Gamma(n+1) = n!$. Luego puede considerarse que la función Gamma es una generalización de la función factorial.
- 6) $\Gamma(1/2) = \sqrt{\pi}$.
- 7) $\Gamma(p) > 0$, para todo p .

- 8) La media y la variancia de la distribución Gamma son, respectivamente:

$$E(X) = \frac{P}{\lambda} \quad (1.17)$$

$$\text{Var}(X) = \frac{P}{\lambda^2} \quad (1.18)$$

- 9) Otra característica importante, considerada por Rothschild y Logothetis (1986), es la llamada propiedad reproductiva: si X_1 y X_2 son variables aleatorias independientes con distribución Gamma $G(p_1, \lambda)$ y $G(p_2, \lambda)$, respectivamente, entonces la variable aleatoria que resulta ($X_1 + X_2$) también tiene una distribución Gamma con parámetros $(p_1 + p_2)$ y λ . Dado que el parámetro $(p_1 + p_2)$ aumenta, la forma de la distribución llega a ser similar a la de la distribución normal (véase apéndice 1, p. 276).
- 10) Respecto al gráfico del apéndice 1 (p. 276), la función Gamma es claramente monótona si $p \leq 1$ y no está acotada cerca del origen. En cambio, para $p > 1$ el gráfico tiene forma de campana.
- 11) Para valores grandes de p hay una fórmula asintótica de la función Gamma, que se conoce como la *fórmula de Stirling*. Esta fórmula y su demostración se puede encontrar en Uspensky (1947).

La distribución Gamma posee un papel primordial en la distribución Beta, puesto que es un caso particular (Johnson y Kotz, 1970). La función Beta se define para todo $p > 0$ y $q > 0$ por la integral de Euler de primera especie:

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx \quad (1.19)$$

Esta distribución es de gran importancia en la estadística Bayesiana, pues permite representar situaciones muy diversas de información sobre una variable aleatoria y, además, es muy flexible.

Los gráficos del apéndice 1 (p. 279) revelan muchas de sus propiedades, según los valores que adopten los parámetros. Por ejemplo, si $p > 1$ y $q > 1$, la distribución Beta es unimodal con la moda en el punto $x = p-1/p+q-2$. Cuando $p=q=1$, la distribución Beta coincide con la distribución uniforme en el intervalo $(0,1)$. Cuando $q=p+1$, la distribución lleva el nombre de distribución generalizada del arco seno y para $p=q=1/2$, distribución del arco seno (Feller, 1978; Koroliuk, 1981). Si p y q aumentan, la variancia decrece y la distribución tiende a la normal estandarizada (Johnson y Kotz, 1970).

Algunas de las propiedades fundamentales de la función Beta (Guillén, 1991; Jambunathan, 1954; Johnson y Kotz, 1970) son las siguientes:

- 1) La integral $B(p,q)$ existe para $p, q > 0$ con valor finito.
- 2) Es simétrica en sus variables, $B(p,q) = B(q,p)$.
- 3) $B(p,q+1) \cdot (p+q) = q \cdot B(p,q)$, $B(p+1,q) \cdot (p+q) = p \cdot B(p,q)$.
- 4) $B(p,q) = \int_0^{\infty} z^{p-1} dz / (1+z)^{p+q}$.
- 5) $B(p,q) = \Gamma(p) \cdot \Gamma(q) / \Gamma(p+q)$.

6) $B(p, 1-p) = \pi / \sin(p\pi)$, en particular $\Gamma(p) \cdot \Gamma(1-p) = \pi / \sin(p\pi)$.

7) $B(p/2, q/2) = 2 \int_0^{\pi/2} \sin^{p-1} t \cdot \cos^{q-1} t \cdot dt$, en particular para $q=1$ es

$$\int_0^{\pi/2} \sin^{p-1} t \cdot dt = \sqrt{\pi} / 2 \cdot \Gamma(p/2) / \Gamma[(p+1)/2].$$

8) La media de la distribución Beta es

$$E(X) = \frac{p}{p+q} \quad (1.20)$$

y la variancia

$$\text{Var}(X) = \frac{pq}{(p+q)^2 (p+q+1)} \quad (1.21)$$

cuya demostración se expone en el siguiente apartado.

Para un desarrollo riguroso de todas esas propiedades, tanto de la función Gamma como de la Beta, sería necesario un nivel de matemáticas superior al que estamos utilizando. Nuestro objetivo ha sido tan sólo describir las principales ideas implicadas. No obstante, todas las demostraciones pueden encontrarse en la mayoría de textos de cálculo integral avanzado.

1.3.2. Relación entre las funciones Gamma y Beta

La conexión existente entre la función Beta y Gamma queda reflejada, principalmente, en la propiedad número 5 de la distribución Beta (Barr y Zehna, 1983; Johnson y Kotz, 1970; Papoulis, 1980):

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \quad (1.22)$$

para toda $p > 0$ y $q > 0$.

La demostración de esta propiedad se encuentra desarrollada en Cramer (1946), Saks y Zygmund (1952) y Tucker (1966).

En orden a determinar la media y la variancia, como características más importantes de la función Beta, se procede de la siguiente forma.

El momento de orden k con respecto al origen viene dado por

$$\begin{aligned} m_k = E(X^k) &= \int_0^1 x^k \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1} dx = \\ &= \frac{1}{B(p, q)} \int_0^1 x^{k+p-1} (1-x)^{q-1} dx = \\ &= \frac{1}{B(p, q)} B(k+p, q) \end{aligned}$$

Si se particulariza para $k=1$, se obtiene la expresión de la media o esperanza matemática

$$E(X) = \frac{1}{B(p, q)} B(p+1, q)$$

Según el teorema expresado en la ecuación 1.22:

$$B(p+1, q) = \frac{\Gamma(p+1)\Gamma(q)}{\Gamma(p+q+1)}$$

y de acuerdo con la propiedad número 5 de la función Gamma:

$$\Gamma(p+1) = p \Gamma(p), \quad \Gamma(p+q+1) = (p+q)\Gamma(p+q)$$

con lo que se infiere:

$$E(X) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \cdot \frac{p \Gamma(p)\Gamma(q)}{(p+q)\Gamma(p+q)} = \frac{p}{p+q}$$

Por otra parte, la variancia cumple la siguiente relación:

$$\text{Var}(X) = E(X^2) - E^2(X)$$

El momento respecto al origen de segundo orden, es decir para $k=2$, es:

$$m_2 = \frac{1}{B(p, q)} B(p+2, q)$$

y, como queda reflejado en la ecuación 1.22,

$$B(p+2, q) = \frac{\Gamma(p+2)\Gamma(q)}{\Gamma(p+q+2)}$$

Ahora bien,

$$\Gamma(p+2) = (p+1)\Gamma(p+1) = (p+1)p\Gamma(p)$$

y

$$\Gamma(p+q+2) = (p+q+1)(p+q)\Gamma(p+q)$$

con lo que

$$\begin{aligned} m_2 &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \cdot \frac{p(p+1)\Gamma(p)\Gamma(q)}{(p+q)(p+q+1)\Gamma(p+q)} = \\ &= \frac{p(p+1)}{(p+q)(p+q+1)} \end{aligned}$$

En consecuencia,

$$\begin{aligned} \text{Var}(X) &= \frac{p(p+1)}{(p+q)(p+q+1)} - \left(\frac{p}{p+q}\right)^2 = \\ &= \frac{p(p+1)(p+q)}{(p+q)^2(p+q+1)} - \frac{p^2(p+q+1)}{(p+q)^2(p+q+1)} = \\ &= \frac{pq}{(p+q)^2(p+q+1)} \end{aligned}$$

Generalizando, podrían calcularse los momentos directamente a partir de la fórmula propuesta por Barr y Zehna (1983) y Johnson y Kotz (1969):

$$m_k = \frac{\Gamma(p+q)\Gamma(p+k)}{\Gamma(p)\Gamma(p+q+k)} = \frac{p(p+1)\dots(p+k-1)}{(p+q)(p+q+1)\dots(p+q+k-1)} \quad (1.23)$$

En este último apartado se ha puesto de manifiesto la importancia que tiene caracterizar una distribución mediante los llamados momentos. Por ello, es conveniente tratar este concepto de una forma más amplia.

1.4. Momentos de una función de distribución

El objetivo de este apartado es introducir un breve análisis sobre los llamados momentos estadísticos. Por momentos se entiende el número más pequeño posible de valores característicos de una distribución dada. Así pues, se puede afirmar que los momentos contienen alguna información acerca de la distribución de la que proceden. Cabe distinguir entre momentos respecto del origen y momentos respecto de la media o momentos centrales (Barr y Zehna, 1983; Beaumont, 1986; Cramer, 1946; Fisz, 1963; Gnedenko, 1976; Johnson y Kotz, 1969; Ostle 1965; Ríos, 1977; Sánchez, 1988b; Tucker 1966; Zoroa y Zoroa, 1991). Los primeros tienen el origen en la propia variable y los segundos, en la media aritmética de la variable. Los momentos respecto del origen también se denominan momentos ordinarios (Fisz, 1963; Zoroa y Zoroa, 1991).

Las definiciones formales de momento ordinario y momento central, respectivamente, son las siguientes:

El k -ésimo momento de una variable aleatoria X (denotado por m_k) es el valor esperado de X a la k -ésima potencia, con $k=1,2,3,\dots$; esto es, $m_k = E(X^k)$ (Larson, 1982; p. 129).

Si k es un entero positivo, se llama momento central de orden k de una variable aleatoria X al valor $\mu_k = E[X-E(X)]^k$, bajo el supuesto que las dos esperanzas implicadas en la definición existan. Al momento central de orden 2 de X , μ_2 , se le llama variancia de X , y se indica por $\text{Var}(X)$ o por $\text{Var } X$ (Tucker, 1966; p. 99).

Al igual que en la segunda definición, en la primera sería conveniente especificar el supuesto de que el valor esperado exista.

A través de estas definiciones se establece que, si X es una variable aleatoria continua con función de densidad $f(x)$, entonces, el momento de orden k (respecto del origen de X) viene expresado por la siguiente integral de Riemman:

$$m_k = E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx \quad (1.24)$$

y el momento central de orden k (respecto de la esperanza de X) viene dado por

$$\mu_k = E[X - E(X)]^k = \int_{-\infty}^{+\infty} [x - E(X)]^k f(x) dx \quad (1.25)$$

El momento de primer orden (respecto del origen), es decir, para $k=1$, corresponde a la esperanza matemática, designada por $E(X)$, y se obtiene, en el caso absolutamente continuo, a partir de la ecuación 1.24:

$$m_1 = E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.26)$$

La esperanza matemática es considerada como la media de la población de la que se ha sacado la muestra, y se acostumbra a simbolizar por la letra griega μ (por convención, se utilizan letras griegas para los parámetros de la población y letras romanas para los parámetros de la muestra). Por lo tanto, $E(X)$ es una generalización del concepto de media aritmética. Por otra parte, la esperanza matemática existe si y sólo si la integral de la ecuación 1.26 es finita o convergente (Cramer, 1946; Fisz, 1963; Kalbfleisch, 1979; Meyer, 1986; Rozanov, 1973; Thomas, 1986).

El momento central de segundo orden es la variancia y se obtiene sustituyendo, en la ecuación 1.25, el valor k por 2

$$\mu_2 = E[X-E(X)]^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx = \text{Var}(X) = \sigma^2 \quad (1.27)$$

La variancia, comúnmente representada por el símbolo σ^2 , es una medida de la dispersión de la población respecto de μ .

La expresión 1.27 puede generalizarse para cualquier variable aleatoria de la forma siguiente: si X es una variable aleatoria con esperanza finita $E(X)$, su variancia es la esperanza de la nueva variable $Y = [X-E(X)]^2$ (Cuadras, 1985; Johnson y Kotz, 1969).

El cálculo de la variancia se simplifica con la ayuda del siguiente teorema, demostrado en Amon (1982), Barr y Zehna (1983), Gnedenko (1976), Kalbfleisch (1979), Meyer (1986), Tucker (1966) y Turner (1974).

Teorema

$$\text{Var}(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$$

Demostración

$$\begin{aligned} E[(X-\mu)^2] &= \int_M (x-\mu)^2 f(x) dx \\ &= \int_M (x^2 - 2\mu x + \mu^2) \cdot f(x) dx \\ &= \int_M x^2 \cdot f(x) dx - 2\mu \int_M x \cdot f(x) dx + \mu^2 \int_M f(x) dx \end{aligned}$$

$$= E(X^2) - 2\mu \cdot \mu + \mu^2 \cdot 1$$

$$= E(X^2) - \mu^2$$

donde M designa el espacio de muestras usado para X .

En la demostración del teorema anterior, puede observarse cómo los momentos centrales se calculan fácilmente con ayuda de los momentos respecto del origen.

En lo referente al momento de orden 0, tanto si es respecto al origen como respecto a la media, existe siempre y es igual a la unidad. Por lo tanto, este valor no da ninguna información acerca de la distribución de la que procede.

Una forma más general es definir los momentos respecto a un punto arbitrario cualquiera (Cramer, 1946; Fisz, 1963; Majó, 1964; Ríos, 1977):

$$E[(X-c)^k] = \int_{-\infty}^{+\infty} (x-c)^k f(x) dx \quad (1.28)$$

Igualando c a cero se obtienen los momentos ordinarios; y haciendo c igual a la media, los centrados.

De igual modo que la media y la variancia miden la localización y dispersión de una distribución, respectivamente, momentos de orden superior miden otras propiedades. Por ejemplo, el tercer momento alrededor de la media determina si una distribución es simétrica o asimétrica y el cuarto, también con respecto a la media, sirve para estudiar el apuntamiento o curtosis.

Así como un sólo momento o algunos momentos proporcionan escasa información sobre la distribución correspondiente, el conjunto de todos los momentos (m_1, m_2, m_3, \dots) determina, en general, la distribución exacta (Chou, 1974). Es decir, si se pudieran conocer todos los momentos, que son infinitos, quedaría demostrado, bajo ciertas condiciones, que existe una sola distribución con tales momentos. Sin embargo, en estadística aplicada no suele importar demasiado el conocimiento de la forma efectiva de la distribución. Pero suele ser necesario conocer, al menos, la posición de la distribución y tener cierta idea de su dispersión.

Aunque existen otros tipos de momentos, tales como los absolutos y los factoriales (Barr y Zehna, 1983; Caridad, 1985; Fisz, 1963; Gnedenko, 1976; Johnson y Kotz, 1969; Ríos, 1977; Zoroa y Zoroa, 1991), no se abordarán por exceder los límites del presente trabajo. Tan sólo destacar, respecto a los factoriales, que tienen una aplicación especial para las variables aleatorias discretas (Larson, 1982).

En suma, el valor medio, la dispersión, la asimetría y el apuntamiento de una distribución vienen dados por los momentos de distintos órdenes, sean centrales o respecto al origen. De ahí se deriva la gran importancia que tienen los momentos dentro de la práctica estadística, en especial para ciertas distribuciones. Por lo tanto, interesa conocer una manera más simplificada de calcular estos momentos, y es a través de las llamadas funciones generatrices de momentos. Las más utilizadas en estadística, formuladas por Zoroa y Zoroa (1991), son las siguientes:

-Función generatriz de probabilidad: $f_x(t) = E(t^X)$

-Función generatriz de momentos: $M_x(t) = E(e^{tX})$

-Función característica: $\varphi_x(t) = E(e^{itX})$

-Función generatriz de momentos

respecto de la media μ : $M_{X-\mu_x}(t) = E[e^{t(X-\mu_x)}]$

-Función generatriz de momentos

factoriales: $h_x(t) = E[(1+t)^X]$

De todas estas funciones, solamente se tratará la función generatriz de momentos (f.g.m.), respecto al origen y a la media, y la función característica (f.c.), consideradas ambas como instrumentos potentes para estudiar diversos aspectos de las distribuciones de probabilidad. Únicamente se detallarán las bases necesarias para comprender las ideas principales, sin entrar en un nivel matemático muy elevado. Esto es, se acepta que ciertas operaciones son correctas.

1.5. Función generatriz de momentos

Las funciones generatrices de momentos respecto del origen son aquellas que generan los momentos de una variable aleatoria (Bendat y Pierson, 1986; Polo, 1985; Quirin, 1978; Stoyanov *et al.*, 1989). Una definición formal del concepto, para el caso continuo, sería la presentada por Meyer (1986):

Sea X una variable aleatoria continua con f.d.p. $f(x)$, la función M_x , llamada *función generadora de momentos de X* , está definida por

$$M_x(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx \quad (1.29)$$

(Meyer, 1986; p. 218).

La expresión anterior es equivalente a

$$M_x(t) = E(e^{tX}) \quad (1.30)$$

donde t representa un parámetro real y X es una variable aleatoria.

La razón por la cual $M_x(t)$ se denomina f.g.m. está muy bien demostrada en Barr y Zehna (1983), Larson (1982), Meyer (1986) y Ross (1980), de manera que las derivadas con respecto a t , centradas en el origen ($t=0$), generan los momentos de X , con lo que

$$M_x^{(k)}(0) = m_k \quad (1.31)$$

Por consiguiente, la primera derivada de la f.g.m. calculada en $t=0$ da el valor esperado de la variable aleatoria $E(X)$, la segunda derivada es igual a $E(X^2)$, y así sucesivamente (Barr y Zehna, 1983; Kalbfleisch, 1979). En efecto, suponiendo que $M_x^{(k)}(0)$ existe, se cumple la siguiente expresión:

$$M_x^{(k)}(0) = E(X^k) \quad (1.32)$$

Esto es, la k -ésima derivada de $M_x(t)$ calculada en $t=0$ da $E(X^k)$. Así, queda demostrado que si se conoce la función M_x , pueden generarse los momentos.

Considérese a continuación, con más detalle, cómo la f.g.m. permite simplificar el cálculo de los momentos ordinarios. Dado que se trata de una variable continua y su amplitud de variabilidad es todo el campo real, se puede sustituir la expresión e^{tX} , en la ecuación 1.29, por su desarrollo en serie potencial:

$$M_X(t) = \int_{-\infty}^{+\infty} \left(1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \frac{t^4 X^4}{4!} + \dots \right) f(x) dx \quad (1.33)$$

Desarrollando la integral en distintos sumandos y aplicando la definición de momento ordinario, se tiene que

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{+\infty} f(x) dx + t \int_{-\infty}^{+\infty} x f(x) dx + \frac{t^2}{2!} \int_{-\infty}^{+\infty} x^2 f(x) dx + \\ &+ \frac{t^3}{3!} \int_{-\infty}^{+\infty} x^3 f(x) dx + \frac{t^4}{4!} \int_{-\infty}^{+\infty} x^4 f(x) dx + \dots = \\ &= 1 + t m_1 + \frac{t^2}{2!} m_2 + \frac{t^3}{3!} m_3 + \frac{t^4}{4!} m_4 + \dots \end{aligned} \quad (1.34)$$

Derivando ahora sucesivamente esta última expresión con respecto a t , se obtiene

$$M_X'(t) = m_1 + t m_2 + \frac{t^2}{2} m_3 + \frac{t^3}{3!} m_4 + \dots$$

$$M_X(t)'' = m_2 + t m_3 + \frac{t^2}{2} m_4 + \dots$$

$$M_X(t)''' = m_3 + t m_4 + \dots$$

$$M_x(t)^{IV} = m_4 + \dots$$

.....

.....

Finalmente, si t se iguala a cero, queda

$$M_x'(0) = m_1$$

$$M_x''(0) = m_2$$

$$M_x'''(0) = m_3$$

$$M_x^{IV}(0) = m_4$$

.....

.....

y, generalizando, se deriva la expresión 1.31, presentada anteriormente.

De esta forma queda demostrado que los momentos respecto al origen pueden obtenerse a partir de la función $M_x(t)$, derivando ésta con respecto a t tantas veces como indique el orden del momento que se quiere hallar y haciendo posteriormente $t=0$. Es lógico preguntarse si no sería más sencillo calcular directamente los momentos de X en lugar de obtener primero la f.g.m. y luego diferenciarla. La respuesta a esta cuestión es que en muchos problemas este último método resulta más simple.

El inconveniente que se presenta es que la integral de la ecuación 1.29 puede no existir para algunas variables aleatorias o para algunos valores de t (Thomas, 1986). Una condición para que no exista la f.g.m. es que no exista $E(X)$ (Polo, 1985).

En el apartado anterior, se vio cómo se calculan los momentos centrales. Por lo tanto, también puede encontrarse la función generatriz de momentos centrales, definida por

$$M_{X-\mu_x}(t) = E[e^{t(X-\mu_x)}] \quad (1.35)$$

donde

$$E[e^{t(X-\mu_x)}] = E(e^{-t\mu_x}e^{tX}) = e^{-t\mu_x}E(e^{tX}) \quad (1.36)$$

de manera que, una vez se conoce la f.g.m. (ecuación 1.30), se multiplica la misma por $e^{-t\mu_x}$ y, así, se obtiene la función generadora de los momentos centrales (Larson, 1982):

$$M_{X-\mu_x}(t) = e^{-t\mu_x}M_X(t) \quad (1.37)$$

En resumen, la f.g.m. permite el cálculo de los momentos y, bajo ciertas condiciones, la sucesión de momentos determina la ley de distribución. De aquí la importancia fundamental que tiene dicha función.

1.6. Función característica

Para $t=0$ se cumple que

$$M_X(0) = \int_{-\infty}^{+\infty} f(x) dx = 1 \quad (1.38)$$

es decir, la integral es convergente. Pero esta convergencia no existe necesariamente para $t \neq 0$, por lo que se ha ideado la denominada función característica, que no es más que

una f.g.m. donde se ha cambiado t por it .

Las funciones características son transformadas complejas de funciones de distribución de gran aplicación en el cálculo de probabilidades. Sin embargo, nos limitaremos a dar una breve descripción de ellas. Para empezar, se presenta su definición:

La función característica φ de una variable aleatoria X es aquella función definida en R por

$$\varphi_X(t) = E(e^{itX}) \quad t \in R \quad (1.39)$$

(Clarke, 1975; p. 134).

Definiciones de este tipo aparecen en la mayoría de textos de teoría de probabilidades que dedican un amplio apartado a la f.c. (ver Bass, 1974; Bendat y Pierson, 1986; Clarke y Disney, 1985; Cramer, 1946; David, 1951; Fabian y Hannan, 1985; Fisz, 1963; Gnedenko, 1976; Gnedenko y Kolmogorov, 1968; Grimmett y Welsh, 1986; Lukacs, 1970; Moran, 1968; Papoulis, 1980; Rozanov, 1973; Stoyanov *et al.*, 1989; Thomas, 1986; Tucker, 1967, entre otros). Todas estas definiciones indican que la f.c. de X es la esperanza matemática de la función compleja e^{itX} , donde $i = \sqrt{-1}$ (la unidad imaginaria).

Aceptando que

$$e^{itx} = \cos tx + i \operatorname{sen} tx \quad t, x \in R \quad (1.40)$$

entonces, la f.c. queda mejor definida como

$$\varphi_X(t) = E(\cos tX) + iE(\operatorname{sen} tX) \quad (1.41)$$

Se trata, por lo tanto, de una función de variable real t que toma valores complejos y que, además, existe siempre (Johnson y Kotz, 1969), ya que

$$|e^{itx}| = |(\cos^2 tx + \operatorname{sen}^2 tx)^{1/2}| = 1 \quad (1.42)$$

Para el caso absolutamente continuo, si $f(x)$ es la función de densidad de X , se obtiene la siguiente expresión matemática:

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx \quad (1.43)$$

A partir de la f.c. podrán también obtenerse los momentos ordinarios, haciendo

$$m_k = \frac{\varphi_X^{(k)}(0)}{i^k} \quad (1.44)$$

Las propiedades básicas de la f.c., extraídas de Cuadras (1985), son:

- 1) Si existen dos variables con la misma f.c., también tendrán la misma distribución.
- 2) Si una variable X tiene todos sus momentos finitos, entonces $\varphi_X(t)$ es desarrollable en serie de Taylor:

$$\varphi_X(t) = 1 + it m_1 + \frac{(it)^2}{2!} m_2 + \dots + \frac{(it)^k}{k!} m_k + \dots \quad (1.45)$$

- 3) La f.c. de la suma de dos variables aleatorias estocásticamente independientes es igual al producto de la f.c. de cada una de ellas:

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) \quad (1.46)$$

- 4) La f.c. existe siempre, es continua y está acotada. Su acotación es:

$$|\varphi_X(t)| \leq \varphi_X(0) = 1 \quad (1.47)$$

Éstas y otras propiedades, los teoremas fundamentales, las condiciones necesarias que debe cumplir toda f.c. y los problemas que presenta, pueden encontrarse con mayor profundidad en Galambos (1988), Lukacs (1970), Lukacs y Laha (1964) y Zigmund (1947).

La primera propiedad permite resolver muchos problemas que requieren obtener la distribución de una determinada variable aleatoria. Para ello, en primer lugar se halla la f.c. de dicha variable, cálculo relativamente sencillo. Si coincide con la f.c. de alguna distribución ya conocida, esto querrá decir que la distribución requerida es idéntica a ésta. En efecto, existe una correspondencia biunívoca entre f.c. y distribuciones.

Se ha demostrado que, tanto la f.g.m., como la f.c. permiten calcular los momentos, si existen, de las distribuciones de forma directa. A pesar de que sea más ventajoso el uso de la f.c. en lugar de la f.g.m., puesto que existe siempre, aparece el problema de que los cálculos que se realizan en la f.c. son con números complejos. Por otra parte, hay que destacar que la f.c. determina los momentos respecto al origen, pero ello no es ningún inconveniente ya que es fácil pasar a los centrales. Por ejemplo, la variancia puede calcularse de la siguiente forma:

$$\text{Var}(X) = m_2 - m_1^2 = \frac{\varphi''(0)}{i^2} - \left[\frac{\varphi'(0)}{i} \right]^2 \quad (1.48)$$

1.7. A modo de resumen

En este primer capítulo se ha llevado a cabo una introducción a la teoría de probabilidades para poder estudiar cualquier tipo de distribución. Se ha presentado una serie de definiciones formales de los conceptos básicos, tales como variable aleatoria (discreta y continua), momentos, función generatriz de momentos y función característica. También se han desarrollado los términos de función de probabilidad y función de distribución, que definen una variable aleatoria, con sus respectivas propiedades. Además, se ha realizando un estudio detallado de las distribuciones Gamma y Beta, por estar indirectamente relacionadas con la distribución C, como se verá en el próximo capítulo.

2. ESTADÍSTICO C: DESCRIPCIÓN

En la introducción, se hizo hincapié en la dificultad de aplicar análisis gráficos y pruebas estadísticas clásicas, así como ASTI a los datos de experimentos de replicación intrasujeto y, por este motivo, aparecieron los procedimientos no paramétricos como análisis alternativos. En las páginas siguientes, se describe una prueba no paramétrica, basada en la media cuadrada de las diferencias sucesivas (MCDS). Se trata del estadístico C de Young (1941) recogido, posteriormente, por Caracciolo, Larcan y Camma (1985), Stanley (1984) y Tryon (1982). De este modo, se promueve el uso de otra prueba de significación estadística para datos secuenciales de una unidad observacional.

En primer lugar, se describirá brevemente la MCDS. A continuación, se expondrá una descripción detallada del estadístico C, incluyendo sus características, momentos y distribución. Llegados a este punto, se detallarán las tablas de la Función $I_x(p, q)$, elaboradas por Pearson (1924), que permiten hallar la significación del estadístico C. Por último, se presenta la relación entre el estadístico C y la prueba de Durbin-Watson.

2.1. Media cuadrada de las diferencias sucesivas

Uno de los principales problemas del análisis de series temporales es decidir si una determinada serie temporal corresponde a un proceso de variables aleatorias independientes o a un proceso con observaciones sucesivas correlacionadas.

Es común que un proceso esté generado por una ecuación de diferencia de primer orden. En tal caso, es apropiado utilizar la fórmula de correlación serial de orden uno, para probar la hipótesis nula de independencia contra la alternativa de que hay una dependencia entre observaciones separadas por una unidad de tiempo.

Queda claro, pues, que una correlación serial es una medida de dependencia serial en una secuencia de observaciones. Otro término alternativo para nombrar a la correlación serial es el de autocorrelación. Así, de ahora en adelante, se utilizarán indistintamente los términos correlación serial, dependencia serial y autocorrelación.

Siguiendo a Anderson (1971), existen varias definiciones de correlación serial dependiendo de que las medias sean conocidas o no. En el caso que sean conocidas, su tratamiento incluye el coeficiente de correlación serial circular y el coeficiente basado en la MCDS. En este apartado desarrollaremos la MCDS (simbolizada por δ^2). Se trata de una prueba de aleatoriedad que permite determinar si una serie de observaciones puede o no considerarse como aleatoria. En efecto, calcula la diferencia entre pares de observaciones consecutivas $(X_{i+1}-X_i)$, elevada al cuadrado con objeto de eliminar el signo $(X_{i+1}-X_i)^2$, y se promedia dividiendo por $n-1$, que son las diferencias sucesivas disponibles. Así, se tiene

$$\delta^2 = \frac{\sum_{i=1}^{n-1} (X_{i+1}-X_i)^2}{n-1} \quad (2.1)$$

Desde un punto de vista práctico, se conoce la importancia de inferir información sobre la población a partir de las observaciones. En efecto, el principal problema de la inferencia estadística es estimar los parámetros de la población (tales como la media o la variancia de la población) a partir de los correspondientes estadísticos muestrales (tales como la media y la variancia de la muestra). Así, a través de la serie empírica se obtienen inferencias de las características o propiedades de la población. Para ello, es preciso que los parámetros estimados se mantengan fijos. Para lograr esto, sería deseable que cada observación procediera del mismo conjunto de variables aleatorias, es decir, con idéntica media y desviación estándar. Sin embargo, hay situaciones en que la desviación estándar se mantiene constante a lo largo de toda la serie, pero la media varía de una observación a la siguiente. En tal caso, si no se corrige dicha variación, la desviación estándar estimada mediante la fórmula convencional

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \quad (2.2)$$

tenderá a ser mayor que la de la población (o proceso). En efecto, la variancia de una serie temporal

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (2.3)$$

aumenta en proporción directa al cambio o tendencia en el valor medio de la serie (von Neumann *et al.* 1941). Dicho de otro modo, es inversamente proporcional a la estacionariedad de la serie.

Un método simple, propuesto por von Neumann *et al.* (1941) para minimizar el efecto de la tendencia, consiste en estimar la variancia mediante el estadístico δ^2 . Este estadístico, al contrario que la expresión dada por la ecuación 2.3, es independiente de los cambios en el valor de la media de la serie temporal (Tryon, 1982), es decir, es independiente de la estacionariedad de la serie.

Consideremos, a modo de ejemplo, las siguientes medidas de una serie que presenta una tendencia: 1, 2, 3, 4, 5. Su media es 3, y, mediante la ecuación 2.3, el cálculo de su variancia resulta 2. Si se toman más registros de esta serie: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, entonces la media resultante es de 5.5 y la variancia de 8.25. Por lo tanto, ante la presencia de una tendencia, la media y la variancia aumentan al incrementar el número de observaciones. Por el contrario, las estimaciones de la variancia dadas por la ecuación 2.1 son igual a 1 para ambas series. Es decir, a través del estadístico δ^2 , se obtiene una variancia corregida de la serie.

Esta forma de estimación de la variancia fue puesta en práctica, por primera vez, por el *Ballistic Research Laboratory* en el *Aberdeen Proving Ground, U.S. Army* (von Neumann *et al.*, 1941). Los estudiosos de balística fueron los primeros en darse cuenta de la utilidad de las diferencias entre observaciones sucesivas, al enfrentarse con el problema de minimizar los efectos debidos a variaciones en el viento, calor y desgaste al medir la dispersión de la distancia recorrida por los proyectiles.

El estadístico δ^2 puede utilizarse para determinar la existencia de una tendencia. En tal caso, deberá conocerse si difiere significativamente de la ecuación 2.3 que calcula la variancia independientemente del orden de las observaciones y, por consiguiente, incluye el efecto de tendencia (von Neumann *et al.* 1941).

Respecto a la distribución de la MCDS, es difícil de hallarla, puesto que las diferencias sucesivas no son independientes. En efecto, la primera diferencia ($d_1 = X_2 - X_1$) y la segunda ($d_2 = X_3 - X_2$) están relacionadas, ya que ambas incluyen el registro X_2 . Es decir, existe una correlación similar entre cada par sucesivo de diferencias entre observaciones (von Neumann *et al.*, 1941). Por este motivo, tan sólo es posible hallar distribuciones aproximadas. Todo lo referente a ellas fue tratado, con detalle, por Anderson (1971), Hart (1942), Hart y von Neumann (1942), von Neumann (1941), von Neumann *et al.* (1941) y Williams (1941). Por otra parte, Young (1941) ha caracterizado la distribución δ^2 dentro del estadístico C.

2.2. Descripción del estadístico C

Por definición, una muestra aleatoria está formada por observaciones que se extraen independientemente de una población común. Si se tienen observaciones de una serie X_1, X_2, \dots, X_n , el estadístico C, dado por la expresión

$$C = 1 - \frac{\sum_{i=1}^{n-1} (X_i - X_{i+1})^2}{2 \sum_{i=1}^n (X_i - \bar{X})^2} \quad (2.4)$$

puede usarse para contrastar la aleatoriedad.

El numerador de la fracción anterior (fórmula 2.4) coincide con el de la ecuación 2.1, y el denominador es dos veces el sumatorio de las desviaciones cuadráticas.

El error estándar (e.e.) del estadístico C depende del número de observaciones de la serie temporal y se calcula fácilmente mediante la siguiente fórmula:

$$e.e. = \sqrt{\frac{n-2}{(n-1)(n+1)}} \quad (2.5)$$

Al dividir el estadístico C por su e.e. se obtiene el estadístico Z que, a partir de 25 observaciones, tiende a aproximarse a la distribución normal, con media cero y variancia uno (Young, 1941)

$$Z = \frac{C}{e.e.} \quad (2.6)$$

Para muestras entre ocho y 25 puntos de observación, Young (1941) calculó los valores críticos del estadístico C a un nivel de significación del 5% y del 1% (tabla 2.1). Con $\alpha=0.05$, el valor crítico gira en torno a 1.65 para cualquier tamaño de muestra, desde $n=8$ hasta $n=\infty$, y con $\alpha=0.01$, es necesario consultar la tabla. Puesto que se trata de una prueba de una cola, sólo se considerarán significativos valores altos de Z (superiores a la tabla 2.1). La regla de decisión es rechazar la hipótesis nula sólo si se obtienen valores grandes de Z y de C, es decir, valores de la cola superior de la distribución nula. En

consecuencia, rechazar la hipótesis nula lleva a concluir que existe algún tipo de tendencia. En general, las tendencias de gran amplitud están asociadas con valores positivos altos de Z y oscilaciones cortas con valores negativos altos (Stanley, 1984).

Tabla 2.1. Valores críticos del estadístico C con diferentes tamaños de muestra (n) y niveles de significación de 0.01 y 0.05. (Young, 1941.)

n	1%	5%
8	2.1664	1.6486
9	2.1826	1.6492
10	2.1958	1.6494
11	2.2068	1.6495
12	2.2161	1.6495
13	2.2241	1.6495
14	2.2310	1.6494
15	2.2369	1.6493
16	2.2423	1.6492
17	2.2470	1.6492
18	2.2513	1.6491
19	2.2550	1.6489
20	2.2585	1.6488
21	2.2616	1.6488
22	2.2647	1.6486
23	2.2676	1.6485
24	2.2700	1.6484
25	2.2717	1.6484
Normal ($n = \infty$)	2.3262	1.6447

A modo de conclusión, el estadístico C es simplemente una prueba de aleatoriedad, diseñada para determinar una forma específica de variabilidad, no aleatoria, en los datos. La cuestión lógica contestada por este estadístico es si la serie temporal contiene o no algún tipo de tendencia (Tryon, 1982), es decir, si existen variaciones sistemáticas que se apartan de la variación aleatoria.

2.2.1. Características

Las características básicas que se desprenden de la fórmula 2.4, descritas por Tryon (1982), son las siguientes:

- 1) El valor de C es cero cuando el sumatorio de las desviaciones cuadráticas es igual a $\frac{1}{2}$ de la suma de las diferencias consecutivas cuadráticas. Este resultado se da con mayor probabilidad cuando los datos no se apartan excesivamente de la media.

- 2) El sumatorio de las desviaciones cuadráticas refleja la presencia de todo tipo de tendencias en la serie. Por lo tanto, ante la presencia de una tendencia, aumentará más rápidamente que el sumatorio de las diferencias sucesivas cuadráticas. Así, el valor de la fracción de la fórmula 2.4 tiende a ser pequeña y, en consecuencia, el estadístico C será mayor. Este resultado ayuda al investigador a detectar la no estacionariedad de la serie.

- 3) La significación estadística de C se evalúa dividiéndolo por su e.e. (ecuación 2.6). Como puede observarse en la ecuación 2.5, el e.e. está en función del tamaño de la muestra. Esto significa que la potencia de la prueba se aproxima a infinito si el tamaño de la muestra también se acerca a infinito. Así, los efectos casi no significativos pueden detectarse siempre que se disponga de un número de datos suficiente.

2.2.2. Momentos de la distribución de C

El presente apartado describe el método seguido por Young (1941) para hallar los momentos de la distribución de C. Antes de pasar a derivar dichos momentos, debe señalarse que si se tiene una secuencia de n observaciones expresada por X_1, X_2, \dots, X_n , entonces la magnitud de C (ecuación 2.4) dependerá del orden en que se registran estas observaciones. Por lo tanto, dada una muestra, C tendrá $n!$ posibles valores; esto es, las permutaciones de n , es decir, todas las variaciones de la secuencia de medidas.

Con la finalidad de facilitar los cálculos, el coeficiente C puede ser expresado como

$$C = \frac{\sum_{i=1}^{n-1} [(X_i - \bar{X}) - (X_{i+1} - \bar{X})]^2}{2 \sum_{i=1}^n (X_i - \bar{X})^2} \quad (2.7)$$

De esta forma, el estadístico C se representa en términos de puntajes de diferencia:

$$C = 1 - \frac{\sum_{i=1}^{n-1} (x_i - x_{i+1})^2}{2 \sum_{i=1}^n x_i^2} \quad (2.8)$$

La ecuación anterior es equivalente a:

$$C = \frac{x_1^2 + x_n^2 + 2 \sum_{i=1}^{n-1} x_i x_{i+1}}{2 \sum_{i=1}^n x_i^2} \quad (2.9)$$

El desarrollo para obtener la fórmula 2.9 a partir de la 2.8 se recoge en el apéndice 2. Su importancia radica en que permite calcular los diferentes momentos a partir del coeficiente C.

Con objeto de hallar el valor medio de C en una muestra dada, es decir, el momento de primer orden con respecto al origen, deben tenerse en cuenta todos los ordenes posibles de las variables aleatorias. Así pues, se calcula el sumatorio sobre las $n!$ permutaciones del numerador de la ecuación 2.9

$$\sum_p \left(x_1^2 + x_n^2 + 2 \sum_{i=1}^{n-1} x_i x_{i+1} \right) = \sum_p x_1^2 + \sum_p x_n^2 + 2 \sum_p \sum_{i=1}^{n-1} x_i x_{i+1} \quad (2.10)$$

En total hay n valores de x_i y $n!$ ordenes distintos. Por ejemplo, cada valor x_i es x_1 en las $(n-1)!$ disposiciones. En consecuencia, se tiene que

$$\sum_p x_1^2 = (n-1)! x_1^2 + (n-1)! x_2^2 + \dots + (n-1)! x_n^2$$

$$\sum_p x_i^2 = (n-1)! \sum_{i=1}^n x_i^2 \quad (2.11)$$

Este mismo razonamiento se aplica para x_n

$$\sum_p x_n^2 = (n-1)! \sum_{i=1}^n x_i^2 \quad (2.12)$$

En cuanto al tercer término de la ecuación 2.10, tenemos que

$$\begin{aligned} 2 \sum_p \sum_{i=1}^{n-1} x_i x_{i+1} &= 2 \sum_p (x_1 x_2 + x_2 x_3 + \dots + x_{n-1} x_n) = \\ &= 2 \left(\sum_p x_1 x_2 + \sum_p x_2 x_3 + \dots + \sum_p x_{n-1} x_n \right) \quad (2.13) \end{aligned}$$

El desarrollo matemático del primer término de la expresión 2.13 es el siguiente:

$$\sum_p x_1 x_2 = (n-2)! x_1 x_2 + (n-2)! x_2 x_1 + \dots + (n-2)! x_1 x_n + (n-2)! x_n x_1 + \dots \quad (2.14)$$

cuya fórmula general es:

$$\sum_p x_1 x_2 = (n-2)! \sum_{i \neq j} x_i x_j \quad (2.15)$$

De igual modo, para el segundo y tercer término de la ecuación 2.13 se obtiene la misma expresión:

$$\sum_p x_2 x_3 = (n-2)! \sum_{i \neq j} x_i x_j \quad (2.16)$$

$$\sum_p x_{n-1} x_n = (n-2)! \sum_{i \neq j} x_i x_j \quad (2.17)$$

Sustituyendo los términos de la ecuación 2.10 por los resultados obtenidos, ésta queda transformada como sigue:

$$\begin{aligned}
 & (n-1)! \sum_{i=1}^n x_i^2 + (n-1)! \sum_{i=1}^n x_i^2 + 2(n-1)(n-2)! \sum_{i \neq j} x_i x_j = \\
 & = (n-1)! \sum_{i=1}^n x_i^2 + (n-1)! \sum_{i=1}^n x_i^2 + 2(n-1)! \sum_{i \neq j} x_i x_j = \\
 & = 2(n-1)! \sum_{i=1}^n x_i^2 + 2(n-1)! \sum_{i \neq j} x_i x_j \quad (2.18)
 \end{aligned}$$

Se conoce que

$$\left(\sum_{i=1}^n x_i \right)^2 = \sum_{i=1}^n x_i^2 + \sum_{i \neq j} x_i x_j \quad (2.19)$$

y como las x_i , al ser variables centradas, tienen la particularidad de que su sumatorio es cero

$$\sum_{i=1}^n x_i = 0 \quad (2.20)$$

entonces,

$$\sum_{i=1}^n x_i^2 + \sum_{i \neq j} x_i x_j = 0 \quad (2.21)$$

$$\sum_{i \neq j} x_i x_j = -\sum_{i=1}^n x_i^2 \quad (2.22)$$

Sustituyendo la ecuación 2.22 en la 2.18, el resultado que se obtiene es cero:

$$2(n-1)! \sum_{i=1}^n x_i^2 - 2(n-1)! \sum_{i=1}^n x_i^2 = 0 \quad (2.23)$$

con lo que se concluye que la esperanza matemática es cero

$$E(X) = 0 \quad (2.24)$$

En el capítulo anterior, se estableció la distinción entre momentos respecto al origen y momentos centrales. Sin embargo, como puede observarse en este caso concreto, ambos momentos coinciden, puesto que la media es cero. Esta coincidencia puede verse con más claridad al observar la ecuación 1.28 del apartado dedicado a los momentos de una función de distribución, ya que en el primer término de la igualdad, c sería igual a cero tanto para los momentos ordinarios como para los centrados.

Elevando al cuadrado el numerador de la expresión 2.9, se obtiene el momento de segundo orden:

$$\begin{aligned} \sum_p \left[x_1^2 + x_n^2 + 2 \sum_{i=1}^{n-1} x_i x_{i+1} \right]^2 &= \\ = \sum_p \left[x_1^4 + x_n^4 + 2x_1^2 x_n^2 + 4x_1^2 \sum_{i=1}^{n-1} x_i x_{i+1} + 4x_n^2 \sum_{i=1}^{n-1} x_i x_{i+1} + 4 \left(\sum_{i=1}^{n-1} x_i x_{i+1} \right)^2 \right] & \quad (2.25) \end{aligned}$$

Realizando el sumatorio \sum_p término a término, se tiene que

$$\frac{\sum_p \left[x_1^2 + x_n^2 + 2 \sum_{i=1}^{n-1} x_i x_{i+1} \right]^2}{n!} = \frac{2(2n-3) \left(\sum_{i=1}^n x_i^2 \right)^2 - 2n \sum_{i=1}^n x_i^4}{n(n-1)} \quad (2.26)$$

Por consiguiente, el segundo momento de C viene dado por

$$M_2 = \frac{2n-3 - \frac{m_4}{m_2^2}}{2n(n-1)} \quad (2.27)$$

donde m_2 y m_4 son los momentos segundo y cuarto respecto a la media.

De forma similar, los momentos tercero y cuarto de la distribución de C se definen por

$$M_3 = \frac{-6 + 4(n-3) \frac{m_3^2}{m_2^3} + 9 \frac{m_4}{m_2^2} - 3 \frac{m_6}{m_2^3}}{4n(n-1)(n-2)} \quad (2.28)$$

$$M_4 = \frac{1}{8n^3(n-1)(n-2)(n-3)} [24n^2(n-3)^2 - 48n(4n-9) \frac{m_3^2}{m_2^3}$$

$$- 24n(3n^2 - 17n + 27) \frac{m_4}{m_2^2} + (8n^3 - 45n^2 - 23n + 210) \frac{m_4^2}{m_2^4}$$

$$+ 16(2n^2 + 5n - 21) \frac{m_5 m_3}{m_2^4} + 4(17n^2 - 37n + 42) \frac{m_6}{m_2^3}$$

$$- (7n^2 + 13n - 6) \frac{m_8}{m_2^4}] \quad (2.29)$$

Estos cuatro primeros momentos de la distribución C para muestras extraídas de una población distribuida normalmente pueden calcularse, siguiendo a Fisher (1929, 1930),

sustituyendo en las fórmulas anteriores $\frac{m_3^2}{m_2^3}$, $\frac{m_4}{m_2^2}$, etc. por

$$\frac{m_3^2}{m_2^3} = \frac{6(n-2)}{(n+1)(n+3)} \quad (2.30)$$

$$\frac{m_4}{m_2^2} = \frac{3(n-1)}{(n+1)} \quad (2.31)$$

$$\frac{m_4^2}{m_2^4} = \frac{3(3n^3+23n^2-63n+45)}{(n+1)(n+3)(n+5)} \quad (2.32)$$

$$\frac{m_5 m_3}{m_2^4} = \frac{60(n-1)(n-2)}{(n+1)(n+3)(n+5)} \quad (2.33)$$

$$\frac{m_6}{m_2^3} = \frac{15(n-1)^2}{(n+1)(n+3)} \quad (2.34)$$

$$\frac{m_8}{m_2^4} = \frac{105(n-1)^3}{(n+1)(n+3)(n+5)} \quad (2.35)$$

Llevando a cabo las correspondientes sustituciones se obtiene que para el segundo momento

$$M_2 = \frac{2n-3 - \frac{3(n-1)}{(n+1)}}{2n(n-1)} =$$

$$\begin{aligned}
 &= \frac{\frac{2n(n+1) - 3(n+1) - 3(n-1)}{(n+1)}}{2n(n-1)} = \\
 &= \frac{2n(n+1) - 3(n+1) - 3(n-1)}{2n(n-1)(n+1)} = \\
 &= \frac{2n^2 + 2n - 3n - 3 - 3n + 3}{2n(n-1)(n+1)} = \\
 &= \frac{2n^2 - 4n}{2n(n-1)(n+1)} = \\
 &= \frac{2n(n-2)}{2n(n-1)(n+1)} = \\
 &= \frac{n-2}{(n-1)(n+1)} \quad (2.36)
 \end{aligned}$$

De igual modo, realizando las sustituciones pertinentes en los momentos tercero y cuarto, se obtiene que

$$M_3 = 0 \quad (2.37)$$

$$M_4 = \frac{3(n^2 + 2n - 12)}{(n-1)(n+1)(n+3)(n+5)} \quad (2.38)$$

Cabe destacar los resultados similares obtenidos por Williams (1941), para muestras sacadas de un universo normal, utilizando otro método.

En este apartado, se han presentado los principales momentos de la distribución que estamos tratando. Sin embargo, el lector habrá podido comprobar que su cálculo no es tan sencillo como podría desprenderse intuitivamente del capítulo anterior. Ello es debido a que nos hemos limitado a seguir, clarificando algunos puntos, la exposición dada por Young en 1941. En su artículo, vemos que se obtienen los momentos a partir de un método propuesto por Fisher en unos trabajos publicados en los años 1929 y 1930. Como es obvio, debido a su antigüedad, sería necesaria una revisión en profundidad para reconstruir el desarrollo de todo el proceso, lo cual desbordaría este capítulo. No obstante, es interesante establecer el concepto de momento de una distribución (reflejado en el capítulo anterior) y, aunque sea en términos generales, el proceso seguido para hallar los momentos de la distribución de C.

2.2.3. Distribución de C

Las funciones de los momentos centrales de $f(x)$ pueden expresarse en términos de β_1 y β_2 , siendo $\beta_1 = M_3^2/M_2^3$ y $\beta_2 = M_4/M_2^2$ (Johnson, Nixon y Amos, 1963). Estos parámetros determinan las curvas de frecuencias $y=f(x)$ del sistema de Karl Pearson (ver apéndice 3).

A partir de los momentos de la distribución C, calculados anteriormente, se tiene que

$$\beta_1 = 0 \quad (2.39)$$

y

$$\beta_2 = \frac{3(n^2+2n-12)(n-1)(n+1)}{(n-2)^2(n+3)(n+5)} \quad (2.40)$$

y como que, para muestras normales, la distribución C está limitada en las dos direcciones $(-a \leq C \leq +a)$ y es simétrica, la distribución Tipo II de Pearson puede ser considerada como representativa (Johnson *et al.*, 1963):

$$y = y_0 \left(1 - \frac{C^2}{a^2}\right)^m \quad (2.41)$$

donde

$$m = \frac{n^4 - n^3 - 13n^2 + 37n - 60}{2(n^3 - 13n + 24)} \quad (2.42)$$

$$a^2 = \frac{(n^2 + 2n - 12)(n - 2)}{(n^3 - 13n + 24)} \quad (2.43)$$

$$y_0 = \frac{\Gamma(2m+2)}{a \cdot 2^{2m+1} [\Gamma(m+1)]^2} \quad (2.44)$$

Los valores del parámetro β_2 para la distribución C se aproximan a los de la normal al aumentar el tamaño de la muestra (tabla 2.2).

Tabla 2.2. Valores β_2 para distintos tamaños de n .
(Young, 1941.)

n	β_2
5	2.300
10	2.570
15	2.684
20	2.750
25	2.793
50	2.833

Obsérvese, en la tabla anterior, que β_2 tiende a 3 al aumentar n , que es el valor que toma β_2 para la normal (véase el gráfico del apéndice 3).

La aproximación a la curva Tipo II no es fiable para muestras con menos de ocho observaciones, puesto que al disminuir el tamaño de la muestra, los extremos de la distribución C se desvían cada vez más de $\pm a$ y, por lo tanto, se trata de una distribución platicúrtica.

Con la finalidad de facilitar el uso de la distribución aproximada para muestras de $n \geq 8$, Young (1941) establece los valores de C asociados a dos niveles de probabilidad (tabla 2.3). La *ratio* de cada uno de estos valores para su respectivo e.e. aparece en la tabla 2.1.

Tabla 2.3. Valores de C para probabilidades de 0.02 y 0.10.
(Young, 1941.)

<i>n</i>	P=0.02	P=0.10
8	0.6686	0.5088
9	0.6456	0.4878
10	0.6242	0.4689
11	0.6044	0.4517
12	0.5860	0.4362
13	0.5691	0.4221
14	0.5534	0.4092
15	0.5389	0.3973
16	0.5254	0.3864
17	0.5128	0.3764
18	0.5011	0.3670
19	0.4900	0.3583
20	0.4797	0.3502
21	0.4700	0.3426
22	0.4609	0.3355
23	0.4521	0.3288
24	0.4440	0.3224
25	0.4361	0.3165

Los niveles de significación registrados en la tabla anterior excluyen el 2% y el 10% del área bajo la curva. En la práctica, corresponden al 1% y al 5%, respectivamente. Esta tabla ha surgido de las tablas de la Función $I_x(p,q)$ (Pearson, 1924), que serán descritas en el siguiente apartado, para $p=m+1$ y $q=0.5$, con la transformación

$$x = 1 - \frac{C^2}{a^2} \quad (2.45)$$

Los valores de $p=m+1$ y $q=0.5$ se obtienen al equiparar la ecuación 2.41 con la Función Beta Incompleta (ver apéndice 4).

2.3. Tablas de la Función $I_x(p,q)$

A modo de breve referencia histórica puede señalarse que, en primer lugar, se publicaron las tablas de la Función Gamma Incompleta (Pearson, 1922). Su principal dificultad fue el rango infinito de x . Posteriormente, aparecieron las tablas de la Función Beta Incompleta (Pearson, 1924) en las que, a diferencia de las anteriores, el rango de x es finito. La función calculada, en estas últimas tablas, es una integral de probabilidad. Esto es, si $B(p, q)$ representa la Función Beta Completa,

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx \quad (2.46)$$

y $B_x(p, q)$ representa la Función Beta Incompleta:

$$B_x(p, q) = \int_0^x t^{p-1} (1-t)^{q-1} dt \quad (2.47)$$

$$0 \leq x \leq 1; \quad p > 0; \quad q > 0,$$

en las tablas aparece la *ratio*:

$$I_x(p, q) = B_x(p, q) / B(p, q) \quad (2.48)$$

Aunque en la literatura existen diversos métodos para el cálculo de la ecuación 2.48 (Aroian, 1941; Camp, 1924; Müller, 1930-31; Soper, 1921; Wishart, 1925a, 1925b, 1927), seguiremos el propuesto por Pearson (1924), puesto que es el punto de partida de las tablas de Young (1941). Así, ya que la Función Beta está relacionada con la Gamma

(ecuación 1.22), como se indicó en el capítulo anterior, la *ratio* calculada en las tablas (apéndice 5) queda expresada de la siguiente forma:

$$I_x(p, q) = B_x(p, q) / B(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1} (1-t)^{q-1} dt \quad (2.49)$$

En la parte superior de cada columna de las tablas de la Función $I_x(p, q)$ aparece el valor de la correspondiente Función Beta Completa (ver apéndice 5). En realidad podríamos decir que se trata de unas tablas indirectas de la Función Beta Incompleta. En efecto, conocidos los valores de $I_x(p, q)$ y $B(p, q)$, se obtiene el valor de $B_x(p, q)$ a partir de la ecuación 2.48.

En un principio, se pensó en proporcionar unas tablas conteniendo valores de p y q , desde 0 hasta 50, con intervalos de 0.1 para p y q menores de 10 y de 0.5 para p y q mayores de 10. Además, sería ideal que los intervalos para la x fueran de 0.005. La publicación de una tabla con tales características se extendería a unas 2000 páginas, por consiguiente tuvo que ser reducida. Finalmente, Pearson (1924) proporcionó unas tablas en donde sólo aparecen los valores de $p \geq q$, con incrementos de 0.5 para valores de p y q menores de 11 e incrementos de la unidad para valores mayores de 11, y con intervalos para la x de 0.01 (véase apéndice 5). De esta forma, las tablas se reducen casi a la mitad.

En el caso que sean necesarias unas tablas para $p < q$, se aplica la siguiente relación de repetición:

$$I_x(p, q) = 1 - I_{1-x}(q, p) = 1 - I_{1-x}(p', q') \quad (2.50)$$

donde $p' = q$ y $q' = p$ y, por lo tanto, $p' > q'$.

Las tablas de Pearson presentan dificultades debido, precisamente, a su reducción. No obstante, su publicación fue una primera tentativa y proporcionó una inestimable ayuda en una extensa variedad de estudios en los que pueden ser aplicadas, directamente o por transformación. La principal dificultad aparece cuando el usuario necesita cubrir aquellas regiones no contempladas. Esto se resuelve, parcialmente, mediante interpolación. Más recientemente, Cran, Martin y Thomas (1977) y Majumder y Bhattacharjee (1973) han construido algoritmos para el cálculo por ordenador de la integral Beta Incompleta y la inversa de la *ratio* de la Función Beta Incompleta.

Para los propósitos del presente estudio, tan sólo se requiere una $q=0.5$ y valores p enteros o medios oscilando entre 4.5 y 12.5. Como puede observarse en el apéndice 5, estos valores están presentes en las tablas con la única excepción de los valores de $p=11.5$, para una $n=22$, y $p=12.5$, para una $n=24$, que aparecen en una tabla adicional incorporada en la edición de Pearson y Johnson (1968). Por consiguiente, la interpolación estará relacionada, principalmente, con la x . En realidad, se trata de una interpolación inversa, es decir, hallar x , dada p , q e $I_x(p,q)$. Existen varios métodos para realizar una interpolación de este tipo en una tabla de triple entrada, como la que se trata. Pearson y Johnson (1968) sugirieron el propuesto por Johnson *et al.* (1963). Sin embargo, este método fue desarrollado con posterioridad al artículo de Young, por lo que deducimos que no se pudo utilizar para obtener los valores de C de la tabla 2.3.

Con objeto de mostrar el proceso de construcción de la tabla 2.3, vamos a ejemplificarlo para una $n=10$ y una q conocida de 0.5. Se calcula, en primer lugar, el valor de m (fórmula 2.42) y, a continuación, el valor de p . En segundo lugar, se calcula el valor de a^2 (fórmula 2.43). Los resultados obtenidos para este tamaño de muestra son los siguientes:

$$m = 4.4798658$$

$$p = 5.4798658 \approx 5.5$$

$$a^2 = 0.9664429$$

Ahora ya es posible entrar en las tablas de Pearson (apéndice 5), puesto que se conoce p , q y el valor del área bajo la curva de la Función $I_x(p,q)$, $P=0.02$ y $P=0.10$.

Para $I_x(p,q)=0.10$, se busca esta área en las tablas de Pearson con $p=5.5$ y $q=0.5$ (apéndice 5, p. 298) y, así, puede obtenerse el valor de x . Obsérvese que el área exacta no existe en las tablas y, por lo tanto, es necesario interpolar. Puesto que Young no especificó el método de interpolación utilizado, llevaremos a cabo una interpolación simple:

$I_x(p,q)$	x
0.1059013 - 0.0972389	----> 0.78 - 0.77
0.1059013 - 0.10	-----> 0.78 - x

de donde $x=0.7731874$.

Hallado el valor de x , se obtiene el de C sustituyendo x y a^2 en la fórmula 2.45 siendo $C=0.4682$. Este resultado es muy similar al obtenido por Young, $C=0.4689$ (tabla 2.3). Por lo tanto, aunque desconocemos el método empleado para la obtención de la tabla 2.3, se puede concluir que los pasos seguidos son los descritos, o bien otros muy similares.

2.4. Analogía entre la prueba de Durbin-Watson y el estadístico C

Casi todas las series de tiempo poseen tendencias significativas; por tanto, los datos de series temporales tienden a estar serialmente correlacionados. Una forma sencilla, pero muy común, de correlación serial es la correlación serial lineal de primer orden o, simplemente, correlación serial de primer orden (Theil, 1971). Como se ha indicado con anterioridad, la correlación de los datos tiene una gran relevancia para una correcta estimación de la variancia. Es decir, las variancias se desvían sustancialmente del valor que poseerían bajo el supuesto de independencia.

Dada la importancia de la dependencia serial, Durbin y Watson (1950, 1951, 1971) han propuesto un estadístico para estimar la autocorrelación de retardo 1 de los errores en muestras pequeñas, desde $n=15$ hasta $n=50$

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (2.51)$$

Las razones críticas del estadístico d no pueden ser tabuladas, ya que su distribución exacta depende de los valores particulares de la variable independiente X . Tan sólo se pueden establecer los límites superior e inferior, d_L y d_U , para los niveles de significación de d . Así, con estos límites se contrasta la hipótesis de autocorrelación nula frente a la hipótesis de autocorrelación positiva de primer orden.

Las principales inferencias que pueden obtenerse a partir del estadístico d se hallan en Arnau (en prensa), Chou (1974), Jobson (1991) y Kendall y Ord (1990), donde quedan resumidas de la siguiente forma:

- 1) Si $d < d_L$, se rechaza la hipótesis según la cual los errores no están correlacionados, deduciéndose que existe una correlación serial positiva significativa.
- 2) Si $d > d_U$, no se rechaza la hipótesis de nulidad.
- 3) Si $d_L \leq d \leq d_U$, la prueba es indecisa.

La razón por la que existe una zona de indeterminación se debe a que es prácticamente imposible tabular todas las distribuciones exactas de d , por lo que sólo se tabulan las distribuciones extremas, es decir, las de d_L y d_U . Para los casos en que d cae dentro de la región de indeterminación, Durbin (1970) ha desarrollado el estadístico d' que posee una distribución igual que d_U . Con este estadístico se puede aplicar una prueba de independencia serial comparando el valor observado de d' con los puntos de significación de d_U de la tabla de Durbin-Watson (Arnau, en prensa). Una mayor aproximación, aunque más compleja, es la basada en las curvas de Pearson (Alí, 1984).

El estadístico de Durbin-Watson fue propuesto, inicialmente, para tamaños muestrales pequeños. Sin embargo, los estudios de Ostrom (1978) y Theil (1971), corroborados por el de Gorsuch (1983), demuestran que es relativamente insensible cuando se tienen pocas observaciones (por ejemplo, $n=20$) y la autocorrelación es moderada.

Tras esta breve explicación de la prueba de Durbin-Watson, podemos ya establecer la relación existente entre ella y el estadístico C, puesto que las dos pruebas se basan en una razón von Neumann generalizada (Vallejo, 1986). Esta relación viene dada por Arnau (1993) quien, siguiendo a Kmenta (1971) y Simonton (1977), propone la aplicación de la regresión generalizada (basada en los mínimos cuadrados generalizados) a DSTI (Gorsuch, 1983; Hibbs, 1974; Johnston, 1972; Theil, 1971). Para estimar los valores de los

parámetros de la regresión que recogen los cambios de una serie temporal, tanto de nivel como de pendiente, presenta una serie de pasos a seguir. Uno de estos pasos es estimar la autocorrelación entre los residuales de cada observación con la precedente, a partir de la siguiente fórmula:

$$\rho \approx 1 - \frac{d}{2} \quad (2.52)$$

donde d es el estadístico de Durbin-Watson.

La relación entre los dos estadísticos, C de Young y d de Durbin-Watson, viene dada por la ecuación anterior. Sustituyendo d en la fórmula 2.52, se obtiene:

$$\rho = 1 - \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{2 \sum_{i=1}^n e_i^2} \quad (2.53)$$

Esta expresión es equivalente al estadístico C (fórmula 2.8), con la particularidad de que la ecuación 2.53 está formulada en términos de errores y el estadístico C en términos de puntajes de diferencia.

2.5. A modo de resumen

En este segundo capítulo, se ha presentado un procedimiento de análisis alternativo a los ASTI. Se trata del estadístico C de Young (1941). En primer lugar, se ha explicado con detalle la MCDS que sirve de base a la prueba C . También se han descrito las características y momentos, así como la distribución del estadístico C .

Otro punto importante que ha sido tratado es el funcionamiento de las tablas de Pearson (1924), necesario para hallar la significación estadística de C . Por último, se expone la analogía existente entre la prueba de Durbin-Watson y el estadístico C .

Con todo lo introducido en el presente capítulo, se pretende dar las bases estadísticas de una nueva alternativa de análisis para poder estudiar, en el siguiente capítulo, su aplicación a DSTI.

3. ESTADÍSTICO C: APLICACIÓN A DISEÑOS DE SERIES TEMPORALES INTERRUMPIDAS

Las técnicas de análisis aplicadas a series cronológicas o temporales y, más concretamente, a los DSTI presentan una gran complejidad, ya que es materialmente imposible dar una ley o un método que sirva para todos los casos. Como destacan Arnau (1986) y Kratochwill *et al.* (1974) la utilización de los análisis y pruebas estadísticas, para la evaluación de los impactos en diseños de $N=1$, depende de las características que presentan los datos que han de ser analizados.

El estadístico C de Young (1941), aplicado a los DSTI por Tryon (1982), es adecuado para evaluar cuantitativamente la presencia de cambios debidos a intervenciones en datos de series temporales, cuando se dispone de un número insuficiente de observaciones para poder identificar un modelo ARIMA tentativo.

En el presente capítulo, se analizan las estrategias propuestas por Tryon (1982) para la aplicación del estadístico C, su uso en DSTI, los campos de investigación en que se ha utilizado, así como sus limitaciones y críticas.

3.1. Estrategias de uso

Con el propósito de analizar los datos de DSTI mediante el estadístico C, Tryon (1982) propuso una serie de estrategias, para diseños A-B, que Blumberg (1984) resume de la siguiente manera:

ESTRATEGIA 1: aplicación del estadístico C para probar si hay o no tendencia dentro de una fase, normalmente la de línea base.

ESTRATEGIA 2: cuando, según la Estrategia 1, la línea base es "estable", entonces el estadístico C se usa para probar si hay una tendencia en los datos al considerar conjuntamente la fase de línea base y la de tratamiento. En el caso de que la prueba C sea estadísticamente significativa, se concluye que el tratamiento es efectivo.

ESTRATEGIA 3: cuando, según la Estrategia 1, la línea base presenta una tendencia estadísticamente significativa, entonces el estadístico C se utiliza para probar si esta tendencia se prolonga en la fase de intervención, o bien se produce un cambio. Por lo tanto, en esta estrategia, la hipótesis nula es que en la fase de tratamiento se observa la misma tendencia (iniciada en la fase de línea base) y la hipótesis alternativa que la tendencia cambia. En el caso de que se acepte la hipótesis alternativa, se concluye que el tratamiento es significativo.

En la Estrategia 2 aparece el término estable entre comillas, dado que el estadístico C sólo es capaz de detectar tendencias o autocorrelaciones, y el hecho de que no sea significativo no implica que exista una estabilidad. Sería más correcto cambiar el término por "estacionario". Una serie temporal estacionaria se define como "una secuencia de datos o valores que no presentan ningún cambio sistemático en la media (la serie no presenta tendencia alguna), ni en la variancia" (Arnau, en prensa).

Para probar la significación del estadístico C, como se indicó en su descripción, hay que transformarlo en el estadístico Z (fórmula 2.6) y luego comparar el valor resultante con la tabla 2.1.

En resumen, el analista, en primer lugar, debe ejecutar la Estrategia 1 y, según el resultado obtenido, pasará a la Estrategia 2 (cuando el valor de Z no sea estadísticamente significativo) o a la Estrategia 3 (cuando el valor de Z sea estadísticamente significativo).

Esquemáticamente, estas tres estrategias quedan representadas en la figura 3.1 (Bono y Arnau, 1993).

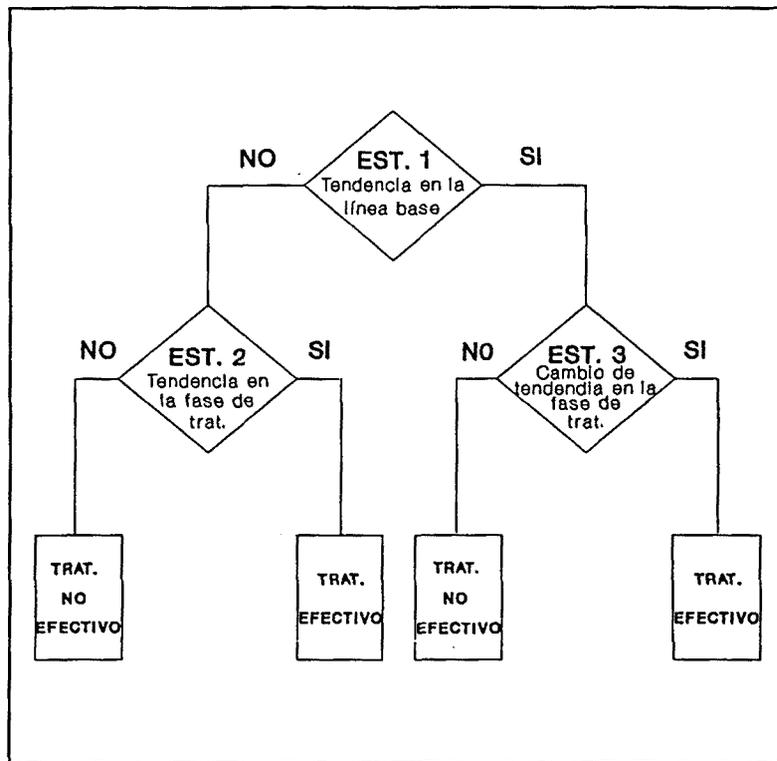


Figura 3.1. Diagrama de flujo de las estrategias de Tryon. (Bono y Arnau, 1993.)

Según Tryon (1982), la aplicación principal del estadístico C es evaluar los datos de la línea base. Estamos de acuerdo en que es preferible que éstos no contengan ningún tipo de tendencia estadísticamente significativa, puesto que se trata de una exigencia básica para la correcta interpretación de los diseños de $N=1$ (Polgan y Thomas, 1993). Pero, muchas veces, en las ciencias psicológicas y sociales, los períodos de pre-intervención son muy breves para que alcancen un estado de equilibrio. Siendo esto así, la fase de línea base puede presentar, de forma aparente, una tendencia. Por otra parte, hay muchos ejemplos de variables en las que hay una clara tendencia en la fase de pre-intervención que puede

ser lineal decreciente, lineal creciente, cuadrática y logarítmica. Algunos de estos ejemplos son, respectivamente: a) cantidad de calorías admitidas por personas con anorexia nerviosa (Agras *et al.*, 1974); b) tiempo y distancia recorrida a pie por un paciente agorafóbico (Agras, Leitenberg y Barlow, 1968); c) asistencia a la escuela (Barber y Kagey, 1977) y, d) puntos ganados por pacientes con depresión (Hersen *et al.*, 1973). En tales casos, para una interpretación válida de los resultados, el analista tiene que tener en cuenta el tipo específico de tendencia que presenta la fase A, y es preferible que ésta sea en sentido opuesto al esperado por la intervención. En el presente trabajo sólo se estudiarán las tendencias lineales

Cuando la línea base presenta una tendencia y, por lo tanto, se pasa a la Estrategia 3, Tryon (1982) sugiere el uso de dos aplicaciones menos potentes del estadístico C. Éstas se basan en la comparación de series que consiste en sustraer el valor de la primera observación de la línea base del valor de la primera observación de la fase de tratamiento, luego se resta el valor correspondiente a la segunda observación de la línea base de la segunda observación de la fase de tratamiento, y así sucesivamente hasta agotar todos los registros de la línea base y/o del tratamiento. Si la serie resultante presenta una C significativa, indica que ésta contiene una tendencia, es decir, que hay un efecto de tratamiento.

La más potente de las dos aplicaciones alternativas es la descrita por Hayes (1981), y consiste en ajustar, en primer lugar, una recta de regresión (o resistente) a los datos de la línea base para, luego, realizar una comparación de series, sustrayendo los valores obtenidos de la línea ajustada a la fase A de los de la fase B. Finalmente, la existencia de una tendencia, en los datos obtenidos de la comparación de series, se prueba a través del estadístico C. No obstante, al aplicar este procedimiento en series temporales cortas, puede ocurrir que valores atípicos afecten gravemente la inclinación de la serie. Por este motivo, Velleman y Hoaglin (1981) proponen ajustar una línea resistente que pase por las medianas de cada tercera parte de los datos.

En el supuesto de que la tendencia de la fase de pre-intervención no sea lineal, entonces quizá sea otro tipo de función (cuadrática, polinómica, trigonométrica, etc.) la que se ajuste a los datos (Daniel y Wood, 1971; Lewis, 1960). En este sentido, han surgido diversos procedimientos que tienen en cuenta explícitamente el tipo de tendencia de la línea base (Bryk, Strenio y Weisberg, 1980; Bryk y Weisberg, 1976; Strenio, Weisberg y Bryk, 1983). Sin embargo, en el próximo capítulo, donde se pondrá a prueba el estadístico C a través de simulación de Monte Carlo, tan sólo se tratarán las tendencias lineales.

La otra alternativa menos potente es muy sencilla. La comparación de series se realiza con las puntuaciones directas, y la serie resultante se evalúa, como en el método anterior, a través del estadístico C. Si éste es estadísticamente significativo, indica que la serie obtenida de la diferencia de las dos fases contiene una tendencia o desviación, lo que significa que el tratamiento es eficaz.

La comparación de series requiere que las dos fases tengan el mismo número de registros. No obstante, es frecuente que haya más observaciones en la fase de intervención que en la de línea base. En tal caso, una extrapolación de la tendencia de la primera fase permite añadir las observaciones necesarias para poder realizar la comparación de series. Además, de esta forma, aumenta la potencia de la prueba.

Estos dos métodos menos potentes de la prueba C tienen en común una limitación, y es que el estadístico C no es significativo cuando ambas fases tienen la misma pendiente y, además, hay un cambio abrupto entre ellas. Es decir, no puede detectar un cambio de nivel cuando la tendencia es igual en las dos fases (Tryon, 1982). En este sentido, el mismo Tryon (1984) propone como solución al respecto evaluar conjuntamente la fase de línea base y la de tratamiento. De este modo, el estadístico C es sensible tanto a cambios en la inclinación como en la intercepción. Sin embargo, no es capaz de detectar cuando la tendencia se prolonga en la fase de tratamiento y no hay ningún cambio de nivel.

Una aplicación del estadístico C, propuesta por Tryon (1982) y seguida por Barlow, Hayes y Nelson (1984), con la que estamos en desacuerdo, consiste en hallar la estabilidad de la respuesta, en el sentido de que la conducta no cambia significativamente o muestra pautas de variación mínimas (Cumming y Schoenfeld, 1960; Killeen, 1978). Según Tryon (1982), un criterio para determinar cuándo una conducta se estabiliza, es decir, cuándo una nueva fase del experimento puede empezar, sería ir recogiendo datos hasta obtener diez puntos consecutivos en que el estadístico C sea no significativo y, en el caso de que la respuesta no llegue a estabilizarse, quizás sería conveniente usar las aplicaciones menos potentes, descritas anteriormente, del estadístico C.

Tryon, en sus artículos de 1982 y 1984, comenta la importancia de la estabilidad de la línea base, ya que, de este modo, se tiene la certeza de que disponemos de una fase de pre-intervención adecuada, a partir de la cual se halla el efecto ejercido por la variable experimental. Dicho en otras palabras, todo cambio que se observe en una conducta estable, al instaurar un tratamiento, puede atribuirse a la intervención.

Es obvio que la línea base ideal debe ser estable si no queremos que oscurezca la interpretación de los datos. Si, por el contrario, es excesivamente variable, cualquier cambio producido en la fase de intervención resultará disfrazado. No obstante, para lograr la estabilidad se requieren, a menudo, un número considerable de observaciones y, aún así, en algunas situaciones la conducta no es estable en absoluto. Contrariamente a lo deseable, en el campo aplicado, las fases de línea base son relativamente cortas y, por lo tanto, es difícil alcanzar un criterio de estabilidad. Por ejemplo, Sharpley (1987) en un estudio de los datos publicados en la revista *JABA*, entre 1981-1985, mostró que de las 1623 series de línea base examinadas, 391 (24.1%) tenían menos de cinco observaciones; 681 (42.1%) tenían entre 5 y 10; 335 (20.6%), entre 10 y 20; 126 (7.7%), entre 20 y 30; 62 (3.8%), entre 30 y 40, y sólo 29 (1.7%) tenían más de 40 observaciones. Además, de las series que tenían cinco o más observaciones, la mayoría (70.3%) se clasificaron como variables (menos del 50% de las observaciones de cada serie en particular tenían el mismo

valor numérico). Sin embargo, pensamos que, al igual que Murray (1978) cuando se refiere a la inspección visual, para demostrar la efectividad de una variable independiente no es necesaria una línea base estable estrictamente definida, siempre y cuando el cambio sea suficiente para ser detectado.

En resumen, el estadístico C no es capaz de determinar la estabilidad de una serie y, en el supuesto de que lo fuera, las fases de pre-intervención son demasiado cortas. No obstante, a partir de la conclusión de Tryon (existencia de estabilidad cuando el estadístico C no es significativo), eminentes metodólogos conductuales, tales como Barlow *et al.* (1984), han recomendado su uso para calcular la estabilidad de la respuesta. Sin embargo, esta aplicación sería inadecuada, puesto que el hecho de que no se presente ningún componente de tendencia no significa que no exista variabilidad en la serie. En palabras de Crosbie (1989), la estabilidad de la respuesta se refiere no sólo a la ausencia de pendiente sino también a la ausencia de variabilidad extrema, característica de autocorrelación negativa.

Con el propósito de ilustrar las estrategias de Tryon y la simplicidad de los cálculos que requiere la prueba C, en el apéndice 6 se reproduce un ejemplo propuesto por este autor en su artículo de 1982, incorporando una serie de diagramas de flujo de los pasos a seguir con objeto de mejorar la comprensión del proceso completo.

3.2. Estadístico C y diseños de series temporales interrumpidas

Antes de establecer en qué DSTI o diseños de replicación intrasujeto la prueba C permite inferir la acción de un tratamiento, presentaremos, a grandes rasgos, las diversas modalidades del diseño experimental de sujeto único o, como destacan Robinson *et al.* (1989), grupo único.

Siguiendo a Arnau (1993) y Hayes (1981), los DSTI se agrupan en tres categorías: a) diseños intra-series (para la evaluación de los cambios en la variable dependiente, a través de las distintas fases), b) diseños entre-series (para la comparación, a lo largo del tiempo, de dos o más tratamientos de forma conjunta) y c) diseños combinados (para la comparación de efectos de tratamiento, tanto a través de las distintas fases como entre series simultáneas en el tiempo). Estos diseños aquí no serán estudiados con detalle, por lo que, para ello, remitimos a los lectores interesados a Arnau (1993).

De entre las categorías generales citadas anteriormente, la primera y la tercera son las más utilizadas. Dentro de los diseños intra-series, la más simple de las estrategias experimentales es el diseño A-B, donde los cambios en la variable dependiente se atribuyen a los efectos del tratamiento, pero con algunas reservas importantes (Arnau, 1986; Arnau, en prensa; Barlow y Hersen, 1973; Campbell, 1969; Campbell y Stanley, 1966; Cook y Campbell, 1979; Hersen, 1982; Kazdin, 1982; Kratochwill, 1978; Risley y Wolf, 1972). Es posible que los cambios en la fase B se hayan producido independientemente de la introducción del tratamiento, o como función de una correlación con alguna variable extraña de confundido. Esto constituye, sin duda, una de las principales limitaciones del diseño A-B que sólo es posible evitar a través de los conocidos diseños de retirada de tratamiento (DRT) -denominados, también, diseños de reversión- que se caracterizan por un regreso a las condiciones iniciales (es decir, se trata de la réplica sucesiva del componente básico A-B). Así, se obtiene un control más enérgico sobre las variables extrañas, en el caso de que se produzca un cambio en la variable de respuesta al introducir el tratamiento y un regreso al nivel de línea base cuando es retirado. No existe ningún criterio sobre el número de réplicas requeridas para obtener resultados aceptables. De este modo, puede haber diseños experimentales de sujeto único con dos réplicas, también llamados A-B-A-B o diseños de retirada de cuatro fases (DR4F), los de retirada de seis fases (DR6F), etc.

El uso de DRT es inadecuado cuando las retiradas o inversiones de tratamiento no son factibles debido a consideraciones éticas (gravedad del trastorno o perjuicios para el propio sujeto u otros) o a que los efectos de los tratamientos son irreversibles (Arnau, 1986; Arnau, 1993; Barlow y Hersen 1973; Birnbrauer, Peterson y Solnik, 1974; Hersen, 1982; Kazdin y Kopel, 1975; Poling y Grosset, 1986). Es evidente que, si se han obtenido resultados favorables en el primer período B, existan reservas en retroceder otra vez a la línea base con el único motivo de establecer relaciones causales. En estos casos se utilizan, cada vez con mayor frecuencia, los diseños de series combinadas, que implican combinaciones tanto intra-series como entre-series. Éstos se denominan diseños de línea base múltiple (DLBM).

En los DLBM se registra un cierto número de respuestas simultáneas para proporcionar dos o más líneas de base. El investigador instauro el mismo tratamiento de forma secuencial en el tiempo, de manera que afecte sólo a una serie de observaciones mientras utiliza las otras como control. En este tipo de diseños, los registros pueden tomarse de conductas diferentes e independientes entre sí de un mismo sujeto (o grupo de sujetos), de la misma conducta en sujetos distintos pero aparejados y que comparten las mismas condiciones ambientales, o bien de la misma conducta de un único sujeto o grupo a través de situaciones distintas e independientes. En esta estrategia de diseño, las variables de registro deben ser sensibles al mismo tratamiento. Es decir, sólo debe aparecer un cambio en la tasa de respuesta después de la intervención, mientras la tasa de conductas no tratadas permanece relativamente constante. En el caso que hubiera covariancia entre las conductas o situaciones, los efectos de control de la variable de tratamiento son cuestionables y no existe ninguna forma clara de interpretar los resultados. De acuerdo con este formato, el experimentador introduce, en primer lugar, el tratamiento para el primer comportamiento, individuo o contexto, y después, de manera secuencial, para los otros hasta que la variable experimental se aplica a todas las conductas, sujetos o situaciones objeto de estudio. En cada caso, normalmente, la variable de tratamiento no se aplica hasta que no se ha conseguido la estabilidad de la línea base. Para un mayor detalle de los

DLBM puede consultarse a Kazdin (1982) y Kazdin y Kopel (1975).

Hasta aquí, se han examinado tres de los diseños que pueden utilizarse para estudiar la conducta de los individuos aisladamente en su situación natural (concretamente los diseños A-B, DRT y DLBM). De hecho, existen diseños más complejos que pueden utilizarse en los estudios de $N=1$. Entre ellos pueden mencionarse los diseños de prueba múltiple (DPM), los diseños de tratamientos alternativos (DTA) y los diseños de tratamientos simultáneos (DTS). Nosotros no trataremos todos estos tipos de diseños, puesto que el número de casos en que se aplican es muy inferior al de los diseños tratados anteriormente. El lector puede consultar para ello a Arnau (1986, 1993) y Barlow y Hersen, (1984).

La aplicación del estadístico C a los datos de diseños A-B, DRT y DLBM es relativamente sencilla. En efecto, se trata de obtener la existencia o no de significación estadística de un cambio de tendencia al pasar de una fase a la siguiente. En cambio, la aplicación del estadístico C a otros diseños más complejos es más dificultosa. Por poner un ejemplo, en los DTA es difícil determinar, a través del estadístico C, qué tratamiento produce un mayor cambio en la conducta. Ello se debe a que se trata de una prueba que detecta los cambios de tendencia, pero que no es capaz, en el caso de que los tratamientos sean significativos, de discriminar el mejor.

En el diseño A-B, el estadístico C muestra la eficacia del tratamiento sólo cuando la tendencia de una conducta en particular cambia al recibir el tratamiento. No obstante, si la variable independiente no ejerce un grado apreciable de control conductual, se puede demostrar su eficacia utilizando un DRT que consiste, como ya se ha indicado, en retirar varias veces la variable de tratamiento en el transcurso de un experimento. De este modo, cuando los cambios de tendencia coinciden con la presencia y retirada del tratamiento, puede afirmarse con seguridad que es la variable independiente la provocadora de los mismos.

Para la aplicación del estadístico C en DLBM, se pueden conceptualizar las fases de línea base y las subsiguientes intervenciones como diseños A-B separados. Dado que los efectos de la variable de tratamiento se infieren a partir de las conductas no tratadas, el analista de los datos comprobará que la intervención es eficaz cuando hay un cambio de tendencia en la respuesta después de la aplicación del tratamiento. Es decir, sólo podrá inferirse la efectividad del tratamiento cuando el cambio se produce escalonadamente en función de la aplicación de las variables independientes a las distintas conductas, individuos o emplazamientos.

3.3. Ámbitos de utilización

El procedimiento descrito por Tryon (1982) para evaluar los efectos de tratamiento se ha utilizado tanto en psicología clínica (Dattilo y Rusch, 1985; Espie y Lindsay, 1987; Neufeld y Fantuzzo, 1987; Salkovskis, 1983; Snyder, 1987; Tarrrier y Barrowclough, 1987) como experimental (Luciano y Polaino-Lorente, 1986; Zettle y Hayes, 1987). Por otra parte, también aparecen en la literatura aplicaciones en el campo de la sociología (Ginsberg, 1987). A partir de la revisión de estos trabajos hemos construido un cuadro resumen en donde aparecen los autores y el año de publicación, la revista, las unidades de observación, la conducta, el tratamiento aplicado y el tipo de diseño (ver apéndice 7). En el capítulo quinto, se retomarán los trabajos de Ginsberg (1987) y Tarrrier y Barrowclough (1987) como ejemplos prácticos de la aplicación de la prueba C.

En términos generales, el estadístico C constituye una estrategia de análisis útil para la investigación social aplicada, tanto en el ámbito clínico, como en el conductual o educativo. Aun cuando ha sido utilizado en estos campos, puede también aplicarse a la investigación de evaluación de programas. Así pues, la prueba C permite el análisis de los efectos de los tratamientos, con pocas observaciones por fase, sobre sujetos, grupos o medidas de resultado de un programa. No obstante, no hay que olvidar que, en aquellos

casos en que la fases son suficientemente prolongadas, es preferible llevar a cabo un análisis de series temporales.

3.4. Limitaciones de aplicación

En el apartado 3.1 ya se mencionó el hecho de que el estadístico C no permite detectar la significación estadística cuando se efectúa una comparación de series en aquellos casos donde se presentan pendientes iguales en la línea base y en la fase de tratamiento y un cambio de nivel. Esta limitación se cumple siempre, aunque el cambio de nivel sea muy grande. Así, para estos casos, la prueba C no puede distinguir entre diferentes tamaños de la magnitud de la respuesta en el punto de intervención (Blumberg, 1984). Por ejemplo, en los gráficos que se muestran a continuación, si se aplica una comparación de series, el estadístico C no será significativo y, en consecuencia, es incapaz de determinar el tamaño del efecto de tratamiento, sea pequeño (figura 3.2) o grande (figura 3.3). Esto no ocurre cuando, junto con el cambio de nivel, aparece un cambio de tendencia en el paso de la fase de línea base a la de tratamiento (figura 3.4).

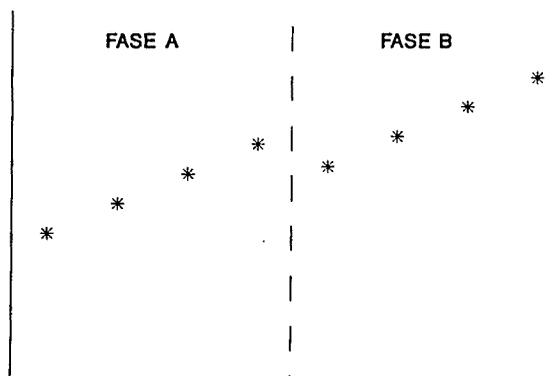


Figura 3.2. Cambio de nivel pequeño como consecuencia de la aplicación del tratamiento y sin cambio de inclinación.

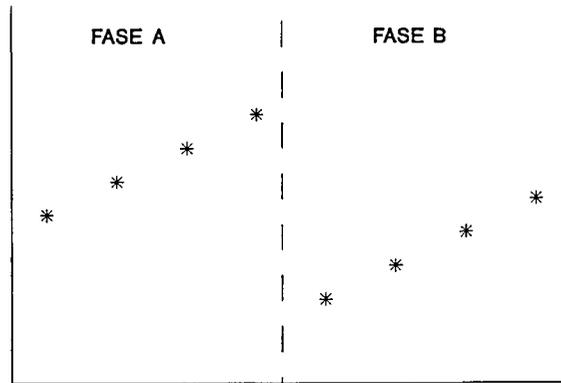


Figura 3.3. Cambio de nivel considerable como consecuencia del tratamiento y sin cambio de inclinación.

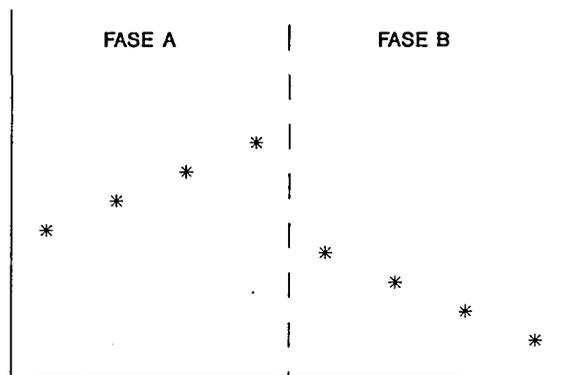


Figura 3.4. Cambio de nivel y tendencia como resultado del efecto del tratamiento.

Una posibilidad para solventar la dificultad planteada por la comparación de series es llevar a cabo un análisis del conjunto de las fases (línea base y tratamiento). Sin embargo, un análisis de este tipo resultará también significativo en el caso de que la tendencia sea la misma en las dos fases y no haya cambio de nivel.

Queda claro, pues, que la mayor diferencia entre las estrategias propuestas por Tryon y el método de análisis ARIMA es que éste puede detectar cambios abruptos en el

nivel de la serie, independientemente de que existan o no cambios en la inclinación. No obstante, la prueba C tiene la ventaja de que puede ser usada con series temporales cortas y no requiere cálculos complejos basados en la construcción del modelo.

Por otra parte, el estadístico C está más en función del número de datos registrados que del grado de inclinación que presenta la serie (Blumberg, 1984). Es decir, su potencia es directamente proporcional al número de observaciones. Por lo tanto, el análisis del conjunto de las fases de línea base y tratamiento es más potente que el análisis de la línea base sola, siempre y cuando la tendencia sea la misma. Si en la Estrategia 1 se concluye que no hay tendencia cuando en realidad sí la hay, debido a un número pequeño de observaciones, entonces, al pasar a la Estrategia 2 la tendencia será significativa únicamente porque ha aumentado el número de observaciones. De esta forma, se llegaría a dos conclusiones erróneas (línea base estacionaria y tratamiento significativo) que serán más probables cuanto menor sea el número de sesiones de observación en la línea base. Lo ideal sería poder estimar *a priori* cuántos puntos de observación en la línea base son necesarios para detectar una posible tendencia.

Por último, un problema importante asociado al estadístico C es que la decisión de usar la Estrategia 2 o la Estrategia 3 depende del resultado obtenido en la Estrategia 1 y, por lo tanto, como que la segunda prueba está condicionada por la primera, los errores Tipo I y Tipo II, a veces, se hallan excesivamente hinchados (Blumberg, 1984).

3.5. Principales críticas

En este apartado se analiza, detalladamente, el artículo de Crosbie publicado en 1989, puesto que hasta el momento constituye la única crítica de la prueba C fundamentada con análisis estadísticos. El contexto general del artículo sugiere que el estadístico C no está formulado para medir la tendencia sino la autocorrelación de retardo 1. Además,

proporciona una medida inadecuada de la estabilidad, ya que no puede detectar excesiva variabilidad en la tasa de respuesta.

Aunque el concepto de estabilidad ya ha sido tratado con anterioridad (ver punto 3.1), resulta procedente mencionarlo otra vez. Cuando un diseño conductual $N=1$ requiere la manipulación de una variable, una línea base estable permite atribuir cualquier desviación de la conducta, en la fase de tratamiento, a la variable independiente. No obstante, en el capítulo 4 se verá con más detalle, a través de simulaciones de Monte Carlo, cómo series extremadamente variables serían consideradas estables, según el uso inapropiado que hace Tryon del concepto de estabilidad.

En el apartado titulado "Características del estadístico C", Crosbie sostiene que Young (1941), en un ejemplo sobre control de calidad, nada infirió acerca de la tendencia en los datos sino que tan sólo concluyó que los datos eran serialmente dependientes. Esto es, que presentaban autocorrelaciones distintas de cero y, en consecuencia, el estadístico C calcula la dependencia serial y no la pendiente o tendencia. Sin embargo, al revisar el artículo de Young, observamos que éste examina una serie de medidas secuenciales con el propósito de hallar la presencia de variabilidad no aleatoria y que, explícitamente, no dice nada acerca de la tendencia de los datos, pero tampoco de la dependencia serial. Tan sólo concluyó que C es significativo y, en consecuencia, los datos están ordenados, o bien no son aleatorios, de lo cual se puede inferir que se refiere tanto a la dependencia serial como a la tendencia. Por lo tanto, existe una mala interpretación del trabajo de Young.

Ante la conclusión anterior, sostenida por Crosbie, estamos totalmente de acuerdo en que el estadístico C mide la autocorrelación. Pero no hay que olvidar que si mide la autocorrelación, también mide la tendencia, en el supuesto de que la haya. En efecto, el mismo Crosbie, después de concluir incorrectamente que la prueba C sólo mide la dependencia serial, afirma que datos con una tendencia frecuentemente tienen un estadístico C significativo (es decir, presentan autocorrelación), pero esto no implica que todos los

datos serialmente dependientes tengan una tendencia. Unas líneas más abajo, comenta que Bennett y Frankling (1954) afirmaron que pruebas basadas en la definición circular o no circular de R_1 (como la autocorrelación) son equivalentes a aquellas basadas en la MCDS (como la prueba C) y que, por lo tanto, C no calcula directamente la tendencia, lo que ésta mide es la dependencia serial que está frecuentemente relacionada con la tendencia, aunque no siempre. En definitiva, puede observarse cómo después de llegar a la conclusión de que C mide la autocorrelación y no la tendencia, inexplicablemente cambia su punto de vista.

Más adelante, Crosbie señala cómo las similitudes entre C y autocorrelación de retardo 1 quedan reflejadas algebraicamente, a través de sus respectivas fórmulas,

$$C = 1 - \frac{\sum_{i=1}^{n-1} (X_i - X_{i+1})^2}{2 \sum_{i=1}^n (X_i - \bar{X})^2} \quad (3.1)$$

$$r_1 = \frac{\sum_{i=1}^{n-1} (X_i - \bar{X})(X_{i+1} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (3.2)$$

A nuestro modo de ver, no existe tal similitud, puesto que el numerador del estadístico C representa la destendenciación (ecuación 3.1), y el numerador de la autocorrelación, la covariancia entre dos puntos de la serie separados por un retardo (ecuación 3.2). Por otra parte, la fórmula 3.2 es para muestras grandes, y para muestras pequeñas se la tendría que multiplicar por algún factor de corrección, como puede ser $n/n-1$ (Arnau, en prensa; Huitema y McKean, 1991).

Crosbie, en la sección siguiente, lleva a cabo un estudio de Monte Carlo, con el objetivo de hallar la relación existente entre tendencia, C, y autocorrelación. En él simula varias series temporales fijando las pendientes (0, 0.5 y 1), longitudes (10, 20, 30 y 40)

y niveles de autocorrelación de retardo 1 (0, 0.2, 0.4, 0.6 y 0.8). Concretamente, los resultados se obtienen a partir de la simulación de 100 series para cada combinación de las condiciones anteriores, y cada serie es generada de acuerdo con la siguiente fórmula:

$$Z_n = L + S_n + \phi Z_{n-1} + e_n \quad (3.3)$$

donde Z_n es el n -ésimo punto en la serie temporal, L es el nivel, S es la pendiente programada, ϕ es el parámetro auto-regresivo y e_n es el n -ésimo error aleatorio distribuido normalmente. Crosbie afirma que esta fórmula es la del proceso auto-regresivo de primer orden formulado por Gottman (1981). Pero una lectura del trabajo de Gottman revela la contradicción de que la expresión que representa el proceso auto-regresivo de primer orden no incluye ningún parámetro que sólo recoja la tendencia. Así, se tiene que

$$x_t = a_1 x_{t-1} + e_t \quad (3.4)$$

donde x_t es la observación definida en términos de desviación de la media, a_1 es el parámetro auto-regresivo de primer orden o el coeficiente de autocorrelación de retardo 1, x_{t-1} es la puntuación para el tiempo $t-1$ y e_t es el error aleatorio. En este mismo sentido, precisamente otro artículo del autor, Crosbie y Sharpley (1989), expresa algebraicamente, al igual que Gottman (1981), una serie auto-regresiva de primer orden sin especificar ningún parámetro de tendencia.

En conclusión, las series generadas por Crosbie, a partir de la ecuación 3.3, incorporan además de un parámetro auto-regresivo, una tendencia y, por lo tanto, no siguen el modelo de Gottman (1981).

En el apéndice 8 se incluyen las tablas resultantes de la simulación llevada a cabo por Crosbie (tablas 1, 2 y 3). Es a partir de ellas que Crosbie concluye que el estadístico C mide la dependencia serial y no la tendencia. Por ejemplo, en la tabla 1, se observa cómo, en una serie sin tendencia y una longitud de 20, C aumenta desde -0.01 a 0.63, al aumentar la autocorrelación. Por lo tanto, para los niveles más altos de autocorrelación,

el estadístico C es estadísticamente significativo aunque no haya tendencia. Esta evidencia está mal interpretada, puesto que autocorrelación y estadístico C son dos conceptos que están muy relacionados (véase sección 2.4). Nótese, pues, cómo una vez más los argumentos de Crosbie son inconsistentes.

Una particularidad con respecto a la tabla 1 (con pendiente programada de cero) es que no coincide el parámetro auto-regresivo de primer orden fijado por el autor y la autocorrelación obtenida. Esto es debido a la distinción existente entre parámetros, ϕ_1 , y estadísticos, r_1 (Huitema, 1988). El hecho de que r_1 sea un estimador preciso de ϕ_1 dependerá del número de observaciones en que se base (Arnau y Bono, 1994; Huitema, 1988; Huitema y McKean, 1991). Si el número de observaciones es grande, el valor r_1 no diferirá excesivamente de ϕ_1 . Por el contrario, si r_1 se basa en la cantidad de observaciones que suele ser común en las ciencias del comportamiento, la igualdad entre ϕ_1 y r_1 será muy baja. Dicho en otras palabras, los estimadores de la autocorrelación son insensibles con tamaños de muestra pequeños. Además, la discrepancia entre la desviación teórica y empírica es mayor con valores positivos altos de ϕ_1 que con valores negativos o positivos cercanos a cero, independientemente del tamaño muestral (Huitema y McKean, 1991; Kendall, 1954). Por otra parte, pueden haber divergencias en los resultados según qué estimador del parámetro de la autocorrelación se haya utilizado (ver Huitema y McKean, 1991), por lo que sería interesante saber cuál ha sido el empleado por Crosbie. Suponemos que es el que aparece en su artículo (ecuación 3.2) por ser el más común. Sin embargo, como ya se ha indicado, esta fórmula es para muestras grandes. Así pues, se desconocen los resultados que se obtendrían al aplicar otras fórmulas. En cuanto a las tablas 2 y 3 del apéndice 8 (con pendientes de 0.5 y 1, respectivamente) la autocorrelación hallada es superior a la fijada debido, precisamente, a la presencia de tendencia.

Otro aspecto a destacar de la tabla 1 del apéndice 8 es que el estadístico C se aproxima más al parámetro auto-regresivo fijado que no la autocorrelación calculada. Este hecho indica, suponiendo que la simulación llevada a cabo por Crosbie sea correcta, que

el estadístico C es más sensible a la autocorrelación, o bien que la prueba C, al ser para muestras pequeñas, detecta mejor la autocorrelación de las series simuladas.

Al aumentar la pendiente a 0.5 y 1 (tablas 2 y 3) se observa que tanto la prueba de la autocorrelación como el estadístico C aumentan considerablemente. Este resultado también lo obtuvieron Busk y Marascuilo (1988). Según estos autores, las autocorrelaciones altas vienen explicadas por la presencia de una tendencia. Otros autores, como Ezequiel y Fox (1959) y Wesolowsky (1976), ya postularon con anterioridad que las tendencias son una causa común de la autocorrelación. Por lo tanto, contrariamente a lo afirmado por Crosbie, tanto la prueba C como la de la autocorrelación detectan la tendencia. Obsérvese en la tabla 2 del apéndice 8 cómo, por ejemplo, para una longitud de 20 y un parámetro auto-regresivo igual a cero, la autocorrelación es de 0.76 ($p < 0.01$) y en la tabla 3, para esta misma combinación, la autocorrelación es de 0.83 ($p < 0.01$). Esto es debido a que la presencia de una pendiente produce una autocorrelación mayor, es decir, toda tendencia implica una autocorrelación distinta de cero. En esta misma línea, Suen (1987) afirma que el coeficiente de autocorrelación es un indicador "ómnibus" de la posible existencia de tendencias, ciclos o cambios.

En resumen, las tablas de Crosbie son poco consistentes, como también la interpretación que hace de ellas. En primer lugar, no se puede fijar, a la vez, un parámetro auto-regresivo igual a cero y una pendiente, puesto que toda pendiente lleva consigo una correlación serial. En segundo lugar, en el caso de que esto fuera lógico, se observa que tanto la prueba de la autocorrelación como el estadístico C, para cada longitud de serie, son significativas (tablas 2 y 3 del apéndice 8). Por lo tanto, a partir de estas tablas se concluiría que el estadístico C, además de la autocorrelación, detecta la pendiente, en cambio Crosbie concluye que C mide la dependencia serial pero no la tendencia.

Otra contradicción, entre los resultados obtenidos y las conclusiones a las que llega Crosbie, queda reflejada en la tabla 4 (ver apéndice 8) que muestra las correlaciones entre

tendencia, tamaño de la muestra, autocorrelación y estadístico C. Obsérvese cómo la prueba C está altamente correlacionada con la autocorrelación de retardo 1 (0.98, $p < 0.01$) pero, también, con la tendencia (0.63, $p < 0.01$). En consecuencia, a partir de esta tabla se inferiría la relación entre C, tendencia y autocorrelación.

En la tabla 5 (véase apéndice 8) se exponen los resultados de dos regresiones múltiples que predicen el valor C a partir de la tendencia y de la autocorrelación de retardo 1 de las series simuladas. El primer análisis de la regresión introduce la tendencia en el primer paso y causa una contribución significativa a la variancia explicada, $R^2=0.40$. El segundo paso introduce la autocorrelación y su contribución es también significativa, cambio en $R^2=0.56$. Estas dos variables juntas explican el 96% de la variancia de C. En el segundo análisis, al entrar en primer lugar la autocorrelación se observa que ésta explica el 96% de la variancia de C y, en consecuencia, ninguna variancia adicional puede ser explicada por la tendencia. La conclusión a la que llega Crosbie es que la correlación entre C y pendiente puede ser explicada por la correlación muy fuerte entre C y autocorrelación de retardo 1. Desde nuestro punto de vista, esto significa que la autocorrelación es un concepto muy amplio que incluye la tendencia. Es decir cuando está presente una tendencia, la autocorrelación la absorbe completamente. Por ejemplo, cuando en una serie temporal existe una tendencia o inclinación, se detecta fácilmente por la inspección visual de los correlogramas de la función de auto-correlación (FAC) y de la función de auto-correlación parcial (FACP). En efecto, si el correlograma de la FAC describe un patrón de decaimiento exponencial y la FACP para el retardo $k=1$ es cercano a la unidad, entonces es necesario diferenciar la serie. Así pues, es evidente que el efecto de la tendencia consiste en elevar el nivel de los coeficientes de auto-correlación de todos los retardos (Arnau, en prensa; Janacek y Swift, 1993).

A partir de las tablas 1-3 de Crosbie (apéndice 8), reproducimos los análisis de la regresión comentados anteriormente y se comprueba que nuestros resultados (tabla 3.1) no coinciden con los de Crosbie (tabla 5 del apéndice 8). El *output* obtenido, así como la

explicación del mismo, está recogido en el apéndice 9. Sin embargo, es ilustrativo considerar la tabla resumen final (tabla 3.1).

Tabla 3.1. Resumen de los análisis de la regresión a partir de las tablas 1, 2 y 3 de Crosbie (1989).

Variabes	R Múltiple	R ²	Cambio R ²	F Cambio	g.l.	Sig.
Tendencia	0.8052	0.6484	0.6484	106.9576	1/58	0.0000
Autocor.	0.9859	0.9721	0.3237	662.3951	2/57	0.0000
Autocor.	0.9811	0.9625	0.9625	1488.0075	1/58	0.0000
Tendencia	0.9859	0.9721	0.0097	19.7588	2/57	0.0000
Interacción	0.9962	0.9924	0.0202	148.5355	3/56	0.0000

g.l. =grados de libertad

Al introducir, en primer lugar, la tendencia y, luego, la autocorrelación (autocor.), la diferencia entre la proporción de variancia explicada por cada variable ($0.6484-0.3237=0.3247$) no es tan grande como cuando se introduce, en primer lugar, la autocorrelación y, a continuación, la tendencia ($0.9625-0.0097=0.9528$) (ver tabla 3.1). A pesar de ello, en los dos análisis, la cantidad de variación debida a cada variable es significativa. Por el contrario, en el segundo análisis, Crosbie halla que la variancia explicada por la tendencia no es significativa (tabla 5 del apéndice 8).

Un tercer paso muy importante, que no ha tenido en cuenta Crosbie, es el análisis de la interacción de las dos variables implicadas. Este paso debería ser el punto de interés inicial, puesto que, como puede observarse en la tabla 3.1, la proporción de variancia aumentada por el componente cruzado es estadísticamente significativa (cambio $R^2=0.0202$, $p<0.01$), lo que corrobora el hecho de que autocorrelación y tendencia son dos términos íntimamente relacionados.

A nuestro entender, dependencia serial y tendencia son dos aspectos altamente implicados. Su equivalencia sería análoga a la correspondencia entre los procesos ARIMA $(p,0,0)$ o AR (p) y ARIMA $(0,0,q)$ o MA (q) . Los análisis de series temporales intentan identificar si un proceso se ajusta a un AR o bien a un MA de primer orden. Sin embargo, ajustando cualquiera de los modelos se llega a ruido blanco, puesto que son dos procesos muy similares. Este punto queda perfectamente demostrado en Arnau (en prensa), Box y Jenkins (1970), Gottman (1981), Granger y Newbold (1977) y Kendall y Ord (1990), en donde se concluye que un proceso AR (1) es equivalente a una suma infinita de impulsos aleatorios pasados, exponencialmente ponderados y, en consecuencia, un proceso auto-regresivo estacionario de orden finito es igual a un proceso de medias móviles de orden infinito. De igual forma, un proceso MA (1) se puede expresar como una suma infinita de las observaciones pasadas, exponencialmente ponderadas. Esto significa que un proceso de media móvil de orden finito es equivalente a un proceso auto-regresivo de orden infinito.

En conclusión, el trabajo de Crosbie intenta demostrar, con argumentos estadísticos y análisis de Monte Carlo, que el estadístico C calcula la dependencia serial, pero no la tendencia. Además, el uso que Tryon hace de C acepta como estables series que presentan excesiva variabilidad (con una autocorrelación de retardo 1 negativa estadísticamente significativa). Estamos en desacuerdo con la primera afirmación, puesto que ha quedado bien claro que el estadístico C es capaz de detectar tanto una tendencia como una autocorrelación, y que siempre que está presente una tendencia lo está una autocorrelación. Respecto a la segunda afirmación, ya se comentó, en el apartado 3.1, el uso indebido que hace Tryon del concepto de estabilidad.

3.6. A modo de resumen

La presencia de dependencia serial en datos de caso único tiene una fuerte implicación en el análisis, tanto visual como estadístico, de tales datos. Con el fin de salvar esta dificultad, y siguiendo a Box y Jenkins (1970), se ha recomendado la utilización de ASTI, mediante modelos auto-regresivos integrados de medias móviles. Pero, para emplear dicha técnica es necesario un número de observaciones, por fase, superior al que aparece en investigaciones conductuales. Como procedimiento alternativo, se ha presentado el estadístico C de Young (1941). Este método es apropiado para evaluar los efectos de intervención en series temporales cortas, con tan pocas como ocho observaciones por fase. Así pues, el propósito del presente capítulo se ha centrado en la aplicación de esta aproximación alternativa a DSTI: estrategias, ámbitos en que se utiliza, limitaciones y críticas primordiales.

4. CÁLCULO DE LA POTENCIA DEL ESTADÍSTICO C: ESTUDIO DE SIMULACIÓN

El objetivo de esta investigación consiste en estudiar y someter a prueba el estadístico C como técnica alternativa de análisis, dentro del contexto conductual aplicado. Con esta finalidad, se estima la potencia de la prueba C para cada una de las estrategias propuestas por Tryon.

En primer lugar, se presentan una serie de consideraciones generales respecto a la teoría de la simulación, puesto que se trata de un estudio de potencia basado en múltiples series temporales simuladas mediante el procedimiento de Monte Carlo. En segundo lugar, se describen las hojas de cálculo como instrumentos útiles para la generación y análisis de datos. En tercer lugar, se explican aquellos aspectos más importantes relacionados con el análisis de potencia. Seguidamente, se describen los objetivos principales de la investigación y, por último, se presentan tres estudios correspondientes a las tres estrategias de Tryon.

4.1. Consideraciones generales en torno a los estudios de simulación

En la actualidad, cada vez es mayor el número de campos y aplicaciones concretas de la simulación. Este hecho se debe a que el experimento real a menudo resulta ser muy complejo, o bien costoso. El método más común de simulación es el de Monte Carlo, basado en la generación de números aleatorios. Las técnicas iniciales se fundamentaban en mecanismos manuales como la ruleta o la lotería (de ahí el nombre genérico de Monte Carlo). Sin embargo, estos procedimientos no fueron muy prácticos en la construcción eficiente de una gran cantidad de números aleatorios. En base a los resultados obtenidos por estas técnicas, existen multitud de tablas de números aleatorios que pueden utilizarse para aquellas aplicaciones llevadas a cabo por medio de procedimientos manuales o con el auxilio de calculadoras convencionales. No obstante, estas tablas no son muy útiles

cuando se requiere el uso del ordenador para resolver el problema planteado.

La consecución de un buen sistema para la generación de números aleatorios, a través de programación, es compleja ya que requiere una fundamentación numérica. De la constatación de este hecho surge la necesidad de la utilización de sistemas generales de simulación por ordenador. La mayoría de paquetes estadísticos utilizan rutinas internas ya prescritas para la obtención rápida de números aleatorios de cualquier clase (en el presente estudio son números aleatorios ajustados a una distribución regular en el intervalo $[0,1]$). De esta forma, no se requiere ningún análisis matemático por parte del usuario. Aunque se ha demostrado que algunos generadores de números aleatorios son más "aleatorios" que otros, no entraremos en esta problemática, por lo que se remite al lector que esté interesado a los trabajos de Jain (1991) y Knuth (1981).

La técnica de Monte Carlo, además de aplicarse en ingeniería, economía y ciencias físicas puede usarse también dentro del ámbito de las ciencias sociales y del comportamiento y, más concretamente, para la investigación de los análisis conductuales. En efecto, hay una evidencia empírica de estudios llevados a cabo a partir de series temporales simuladas mediante procedimientos de Monte Carlo. Por citar algunos, están los trabajos de Crosbie (1989), Gottman (1981), Greenwood y Matyas (1990), Harrop y Velicer (1985), Huitema y Mckean (1991), Matyas y Greenwood (1991) y Vallejo (1993). Por otro lado, recientemente se ha producido una proliferación de programas de ordenador que perfeccionan las técnicas de generación de números aleatorios (uno de los más recientes es el *Mathematica*, 1993), así como un conjunto de manuales sobre análisis de series temporales que presentan ejemplos con datos simulados y que, además, proporcionan los algoritmos para la generación de las series (Brockwell y Davis, 1990; Dielman, 1989; Janacek y Swift, 1993; Lutkepohl, 1991).

Centrándonos, pues, en la simulación de series temporales, se tiene que para generar una realización de una serie a partir de un modelo, es necesario una forma de

producir una o más series de números que tengan las mismas propiedades que el modelo. Como se verá en los apartados dedicados a los métodos de cada estudio (4.5.1, 4.6.1, 4.7.1), para obtener las puntuaciones de cada serie se necesita un número que sea la realización del término aleatorio e_t . A tal fin se establecen algunos supuestos respecto a la distribución de e_t , siendo lo común asumir la normalidad. La mayor parte de los sistemas de ordenador actuales incluyen una función uniforme de números aleatorios, que permite crear una función que se ajuste a la normal. De este modo, se pueden simular modelos de series temporales siempre que hayan números aleatorios disponibles.

Antes de proseguir, advertimos que aquellos autores que simulan modelos que incluyen valores anteriores a Y_t (variable de sujeto), como por ejemplo los procesos ARMA, deben tener mucho cuidado, puesto que es necesario establecer estos valores de partida para poder llevar a cabo la simulación. Si el modelo es estacionario, los efectos de tales valores desaparecen con el tiempo (demostrado en Janacek y Swift, 1993). Por esta razón, se suelen ignorar las primeras 50 o 100 observaciones simuladas. Así, se elimina lo que algunos autores han llamado efecto *start-up* (Janacek y Swift, 1993; Matyas y Greenwood, 1991). En el presente estudio, este tipo de efecto no interviene, ya que se simulan modelos lineales de la regresión y, por lo tanto, no aparece ningún valor precedente a Y_t .

4.2. Hojas de cálculo: generación y análisis de datos

Algunos de los programas de ordenador que generan números aleatorios son las ya conocidas hojas de cálculo, con la ventaja de que permiten realizar, de forma sencilla, cálculos con los datos generados. Ejemplos de estudios que utilizan este tipo de programas son los de Borenstein *et al.* (1992) y Eamon (1992). El primero trata de un análisis de la potencia estadística de la prueba t a partir de simulaciones de Monte Carlo. Esta investigación tiene un gran interés para nuestros propósitos, puesto que muestra las

ventajas de las hojas de cálculo para realizar simulaciones y, a su vez, hallar la potencia de un estadístico objeto de estudio. El segundo trabajo trata de la utilidad que tienen las hojas de cálculo para la generación y análisis de datos simples.

Hewett (1985) destaca dos importantes ventajas de la aplicación de las hojas de cálculo: a) bajo coste en términos de tiempo y esfuerzo y b) gran flexibilidad. No obstante, existe un problema adicional, al generar datos para el análisis estadístico, y es que las hojas de cálculo poseen un formato relativamente fijo. Por ejemplo, una hoja de cálculo con ocho filas (observaciones), normalmente no podrá ser utilizada para generar o analizar series de 15 filas. Por supuesto, se pueden copiar las fórmulas de fila a fila, pero hacerlo es un tanto laborioso, ya que al insertar o eliminar filas -o bien columnas- se requiere que las fórmulas, que hacen referencia a otras celdas, sean modificadas. Sin embargo, este problema se resuelve con las modernas hojas de cálculo, tales como el *LOTUS 1-2-3* (1990). En efecto, en lugar de referencias a celdas concretas (por ejemplo, *@MEDIA(A12..A22)*) se utilizan rangos especificados de antemano (por ejemplo, *@MEDIA(SERIE)*). De esta forma, las filas y/o columnas pueden añadirse o eliminarse dentro de los rangos, sin causar ningún error. Estos rangos son de gran utilidad para crear macros. Una macro es un conjunto de operaciones y mandatos (macroinstrucciones) de la hoja de cálculo que puede ser ejecutada con un par de pulsaciones que definen el nombre de la macro, y posee un lenguaje de programación propio (lineal, tipo BASIC). Los lenguajes macro son muy potentes para manipular la forma y función de las hojas de cálculo y, así, dar una mayor flexibilidad.

4.3. Potencia de una prueba estadística

Debido a la creciente importancia del análisis de la potencia en ciencias del comportamiento y a la existencia de una literatura escasa y ambigua durante décadas (Fischer, 1978; Prather y Gibson, 1977; Rossi y Wright, 1984), el objetivo de este

apartado es presentar el marco conceptual básico del análisis estadístico de la potencia. Así, se discutirá lo más relevante acerca de aquellos elementos que afectan directamente a la potencia: tamaño muestral, nivel de significación y tamaño del efecto. Sin embargo, una discusión en profundidad excede los límites de este trabajo, por lo que se recomienda, para una amplia revisión del tema, consultar a Cohen (1988) y Lipsey (1990).

4.3.1. Desarrollo histórico

Históricamente, el concepto de potencia se atribuye a Neyman y Pearson (1928, 1933). A partir de entonces, aparecen una serie de autores que tienen en cuenta en sus estudios la potencia estadística (Cox, 1948; McNemar, 1960; Sterling, 1959; Tukey, 1960; Tullock, 1959) y, dentro de la literatura de las ciencias sociales, es de destacar Mosteller y Bush (1954). Sin embargo, no es hasta 1962 cuando aparece por primera vez, en el ámbito de las ciencias sociales, un estudio sistemático de la potencia estadística. Se trata del trabajo realizado por Cohen (*The statistical power of abnormal-social psychological research: A review*, 1962).

La investigación llevada a cabo por Cohen (1962) establece, en primer lugar, la importancia de la potencia estadística dentro de la investigación experimental y, en segundo lugar, proporciona una serie de pautas para llevar a cabo un análisis de potencia. Este estudio alentó a los investigadores a prestar mayor atención a la potencia de las pruebas y no centrarse, exclusivamente, en el nivel de significación. En 1965, Cohen publica una monografía que contiene todas las recomendaciones para tratar con eficacia el problema de la potencia. Más adelante, sigue esta discusión en su libro de 1969, revisado en una segunda edición (1988) donde presenta unas tablas muy útiles para el cálculo de la potencia. Sin duda, dicha publicación inspira docenas de estudios sobre la potencia y el tamaño del efecto en muchas áreas de las ciencias sociales, así como diversos programas de ordenador (examinados en Goldstein, 1989). Otros investigadores, que también

ayudaron a promover el interés por el tema, fueron los sociólogos Labovitz (1968) y Skipper, Guenther y Nass (1967) y los psicólogos Bakan (1966), Lykken (1968) y Runyon (1969). Sin embargo, esta proliferación de estudios no ha tenido un impacto en investigaciones posteriores. En efecto, de 54 artículos publicados en el volumen de 1984 del *Journal of Abnormal Psychology*, sólo dos mencionaban la potencia y ninguno la estimaba (Sedlmeier y Gigerenzer, 1989). Además, si se toman como evidencia manuales de estadística, se observa que con frecuencia la potencia no es tratada.

No está claro el porqué los investigadores ignoran el análisis de la potencia. Según Cohen (1992), una buena parte se debe a la poca importancia que se da al tamaño del efecto, como si la principal preocupación, en la mayor parte de la investigación psicológica, fuese los resultados estadísticos y los valores de p , sin tener en cuenta el fenómeno psicológico que se está estudiando. Chase y Tucker (1976) y Sedlmeier y Gigerenzer (1989) atribuyen este hecho al predominio de la teoría fisheriana. La escuela de Fisher interpreta las pruebas estadísticas como pruebas de significación (Fisher, 1935, 1950, 1955), mientras que la escuela de Neyman-Pearson las caracteriza como pruebas de decisión (Neyman y Pearson, 1928, 1933, 1936).

Según Chase y Tucker (1976), esta controversia entre tradiciones puede reducirse al hecho de que para detectar una diferencia significativa, un defensor del enfoque fisheriano concluiría que la hipótesis nula (H_0) no es válida (con lo que se prueba la existencia del fenómeno que se esté estudiando). De este modo, se está condicionado a la veracidad de la H_0 . En cambio, un investigador de la tradición de Neyman-Pearson, simplemente rechazaría la H_0 para esta ocasión particular.

Otra distinción importante entre estas dos aproximaciones es la que concierne a si el nivel de significación empleado en el análisis debe fijarse anticipadamente. Aunque Fisher daba prioridad a un nivel de significación de 0.05, nunca prescribió que tal nivel debiera mantenerse fijo o que debiera establecerse antes de llevar a cabo el experimento.

Por contra, la posición de Neyman-Pearson requiere que el nivel de significación se determine antes de cualquier análisis estadístico y que el investigador se adhiera a él para todas las decisiones estadísticas. En este sentido, si el nivel de significación no se estableciera con anterioridad, se podría caer en el error de fijar un valor α basado más en los resultados que en una estimación de la probabilidad.

Un tercer desacuerdo entre las dos teorías se refiere a la interpretación de resultados positivos y negativos en la investigación experimental. Si se sigue a Fisher, se puede afirmar que el efecto no es cero cuando se rechaza la H_0 , pero no se puede concluir que sea cero cuando se acepta. Lo que realmente significa la falsa afirmación de que la H_0 es cierta no es que el efecto sea cero, sino que es despreciable o trivial (Cohen, 1988). Por el contrario, la formulación de Neyman-Pearson para la hipótesis alternativa (H_1) postula que existe una alternativa exacta para el tamaño del efecto. Esta afirmación llevó a Neyman y Pearson (1928, 1933) al concepto de error de Tipo II (aceptar la H_0 cuando es falsa), interrelacionado con el de potencia. En cuanto al concepto de error de Tipo I, ya era conocido en la aproximación fisheriana; no como el error de rechazar la H_0 cuando es verdadera sino como la probabilidad de error que un investigador está dispuesto a aceptar cuando interpreta los resultados de una prueba estadística.

En conclusión, la razón de porque los psicólogos descuidan la potencia se debe a hechos históricos. Los manuales de psicología y educación transmitieron, primero, el mensaje fisheriano y, después de la II Guerra Mundial, los autores comenzaron a darse cuenta del impacto de la aproximación de Neyman-Pearson y a sustituir la teoría anterior. Sin embargo, los conceptos de error Tipo II y potencia fueron incluidos por la mayoría de investigadores, dentro del marco fisheriano. El resultado fue una teoría híbrida que presentaba ideas desde dos puntos contrapuestos como un cuerpo monolítico único (prueba de la H_0 , siguiendo a Fisher, y error de Tipo II y especificación del error de Tipo I antes de la recogida de datos, siguiendo a Neyman y Pearson), lo que conlleva una confusión del significado de conceptos básicos (Bakan, 1966; Oakes, 1986) y, en algún

grado, una explicación del continuo descuido acerca de la potencia. Esta estadística híbrida se enseña, normalmente, como estadística *per se*, sin mención de los originadores de las respectivas ideas (Gigerenzer y Murray, 1987).

Según Sedlmeier y Gigerenzer (1989), este continuado desinterés por la potencia de las pruebas estadísticas sólo cambiará cuando los editores de las principales revistas exijan, dentro de su política editorial, que los autores estimen la potencia de sus pruebas de significación. De todos modos, en los últimos años han aparecido una serie de textos básicos sobre potencia estadística (Cohen, 1988; Kraemer y Thiemann, 1987; Lipsey, 1990, entre los más destacados). Desafortunadamente, la mayor parte de estas fuentes aún no contienen mucha información sobre cómo calcular la potencia o están limitadas a la consideración de pruebas simples, tales como *t* y *F*. Por otro lado, están disponibles programas informáticos que calculan la potencia del análisis de la variancia unidireccional (Borenstein *et al.*, 1990), de la regresión múltiple (Rothstein *et al.*, 1990), y de correlaciones y diferencias entre proporciones (Borenstein y Cohen, 1988).

Después de esta breve referencia histórica, es interesante revisar la situación actual dentro de los diseños conductuales. Sharpley (1987), por ejemplo, da una gran importancia a las cuestiones referidas a la probabilidad de cometer errores de Tipo I y Tipo II en los ASTI, y sugiere que la potencia debería recibir más atención por parte de los investigadores. También Gottman (1981) se ha referido con detalle a este tema. Suen (1987) y Suen y Ary (1987) cuestionan, a través de un análisis de la potencia de la autocorrelación, la conclusión de Huitema (1985) de autocorrelación nula en datos conductuales. Busk y Marascuilo (1988), al polemizar los resultados de Huitema, se basan en la relación existente entre potencia y tamaño muestral de los coeficientes de autocorrelación (es decir, para detectar las autocorrelaciones como estadísticamente significativas son necesarias muestras grandes). Más recientemente, Vallejo (1993) proporciona estimaciones empíricas concernientes a la probabilidad de cometer errores de Tipo I y a la potencia de tres procedimientos analíticos -modelo correctamente identificado

y enfoques asumidos de antemano, AR(1) y AR(5)- para diferentes modelos ARIMA simulados.

4.3.2. Concepto de potencia: su relación con errores de Tipo I y Tipo II

Es sabido que al tomar una decisión estadística (rechazar o aceptar la H_0) se corre el riesgo de cometer un error, debido a que los problemas rara vez se presentan de manera que sea posible discriminar con certeza entre las hipótesis falsas y las verdaderas. Estos errores pueden ser de dos clases: a) error de Tipo I y b) error de Tipo II. Los dos tipos tienen diferentes consecuencias y costos, por lo que la probabilidad de cometer ambos errores se intenta mantener tan baja como sea posible.

El error de Tipo I o nivel de significación (α) es el más familiar para la mayoría de investigadores y, por lo común, está fijado en 0.05. Al aplicar una prueba estadística también existe la posibilidad de cometer un error de Tipo II, denotado por β , y es la probabilidad de aceptar la H_0 cuando es falsa. Los dos tipos de errores están inversamente relacionados. En efecto, al controlar el error de Tipo I, el error de Tipo II queda determinado (Cohen, 1988; Rossi, 1990; Stevens, 1986). Así, dada una muestra, al disminuir α aumenta β y si se desea disminuir ambos errores, la única forma consiste en aumentar el número de observaciones, que no siempre es posible. En consecuencia, el problema que surge al investigador es el de alcanzar un equilibrio entre los dos tipos de errores. Un análisis de los costes, beneficios y otros factores relevantes asociados a los errores Tipo I y II se halla en Brown (1983), Cascio y Zedeck (1983), Nagel y Neef (1977) y Schneider y Darcy (1984).

La probabilidad de un error Tipo I viene designada de antemano por α . Así, se controlan las probabilidades de rechazar H_0 verdaderas. Por contra, la tasa de error Tipo II no se puede determinar hasta que se conozca H_1 , o sea, hasta que el investigador

sepa el efecto concreto existente. Por lo tanto, el experimentador deberá basarse en estimaciones.

En la teoría de Neyman-Pearson, la potencia de una prueba estadística es la probabilidad de resultados significativos, esto es la aceptación de la H_1 cuando es verdadera (Cohen, 1988; Lipsey, 1990). En otras palabras, la potencia de un estadístico viene dada por su capacidad de rechazar la H_0 correctamente, de modo que está determinada por la probabilidad de cometer errores de Tipo II. Así, la potencia de una prueba es el complemento de la probabilidad de un error Tipo II ($1-\beta$). Cuando β es pequeño, entonces la potencia es grande y viceversa.

Cohen (1992) propone como convención de uso general una potencia de 0.80 (o sea, $\beta=0.20$). Un valor sustancialmente inferior a 0.80 implicaría un gran riesgo de incurrir en un error Tipo II, y un valor superior exigiría una muestra muy grande, fuera de los recursos del investigador. Si se toma un valor α convencional de 0.05 y una potencia de 0.80, la *ratio* resultante de $\beta:\alpha$ es 4:1 (0.20 a 0.05).

Según Strube (1985), el hecho de rechazar la H_1 debido a una potencia baja lleva, en un nivel de meta-análisis, a abandonar prematuramente una área de investigación prometedora. Por ello, dada la influencia que tienen las revisiones de estudios anteriores para una investigación futura, es indispensable que las decisiones concernientes a las hipótesis sean correctas.

Por último, antes de presentar los distintos elementos que determinan la potencia, se detallan a continuación las consecuencias positivas que tiene la consideración de la potencia estadística (Rossi, 1990):

- 1) La potencia de una prueba es la probabilidad de obtener un resultado estadísticamente significativo. Así, si las estimaciones de potencia obtenidas *a priori* son bajas, el investigador puede elegir entre aumentar la potencia o abandonar la investigación si los costes de incrementar la potencia son demasiado altos, o si la baja potencia no justifica el tiempo, el esfuerzo y la cantidad de recursos que requiere la investigación que se lleva a cabo.

- 2) Conocer la potencia de una prueba estadística facilita la interpretación de los resultados nulos. La incapacidad de rechazar la H_0 no significa que sea cierta, sino que no existen suficientes evidencias para rechazarla. Si la potencia es baja, es razonable sugerir, *a priori*, que no hay una razón favorable para rechazar la H_0 . En cambio, si la potencia es alta, el no poder rechazar la H_0 puede considerarse -dentro de unos límites- como una afirmación de dicha hipótesis, puesto que la probabilidad de un error de Tipo II será baja.

- 3) Cuando la potencia estadística media es baja, se cuestiona la validez de los resultados estadísticamente significativos. En este sentido, una proporción sustancial de resultados significativos publicados pueden ser errores de Tipo I. Por otro lado, cuando la potencia es marginal (aproximadamente 0.50) es muy posible obtener un patrón de resultados inconsistentes, que en algunos casos dará lugar a resultados significativos y en otros no (Kazdin y Bass, 1989).

4.3.3. Elementos de la potencia estadística

Las estimaciones de la potencia estadística quedan determinadas por tres elementos: a) tamaño muestral, cuanto más grande sea la muestra, mayor será la potencia; b) nivel de significación, al incrementarse la probabilidad de error Tipo I, la potencia también aumenta, y c) tamaño del efecto o grado en que el fenómeno de interés está presente, cuando los efectos son grandes, la potencia incrementa.

A continuación, se presenta una breve exposición de cada uno de estos elementos (para una explicación más amplia, puede consultarse a Cohen, 1988, 1992); sin embargo, el tamaño del efecto merece una consideración más detallada debido a que, por un lado, es una de las formas más importantes para determinar la potencia o el tamaño muestral requerido -o ambos- y, por otro, es el concepto, relacionado con la inferencia estadística, menos familiar a los investigadores de las ciencias del comportamiento.

4.3.3.1. Tamaño muestral

La potencia depende del número de observaciones. Ésta aumentará cuando así lo haga el tamaño de la muestra, manteniéndose constantes los demás elementos (nivel de significación y tamaño del efecto). Este hecho se debe a que al aumentar la muestra, será menor el error aleatorio de medida y mayor la fiabilidad o precisión de los resultados y, por consiguiente, mayor la probabilidad de manifestarse el fenómeno bajo prueba (Cohen, 1988; Lipsey, 1990). Así, el excluir fuentes de variabilidad irrelevantes para el conocimiento del fenómeno que se esté estudiando es útil para incrementar la potencia.

Cuando se planifica una investigación, es conveniente conocer, para un error α determinado y un tamaño del efecto hipotético, la muestra necesaria para obtener la potencia deseada. Es preciso incrementar el tamaño muestral al aumentar la potencia y

disminuir el tamaño del efecto y α (Cohen, 1992).

4.3.3.2. Nivel de significación

El nivel de significación representa un estadístico de prueba en virtud del cual existe el fenómeno, o el riesgo de rechazar erróneamente la H_0 . Para hallar la potencia, de acuerdo con la teoría de Neyman-Pearson, el valor de α debe establecerse por anticipado. A no ser que se indique lo contrario, se toma $\alpha=0.05$ (es parte del legado fisheriano: Cohen, 1990); aunque evidentemente pueden seleccionarse otros valores.

Asumiendo que un fenómeno concreto existe en un determinado grado en la población y manteniéndose los demás elementos constantes, cuanto más estricto sea el estándar de prueba (α), menor será la probabilidad de que la muestra lleve a resultados que se ajusten a dicho estándar; o sea, la potencia será pobre (Cohen, 1988; Lipsey, 1990; Sedlmeier y Gigerenzer, 1989). Así pues, la práctica de tomar valores de α pequeños -cuanto más pequeños mejor- conlleva a que la potencia sea baja. Además, el complemento de la potencia (β) constituye también un error (el error de Tipo II), por lo que la inferencia estadística puede ser considerada como una ponderación de los dos tipos de error (Cohen, 1962). Un investigador puede establecer el riesgo de rechazar equivocadamente la H_0 a un nivel de, por ejemplo, $\alpha=0.001$ y, con ello, reducir la potencia de la prueba (por consiguiente, aumentar el valor de β). Por otro lado, es posible incrementar la potencia de la prueba hasta un valor aceptable usando un nivel de significación menos riguroso. No obstante, de esta forma se incrementa la probabilidad de cometer un error Tipo I. Por lo tanto, debe equilibrarse la probabilidad de incurrir en un error de Tipo I con la probabilidad de caer en un error de Tipo II, puesto que, como ya se ha comentado anteriormente, no puede reducirse la probabilidad de cometer un tipo de error sin producir un incremento en la otra.

4.3.3.3. Tamaño del efecto

Por tamaño del efecto (TE) se entiende el grado en que el fenómeno bajo estudio está presente en la población, o el grado en que la H_0 es falsa (Cohen, 1988), es decir, la discrepancia entre H_0 y H_1 . Así, el TE puede ser tratado como un factor que toma el valor cero cuando la H_0 es verdadera, y cualquier otro valor distinto de cero cuando es falsa. Por lo tanto, se trata de un índice del grado de distanciamiento de la H_0 (Lipsey, 1990). Lo ideal sería que este índice fuera aplicable a todos los ámbitos de investigación; sin embargo, la unidad de medida debe ser apropiada a los datos y a la prueba de significación empleada. De todas formas, las tablas de potencia existentes para determinadas pruebas tienen un índice libre de unidades, ya que es inviable preparar tablas para cada nueva unidad de medida con la que se trabaje.

Los investigadores del comportamiento encuentran que especificar el TE es la parte más difícil del análisis de potencia. La dificultad se debe, al menos en parte, a un bajo nivel de conciencia sobre la magnitud de los fenómenos y a la teoría fisheriana que considera el TE como cualquier valor distinto de cero. Una solución al respecto es buscar ayuda en el marco teórico o en investigaciones anteriores dentro del área de estudio. Si esto no es suficiente, un recurso adicional es especificar los TE en términos de "pequeño", "mediano" y "grande" (Cohen, 1988). Se trata, al igual que el nivel de significación de 0.05, de una convención que puede ser mal utilizada si se abusa de su condición de convencionalidad. Aunque sus definiciones son arbitrarias, deben ser razonables (Cohen, 1988). Así, por ejemplo, los TE pequeños han de ser lo suficientemente grandes para que su búsqueda no sea una tarea sin sentido; pero no tan grandes como para que sean fácilmente perceptibles. Cohen (1988, 1992) establece los valores convencionales pequeños, medianos y grandes para varias pruebas estadísticas distintas (prueba t , análisis de la variancia unidireccional, pruebas de correlación, diferencias entre correlaciones, prueba de los signos, diferencias entre proporciones, prueba ji-cuadrado, y correlación parcial y múltiple), de manera que los TE medios representan un efecto que el observador

cuidadoso puede observar directamente, los TE pequeños son bastante más pequeños que los medios, pero no tanto como para ser triviales, y los TE grandes están a la misma distancia por encima de los medios que los pequeños lo están por debajo.

Debido a que la estimación del TE es un tema relativamente nuevo en ciencias conductuales, la interpretación está cargada de dificultades. La caracterización de Cohen (1988) de los TE como pequeños, medianos y grandes es el primer paso tentativo en esta dirección. Estas convenciones han quedado fijadas y su uso se ha convertido en general, no sólo en el análisis de potencia sino también en el meta-análisis. No obstante, aun así, existe un gran riesgo de malas interpretaciones.

El TE es un elemento crucial en el análisis de potencia, ya que la potencia depende de la magnitud del efecto que se quiere detectar. Cuanto mayor sea el TE postulado, manteniéndose lo demás constante (nivel de significación y tamaño muestral), mayor será la potencia estadística. Cuanto mayor sea el TE, manteniéndose el nivel de significación y la potencia deseada constante, menor será la muestra necesaria para detectarlo. No obstante, la mayoría de efectos buscados en investigaciones de la personalidad, clínico-psicológicas y sociales son pequeños o moderados (Cohen, 1988; Cooper, 1981; Cooper y Findley, 1982; Stevens, 1986). Los efectos grandes son, con frecuencia, un tema de discusión en ámbitos como la economía, la psicología experimental y la fisiológica, campos caracterizados por el estudio de variables potentes o un buen control experimental.

Debido a la falta de estudios sobre TE en las ciencias del comportamiento, sería interesante que los investigadores informarán sobre los índices del TE de sus trabajos para que, así, en investigaciones futuras, haya una base de datos que permita estimar la potencia a través de, por ejemplo, procedimientos meta-analíticos (véase Cordray y Sonnefeld, 1985; Glass, McGaw y Simth, 1981; Hedges y Olkin, 1985; Holmes, 1984; Hunter, Schmidt y Jackson, 1982; Rosenthal, 1984; Rossi, 1990). Sin embargo, debe considerarse el hecho de que los estudios publicados no son representativos de todas las investigaciones

llevadas a cabo (Lane y Dunlap, 1978). En efecto, aquéllas que rechazan la H_1 (debido frecuentemente a una baja potencia), la mayoría de veces no son admitidas para su publicación (Greenwald, 1975; Rosenthal, 1979).

4.3.4. Tipos de análisis de potencia

Hay dos formas de estimar la potencia: una, *a priori* y otra, *a posteriori*. La primera alerta al investigador de la cantidad de sujetos por grupo necesarios para una potencia adecuada, y la segunda es importante en la interpretación de los resultados de estudios ya acabados, especialmente cuando son negativos (Rossi, 1990; Stevens, 1986). La mayoría de investigadores que no están familiarizados con la potencia, suelen interpretar los resultados no significativos en términos de que no existe efecto alguno. Sin embargo, puede ser que el experimento posea una potencia pobre (debida a un tamaño de muestra y/o efecto pequeño) incapaz de detectar la significación estadística (Cronbach y Snow, 1977).

En los apartados anteriores se han descrito cuatro elementos de la inferencia estadística: potencia, nivel de significación (α), tamaño de la muestra (n) y tamaño del efecto (TE). Todos ellos están relacionados de manera que cada uno es función de los otros tres. Esta relación hace posible cuatro tipos de análisis de potencia (Cohen, 1965; 1988): a) potencia en función de α , TE y n ; b) n en función de α , TE y potencia; c) TE en función de α , n y potencia, y d) α en función de n , TE y potencia.

En el primer caso se determina la potencia, dado α , TE y n . De este modo, el investigador puede decidir cambiar las especificaciones iniciales con el fin de aumentar la potencia. En el segundo tipo de análisis, el experimentador estima un TE, establece un nivel de significación α y la cantidad de potencia deseada, con lo que queda determinada la n necesaria. Esta segunda clase de análisis es la base para la decisión del tamaño

muestral a utilizar en una investigación concreta. El tercer tipo de estudios de potencia permite encontrar el TE, dados los restantes elementos. Esta clase de análisis es de menor utilidad que las anteriores; no obstante, puede ser muy favorable en determinadas circunstancias (Cohen, 1970). La última forma de análisis de potencia responde a la pregunta: ¿qué α es necesario para detectar un TE, con una probabilidad específica (potencia) y una n fijada?. Este tipo de análisis es muy poco común debido a la existencia de una convención de $\alpha=0.05$ que hace que los investigadores estén poco dispuestos a considerar valores de α mayores.

En el presente apartado se han descrito cuatro tipos distintos de análisis de potencia estadística. Quizá el procedimiento más frecuente sea determinar los tamaños muestrales de las investigaciones que se desean llevar a cabo. Las tablas de Cohen (1988) permiten obtener, para algunas pruebas, los dos primeros tipos de análisis y también, aunque no con tanta facilidad, los otros dos, a través de una interpolación inversa.

Los aspectos generales respecto a la potencia, esbozados hasta aquí, sirven para el análisis de la potencia del estadístico C, dados α , TE y n . Así, al planificar una investigación, será posible estimar la potencia y determinar el tamaño de muestra apropiado, o bien interpretar estudios ya terminados.

4.4. Objetivos

El propósito de esta investigación es estudiar el estadístico C como técnica alternativa de análisis de series temporales breves o cortas, dentro del contexto conductual aplicado. A tal efecto se lleva a cabo un estudio de simulación a partir de modelos lineales simples, variando la longitud de las series de diseños conductuales A-B, así como el nivel de significación y los cambios de pendiente y/o nivel. Con ello, se pretende hallar una

solución a la inferencia estadística del efecto de los tratamientos que tenga en cuenta tanto la dependencia serial como la brevedad de la serie.

Si bien los modelos ARIMA resuelven uno de los principales problemas de las series temporales (la dependencia serial de los datos), fracasan en su aplicación a diseños de caso único o $N=1$, que son los más ampliamente utilizados dentro del ámbito conductual aplicado. Ello se debe a que, por lo general, los puntos de observación de cada fase son escasos, lo que genera graves problemas en el momento de identificar la estructura subyacente de la serie.

Ante la dificultad planteada por los ASTI cuando se dispone de pocas observaciones por fase, Tryon (1982) propone, como técnica alternativa de análisis, el estadístico C de Young. En este sentido plantea tres estrategias (explicadas en el punto 3.1), sosteniendo que el orden en que disminuye la potencia se ajusta a la siguiente secuencia: *a*) bajo el supuesto de línea base no significativa, se analiza el total de la serie (fase A más fase B); *b*) cuando, por el contrario, en la línea base hay tendencia se comparan las series previo ajuste de una recta de regresión, *c*) por último, como alternativa a *b*, cabe también la posibilidad de una comparación de series con datos directos. Por otra parte, Young (1941) apuntó la necesidad de un estudio de potencia de la prueba C para varios tipos de no aleatoriedad. Siguiendo a estos autores, los objetivos particulares del presente trabajo son:

- 1) Profundizar en la temática del estadístico C (descripción estadística y aplicación a DSTI) como alternativa de análisis para series temporales cortas. Este aspecto ha sido tratado en el segundo y tercer capítulo.

- 2) Hallar, a través de simulación de Monte Carlo, la potencia del estadístico C para cada una de las estrategias propuestas por Tryon combinando, simultáneamente, cambios de pendiente y/o nivel, longitudes de fase y niveles de significación en diseños conductuales A-B.
- 3) Estudio de las limitaciones que presenta el estadístico C cuando se ejecuta la tercera estrategia de Tryon.
- 4) Establecer una comparación de las tres alternativas de análisis posibles en la tercera estrategia de Tryon (comparación de series utilizando la técnica de la regresión, comparación de series con datos directos y análisis del conjunto de las fases), con el fin de comprobar la efectividad de cada una de ellas para las diferentes combinaciones de las condiciones experimentales.

De este modo, se considera la posibilidad de aportar una alternativa de análisis a datos de series temporales cortas, propios de diseños conductuales. Con ello pensamos contribuir al desarrollo de nuevas técnicas de análisis apropiadas a esta clase de datos que, hasta el momento, no se han probado suficientemente. De esta manera, se pretende resolver, en parte, uno de los principales problemas que afecta a la investigación conductual aplicada: la correcta inferencia del efecto de los tratamientos. Siguiendo con esta línea de trabajo se ha planificado una investigación dividida en tres experimentos que corresponden a cada una de las estrategias de uso del estadístico C.

4.5. Estudio 1

4.5.1. Método

Con el propósito de evaluar la primera estrategia de Tryon se ha diseñado un experimento de simulación de Monte Carlo. La generación de los datos, así como todos los cálculos posteriores, se ha llevado a cabo a través de la hoja de cálculo *LOTUS 1-2-3* para PC (1990; versión 3.1+).

Mediante simulación se han generado fases de línea base con longitudes desde $n_1=8$ a $n_1=15$ para tres modelos distintos: línea base estable, línea base con variabilidad (baja, media y alta) y línea base con tendencia lineal programada, cuyos valores van de -0.1 a -1 aumentando de décima en décima.

Tanto las longitudes de fase como las tendencias fijadas se ajustan a los objetivos propuestos, ya que aumentan lentamente y, por lo tanto, permiten hallar cuándo el estadístico C empieza a poseer una potencia estadística aceptable. En total se han generado 22400 fases de pre-intervención, 200 para cada una de las combinaciones.

Cada serie de línea base con ausencia de tendencia ha sido generada de acuerdo a la siguiente fórmula:

$$Y_t = b_0 + e_t \quad (4.1)$$

donde Y_t es la observación obtenida en el punto t ; b_0 es la constante, y e_t es el término de error.

Las medidas de variabilidad presentes en las fases, es decir, las dispersiones de las distintas puntuaciones alrededor de su tendencia central se han fijado en términos de amplitudes totales, puesto que éstas son útiles cuando se quiere hacer una comparación de dos o más grupos en cuanto a su variabilidad (Garret, 1974). Para ello se establece, arbitrariamente, una media de 10 y unas puntuaciones que se extienden de 8 a 12 (amplitud 4) para una variabilidad baja; de 5 a 15 (amplitud 10) para una variabilidad media, y de 0 a 20 (amplitud 20) para una variabilidad alta. Así, se tiene, para cada tamaño muestral, tres distribuciones de frecuencias con la misma tendencia central, pero de muy distinta variabilidad.

Para períodos de no intervención con tendencia, los datos se han generado a partir de la siguiente función matemática:

$$Y_t = b_0 + b_1T + e_t \quad (4.2)$$

donde b_1 simboliza la tendencia en la fase de pre-intervención, y T es una unidad de tiempo particular en la secuencia observacional.

El error aleatorio que se añade a las diferentes series temporales, tanto en este estudio como en los siguientes, ha sido generado de forma que sea normal e independientemente distribuido con media cero y variancia uno. Para ello se ha usado una función matemática (@ALEAT) propia del *LOTUS 1-2-3* que proporciona desviaciones aleatorias uniformes y, posteriormente, se ha procedido a la transformación de estas desviaciones a otras normalmente distribuidas aplicando el siguiente algoritmo:

$$e_t = \frac{\ln \frac{@ALEAT}{1 - @ALEAT}}{1.7} \quad (4.3)$$

Esta expresión matemática es una función logística que, gracias al uso de una constante adicional ($D=1.7$), logra que sus valores se aproximen notablemente a los de la curva normal (Muñiz, 1990). El hecho de que se haya optado por esta función al generar los errores se debe a que, respecto a la normal, con ella es más fácil de operar matemáticamente.

El programa completo calcula, de cada serie temporal simulada (en este caso fase A), la media, la variancia, el estadístico MCDS, el e.e., los estadísticos C y Z, la significación estadística ($p < 0.05$ o $p < 0.01$) y, finalmente, la proporción de significaciones positivas de cada combinación de las distintas condiciones experimentales, para $\alpha=0.05$ y $\alpha=0.01$.

Con la finalidad de ir generando las series, se ha especificado, en la hoja de cálculo, el comando *Hoja Global Recálculo (/HGR)* y una opción manual que computa nuevamente todas las fórmulas al pulsar la tecla [F9], con lo que para generar una nueva serie, con sus correspondientes cálculos, se tendría que pulsar cada vez [F9], lo cual resultaría muy lento y repetitivo. Con objeto de agilizar el trabajo y ahorrar tiempo se ha creado una macro que permite generar automáticamente, para cada modelo lineal de la regresión, todas las simulaciones especificando únicamente la longitud de la serie.

Los análisis de la variancia (AVAR), pertinentes a cada estudio, se han llevado a cabo mediante el paquete estadístico *SPSS/PC+*, versión 4.0 (Norusis, 1990).

4.5.2. Resultados y discusión

En este apartado, siguiendo la afirmación de Tryon (1982) según la cual una serie es estable cuando el estadístico C es no significativo, se muestran las tasas de error Tipo I cuando la línea base es estable (tablas 4.1), y de error Tipo II cuando la fase de

pre-intervención presenta distintas amplitudes de variabilidad aleatoria no sistemática (tabla 4.2). Estos resultados se han obtenido calculando, respectivamente, el porcentaje de rechazos erróneos y aceptaciones incorrectas de la H_0 , con niveles de significación 0.05 y 0.01, tras aplicarse el estadístico C en las series simuladas.

Tabla 4.1. Tasas de error Tipo I obtenidas mediante el estadístico C: series de línea base estable.

n_1	$\alpha=0.05$	$\alpha=0.01$
8	0.0500	0.0075
9	0.0575	0.0100
10	0.0275	0.0050
11	0.0525	0.0075
12	0.0425	0.0100
13	0.0325	0.0025
14	0.0675	0.0125
15	0.0375	0.0075

Tabla 4.2. Tasas de error Tipo II obtenidas mediante el estadístico C: series de línea base variable.

n_1	Variabilidad baja (amplitud 4)		Variabilidad media (amplitud 10)		Variabilidad alta (amplitud 20)	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$
8	0.9350	0.9950	0.9375	0.9850	0.9475	0.9850
9	0.9425	0.9925	0.9475	0.9975	0.9500	0.9925
10	0.9575	0.9950	0.9400	0.9775	0.9550	0.9950
11	0.9600	0.9925	0.9450	0.9775	0.9600	0.9925
12	0.9325	0.9825	0.9350	0.9825	0.9600	0.9875
13	0.9450	0.9925	0.9600	0.9825	0.9600	0.9950
14	0.9325	0.9925	0.9375	0.9850	0.9775	0.9900
15	0.9375	0.9775	0.9425	0.9925	0.9425	0.9950

Como se puede apreciar en la tabla 4.1, los valores empíricos de α son estimaciones de los correspondientes valores nominales, siendo las pequeñas fluctuaciones consecuencia del azar (por ejemplo, para $n_1=8$ y $\alpha=0.05$, $\hat{\alpha} =0.05$ y para $\alpha=0.01$, $\hat{\alpha} =0.0075$). Los resultados hallados indican que cuando las series son estables, el estadístico C no es significativo. Por otro lado, en la tabla 4.2 se observa que, si se utiliza el concepto de estabilidad en sentido estricto, la aplicación de la prueba C produce tasas de error Tipo II elevadas cuando en la serie hay variabilidad ($\beta \geq 0.9325$ para $\alpha=0.05$ y $\beta \geq 0.9775$ para $\alpha=0.01$), independientemente de si la variabilidad es baja, media o alta. Consecuentemente, el estadístico C no discrimina entre una fase estable y una variable. Este hecho se constata al observar la baja potencia de C para fases variables, en ausencia de tendencia (tabla 4.3): para una variabilidad baja y $\alpha=0.05$, la potencia estadística oscila entre 0.0400 y 0.0675; para una variabilidad baja y $\alpha=0.01$, entre 0.0050 y 0.0225; para una variabilidad media y $\alpha=0.05$, entre 0.0400 y 0.0650; para una variabilidad media y $\alpha=0.01$, entre 0.0025 y 0.0225; para una variabilidad alta y $\alpha=0.05$, entre 0.0225 y 0.0575, y para una variabilidad alta y $\alpha=0.01$, entre 0.0050 y 0.0150.

Tabla 4.3. Potencia del estadístico C: series de línea base variables.

n ₁	Variabilidad baja (amplitud 4)		Variabilidad media (amplitud 10)		Variabilidad alta (amplitud 20)	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$
8	0.0650	0.0050	0.0625	0.0150	0.0525	0.0150
9	0.0575	0.0075	0.0525	0.0025	0.0500	0.0075
10	0.0425	0.0050	0.0600	0.0225	0.0450	0.0050
11	0.0400	0.0075	0.0550	0.0225	0.0400	0.0075
12	0.0675	0.0175	0.0650	0.0175	0.0400	0.0125
13	0.0550	0.0075	0.0400	0.0175	0.0400	0.0050
14	0.0675	0.0075	0.0625	0.0150	0.0225	0.0100
15	0.0625	0.0225	0.0575	0.0075	0.0575	0.0050

A la vista de los resultados obtenidos se concluye que cuando el estadístico C es no significativo, no resulta adecuado afirmar que la serie es estable; en todo caso, tan sólo es posible sostener que un conjunto de datos ordenados se comporta de acuerdo con una expectativa al azar. Un intento de explicar estos resultados podría ser el siguiente: sea la expresión matemática del estadístico C (fórmula 2.4; p. 51), el denominador (sumatorio de las desviaciones cuadráticas multiplicado por dos) recoge tanto la presencia de tendencia como de variabilidad en la serie. Ante una tendencia, el denominador aumenta más rápidamente que el numerador (sumatorio de las diferencias sucesivas cuadráticas), con lo que el valor del estadístico C tiende a ser alto. En cambio, en caso de existir variabilidad en la serie, tanto el denominador como el numerador aumentan rápidamente, con lo que el valor de la fracción de la ecuación 2.4 es próximo a la unidad y, en consecuencia, el estadístico C es pequeño. Este hecho explica el porqué C no es capaz de detectar variabilidad en la serie.

En resumen, la no significación del estadístico C puede ser debida tanto a la estabilidad de la serie como a la existencia de variabilidad. Por este motivo, ante la ausencia de significación estadística, antes de afirmar que la serie es estable, se recomienda tener en cuenta alguna medida de variabilidad como puede ser la desviación media, la desviación estándar, la variancia, o bien el estadístico δ^2 (MCDS). De este modo, se podría saber si la no significación de C se debe a que la serie es estable o, por el contrario, variable.

Con objeto de hallar en qué medida el estadístico C es capaz de detectar una tendencia en la serie, se han construido las tablas 4.4 y 4.5. En ellas se muestra (para $\alpha=0.05$ y $\alpha=0.01$, respectivamente) el límite -línea discontinua- a partir del cual la potencia del estadístico C es superior o igual a 0.80 (cifra recomendada por la mayoría de metodólogos). Lo que básicamente aportan estas tablas es el número de registros necesarios para una potencia dada. Por ejemplo, para $\alpha=0.05$ y un TE medio (como puede ser una tendencia de -0.4) se requiere un tamaño muestral de 12 (tabla 4.4), y para $\alpha=0.01$, la muestra necesaria es de 14 observaciones (tabla 4.5). De esta forma, se establece que la relación entre tamaño muestral y potencia de la prueba C es determinista (a mayor número de observaciones, mayor potencia).

Los resultados obtenidos no son alentadores para efectos pequeños. En concreto, en la tabla 4.4 se tiene que para pendientes de -0.1 y -0.2 no existe ni un 60% de posibilidades de detectar tal tendencia, cualquiera que sea la muestra. Con tamaños del efecto grandes (pendientes de -0.7, -0.8, -0.9 y -1), son suficientes ocho registros para proporcionar una potencia igual o mayor de 0.80.

La tabla 4.5 muestra los resultados para $\alpha=0.01$ y se observa una reducción de la potencia estadística de C. Considérese, por ejemplo, para $n_1=9$ y pendiente de -0.2: $\beta/\alpha=0.95/0.01=95$, es decir, el riesgo de rechazar equivocadamente la H_0 en las condiciones establecidas es 95 veces más importante que el de aceptarla erróneamente. En

cambio, para $\alpha=0.05$ la potencia es igual a 0.175 y $\beta/\alpha = 0.825/0.05 = 16.5$, con lo que la posibilidad de rechazar la H_0 cuando es verdadera es 16.5 veces de mayor importancia que aceptarla cuando es falsa.

Un aspecto importante a destacar es que cuando la potencia estadística es baja, la probabilidad de rechazar una H_0 verdadera puede ser tan sólo un poco más pequeña que la probabilidad de rechazar la H_0 cuando la H_1 es verdadera. Por ejemplo, en la tabla 4.4 se observa que para una tendencia de -0.1, $n_1=10$ y $\alpha=0.05$, la potencia es de 0.085. Esto significa que la tasa de error Tipo I con respecto a la potencia es muy grande y, en consecuencia, una proporción sustancial de resultados significativos se puede deber a falsos rechazos de la H_0 . Por este motivo, la baja potencia estadística pone en duda tanto los resultados estadísticamente significativos como los no significativos.

Tabla 4.4. Potencia del estadístico C: series de línea base con tendencia y $\alpha=0.05$.

n_1	Tendencia									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8	0.050	0.130	0.200	0.395	0.580	0.720	0.800	0.875	0.930	0.925
9	0.055	0.175	0.290	0.495	0.690	0.830	0.895	0.945	0.970	1.000
10	0.085	0.165	0.430	0.615	0.770	0.915	0.960	0.990	0.985	1.000
11	0.100	0.245	0.500	0.730	0.880	0.950	1.000	1.000	0.995	1.000
12	0.115	0.295	0.640	0.885	0.935	0.985	0.990	1.000	1.000	1.000
13	0.140	0.335	0.650	0.935	0.975	0.980	1.000	1.000	1.000	1.000
14	0.150	0.405	0.755	0.920	0.990	0.995	1.000	1.000	1.000	1.000
15	0.155	0.560	0.835	0.975	1.000	1.000	1.000	1.000	1.000	1.000

Tabla 4.5. Potencia del estadístico C: series de línea base con tendencia y $\alpha=0.01$.

n_t	Tendencia									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8	0.010	0.055	0.055	0.205	0.270	0.440	0.525	0.660	0.720	0.815
9	0.020	0.050	0.050	0.305	0.445	0.520	0.715	0.815	0.855	0.920
10	0.025	0.075	0.200	0.365	0.545	0.735	0.845	0.910	0.950	0.970
11	0.010	0.065	0.290	0.500	0.650	0.815	0.965	0.985	0.980	1.000
12	0.030	0.115	0.355	0.665	0.765	0.930	0.980	0.985	1.000	1.000
13	0.055	0.190	0.385	0.710	0.900	0.960	1.000	0.995	1.000	1.000
14	0.050	0.240	0.510	0.830	0.955	0.980	1.000	1.000	1.000	1.000
15	0.045	0.275	0.625	0.870	0.980	1.000	1.000	1.000	1.000	1.000

Para corroborar estadísticamente la relación funcional positiva entre el incremento de la potencia al aumentar la pendiente y el valor de α , es necesario bloquear una de las tres variables implicadas en el estudio. Esto se debe a que se trata de un diseño factorial con una sola unidad por casilla o condición experimental. La variable de bloqueo por la que se ha optado ha sido el tamaño muestral. Esta decisión es lógica, puesto que se conoce cómo influye el número de observaciones sobre la potencia y, por lo tanto, su estudio no resulta imprescindible. Por otra parte, el aspecto más importante en la presente investigación es el efecto de la tendencia sobre la potencia estadística de C. En total se han formado ocho bloques de acuerdo con el número de observaciones, y con el propósito de simplificar los cálculos se han seleccionado, dentro de cada bloque, cinco niveles de la variable tendencia (-0.1, -0.3, -0.5, -0.7 y -0.9).

Antes de iniciar el análisis de los resultados obtenidos, es importante destacar que la estrategia de diseño descrita anteriormente será adoptada, en este estudio y en los siguientes, de acuerdo con un modelo aditivo, es decir, ausencia de interacción bloques x tratamientos. Por lo tanto, el término de contraste será la variancia residual global. En todo caso, si los datos del experimento se ajustasen a un modelo no aditivo, se produciría una sobreestimación de la variancia del error, con lo cual el diseño perdería sensibilidad, dado

que la variancia de los tratamientos se contrastaría con una estimación positivamente sesgada del error. En consecuencia, se obtendría una razón F negativamente sesgada (Arnau, 1988). Este hecho influiría si el efecto de los tratamientos no fuera significativo. En tal caso sería preciso aplicar una prueba de no aditividad (Tukey, 1949), o bien una transformación de los datos que elimine los elementos no aditivos (Box, Hunter y Hunter, 1988). Este problema, propio de los diseños de bloques con una sola unidad por casilla, no se plantea en el presente trabajo. En efecto, como se verá a continuación, los tratamientos y sus interacciones son significativos.

Los resultados del AVAR revelan que tanto el nivel de significación como la tendencia son significativos: $F(1,63)=32.07$, $p<0.001$; $F(4,63)=262.30$, $p<0.001$, respectivamente. También resulta ser significativa la interacción entre estas dos variables: $F(4,63)=2.52$, $p<0.05$. En caso de que los resultados no fueran significativos y en el supuesto de la existencia de interacción bloques x tratamientos, una posible solución sería llevar a cabo la prueba de no aditividad de Tukey, y si ésta es positiva sería imprescindible calcular nuevamente las F con una variancia residual ajustada, es decir, se utilizarían términos de contraste parciales (por ejemplo, bloques x A, bloques x B, bloques x A x B). Aunque en el presente estudio no se da tal situación, también se ha llevado a cabo un análisis como si se tratase de un modelo no aditivo. Los resultados obtenidos son los siguientes: para la variable nivel de significación, $F(1,7)=58.61$, $p<0.001$; para la variable tendencia, $F(4,28)=141.27$, $p<0.001$, y para la interacción, $F(4,28)=8.97$, $p<0.001$. De esta forma se comprueba que cuando se asume un modelo aditivo y los resultados son significativos, no es necesario pasar a un modelo no aditivo, aun cuando se sospecha de la interacción bloques x tratamientos, ya que este último modelo también producirá resultados significativos. En los estudios posteriores sólo se analizará el modelo no aditivo cuando el aditivo no sea significativo.

Siguiendo con el análisis de los resultados obtenidos y a fin de interpretar el efecto de interacción, se transforman los valores de las tablas 4.4 y 4.5 en una tabla de potencias medias (tabla 4.6).

Tabla 4.6. Potencias medias de la interacción nivel de significación x tendencia.

Tendencia	Nivel de significación	
	Alfa=0.05	Alfa=0.01
-0.1	0.106	0.031
-0.3	0.538	0.309
-0.5	0.853	0.689
-0.7	0.956	0.879
-0.9	0.985	0.938

En la figura 4.1 se constata que, para un nivel de significación del 5%, se da una relación positiva entre potencia y tendencia. La línea inferior ($\alpha=0.01$) muestra el mismo tipo de relación. Esto significa que con pendientes altas la potencia es mayor. Por otra parte, cuando pendientes de -0.1, -0.7 y -0.9 se combinan con los niveles de significación, se produce una diferencia menor en la potencia media que cuando las tendencias toman valores de -0.3 y -0.5. En conclusión, la potencia para valores altos de tendencia es próxima a la unidad y para valores bajos es insignificante, independientemente del nivel de significación. En cambio, para tendencias intermedias la potencia es mayor con un error $\alpha=0.05$.

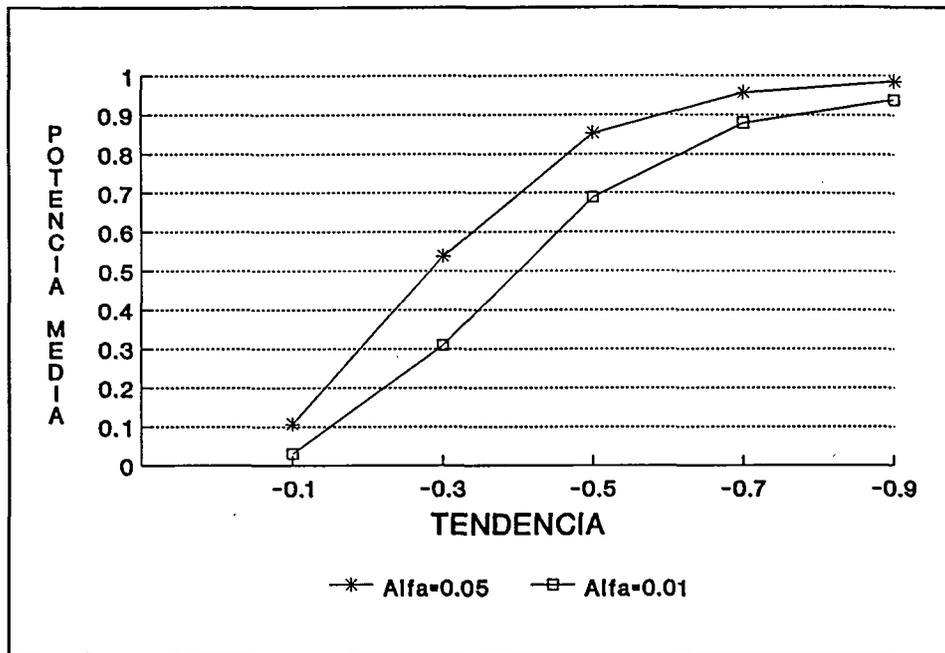


Figura 4.1. Efecto de α en función de los niveles de la tendencia.

4.6. Estudio 2

4.6.1. Método

Siguiendo el mismo proceso de simulación descrito en el estudio 1, se pone a prueba el estadístico C con objeto de evaluar la segunda estrategia de Tryon, considerada como la más fiable. Para ello, se han combinado, en diseños conductuales A-B, distintas longitudes de fase, pendientes en la fase de intervención y cambios de nivel.

Las series temporales generadas tienen una longitud que oscila entre $n_1=8$ y $n_2=8$ hasta $n_1=15$ y $n_2=15$, línea base estable, pendiente negativa en la fase de tratamiento que va desde -0.1 a -1, con incrementos de una décima, y cambios de nivel de cero y una desviación estándar ($\sigma=0$ y $\sigma=1$). El tamaño de la intervención de 1σ es muy común en ciencias socio-comportamentales, puesto que se conoce que no suele ser muy grande (Vallejo, 1993). Bajo estas condiciones se han generado un total de 32000 series de tiempo (8 longitudes de serie x 10 niveles de pendiente x 2 magnitudes de cambio de nivel x 200 repeticiones). Además, para la condición de no cambio de nivel, se han generado otras 32000, con la diferencia de que el número de réplicas por combinación es de 400 simulaciones. Con ello, al tratarse de un estudio con un número reducido de simulaciones, en comparación a otras investigaciones (Gottman, 1981; Matyas y Greenwood, 1991; Vallejo, 1993), se pretende ver la posible influencia de la cantidad de réplicas.

En suma, en este estudio se tratan los dos patrones de cambio que pueden darse, al pasar de una fase a la siguiente, en datos ordenados temporalmente con línea base estacionaria y tendencias de diferentes inclinaciones en la fase de intervención: a) aquellos casos caracterizados por ausencia de cambio de nivel, y b) aquellos casos donde se observa un cambio abrupto en el nivel de la serie. Las ecuaciones lineales correspondientes a los diferentes modelos de la regresión son las que se detallan a continuación.

Para aquellas situaciones con línea base estable y tendencia en la fase de intervención, el modelo estadístico es:

$$Y_t = b_0 + b_2X_tT + e_t \quad (4.4)$$

donde b_0 estima el punto de corte con el eje Y_t antes de la intervención; b_2 simboliza la tendencia de la fase de post-intervención; X_t es una variable *dummy* codificada dicotómicamente con 0 antes de la intervención y 1 después de la misma, y T define la unidad temporal concreta dentro de la secuencia de observaciones.

Cuando además de la tendencia en la fase de tratamiento existe un cambio de nivel, el modelo queda representado por la siguiente ecuación:

$$Y_t = b_0 + b_2X_tT + b_3X_t + e_t \quad (4.5)$$

donde b_3 simboliza el cambio de nivel, es decir, la diferencia entre el punto de corte con el eje Y_t anterior a la intervención y el posterior.

4.6.2. Resultados y discusión

El presente modelo de la regresión ha sido estudiado con anterioridad en una investigación donde las longitudes de serie oscilaban desde $n_1-n_2=4$ a $n_1-n_2=12$ (Bono y Arnau, 1993), obteniéndose resultados similares a los que se presentan a continuación.

Las tablas de potencia (tablas 4.7 y 4.8) ponen de relieve, al igual que en el estudio anterior, que las identificaciones correctas aumentan a medida que incrementa el número de observaciones, así como la pendiente de la fase de tratamiento.

Tabla 4.7. Potencia del estadístico C: series con línea base estable, pendiente negativa en la fase de tratamiento, $\sigma=0$ y $\alpha=0.05$.

n_1-n_2	Tendencia en la fase de tratamiento									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8-8	0.140	0.185	0.365	0.575	0.785	0.885	0.985	0.990	0.995	1.000
9-9	0.090	0.210	0.555	0.770	0.875	0.990	1.000	1.000	1.000	1.000
10-10	0.140	0.325	0.695	0.855	0.980	0.995	1.000	1.000	1.000	1.000
11-11	0.090	0.475	0.715	0.925	0.995	1.000	1.000	1.000	1.000	1.000
12-12	0.185	0.475	0.865	0.975	0.995	1.000	1.000	1.000	1.000	1.000
13-13	0.175	0.590	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14-14	0.175	0.705	0.985	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15-15	0.205	0.765	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tabla 4.8. Potencia del estadístico C: series con línea base estable, pendiente negativa en la fase de tratamiento, $\sigma=0$ y $\alpha=0.01$.

n_1-n_2	Tendencia en la fase de tratamiento									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8-8	0.050	0.045	0.195	0.335	0.585	0.760	0.910	0.910	0.985	0.985
9-9	0.015	0.105	0.295	0.520	0.720	0.945	0.935	0.985	1.000	1.000
10-10	0.055	0.185	0.425	0.630	0.885	0.980	1.000	0.995	1.000	1.000
11-11	0.030	0.235	0.500	0.820	0.950	0.990	1.000	1.000	1.000	1.000
12-12	0.045	0.245	0.655	0.875	0.990	1.000	1.000	1.000	1.000	1.000
13-13	0.055	0.355	0.745	0.960	1.000	1.000	1.000	1.000	1.000	1.000
14-14	0.045	0.395	0.875	0.995	1.000	1.000	1.000	1.000	1.000	1.000
15-15	0.085	0.515	0.925	0.990	1.000	1.000	1.000	1.000	1.000	1.000

En la tabla 4.7 se observa que, con una línea base estable, una pendiente negativa en la fase de tratamiento, sin cambio de nivel, $\alpha=0.05$ y una muestra de ocho observaciones por fase, la prueba C empieza a ser bastante potente con una tendencia de -0.5 (78.5% de confianza de que se ha adoptado la decisión correcta). Con un tamaño muestral de $n_1-n_2 = 9-9$, la tendencia a partir de la cual el estadístico C posee una potencia superior a 0.80 es -0.5; con $n_1-n_2 = 10-10$ y $n_1-n_2 = 11-11$ la tendencia es de -0.4 y para el resto de tamaños de muestra es de -0.3. En la tabla 4.8 el área en la cual es apropiado

aplicar el estadístico C, con una tasa de error Tipo II menor o igual al 20%, es aún más reducida, por ser el nivel de significación fijado del 1%.

Cuando, además, se añade un cambio de nivel de una desviación estándar (tablas 4.9 y 4.10), la potencia aumenta notablemente. Si, por ejemplo, comparamos la tabla 4.7 con la 4.9, se observa que esta última presenta una potencia superior a 0.80, cualquiera que sea su tamaño de muestra o pendiente en la fase de tratamiento.

Tabla 4.9. Potencia del estadístico C: series con línea base estable, pendiente negativa en la fase de tratamiento, $\sigma=1$ y $\alpha=0.05$.

n_1-n_2	Tendencia en la fase de tratamiento									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8-8	0.805	0.820	0.955	0.985	0.990	0.995	1.000	1.000	1.000	1.000
9-9	0.795	0.920	0.955	0.995	0.995	1.000	1.000	1.000	1.000	1.000
10-10	0.815	0.955	0.970	1.000	1.000	1.000	1.000	1.000	1.000	1.000
11-11	0.865	0.990	0.995	0.995	1.000	1.000	1.000	1.000	1.000	1.000
12-12	0.945	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
13-13	0.940	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14-14	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15-15	0.990	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tabla 4.10. Potencia del estadístico C: series con línea base estable, pendiente negativa en la fase de tratamiento, $\sigma=1$ y $\alpha=0.01$.

n_1-n_2	Tendencia en la fase de tratamiento									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8-8	0.535	0.660	0.820	0.900	0.950	0.970	0.990	0.995	1.000	1.000
9-9	0.590	0.775	0.855	0.955	0.980	0.990	1.000	1.000	1.000	1.000
10-10	0.615	0.875	0.905	0.990	1.000	1.000	1.000	1.000	1.000	1.000
11-11	0.690	0.995	0.980	0.995	1.000	1.000	1.000	1.000	1.000	1.000
12-12	0.800	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
13-13	0.820	0.965	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14-14	0.830	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15-15	0.960	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Una vez descritas las tablas de potencia, nos hallamos en la necesidad de corroborar estadísticamente las afirmaciones hechas anteriormente sobre el incremento de la potencia al aumentar la pendiente en la fase de tratamiento y/o al introducir un cambio de nivel. Con este propósito se ha llevado a cabo un AVAR con un bloqueo del tamaño muestral y, al igual que en el primer estudio, se han seleccionado cinco niveles de tendencia (-0.1, -0.3, -0.5, -0.7 y -0.9) para simplificar los cálculos.

Los resultados del AVAR indican que los efectos principales de cada variable (nivel de significación, tendencia en la fase de tratamiento y cambio de nivel) son significativos, con una probabilidad de error menor de 0.001. De todas las interacciones posibles, la más importante para nuestros propósitos es el efecto de σ x tendencia [$F(4,133)=106.4$, $p<0.001$]. Las representaciones gráficas de esta interacción -para $\alpha=0.05$ y $\alpha=0.01$ - obtenidas a partir de las tablas de potencias medias (tablas 4.11 y 4.12) se muestran en las figuras 4.2 y 4.3.

Tabla 4.11. Potencias medias de la interacción cambio de nivel x tendencia del tratamiento ($\alpha=0.05$).

Tendencia tratamiento	Cambio de nivel	
	Sigma=0	Sigma=1
-0.1	0.150	0.888
-0.3	0.770	0.984
-0.5	0.954	0.998
-0.7	0.998	1.000
-0.9	0.999	1.000

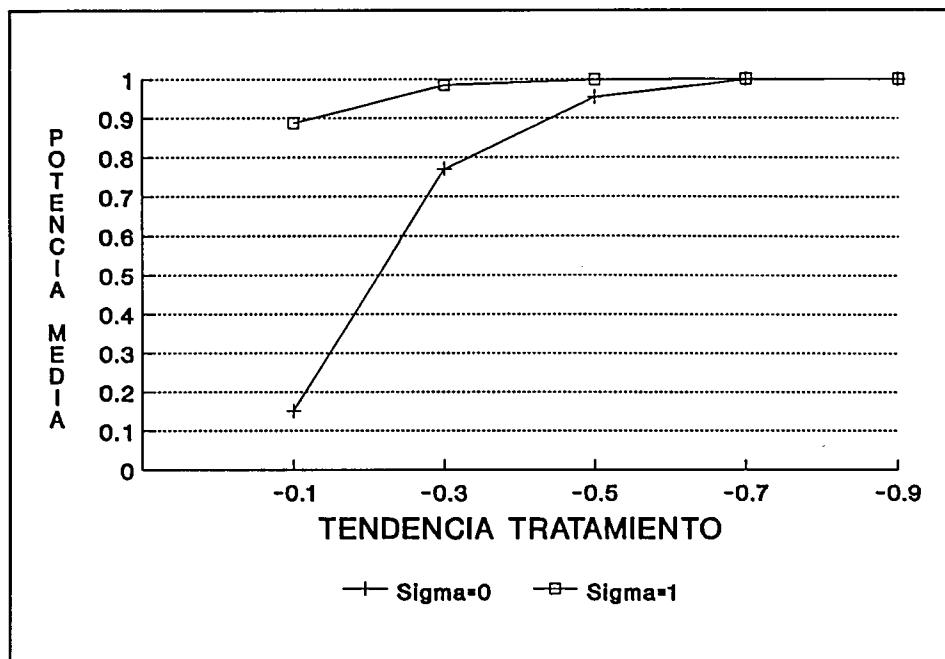


Figura 4.2. Interacción entre σ y tendencia en la fase de tratamiento ($\alpha=0.05$).

Tabla 4.12. Potencias medias de la interacción cambio de nivel x tendencia del tratamiento ($\alpha=0.01$).

Tendencia tratamiento	Cambio de nivel	
	Sigma=0	Sigma=1
-0.1	0.048	0.730
-0.3	0.577	0.944
-0.5	0.891	0.991
-0.7	0.981	0.999
-0.9	0.999	1.000

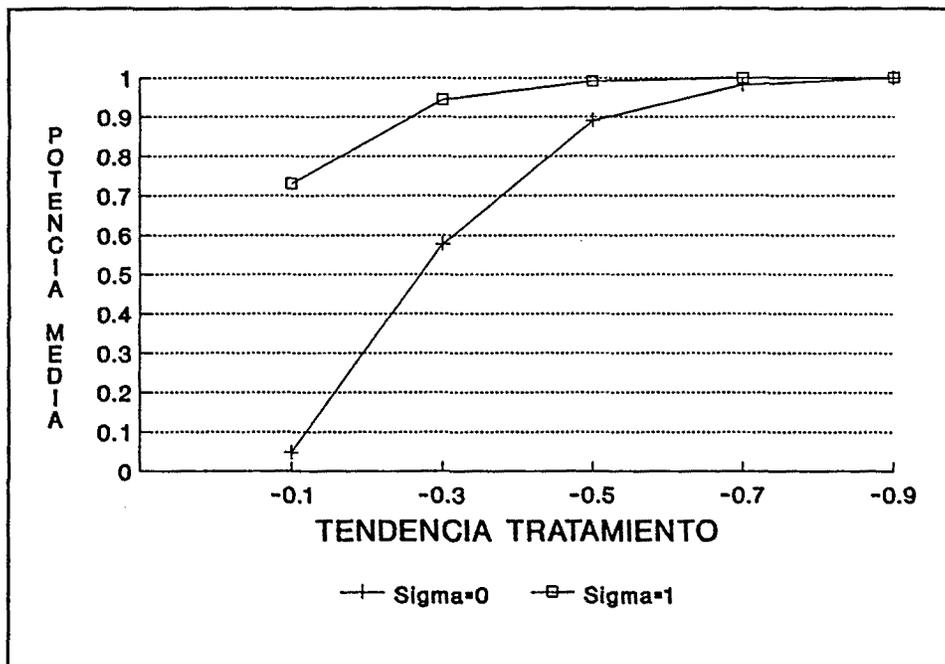


Figura 4.3. Interacción entre σ y tendencia en la fase de tratamiento ($\alpha=0.01$).

De la representación gráfica de las interacciones (figuras 4.2 y 4.3) se desprende una interacción negativa en el sentido de que cuando σ se combina con niveles bajos de tendencia (-0.1 y -0.3), se produce una mayor diferencia en la potencia media que cuando se combina con niveles altos (-0.7 y -0.9). Por lo tanto, se concluye que, para aquellas condiciones en que hay una fase de pre-intervención estable y una tendencia negativa (baja o media) en la fase de tratamiento, el estadístico C es sensible a los cambios de nivel; aunque, incluso estos sean relativamente pequeños ($\sigma=1$). La explicación que se da a este hecho es que al aumentar el tamaño del efecto la potencia incrementa notablemente.

Para finalizar, debido a que el presente estudio posee un número de simulaciones un tanto reducido (200 por combinación), en comparación con otros de características similares, se ha duplicado el número de simulaciones para las condiciones simuladas en las tablas 4.7 y 4.8 (ver tablas 4.13 y 4.14). De este modo, se analizan la diferencias significativas en función de la cantidad de simulaciones.

Tabla 4.13. Potencia del estadístico C: series con línea base estable, pendiente negativa en la fase de tratamiento, $\sigma=0$ y $\alpha=0.05$ (400 simulaciones por combinación).

n_1-n_2	Tendencia en la fase de tratamiento									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8-8	0.0750	0.1825	0.3725	0.5675	0.7825	0.8875	0.9475	0.9775	1.0000	0.9975
9-9	0.0800	0.2250	0.4350	0.7450	0.9275	0.9750	0.9950	1.0000	1.0000	1.0000
10-10	0.1175	0.3075	0.6175	0.8550	0.9575	1.0000	1.0000	1.0000	1.0000	1.0000
11-11	0.1375	0.3900	0.7550	0.9125	0.9975	0.9975	1.0000	1.0000	1.0000	1.0000
12-12	0.1600	0.4375	0.8400	0.9700	0.9975	1.0000	1.0000	1.0000	1.0000	1.0000
13-13	0.1875	0.5950	0.9300	0.9950	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14-14	0.1975	0.6775	0.9700	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15-15	0.2550	0.7725	0.9850	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tabla 4.14. Potencia del estadístico C: series con línea base estable, pendiente negativa en la fase de tratamiento, $\sigma=0$ y $\alpha=0.01$ (400 simulaciones por combinación).

n_1-n_2	Tendencia en la fase de tratamiento									
	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.8	-0.9	-1
8-8	0.0250	0.0825	0.1950	0.3600	0.5725	0.7600	0.8600	0.9050	0.9775	0.9825
9-9	0.0200	0.0700	0.2175	0.4825	0.7500	0.9125	0.9725	0.9925	0.9975	0.9975
10-10	0.0425	0.1525	0.3675	0.6375	0.8425	0.9800	0.9900	1.0000	1.0000	1.0000
11-11	0.0375	0.1925	0.5125	0.8075	0.9650	0.9875	1.0000	1.0000	1.0000	1.0000
12-12	0.0500	0.2250	0.6675	0.8925	0.9850	0.9975	1.0000	1.0000	1.0000	1.0000
13-13	0.0625	0.3200	0.7825	0.9600	0.9975	1.0000	1.0000	1.0000	1.0000	1.0000
14-14	0.0800	0.4175	0.8950	0.9825	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15-15	0.0800	0.5400	0.9350	0.9950	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Un análisis de los efectos simples, o sea, la acción que ejerce la cantidad de simulaciones sobre la potencia, para cada tamaño muestral (tablas 4.15 y 4.17 con valores de $\alpha=0.05$ y $\alpha=0.01$, respectivamente) y para cada valor que toma la tendencia en la fase de tratamiento (tablas 4.16 y 4.18 con valores de $\alpha=0.05$ y $\alpha=0.01$, respectivamente), demuestra que las tablas correspondientes a 200 simulaciones (tablas 4.7 y 4.8) no difieren estadísticamente de las tablas correspondientes a 400 simulaciones (tablas

4.13 y 4.14), tanto si se fija la pendiente (tablas 4.15 y 4.17) como el tamaño muestral (tablas 4.16 y 4.18). Así pues, se concluye que el número de réplicas no afecta a la potencia resultante.

Tabla 4.15. Significación de los efectos simples de VC (tablas 4.7 y 4.13) para cada nivel de VA.

Fuente de variación	Sig. de F
VC DENTRO VA(1)	0.933
VC DENTRO VA(2)	0.938
VC DENTRO VA(3)	0.921
VC DENTRO VA(4)	0.994
VC DENTRO VA(5)	0.948
VC DENTRO VA(6)	0.972
VC DENTRO VA(7)	0.988
VC DENTRO VA(8)	0.969

VA = tamaño muestral.

VC = cantidad de simulaciones.

$\alpha=0.05$

Tabla 4.16. Significación de los efectos simples de VC (tablas 4.7 y 4.13) para cada nivel de VB.

Fuente de variación	Sig. de F
VC DENTRO VB(1)	0.994
VC DENTRO VB(2)	0.908
VC DENTRO VB(3)	0.836
VC DENTRO VB(4)	0.964
VC DENTRO VB(5)	0.979
VC DENTRO VB(6)	0.994
VC DENTRO VB(7)	0.973
VC DENTRO VB(8)	0.992
VC DENTRO VB(9)	0.997
VC DENTRO VB(10)	0.998

VB = tendencia en la fase de tratamiento.
 VC = cantidad de simulaciones.
 $\alpha=0.05$

Tabla 4.17. Significación de los efectos simples de VC (tablas 4.8 y 4.14) para cada nivel de VA.

Fuente de variación	Sig. de F
VC DENTRO VA(1)	0.980
VC DENTRO VA(2)	0.946
VC DENTRO VA(3)	0.929
VC DENTRO VA(4)	0.854
VC DENTRO VA(5)	0.774
VC DENTRO VA(6)	0.996
VC DENTRO VA(7)	0.967
VC DENTRO VA(8)	0.982

VA = tamaño muestral.
 VC = cantidad de simulaciones.
 $\alpha=0.01$

Tabla 4.18. Significación de los efectos simples de VC (tablas 4.8 y 4.14) para cada nivel de VB.

Fuente de Variación	Sig. de F
VC DENTRO VB(1)	0.890
VC DENTRO VB(2)	0.955
VC DENTRO VB(3)	0.976
VC DENTRO VB(4)	0.996
VC DENTRO VB(5)	0.990
VC DENTRO VB(6)	0.979
VC DENTRO VB(7)	0.987
VC DENTRO VB(8)	0.996
VC DENTRO VB(9)	0.994
VC DENTRO VB(10)	0.997

VB = tendencia en la fase de tratamiento.

VC = cantidad de simulaciones.

$\alpha=0.01$

4.7. Estudio 3

4.7.1. Método

Con objeto de evaluar la tercera estrategia de Tryon se ha seguido el mismo proceso de simulación aplicado en los estudios precedentes.

Mediante simulación de Monte Carlo, se han generado series temporales interrumpidas con longitudes desde $n_1=8$ y $n_2=8$ a $n_1=15$ y $n_2=15$, a las que se ha aplicado los tres métodos de análisis propuestos por Tryon (comparación de series utilizando la técnica de la regresión, comparación de series con datos directos y análisis del conjunto de las fases) de acuerdo a las siguientes combinaciones (200 réplicas para cada una de ellas):

- 1) Longitud de la serie (ocho niveles), tendencia lineal en la línea base y en la fase de tratamiento con pendiente igual $(-0.1(0.10)-1)$, y cambio de nivel en el punto de intervención ($\sigma=0$ y $\sigma=1$).
- 2) Longitud de la serie (ocho niveles), tendencia en la línea base (0.2, 0.4, 0.6, y 0.8), cambio de tendencia en la fase de tratamiento $(-0.1(0.10)-1)$, y cambio de nivel ($\sigma=0$ y $\sigma=1$).
- 3) Longitud de la serie (ocho niveles), tendencia en la fase de pre-intervención de -0.1, cambio de tendencia en la fase de tratamiento $(-0.2(0.10)-1)$, y sin cambio de nivel.

- 4) Longitud de la serie (ocho niveles), tendencia en la fase de línea base de -0.2, cambio de tendencia en la fase de intervención (-0.3(0.10)-1), y sin cambio de nivel.

- 5) Longitud de la serie (ocho niveles), tendencia en la fase de línea base de -0.3, cambio de tendencia en la fase de tratamiento (-0.4(0.10)-1), y sin cambio de nivel.

El número de simulaciones obtenidas del primer grupo de combinaciones experimentales es de 96000; 384000 del segundo grupo, y un total de 115200 de los restantes. Los patrones de cambio presentados en este experimento pueden reducirse, básicamente, a cambios de tendencia y cambios de nivel, en aquellos casos donde la línea conductual de pre-intervención es no estacionaria. Dentro del primer grupo de combinaciones se estudian aquellas situaciones en que la línea experimental sigue la misma dirección y pendiente iniciada en la fase previa, y aquellas en que se añade, además, un cambio de nivel. En un segundo grupo, se analizan aquellos patrones de cambio en los que las observaciones reflejan un cambio de pendiente en sentido opuesto, antes y después de la intervención, y aquellos donde se produce simultáneamente un cambio de dirección de la tendencia, en sentido opuesto, y un cambio abrupto de nivel. Por último, se estudian una serie de situaciones donde la línea de pre-intervención presenta una moderada pendiente que luego, después de la aplicación del tratamiento, pasa a ser más acusada. En este último caso no se analiza ningún cambio de nivel, puesto que este tipo de patrones de cambio lo que hacen es acelerar la tendencia iniciada en la fase de línea base (Arnau, 1986).

Los modelos lineales de la regresión seleccionados pueden representarse a través de la siguiente ecuación general, que contempla tanto el cambio de nivel como el cambio de tendencia

$$Y_t = b_0 + b_1T + b_2X_tT + b_3X_t + e_t \quad (4.6)$$

donde T simboliza la unidad temporal de cada observación; X_t es una variable ficticia dicotómica con valores 0 para las observaciones anteriores a la intervención y 1 para las observaciones posteriores. De acuerdo con el modelo de la ecuación 4.6, b_0 estima el punto de corte con el eje Y_t antes de la intervención, b_1 representa la tendencia de la fase de pre-intervención, b_2 es un parámetro que simboliza el cambio de tendencia durante el período post-intervención y b_3 representa el cambio de nivel (figura 4.4).

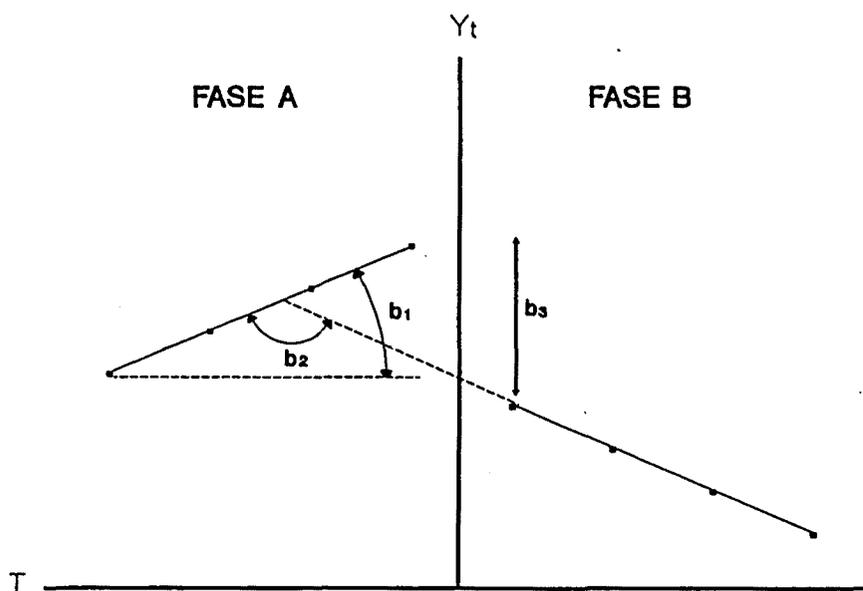


Figura 4.4. Representación gráfica de la ecuación 4.6, donde b_1 = Tendencia de la línea base; b_2 = Cambio de tendencia, y b_3 = Cambio de nivel. (Modificada de Simonton, 1977; p. 493.)

4.7.2. Resultados y discusión

En este estudio se han analizado tres patrones de cambio que se ajustan a la tercera estrategia de Tryon. Para la descripción de los resultados se seguirá el orden establecido en el método. Así, en primer lugar, se estudian aquellos patrones que presentan en la fase de tratamiento la misma pendiente que se inicia en la fase anterior y, en segundo lugar, aquellos en los que las tendencias tienen inclinaciones opuestas. Por último, se lleva a cabo el análisis de un patrón con una moderada pendiente en la línea base que luego pasa a ser más acusada, después de la intervención.

El número total de tablas de resultados es de 78 por lo que no se incluyen en este apartado sino que se hará referencia, a lo largo de la discusión, al apéndice 10 que recoge el conjunto de todas las tablas. De estas tablas tan sólo se analizan las que corresponden a una $\alpha=0.05$. El motivo de ello, aparte del de reducir los cálculos estadísticos, es que en la mayoría de situaciones es más prudente no utilizar un nivel de significación del 1% y optar por un nivel más liberal, a fin de mejorar la potencia. Naturalmente, se tiene un riesgo mayor de rechazar falsamente la H_0 , pero este riesgo está más equilibrado con un error de Tipo II.

a) *Igual tendencia entre fases, con y sin cambio de nivel.*

Como se ha indicado en el capítulo anterior, el estadístico C no detecta un cambio de nivel cuando se lleva a cabo una comparación de series en aquellos casos donde tanto la fase de línea base como la de tratamiento poseen la misma tendencia. La solución que propone Tryon al respecto es aplicar el estadístico C al conjunto de las fases. Sin embargo, este método de análisis presenta un inconveniente y es que la prueba C resulta significativa cuando la tendencia se prolonga en la fase de tratamiento y no existe ningún cambio de nivel en el punto de intervención. Para el estudio de esta limitación se han obtenido las

identificaciones del estadístico C para los tres métodos de análisis (comparación de series utilizando la técnica de la regresión, comparación de series con datos directos y análisis del conjunto de las fases), en presencia y ausencia de cambio de nivel (tablas 1-12; apéndice 10). A fin de simplificar el análisis de los resultados, se han seleccionado pendientes de -0.1, -0.3, -0.5, -0.7 y -0.9.

Antes de comentar los resultados, debe advertirse que, cuando $\sigma=0$, los porcentajes de significaciones del estadístico C corresponden a errores de Tipo I y cuando $\sigma=1$, corresponden a la potencia de la prueba C. Estos porcentajes promediados por las 200 replicaciones por combinación se muestran en la tabla 4.19. Se observa que el análisis del conjunto de las fases (A_3) presenta un elevado porcentaje de identificaciones correctas cuando $\sigma=1$ (99.1%), pero cuando $\sigma=0$ la tasa de error Tipo I es del 87.2%. Las mayores dificultades surgen cuando los métodos de comparación de series (A_1 y A_2) se combinan con $\sigma=1$.

Tabla 4.19. Porcentajes medios de pruebas C significativas de la interacción método de análisis x cambio de nivel (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Cambio de nivel	Método de análisis		
	A_1	A_2	A_3
Sigma=0	4.9	5.0	87.2
Sigma=1	5.1	5.2	99.1

El AVAR muestra la existencia de un efecto significativo en la interacción método de análisis x σ [$F(2,203)=87$, $p<0.001$]. La representación gráfica de tal interacción se presenta en la figura 4.5.

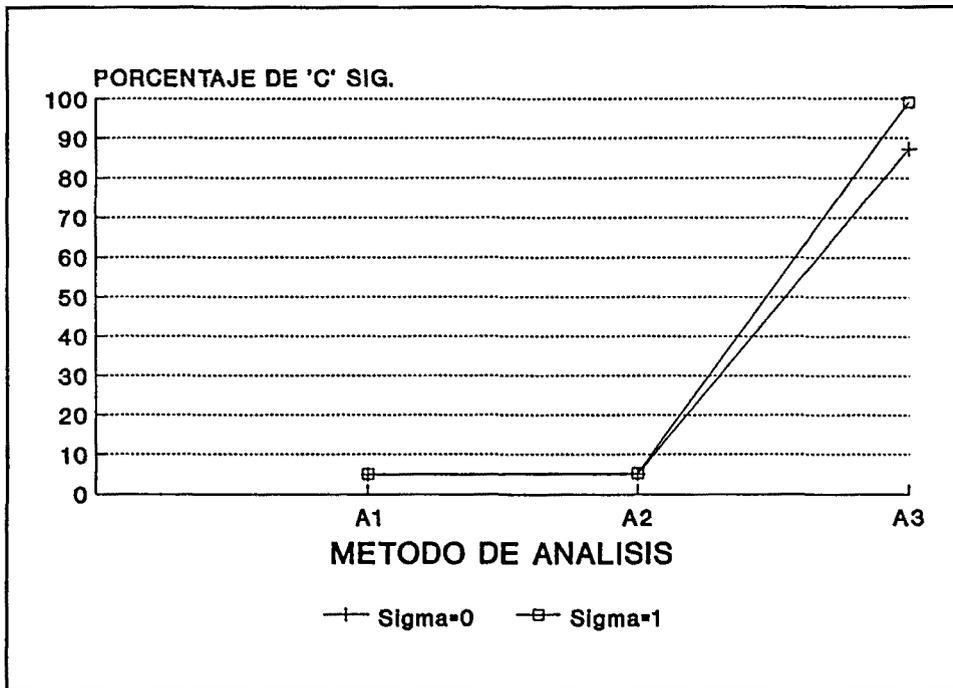


Figura 4.5. Interacción entre σ y método de análisis (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Los dos métodos de análisis de comparación de series (A_1 y A_2), independientemente del valor de σ , son similares tanto si se sigue un modelo aditivo [$F(1,133)=0.266$, $p > 0.05$] como no aditivo [$F(1,17)=0.001$, $p > 0.05$]. En la figura 4.5 puede observarse que tanto la potencia como el error de Tipo I tienen valores muy bajos cuando se aplica alguna técnica de comparación de series. En cambio, el error de Tipo I aumenta considerablemente con un análisis del conjunto de las fases (A_3) y, en mayor medida, la potencia.

A partir de las tablas del apéndice 10, se obtiene que el valor medio de error Tipo I (caso $\sigma=0$) para las técnicas de comparación de series es de 0.049 (tablas 1 y 5) y la potencia media (caso $\sigma=1$) es de 0.053 (tablas 3 y 7). Por otra parte, el análisis del conjunto de las fases identifica correctamente el efecto de tratamiento cuando hay un

cambio abrupto de nivel, con una potencia media de 0.995 (tabla 11). Pero si no existe ningún cambio de nivel, la tasa de error Tipo I tiene una media de 0.92 (tabla 9). Por lo tanto, es preciso tener mucha cautela al inferir el efecto de tratamiento, puesto que éste dependerá de si el método de análisis adoptado es el correcto, dependiendo del valor de σ .

b) *Tendencias opuestas entre fases, con y sin cambio de nivel.*

La potencia de este segundo patrón de cambio, para cada una de las combinaciones, se recoge en el apéndice 10 (tablas 13-60). Al igual que en los estudios anteriores, ésta se basa en el rechazo de la H_0 en favor de la H_1 .

Bajo el supuesto de que se ajuste una recta de regresión a los datos de línea base y las tendencias sean pequeñas, tanto las de la línea base como las de la fase de tratamiento, se observa que para un valor $\sigma=0$ la muestra necesaria para obtener una potencia grande (superior o igual a 0.80) oscila entre 12 y 15 observaciones por fase. Por ejemplo, para una pendiente de línea base de 0.2 y una de tratamiento de -0.1 (tabla 13) se requiere una muestra de 15 registros por fase (167 pruebas de 200, 83.5%). Con una pendiente en la fase de intervención de -0.2, son suficientes 12 observaciones por fase (tabla 13). Respecto al resto de tendencias en la línea base (0.4, 0.6 y 0.8) la potencia aumenta considerablemente, sobre todo cuando la tendencia en la fase de tratamiento es inferior a -0.5 (ver tablas 15, 17 y 19). Obsérvese que en las tablas 17 y 19 del apéndice 10 la potencia es mayor o igual a 0.80, cualquiera que sea la tendencia en la fase de tratamiento.

En cuanto a los métodos de análisis restantes, cabe destacar que cuando se lleva a cabo una comparación de series con datos directos (tablas 21-28) se necesita una n de mayor tamaño que en el caso anterior. Las tablas correspondientes al análisis del conjunto

de las fases (tablas 29-36) presentan una potencia para una tendencia de 0.2 en la línea base e inclinaciones en la fase de post-intervención de -0.1, -0.2 y -0.3 inferior a la hallada por los otros dos métodos y, a partir de -0.5, se observa un cambio en el sentido que las diferencias van aminorando. Además, al aumentar el valor de la tendencia en la de fase de pre-tratamiento se detecta que la potencia también incrementa considerablemente, llegando a ser superior a las otras técnicas de análisis cuando la tendencia es de 0.8.

Hasta aquí se han descrito los resultados obtenidos cuando $\sigma=0$, considérese a continuación qué ocurre cuando aparece un cambio de nivel de una desviación estándar al introducir el tratamiento. En este último caso, se constatan resultados similares a los anteriores en los dos métodos de comparación de series. En cambio, para el análisis del conjunto de las fases la potencia presenta un notable incremento cuando la muestra es pequeña y la tendencia en la línea base es de 0.2. Para el resto de tendencias en la fase de pre-intervención (0.4, 0.6 y 0.8) es más eficaz la comparación de series utilizando la técnica de la regresión (ver tablas 37-60, apéndice 10).

Antes de continuar queremos señalar que el análisis estadístico de los resultados se ciñe a tendencias de línea base de 0.2 y 0.6 y pendientes de tratamiento de -0.1, -0.3, -0.5, -0.7 y -0.9, pues si bien se han generado las series para un mayor número de tendencias, de esta forma se simplifican notablemente los análisis dando una idea más clara de los resultados de la investigación. Por otra parte, recordamos que sólo nos centraremos en las tablas resultantes de un valor $\alpha=0.05$.

A partir de los resultados del AVAR se comprueba que el efecto de las variables método de análisis y cambio de nivel es significativo: $F(2,413)=146.37$, $p<0.001$; $F(1,413)=15.68$, $p<0.001$, respectivamente. Por lo que respecta a la interacción, el efecto también resulta significativo: $F(2,413)=20.91$, $p<0.001$. En la interpretación de este efecto secundario es importante resaltar que cuando se analizan únicamente los métodos de comparación de series, éstos no difieren significativamente para el modelo aditivo

[$F(1,273)=0.18$, $p>0.05$] ni para el modelo no aditivo [$F(1,7)=0.57$, $p>0.05$]. Por lo tanto, las diferencias significativas se deben al análisis del conjunto de las fases.

En la tabla 4.20 se presentan las potencias medias de la interacción método de análisis x cambio de nivel infiriéndose que los niveles de potencia son mayores para el método de la regresión (A_1). Para el análisis del conjunto de las fases la potencia aumenta al presentarse un cambio de nivel de una desviación estándar.

Tabla 4.20. Potencias medias de la interacción método de análisis x cambio de nivel (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Cambio de nivel	Método de análisis		
	A_1	A_2	A_3
Sigma=0	0.924	0.845	0.872
Sigma=1	0.926	0.844	0.924

De acuerdo con los valores promedio de la tabla anterior, la figura 4.6 representa la interacción existente entre método de análisis y cambio de nivel. En ella puede verificarse que, en general, la comparación de series con datos directos (A_2) tiene más dificultad en identificar una prueba C significativa y que la diferencia entre $\sigma=0$ y $\sigma=1$ se produce en el análisis del conjunto de las fases (A_3).

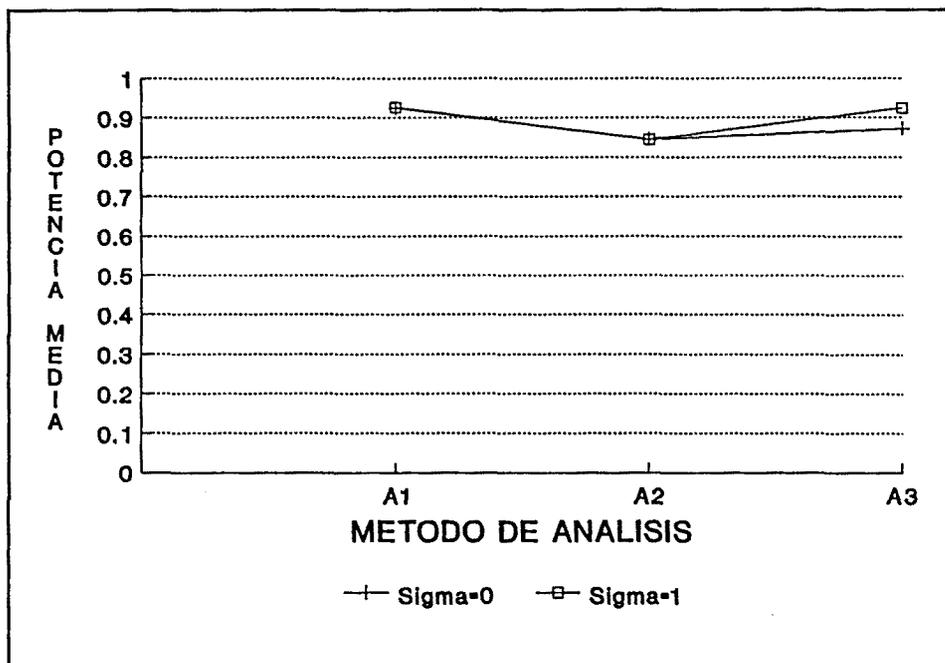


Figura 4.6. Interacción entre método de análisis y cambio de nivel (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Por lo que respecta a la interacción entre método de análisis y tendencia en la línea base ha resultado ser significativa [$F(2,273)=12.49$, $p<0.001$]. La tabla 4.21 pone de relieve el efecto de la pendiente en la fase de pre-intervención. También se observa que el método de análisis más poderoso para detectar la significación del estadístico C es el de la regresión (A_1).

Tabla 4.21. Potencias medias de la interacción método de análisis x tendencia en la línea base (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Tendencia línea base	Método de análisis		
	A_1	A_2	A_3
0.2	0.863	0.741	0.831
0.6	0.986	0.947	0.964

En la figura 4.7 se muestra la representación geométrica de las potencias medias resultantes. De la observación de dicha figura se infiere que, en general, a mayor pendiente en la línea base (B_2), mayor potencia. Cabe destacar la menor efectividad del factor A_2 (comparación de series con datos directos) cuando la tendencia de la fase de pre-intervención es de 0.2 (B_1).

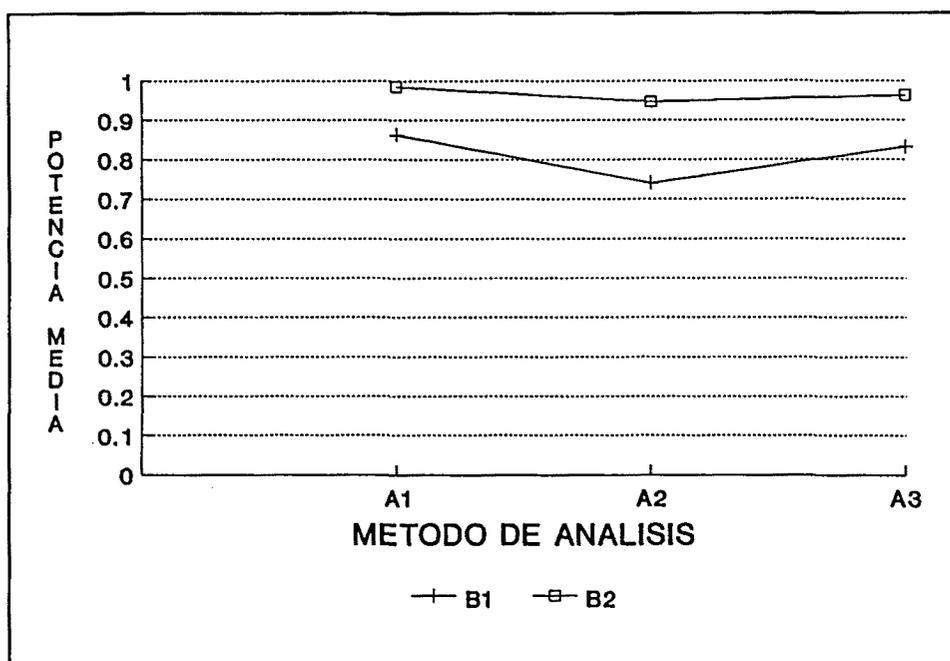


Figura 4.7. Interacción entre método de análisis y tendencia en la línea base (A₁=Comparación de series utilizando la técnica de la regresión; A₂=Comparación de series con datos directos; A₃=Análisis del conjunto de las fases; B₁=Tendencia línea base de 0.2; B₂=Tendencia línea base de 0.6).

Por lo que respecta a la interacción método de análisis x tendencia en la fase de tratamiento, el AVAR genera un efecto significativo: $F(8,273)=3.57$, $p<0.001$. Con objeto de verificar dicha interacción, en la tabla 4.22 se muestran las potencias medias y se representan gráficamente en la figura 4.8.

Tabla 4.22. Potencias medias de la interacción método de análisis x tendencia en la fase de tratamiento (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Tendencia tratamiento	Método de análisis		
	A_1	A_2	A_3
-0.1	0.745	0.574	0.673
-0.3	0.916	0.773	0.861
-0.5	0.973	0.903	0.954
-0.7	0.994	0.962	0.994
-0.9	0.999	0.986	0.999

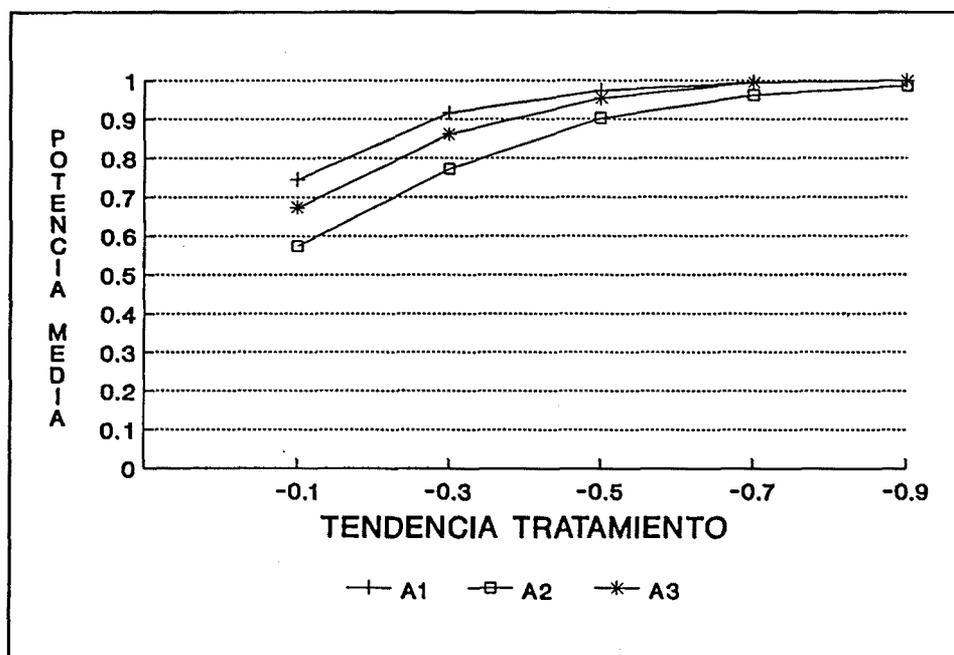


Figura 4.8. Interacción entre método de análisis y tendencia en la fase de tratamiento (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Al observar la figura 4.8 se comprueba que la comparación de series utilizando la técnica de la regresión (A_1) es más eficaz que el análisis del conjunto de las fases (A_3) y que la comparación de series con datos directos (A_2). Adicionalmente, la potencia de la prueba también se ve favorecida por el incremento de la tendencia en la fase de tratamiento, cualquiera que sea el método de análisis. En efecto, los resultados obtenidos muestran que las diferencias entre los tres métodos de análisis se van estrechando a medida que la magnitud de las pendientes en la fase de tratamiento aumenta.

En la tabla 4.23 se observa que la potencia estadística media aumenta con el tamaño del efecto (cambio de nivel=1 y tendencia en la línea base=0.6). El AVAR resultante produce una interacción significativa entre σ y tendencia en la línea base, $F(1,273)=17.84$, $p<0.001$.

Tabla 4.23. Potencias medias de la interacción tendencia en la línea base x cambio de nivel

Cambio de nivel	Tendencia línea base	
	0.2	0.6
Sigma=0	0.789	0.971
Sigma=1	0.834	0.962

La representación gráfica de la tabla 4.23 (figura 4.9) ofrece un tipo de interacción positiva, en el sentido de que las diferencias de la variable cambio de nivel son relativamente mayores al pasar de una tendencia de línea base de 0.6 a una de 0.2.

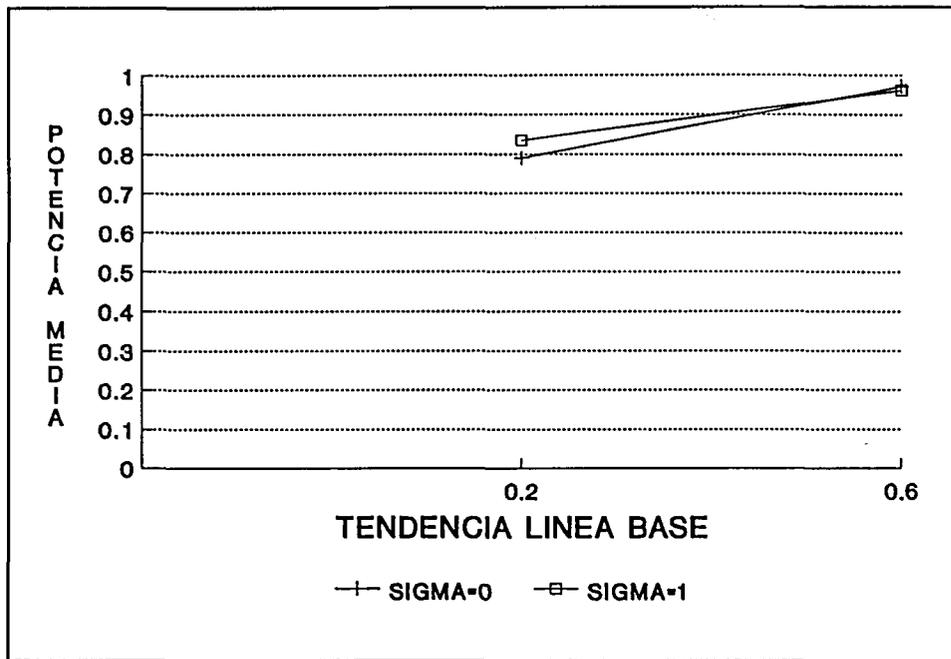


Figura 4.9. Interacción entre cambio de nivel y tendencia en la fase de línea base.

A partir de las potencias medias de la interacción entre σ y tendencia en la fase de tratamiento (tabla 4.24) se infiere que la dificultad de hallar una identificación correcta (estadístico C significativo) disminuye conforme aumenta el tamaño del efecto, sea el cambio de nivel o bien la tendencia en la fase de intervención. En efecto, $F(4,273)=2.50$, $p < 0.05$.

Tabla 4.24. Potencias medias de la interacción cambio de nivel x tendencia en la fase de tratamiento.

Tendencia tratamiento	Cambio de nivel	
	Sigma=0	Sigma=1
-0.1	0.654	0.675
-0.3	0.834	0.882
-0.5	0.938	0.954
-0.7	0.981	0.985
-0.9	0.994	0.995

La representación gráfica de la interacción σ x tendencia en la fase de tratamiento, mostrada en la figura 4.10, puede interpretarse como sigue: los valores altos de tendencia en la fase de intervención (-0.7 y -0.9) producen una potencia de gran tamaño, independientemente de la existencia o no de cambio de nivel. Por el contrario, para pendientes inferiores se observa que la magnitud de σ influye ligeramente sobre la potencia.

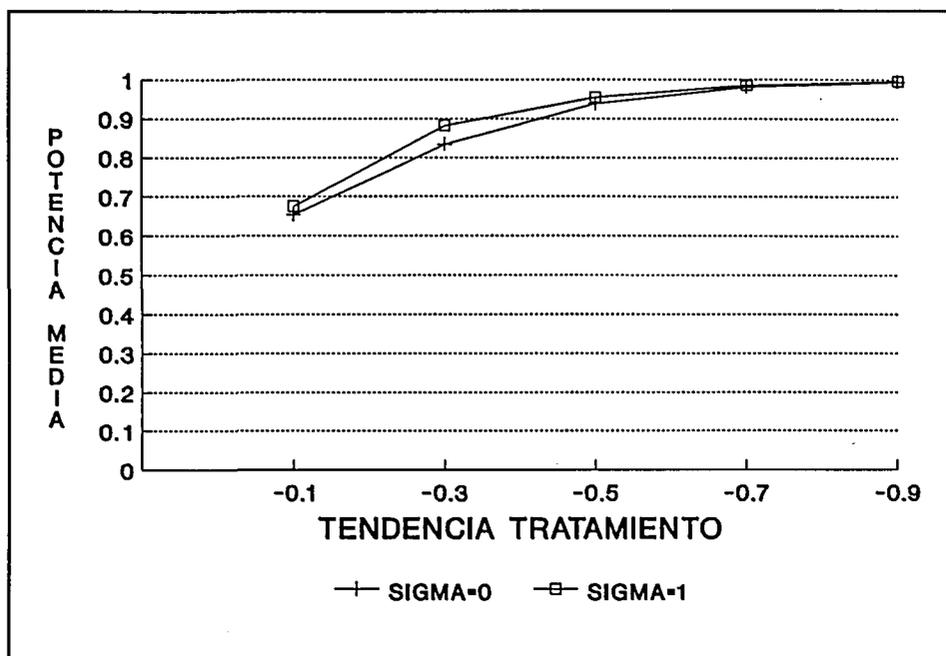


Figura 4.10. Interacción entre cambio de nivel y tendencia en la fase de tratamiento.

Por último, la interacción entre tendencia en la fase de tratamiento y tendencia en la línea base es también estadísticamente significativa: $F(4,277)=115.98$, $p<0.001$. Similarmente a las tablas 4.23 y 4.24 y a sus respectivas representaciones gráficas, la tabla 4.25, junto con la figura 4.11, pone de relieve un aspecto muy importante y es que el incremento del tamaño del efecto, en este caso tendencias en la línea base y en la fase de tratamiento, va parejo a una disminución de la dificultad de identificación correcta de la significación estadística de la prueba C. Así, por ejemplo, la mayor dificultad de identificación de una C significativa ocurre bajo pendientes pequeñas (tendencia línea base=0.2 y tendencia tratamiento=-0.1) con una potencia media de 0.427.

Tabla 4.25. Potencias medias de la interacción tendencia en la línea base x tendencia en la fase de tratamiento.

Tendencia tratamiento	Tendencia línea base	
	0.2	0.6
-0.1	0.427	0.903
-0.3	0.761	0.955
-0.5	0.910	0.981
-0.7	0.973	0.995
-0.9	0.991	0.998

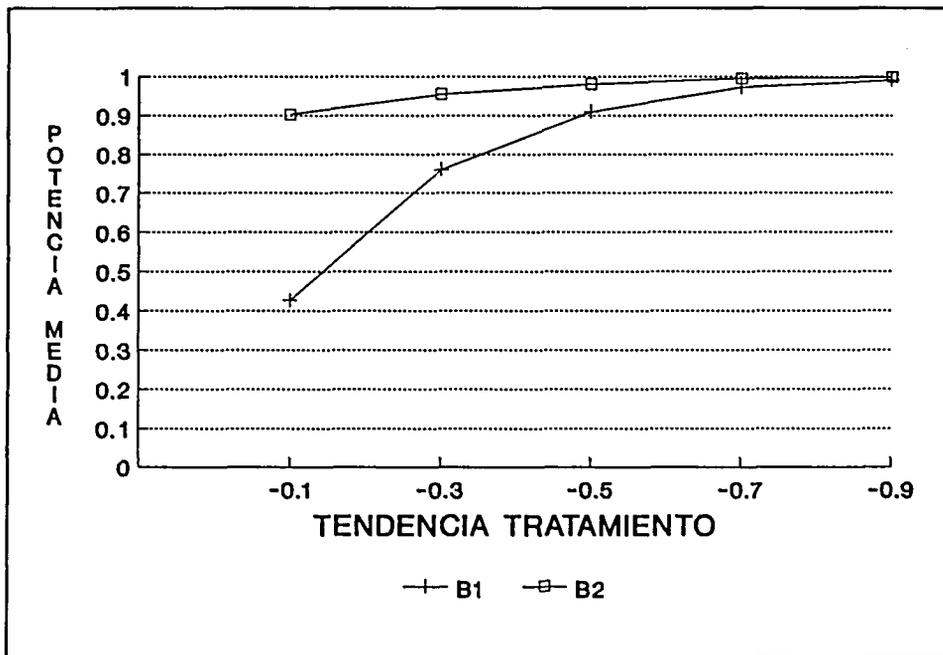


Figura 4.11. Interacción entre tendencia en la línea base y tendencia en la fase de tratamiento (B_1 =Tendencia línea base de 0.2; B_2 =Tendencia línea base de 0.6).

c) *Tendencias iguales entre fases (más pronunciadas en la fase de tratamiento) y sin cambio de nivel.*

Con objeto de simplificar los cálculos se han seleccionado, en este tercer patrón de cambio, tendencias en la fase de tratamiento de -0.2, -0.4, -0.6, -0.8 y -1 con una pendiente en la línea base de -0.1; tendencias en la fase de intervención de -0.3, -0.5, -0.7 y -0.9 con una pendiente en la línea base de -0.2, y tendencias en la fase de tratamiento de -0.4, -0.6, -0.8 y -1 con una pendiente en la línea base de -0.3.

Los gráficos correspondientes a la interacción método de análisis x tendencia en la fase de tratamiento (figuras 4.12, 4.13 y 4.14), cuyas potencias medias se muestran en las tablas 4.26, 4.27 y 4.28 presentan, en términos generales, un mismo patrón, siendo el grado de dificultad de identificación correcta de la significación estadística de C de mayor a menor para los distintos métodos de análisis: comparación de series con datos directos; comparación de series usando la técnica de la regresión, y análisis del conjunto de las fases. Además, la potencia media aumenta conforme incrementa la tendencia en la fase de intervención.

En el presente patrón de cambio la interacción de las dos variables analizadas es significativa para cualquier valor de la tendencia en la fase de línea base: -0.1 [$F(8,98)=19.83$, $p<0.001$], -0.2 [$F(6,77)=36.02$, $p<0.001$] y -0.3 [$F(6,77)=45.22$, $p<0.001$].

A partir de las figuras que se muestran a continuación (4.12, 4.13 y 4.14) se observa que la mayor diferencia entre los métodos de comparación de series y el análisis del conjunto de las fases está asociada con valores bajos de tendencia en la fase de tratamiento, y esta diferencia disminuye a medida que incrementa dicha tendencia. Por otro lado, los resultados obtenidos al aplicar el método de comparación de series utilizando la técnica de la regresión y el método de comparación de series con datos directos no varían

en función del valor de la tendencia en la fase de línea base, $F(2,119)=0.48$, $p>0.05$ (para el modelo aditivo) y $F(2,14)=2.19$, $p>0.05$ (para el modelo no aditivo). Esto es obvio, puesto que al llevar a cabo una comparación de series, las tendencias obtenidas son idénticas para cada valor de línea base, ya que los cambios de tendencia simulados son iguales; es decir, forman los mismo ángulos. En otras palabras, conforme disminuye la pendiente en la fase de línea base una décima, también se produce un decremento del mismo valor de las tendencias en la fase de tratamiento.

Tabla 4.26. Potencias medias de la interacción método de análisis x tendencia en la fase de tratamiento, para una pendiente en la línea base de -0.1 (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Tendencia tratamiento	Método de análisis		
	A_1	A_2	A_3
-0.2	0.107	0.079	0.671
-0.4	0.529	0.312	0.935
-0.6	0.840	0.652	0.992
-0.8	0.959	0.845	0.998
-1.0	0.989	0.934	1.000

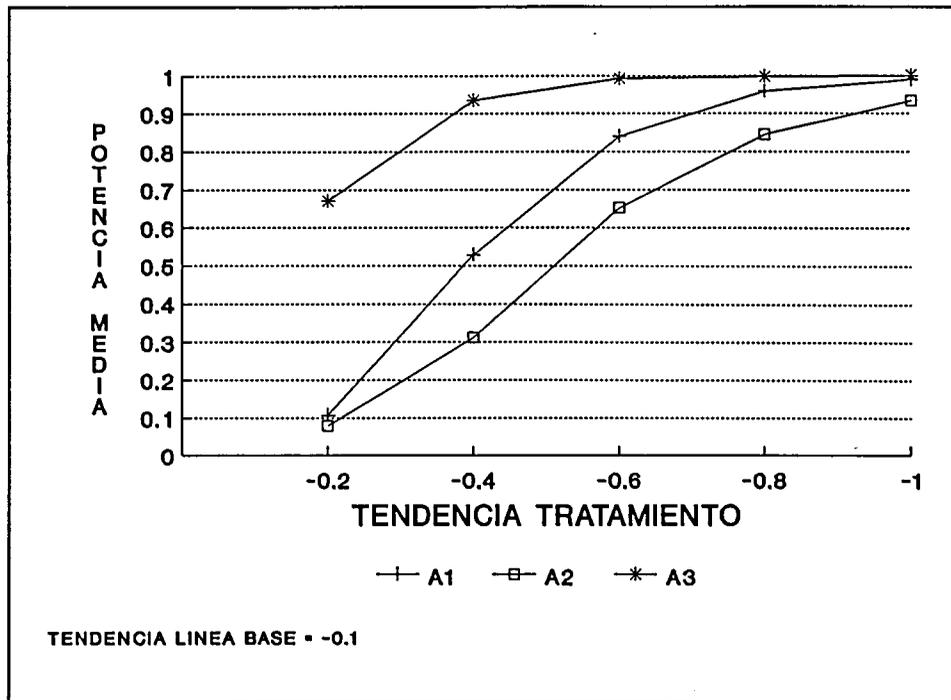


Figura 4.12. Interacción entre método de análisis y tendencia en la fase de tratamiento, para una pendiente en la línea base de -0.1 (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Tabla 4.27. Potencias medias de la interacción método de análisis x tendencia en la fase de tratamiento, para una pendiente en la línea base de -0.2 (A_1 =Comparación de series utilizando la técnica de la regresión; A_2 =Comparación de series con datos directos; A_3 =Análisis del conjunto de las fases).

Tendencia tratamiento	Método de análisis		
	A_1	A_2	A_3
-0.3	0.109	0.079	0.923
-0.5	0.549	0.332	0.989
-0.7	0.852	0.633	0.998
-0.9	0.956	0.834	1.000

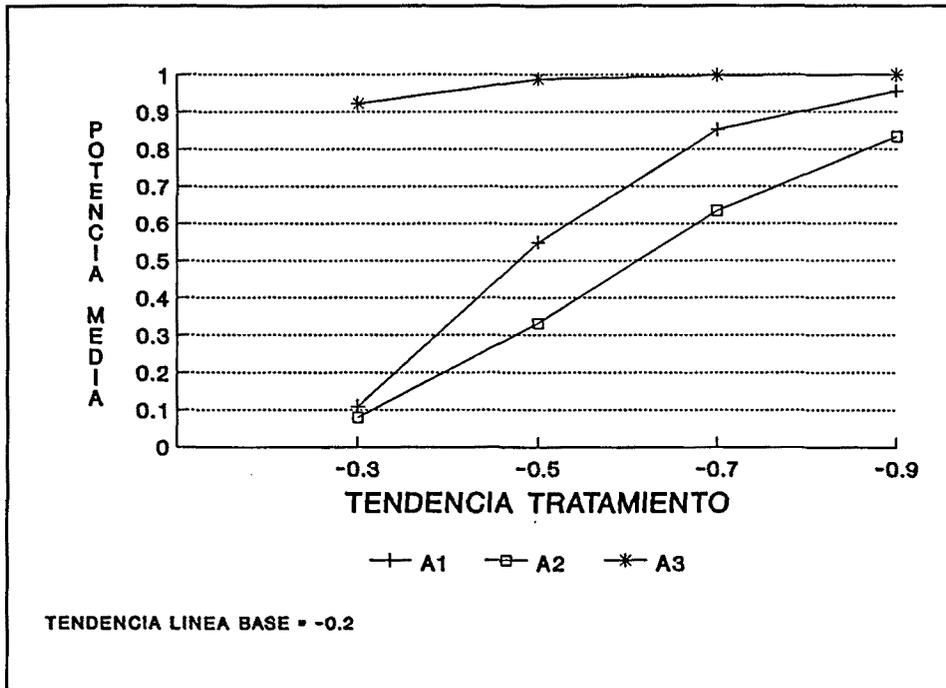


Figura 4.13. Interacción entre método de análisis y tendencia en la fase de tratamiento, para una pendiente en la línea base de -0.2 (A₁=Comparación de series utilizando la técnica de la regresión; A₂=Comparación de series con datos directos; A₃=Análisis del conjunto de las fases).

Tabla 4.28. Potencias medias de la interacción método de análisis x tendencia en la fase de tratamiento, para una pendiente en la línea base de -0.3 (A₁=Comparación de series utilizando la técnica de la regresión; A₂=Comparación de series con datos directos; A₃=Análisis del conjunto de las fases).

Tendencia tratamiento	Método de análisis		
	A ₁	A ₂	A ₃
-0.4	0.120	0.078	0.991
-0.6	0.544	0.307	0.999
-0.8	0.853	0.640	1.000
-1.0	0.945	0.857	1.000

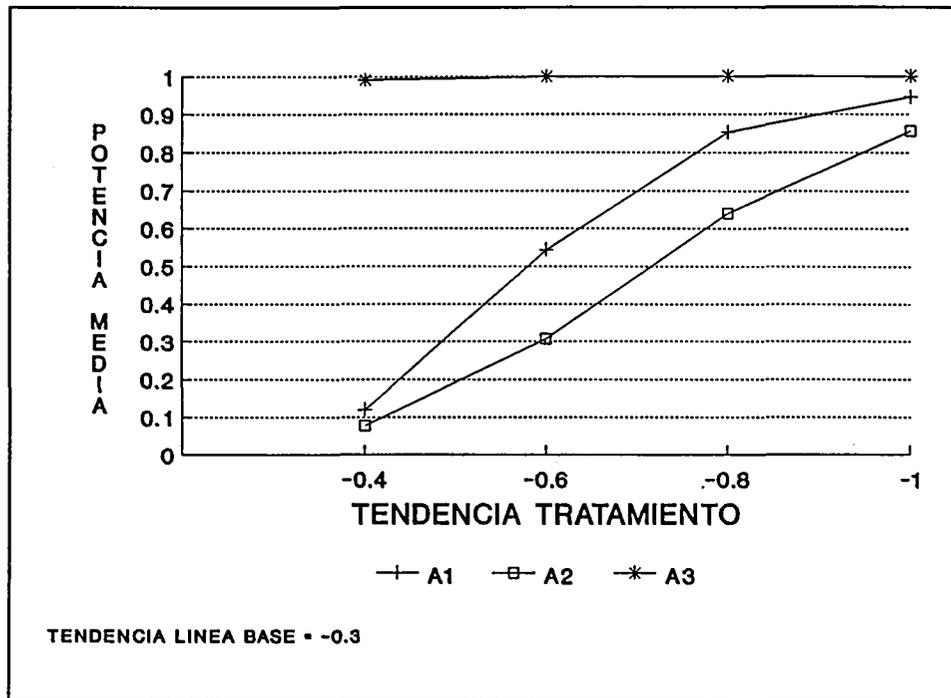


Figura 4.14. Interacción entre método de análisis y tendencia en la fase de tratamiento, para una pendiente en la línea base de -0.3 (A₁=Comparación de series utilizando la técnica de la regresión; A₂=Comparación de series con datos directos; A₃=Análisis del conjunto de las fases).

4.8. Discusión general

El objetivo de este trabajo de investigación ha consistido en valorar, tanto metodológica como estadísticamente, las estrategias de análisis propuestas por Tryon. Al mismo tiempo, mediante el cálculo de la potencia del estadístico C, se ha obtenido la probabilidad de inferencias válidas sobre el impacto de los tratamientos, en función de cambios producidos en la serie.

Los análisis del primer estudio han puesto de manifiesto que el estadístico C detecta si una serie temporal sigue un proceso aleatorio o bien presenta una tendencia. Por este motivo, la prueba C no es capaz de discriminar entre estabilidad y variabilidad de la serie.

Una forma de abordar este problema sería a través del cálculo de alguna medida estadística sobre variabilidad junto con algún análisis gráfico de los datos, puesto que ayuda a la descripción de los registros seriales (véase Arnau, 1993, 1994; Morley y Adams, 1989, 1991). En caso de que se detectara una gran variabilidad entre los distintos puntos de la línea base, una solución sería suavizar el conjunto de datos mediante el cálculo de las medianas con distintos recorridos (Arnau, 1993; Cleveland, 1985; Morley y Adams, 1991), y una vez se hayan eliminado los puntos atípicos, pasar a la estrategia 2 o 3 y aplicar el estadístico C (véase apartado 3.1).

Los resultados obtenidos del análisis de línea base revelan que cuando está presente una tendencia de gran tamaño (entre -0.8 y -1), la potencia se acerca considerablemente a la unidad, y cuando la pendiente es despreciable la potencia es casi nula, tanto para $\alpha=0.05$ como $\alpha=0.01$. Por el contrario, para tendencias medias la potencia es mayor con un valor $\alpha=0.05$. Estos resultados evidencian que es mejor optar por un nivel del 5%, sobre todo cuando el tamaño del efecto es medio.

En el segundo estudio se ha demostrado que para una línea base estable y tendencia en la fase de tratamiento, el estadístico C detecta mejor una intervención significativa a medida que aumenta la pendiente y el tamaño de la muestra. De esta forma, cuando las pendientes son más acusadas la prueba C es significativa con menos observaciones, y viceversa. En caso de que la aplicación del tratamiento produzca un cambio brusco de nivel de una desviación estándar en el punto de intervención, la potencia del estadístico C aumenta considerablemente. El AVAR resultante de este segundo estudio confirma una interacción significativa entre tendencia y cambio de nivel de modo que la mayor diferencia se produce cuando σ se combina con tendencias bajas.

Con el fin de facilitar la interpretación de los resultados, se ha subdividido la tercera estrategia de Tryon en tres partes. En la primera se analiza una de las dificultades principales del estadístico C: identificar un cambio de nivel cuando la fase de línea base

y la de tratamiento presentan la misma tendencia. Los resultados demuestran que se debe utilizar un análisis del conjunto de las fases cuando $\sigma=1$. Sin embargo, en ausencia de cambio de nivel este análisis llevaría a la afirmación de la existencia de un efecto significativo cuando en realidad no hay ninguno. Por lo tanto, en este último caso, sería conveniente aplicar un procedimiento de comparación de series. En resumen, se concluye que la efectividad del tratamiento queda en entredicho cuando la fase de intervención sigue la misma pendiente iniciada en la fase de línea base. Con objeto de solventar una posible inferencia errónea, sería conveniente ayudarse de la inspección visual.

En un segundo apartado se evalúan series que presentan un cambio de tendencia en sentido opuesto, antes y después del tratamiento. A partir de los resultados, se observa que la potencia es mayor cuando los efectos del tratamiento (cambio de nivel y cambio de tendencia) son grandes y que las posibilidades de detectar efectos pequeños son escasas (véase apéndice 10). Aquí es donde radica uno de los principales problemas, puesto que en psicología es muy difícil conseguir que los efectos sean grandes, especialmente para investigaciones llevadas a cabo fuera del laboratorio. Los resultados indican que el método de análisis más potente es el de la comparación de series utilizando la técnica de la regresión, seguido del análisis del conjunto de las fases y, finalmente, del método de comparación de series con datos directos. Estas diferencias entre los tres métodos de análisis se van diluyendo a medida que aumenta el tamaño del efecto. Por otra parte, los datos del estudio ponen de manifiesto que únicamente el análisis del conjunto de las fases detecta un cambio de nivel.

Para finalizar el análisis de la tercera estrategia de Tryon, se estudia un patrón de cambio donde la línea base presenta una moderada tendencia que pasa a ser más acusada después de la intervención. En este caso, el orden en que aumenta la potencia del estadístico C, según los distintos métodos de análisis, es el siguiente: comparación de series con datos directos, comparación de series usando la técnica de la regresión y análisis del conjunto de las fases. Las mayores diferencias se producen entre los métodos de

comparación de series y el análisis del conjunto de las fases, y van disminuyendo a medida que aumenta la tendencia en la fase de tratamiento.

La presente investigación, no sólo ha permitido poner a prueba las estrategias de Tryon, sino también la elaboración de las tablas de potencia del estadístico C. Éstas son instrumentos útiles para la planificación y valoración de investigaciones. En efecto, las tablas de potencia halladas permiten determinar la potencia y los tamaños muestrales de estudios que se deseen llevar a cabo, o bien interpretar resultados no significativos de trabajos ya finalizados. Sin embargo, ni la determinación de la potencia ni la del tamaño muestral necesario puede tener lugar si el investigador no tiene una idea del grado en que la H_0 puede ser falsa. Por lo general, se formulan hipótesis teóricas sobre el tamaño del efecto. Si tal efecto es significativo y, por lo tanto, la hipótesis *a priori* se ve confirmada, entonces la potencia no es ningún problema, aunque los errores de Tipo I son los más temidos en las ciencias del comportamiento.

La potencia merece una atención especial en la decisión de si un experimento va a ser llevado a cabo o no. Si ya se ha decidido realizar el experimento, la potencia es importante pero no fundamental. Por otra parte, si al que hace el experimento le interesa descubrir una desviación de una cierta magnitud respecto de la hipótesis nula, entonces es imprescindible conocer la potencia con que se está trabajando. Si, por ejemplo, sólo hay una probabilidad del 15 por ciento de obtener tal diferencia, no interesa hacer el experimento. En tal caso, el investigador quizá desee usar sus recursos en una investigación donde tenga más seguridad de llegar a una conclusión correcta. El descubrimiento, durante la fase de planificación de una investigación, de que la potencia del estadístico C es baja llevaría a una revisión del planteamiento. Como ejemplo, considérese que un experimento terminado acepta como verdadera la H_0 , pero la potencia estadística es baja. Ante esta situación se interpreta el resultado negativo como ambiguo, ya que el no poder rechazar la H_0 no tiene un gran significado cuando la probabilidad *a priori* de rechazarla es baja.

A modo de resumen, las tablas de potencia del estadístico C son útiles para planear investigaciones y comprobar las ya terminadas. Así, con un análisis *a priori* se determina la potencia y el tamaño de muestra apropiado y con un análisis *post hoc*, si las investigaciones que producen resultados negativos poseen una potencia adecuada. Por último, es importante no olvidar que tamaños del efecto moderados o pequeños son muy comunes en ciencias del comportamiento y que, por lo tanto, a veces, se requieren tamaños de muestra grandes para detectar tales efectos. Si en las tablas del apéndice 10, para una determinada combinación de factores, no se puede determinar la muestra apropiada, entonces el tamaño muestral necesario es superior a 15 observaciones por fase. En tal caso es lícito pensar, si el campo de aplicación permite obtener un número elevado de registros, en un análisis de series temporales.

