

EXPLORACION DE LOS DATOS EN EXPERIMENTOS DE SERIES TEMPORALES¹

M.V. GARCÍA JIMÉNEZ*, ; E.M. SEBASTIÁN VILLARIÑOS*
Y A. DOMENECH MARTÍN*

* Dpto. Metodología de las C.C. del Comportamiento. Fac. Psicología. UCM.

* Instituto Pluridisciplinar. Unidad de R.M.N. UCM.

Resumen

El objetivo de este trabajo es estudiar la congruencia, entre el análisis visual y el estadístico en experimentos de series temporales, ampliamente debatido por numerosos autores.

En primer lugar, hemos realizado una revisión de los artículos publicados entre 1984 y 1993, en la Revista «Análisis y Modificación de Conducta», por recogerse en ella la mayor parte de los estudios de caso único. El número de artículos que utilizan el análisis estadístico en este periodo es escaso.

A continuación, mostramos los resultados de 12 nuevos estudios de caso único realizados por nosotros, comparando los resultados del análisis visual (cambios de nivel y tendencia) con los del análisis estadístico (pruebas paramétricas, no paramétricas y ARIMA). En todos los casos hemos encontrado congruencia entre los resultados provenientes de ambos tipos de análisis

Palabras-Clave: series temporales, análisis visual, análisis estadístico, nivel, tendencia

Abstract

The purpose of this paper is to study the agreement between the visual and statistical analysis in experiments of time series. This topic has been controverted by many authors.

We have done a review of the articles published in the Journal "Análisis y Modificación de Conducta" between 1984 and 1993. We have chosen this journal because it include most of the studies of single case. Nevertheless we have found few papers where statistical analysis is used.

In this paper we show our results of 12 studies of single case. We compare the results of visual analysis (change of level and tendency) with those of statistical analysis (parametric and non-parametric tests and ARIMA). Always we have find agreement between results coming form both type of analysis.

Keywords: time series, visual analysis, statistical analysis, level, tendency

1. Introducción

Los diseños de series temporales han sido, desde sus comienzos, los más utilizados en Modificación de Conducta. Su característica principal es la de tomar una serie de medidas del

¹ Este estudio, en parte, ha sido posible gracias a los recursos aportados por el Ministerio de Educación y Ciencia, DGICYT, proyecto nº PB92-0207

sujeto a lo largo del tiempo, generalmente antes, durante y después del tratamiento. Para la comprobación de la efectividad de éste se comparan las diferencias existentes entre las medidas tomadas antes y después con las tomadas durante el tratamiento.

Esta comparación se ha ido realizando a partir de distintas técnicas, en diferentes momentos. Así, antes de la década de los 70, se ha realizado generalmente a partir del análisis visual, siendo relativamente escaso el empleo de técnicas de análisis de datos estadísticos. (Sidman, 1960).

A partir de dicha época, y con el objetivo de hacer una Psicología cada vez más científica, se da mayor importancia a la utilización de los análisis estadísticos, no como sustituto, sino como complemento de la inspección visual, ya que en muchas ocasiones resulta difícil la interpretación de los efectos del tratamiento, a partir únicamente del estudio detenido de los gráficos, como por ejemplo, cuando surgen dificultades para conseguir la estabilidad de la línea base, o cuando se realizan investigaciones de carácter exploratorio (Kazdin, 1976).

Sin embargo, la aplicación a estos diseños de las *técnicas clásicas* de análisis de datos, como las pruebas de «t» y «F», han tenido sus críticas, ya que en los diseños conductuales, no suele darse el requisito de independencia de las observaciones, por lo que su aplicación puede llevar a interpretaciones incorrectas. (Scheffe, 1959 y Glass et al., 1975). Para evitar esto, algunos autores han propuesto distintas modificaciones de las pruebas tradicionales tales como tratar las fases del diseño (línea base-intervención) como si fuesen las fuentes de variación intergrupo y las tasas de conducta observadas en cada fase, como varianza intragrupo del análisis de varianza clásico (Shine and Bower, 1971; Gentile et al., 1972). Estos últimos, conscientes de no haber conseguido el supuesto de independencia de las observaciones, proponen la disminución de la autocorrelación de las mismas entre fases adyacentes, combinando las observaciones de fases no adyacentes que empleen el mismo tratamiento. Otros autores han intentado distintas modificaciones, pero todas ellas siguen sin solucionar el problema de la autocorrelación (Kesselman and Leventhal, 1974), por lo que han recibido continuas críticas (Hartmann, 1974; Michael, 1974).

Para solucionar este problema, se propone la utilización del *análisis de series temporales*. Este tiene en cuenta la dependencia serial de los datos y permite probar las diferencias de nivel y tendencia de los mismos, así como los valores de la pendiente, aunque por su complejidad, su uso todavía no es muy frecuente en los experimentos y cuasiexperimentos de Modificación de Conducta.

Para el estudio de los efectos de la intervención, se consideran adecuados los métodos estadísticos de series temporales derivados de la línea de Box (Box and Tiao, 1965; Box and Jenkins, 1970) y de Glass y cols. (Glass et al., 1975; Gottman and Glass, 1978).

El análisis de series temporales se basa en la transformación de las puntuaciones brutas de las series, en puntuaciones serialmente independientes o no correlacionadas y, una vez realizada la transformación, establece comparaciones mediante los procedimientos estadísticos convencionales, entre las puntuaciones transformadas de las diferentes fases. El modelo más utilizado en Psicología es *el modelo ARIMA* desarrollado fundamentalmente por Box y Jenkins (1970).

A pesar de sus ventajas, estas técnicas no resuelven todos los problemas del análisis de datos en estos diseños, ya que también presentan una serie de limitaciones, tales como el nº de observaciones mínimas requeridas en cada fase para poder detectar la autocorrelación que debe ser considerablemente grande y, aunque no se puede establecer a priori pues dependerá de la naturaleza de los datos de cada investigación concreta, puede oscilar entre las 100 consideradas por Box and Jenkins (1970), las 50 de Glass y cols. (1975) y las 10 de Jones y cols. (1977). El otro problema hace referencia a la inconsistencia entre el análisis visual y el de series temporales de los mismos datos. El estudio de Jones y cols. (1978) replicado por De Prospero y Cohen (1979) mostrando resultados análogos, obtuvo un grado de acuerdo medio entre las dos técnicas en torno al 0.60, aumentando la disparidad cuando la dependencia serial

es mayor y cuando el análisis estadístico detecta cambios significativos. Por ello puede plantearse la duda de cual es el análisis que proporciona los resultados más fiables, puesto que también el análisis estadístico parte de la subjetividad del investigador a la hora de elegir el modelo adecuado.

Hartmann y cols. (1980), proponen que cuando el efecto del tratamiento aparece como significativo en el análisis de series temporales, pero no en el visual, probablemente el tratamiento resulte poco efectivo para un paciente concreto, aunque los resultados pueden ser útiles para la investigación básica. En el caso contrario, si los efectos se aprecian en el análisis visual pero no en el estadístico, la interpretación sería mucho más conflictiva pues no existe un acuerdo unánime sobre si se debe guiar el investigador por uno u otro criterio.

Otro procedimiento estadístico aplicable a estos diseños, son las pruebas de aleatorización que sirven para determinar directamente la significación estadística de los resultados experimentales. Se basan en las permutaciones de los resultados entre los tratamientos, y se calcula después un estadístico de contraste para cada permutación. (Fisher, 1935; Eddington, 1967,1969,1975,1980; Levin et al., 1978). La validez de estas pruebas exige la asignación aleatoria de los tratamientos a las diferentes fases, por lo que solamente serán aplicables a diseños que permitan este requisito. Solucionan el problema de la dependencia serial, pues mediante este procedimiento queda balanceada en los diferentes tratamientos. Sin embargo, también tiene ciertas limitaciones como la exigencia de diseños que permitan la asignación aleatoria de los tratamientos a las fases, la imposibilidad de aplicación cuando los efectos del tratamiento son irreversibles, y la laboriosidad que puede suponer calcular estadísticos de contraste para todas las permutaciones, dado el número elevado de ellas que se obtienen, incluso con muestras pequeñas de datos.

Otras técnicas estadísticas que se han utilizado en este tipo de diseños son el estadístico *Rn* (Revusky, 1967) para diseños de línea base múltiple, y que supone la aplicación del tratamiento en cada serie en un orden aleatorio y requiere un número mínimo de cuatro series., o las cadenas de Markov para experimentos temporales con finalidad exploratoria, que permiten predecir la probabilidad conocida en otro momento anterior.

Así pues, a pesar de la recomendación hecha por numerosos autores acerca del empleo de las técnicas estadísticas en este tipo de diseños, hoy por hoy, como puede deducirse a partir de lo expuesto, no dejan de presentar serias dificultades a la hora de su utilización ante problemas concretos.

En este trabajo nos proponemos estudiar la congruencia entre el análisis visual y el estadístico en experimentos de series temporales.

Para ello se realiza una revisión bibliográfica de 10 años (1984-1993) en la Revista «Análisis y Modificación de Conducta», por ser una de las que más investigaciones recogen utilizando este tipo de diseño, para conocer la cuantía en que son utilizadas ambas técnicas de análisis de datos (visual y estadística).

Por último se llevan a cabo 12 nuevos estudios (11 experimentales y 1 cuasiexperimental) de series temporales, para comparar los resultados con ambos tipos de análisis y estudiar su congruencia. En el último caso estudiado (diseño cuasiexperimental AB), además de los estadísticos clásicos se utiliza un análisis ARIMA a pesar de las dificultades que presenta su aplicación, para comprobar si añade alguna información a la obtenida mediante los demás análisis.

2. Revisión bibliográfica

Se ha llevado a cabo una revisión bibliográfica de los 263 artículos publicados durante el periodo comprendido entre 1984 y 1993 en la revista española «Análisis y Modificación de Conducta».

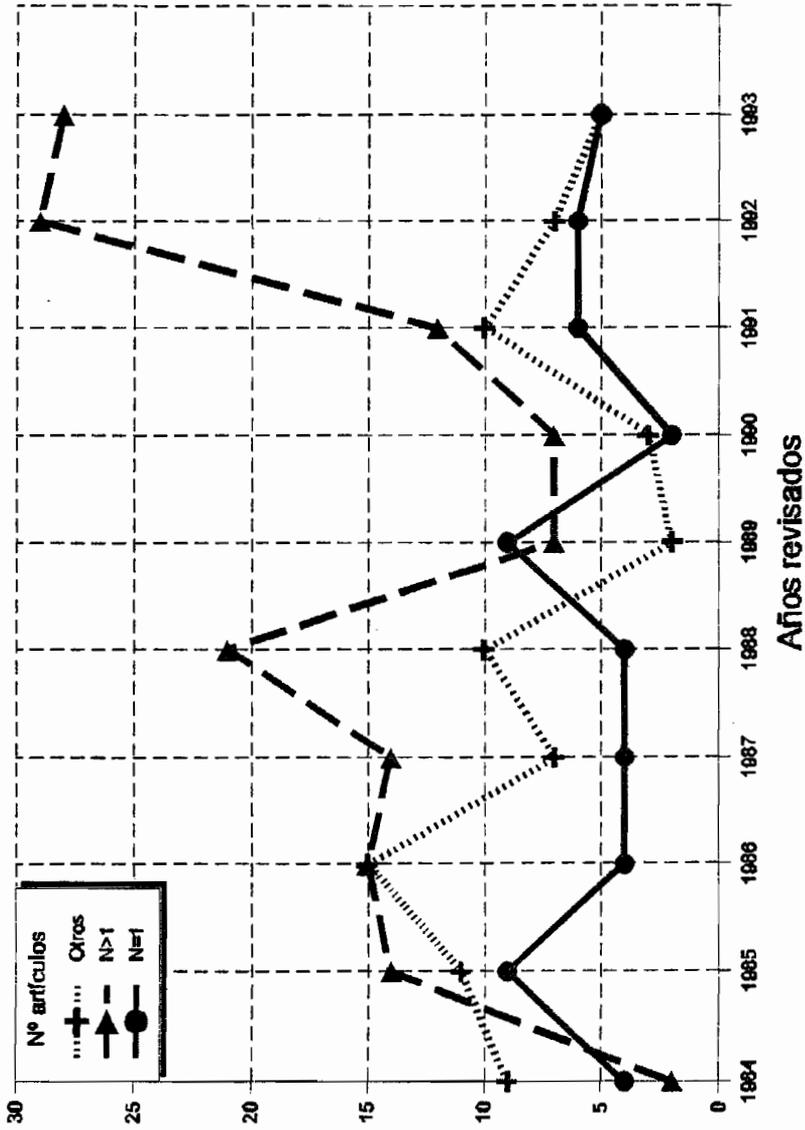


Figura 1.- Evolución del número de artículos cuyo estudio se dedica a casos únicos, casos de más de un sujeto y otros (teóricos y revisionistas) en el período 1984-1993

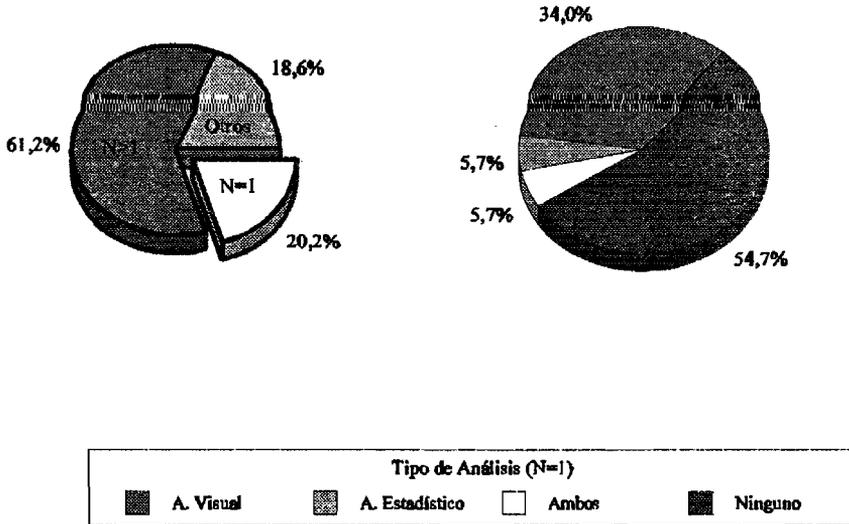


Figura 2.- Distribución de los distintos tipos de diseños (izquierda). Tipo de análisis realizado para los diseños N=1 (derecha)

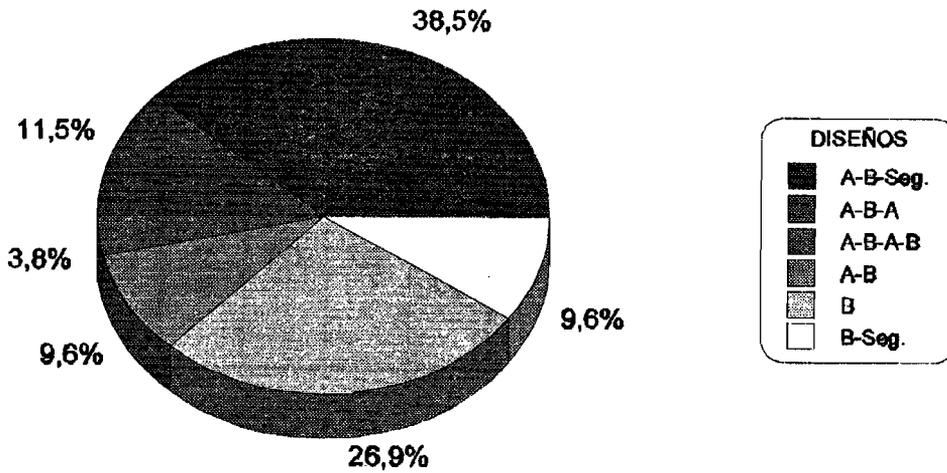


Figura 3.- Distribución del tipo de diseño empleado en experimentos N=1

	Total Art.	N>1	N=1	Sólo A. Visual	Sólo A. Estadístico	Ambos análisis	Ningún análisis
1984	15	2	4	2	0	0	2
1985	34	14	9	2	1	0	6
1986	34	15	4	1	0	0	3
1987	25	14	4	2	0	0	2
1988	35	21	4	1	0	0	3
1989	18	7	9	2	0	3	4
1990	12	7	2	0	1	0	1
1991	40	24	6	6	0	0	0
1992	42	29	6	1	1	0	4
1993	38	28	5	1	0	0	4

Tabla 1.- Número de artículos dedicados a casos múltiples y número dedicado a caso único; de estos últimos, frecuencia con que se realiza análisis visual, estadístico, ambos o ninguno

Debemos diferenciar entre todos estos artículos, aquellos cuyo estudio está dedicado: a los casos únicos ($N=1$), a los casos de más de un sujeto ($N>1$) y a otros trabajos cuya dedicación es más bien teórica y revisionista.

En la Figura 1, se muestra la evolución de la frecuencia de los diferentes tipos de artículos a lo largo de estos últimos años, advirtiéndose que entre 1984-1988 los trabajos basados en $N>1$ muestran un importante incremento respecto a los otros tipos de estudios; de 1988 a 1990 sufren un marcado decremento, para luego volver a recuperarse con mayor fuerza hasta 1993. Con respecto a los artículos $N=1$, no parecen existir incrementos o decrementos significativos a lo largo de su evolución.

Como podemos observar, en la Figura 2, el porcentaje de artículos cuyo desarrollo está basado en $N>1$ es del 61,2%, notablemente superior al 20,2% de los artículos de $N=1$, quedando un porcentaje muy similar al anterior (18,6%) de trabajos con formato teórico. En cuanto a los 53 artículos de caso único, ($N=1$), que corresponden al 20,2% del total, sólo realizan análisis visual el 34% de estos, pero en ningún caso se analiza la presencia o ausencia de cambios de nivel y/o tendencia, por lo que podemos considerarlo como un análisis más bien observacional. Respecto al análisis estadístico solamente el 5,7% de los artículos trabajan con él, siendo este análisis muy somero, encontrándose solamente en 1989 un artículo que utiliza ambos tipos de análisis en tres casos de $N=1$ (5,7%). Por último, un 54% de estudios no utilizan ningún tipo de análisis.

Respecto al tipo de diseño empleado, como puede observarse en la Figura 3, diremos que predominan los diseños A-B con seguimiento, seguidos de los diseños tipo B, en los que no se puede probar el efecto de la intervención ya que carecen de línea base.

En conclusión, y como se muestra en la Tabla I, que recoge las frecuencias encontradas, son escasos los estudios dedicados a $N=1$ en comparación con los estudios de $N>1$ en la Revista «Análisis y Modificación de Conducta», pareciendo la tendencia en el tiempo a aumentar los artículos de $N>1$. También hay que llamar la atención acerca de que es más frecuente el análisis visual que el estadístico, siendo en la mayoría de las ocasiones la subjetividad del terapeuta la que determina la efectividad o no del tratamiento. Por nuestra parte hemos encontrado en estudios anteriores (García Jiménez et al, 1994a, 1994b, 1995a, 1995b y 1995c) la congruencia entre el análisis visual y el estadístico en experimentos, que utilizan diferentes diseños. Por lo que creemos que la tendencia a utilizar solamente el análisis visual es más bien una costumbre adoptada por los modificadores de conducta que una dificultad atribuida a la incongruencia entre los diferentes tipos de análisis.

3. Experimentos (Diseños y Tratamientos)

Como hemos visto en la revisión bibliográfica, muy pocos autores hacen análisis estadístico en este tipo de diseños, algunos (únicamente 18) lo hacen visual, y la mayoría no hace ningún tipo de análisis. Esto quizás se deba a la gran controversia que hay, ya que no está claro que sea mejor utilizar uno u otro método y algunos autores sugieren que si se utilizan los dos se producen incongruencias en los resultados. Por ello en éste y en anteriores trabajos (García Jiménez et al, 1994a, 1994b, 1995a, 1995b y 1995c) se analiza la consistencia o no entre los resultados del análisis estadístico y el visual.

Hemos realizado algunos estudios en los que se utilizan distintos tipos de diseños de series temporales, analizándolos mediante estadísticos clásicos, tanto paramétricos como no paramétricos, y mediante análisis visual en el que se han estudiado los cambios de tendencia y de nivel a partir de las líneas de regresión de las puntuaciones en cada fase.

Para este estudio hemos seleccionado los tipos de diseños de series temporales siguientes:

3.1.- Diseños de reversión

En este tipo de diseños la característica principal es que hay una vuelta a la línea base o retirada del tratamiento, por lo que se consigue una inferencia más clara de la eficacia del mismo.

a) *Enuresis nocturna*. (A-B-A₁). En este caso se ha elegido una vuelta a la línea base. La variable dependiente se midió por el número de días en que la enuresis aparece durante el tiempo estudiado. El tratamiento consistió en mandar al niño al baño antes de acostarse, acostumbrarlo a retener la orina dándole refuerzo verbal, ver la T.V. y jugar con sus amigos. (ver fig.4)

b) *Hiperactividad*. (A-B-A₁-B₁). Tenemos aquí un diseño de reversión del tratamiento. La variable dependiente es el número de veces que se repite una conducta que presupone Hiperactividad: «Levantarse del asiento», que se quiere extinguir, y el número de veces que se repite la conducta «Levantar la mano para preguntar», que pretendemos reforzar. Para el tratamiento se aplicó refuerzo diferencial de conductas incompatibles con la problemática, utilizándose tanto reforzadores escolares (borrar la pizarra, jugar con plastelina, usar las pinturas y pinceles...) como reforzadores materiales (chicles, galletas, pegatinas, canicas...). Se elogió la conducta a adquirir y se retiró la atención cuando se emitía la conducta problema. (ver fig. 5)

c) *Tartamudez* (A-B-B₁). En este diseño se deja trascurrir un tiempo entre B y B₁ por lo que se puede considerar de reversión. La variable dependiente se mide por el tiempo que el sujeto tartamudea durante una conversación regular. El tratamiento se establece con distintas técnicas combinadas en las diferentes fases: Condicionamiento con un simulador rítmico, entrenamiento en eco, relajación, entrenamiento en asertividad y reestructuración cognitiva. (ver fig. 6)

d) *Hora de irse a dormir* (A-B-A₁-B₁). Se estudian 3 casos (Carlos, Ana y David). Este diseño es de reversión del tratamiento. La variable independiente es el refuerzo económico (100 ptas.) cuando la conducta es la adecuada. La variable dependiente, sería la hora de irse a dormir. (ver fig. 7)

3.2.- Diseños con cambios de criterio

Este tipo de diseños tiene como característica principal, que al sujeto se le obliga a alcanzar una determinada meta, cuando ésta es alcanzada, se pasa a la siguiente fase en la cual el sujeto tiene que alcanzar una meta superior, y así sucesivamente, por lo que es normal que se establezcan diferencias entre unas fases y otras, y ésta diferencia vendrá determinada por la diferencia existente entre el criterio a alcanzar de una fase a otra.

a) *Obesidad*. (A-B₁-B₂-B₃). La variable dependiente sería el tiempo dedicado a comer y a cenar. El tratamiento consistió en que la paciente dedicase al menos 15, 25 y 30 minutos en comer y cenar en cada fase. (ver fig. 8)

b) *Tabaquismo*. (A-B₁-B₂-...-B₉) La variable independiente será el número de cigarrillos permitidos por día y la dependiente el número de cigarrillos fumados al día. El tratamiento consistió en que el paciente debía pagar 2.000 ptas. si no cumplía con el criterio. El criterio reducía cada 5 días el número de cigarrillos en dos. (ver fig. 9)

c) *Hábitos de estudio I* (A-B₁-B₂-...-B₅). La variable dependiente es el tiempo establecido de estudio, la dependiente, el tiempo real de estudio. En cuanto al tratamiento, se estableció un tiempo determinado para el estudio de la asignatura de Ciencias Naturales y si se alcanzaba este tiempo, entonces el sujeto era reforzado económicamente. (ver fig. 10)

d) *Hábitos de estudio II* (A-B₁-B₂-...-B₇). Es un diseño igual que el anterior, pero con mayor número de fases, utilizando la misma variable dependiente (el tiempo establecido para el estudio) y la independiente (el tiempo dedicado al mismo). El tratamiento consistió en dar una hora al día para jugar, de la cual se descontarían 5 minutos por cada uno que el criterio no fuera superado y 2 minutos más de juego, por cada uno que pasara del criterio. (ver fig 11).

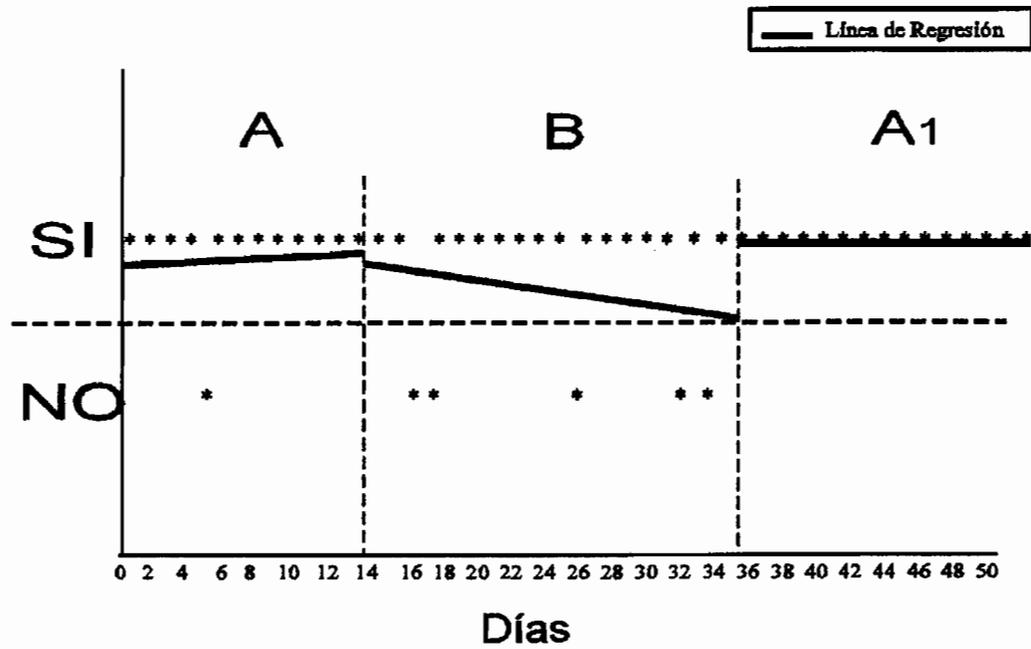


Figura 4.- Enuresis nocturna (A-B-A₁)

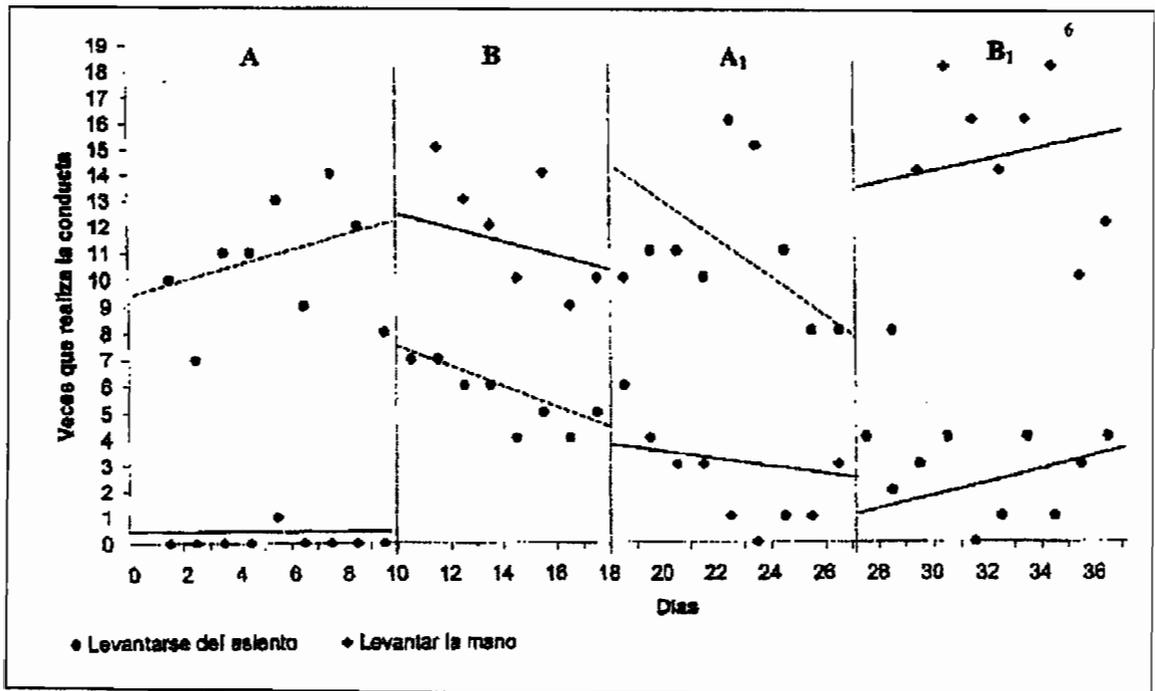


Figura 5.- Hiperactividad (A-B-A₁-B₁)

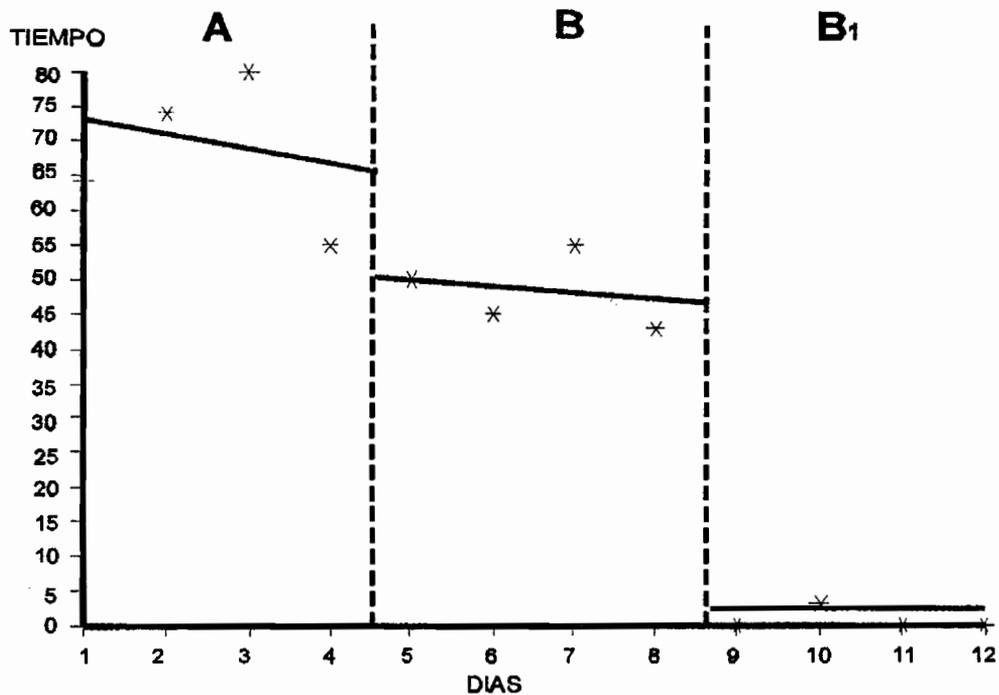


Figura 6.- Tartamudez (A-B-B₁)

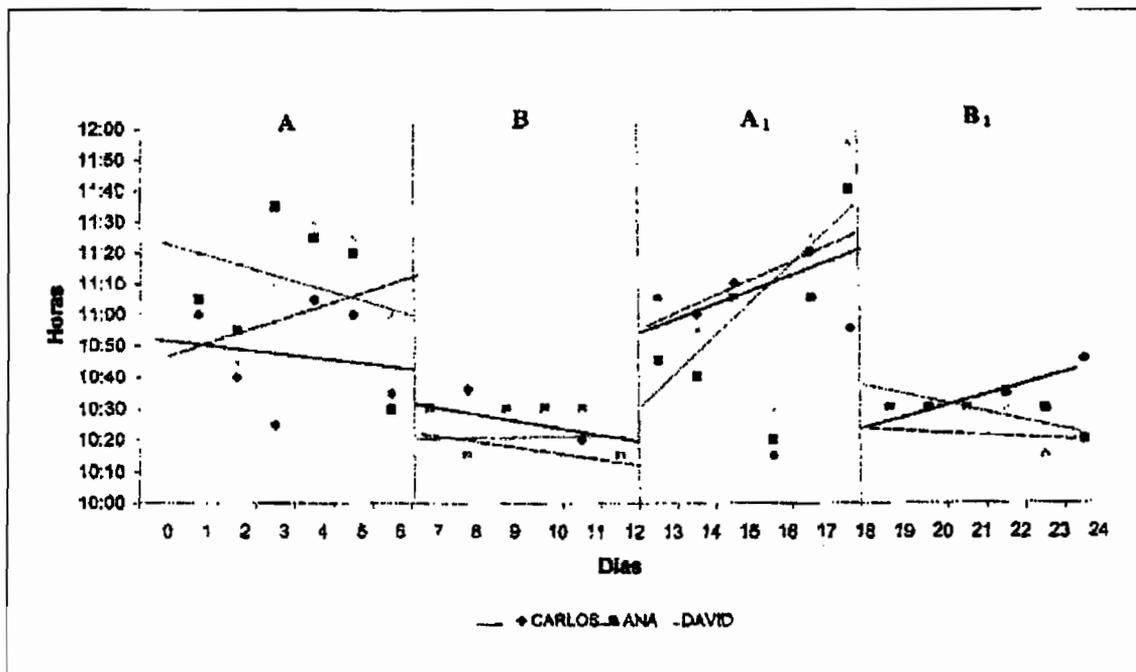


Figura 7.- Hora de irse a dormir (A-B-A₁-B₁)

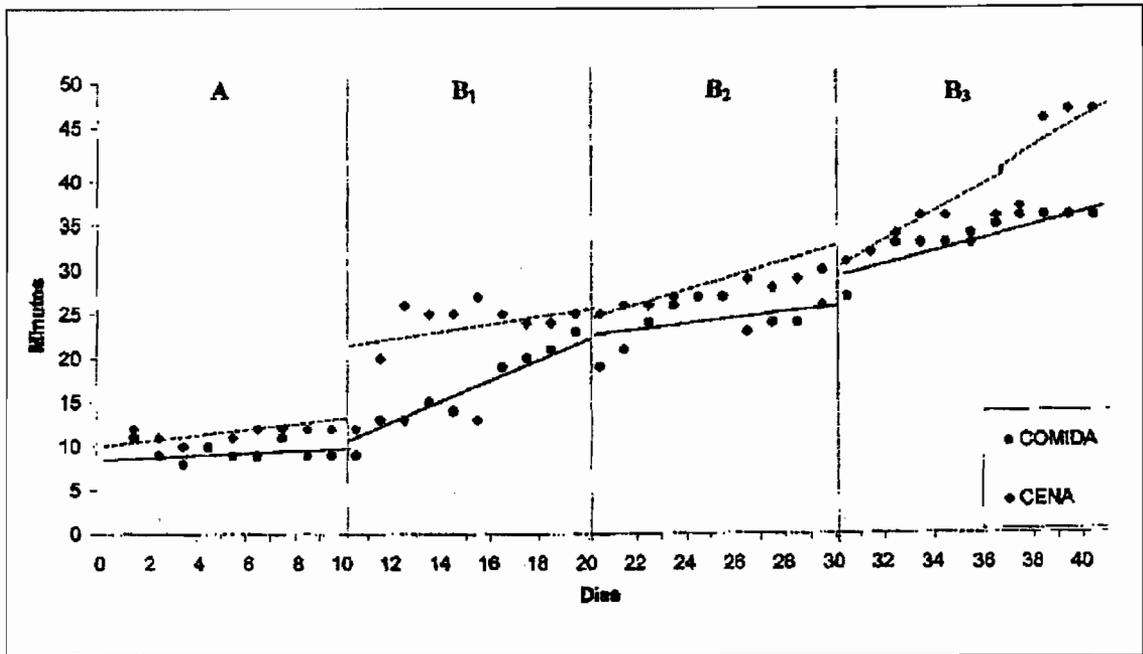


Figura 8.- Obesidad (A-B₁-B₂-B₃)

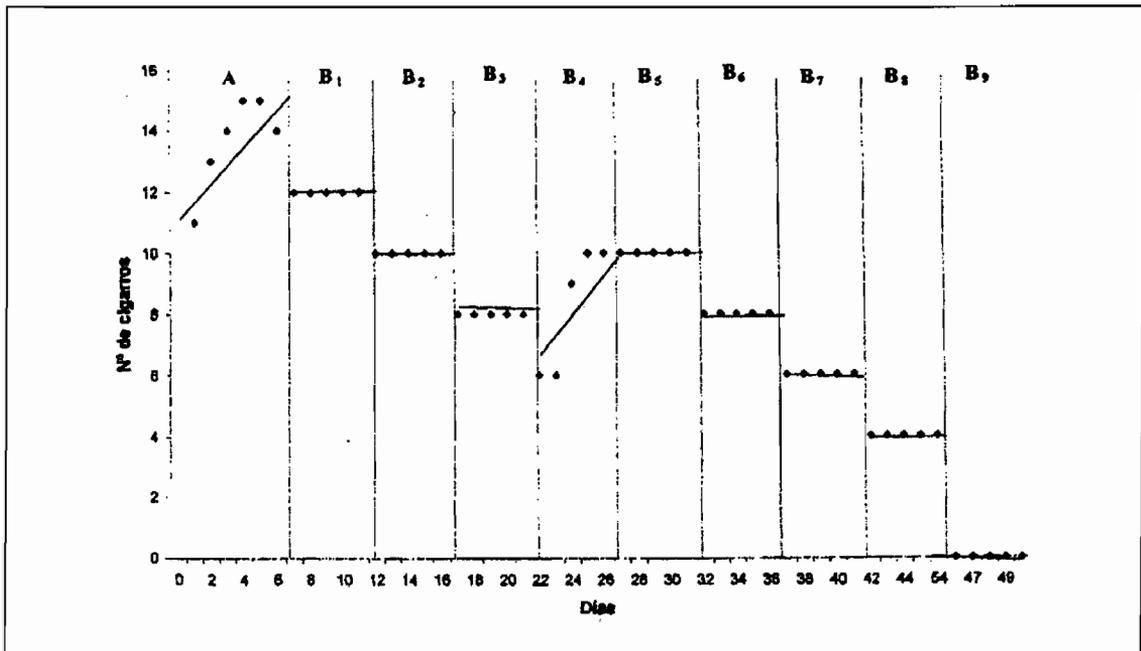


Figura 9.- Dependencia del tabaco (A-B₁-B₂-B₃-B₄-B₅-B₆-B₇-B₈-B₉)

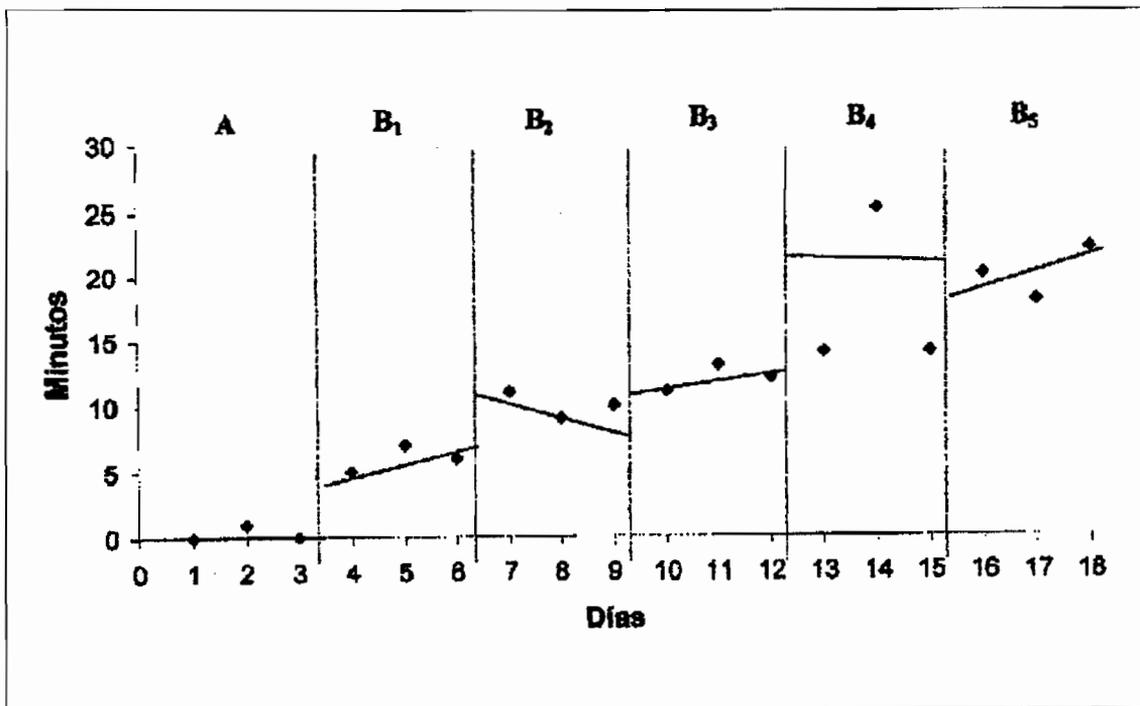


Figura 10.- Hábitos de estudio (I) (A-B₁-B₂-B₃-B₄-B₅)

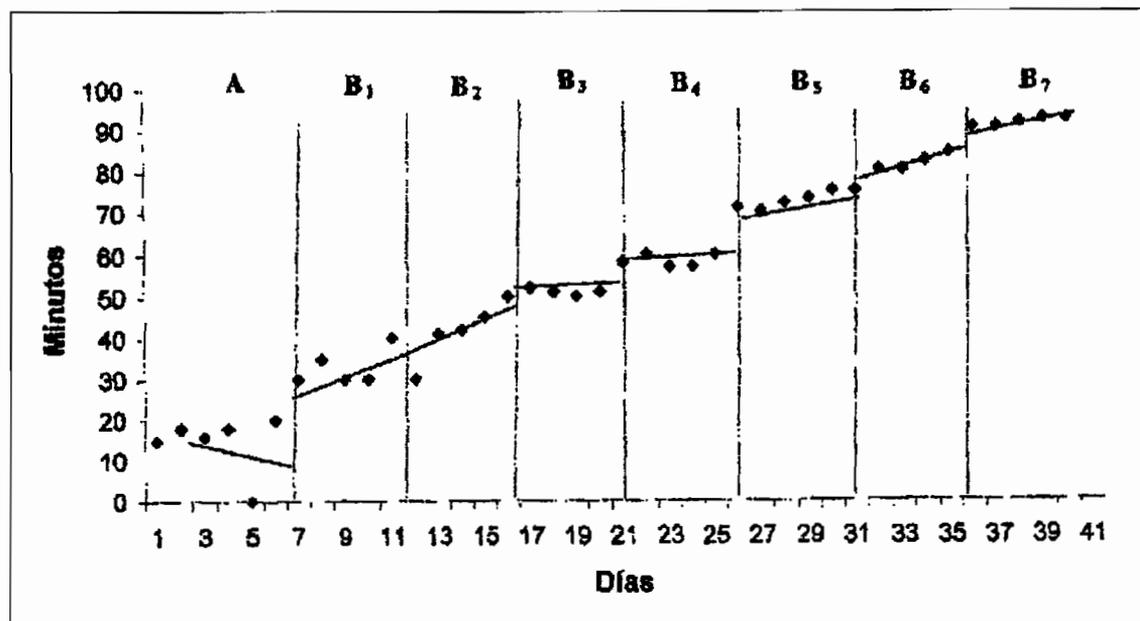


Figura 11.- Hábitos de estudio (II) (A-B₁-B₂-B₃-B₄-B₅-B₆-B₇)

3.3.- Diseños de interacción

Este tipo de diseños tiene como característica principal, que al sujeto se le obliga a alcanzar una determinada meta, con refuerzo positivo en una fase del tratamiento y negativo en otra, se vuelve a línea base, y luego se da una fase de tratamiento en la que se combinan las dos fases de tratamiento anteriores, volviéndose por último a la línea base.

a) *Hábitos de estudio III* (A-B-A-C-A-BC-A). La variable independiente sería el refuerzo económico (75 pts.) si se alcanza y supera el criterio en la fase B, y castigar al sujeto a no poder oír música si no alcanza el criterio en la fase C. El criterio a cubrir, sería estudiar al menos 20 minutos. La variable dependiente sería el tiempo dedicado al estudio de latín. (ver fig. 12)

3.4.- Diseños con línea base múltiple

Estos diseños se caracterizan por tratar conductas diferentes (independientes entre sí) con un mismo tratamiento, pero en tiempos diferentes. De tal manera que al afectar el tratamiento a una de ellas, no se vean afectadas las restantes.

a) *Modificación de tres hábitos de conducta*: tiempo de estudio diario, tiempo tardado en levantarse y tiempo dedicado al deporte ($A_1-B_1-B_1-B_1$; $A_2-A_2-B_2-B_2$; $A_3-A_3-A_3-B_3$). La variable independiente será castigar al sujeto si el criterio no es alcanzado. La variable dependiente serán los minutos dedicados a cada actividad. El tratamiento establece que el sujeto tiene que dedicar al menos 90 minutos al estudio, no tardar más de 5 minutos en levantarse de la cama y dedicar al menos 10 minutos a hacer deporte. En caso de no cumplir estos criterios, deberá escribir en una hoja el porqué de no haberlos alcanzado. (ver fig. 13)

b) *Modificación de cuatro hábitos de conducta*: ordenar el recibidor, hacer la cama, recoger la habitación y lavar los platos. ($A_1-B_1-B_1-B_1-B_1$; $A_2-A_2-B_2-B_2-B_2$; $A_3-A_3-A_3-B_3-B_3$; $A_4-A_4-A_4-A_4-B_4$) La variable independiente será un refuerzo económico de mil pesetas a la semana si realiza, al principio, una de las tareas al día, luego dos, luego tres, y por último las cuatro a la vez, si no, tendrá que pagar 200 pts. La variable dependiente será el cumplimiento o no de las tareas. (ver fig. 14)

3.5.- Diseño simple de no reversión

Estos diseños son llamados así por la no vuelta a la línea base o retirada del tratamiento.

a) *Hábitos de estudio IV*. (A-B). Se trata este de un diseño cuasiexperimental, por no poderse comprobar, al no haber segunda fase de línea base, si actúan o no la madurez y la historia con el paso del tiempo; sin embargo lo consideramos adecuado para realizar el ARIMA. La variable independiente fue el refuerzo verbal positivo cada vez que el sujeto dedicaba tiempo al estudio. La variable dependiente fue el tiempo empleado de estudio al día.

4. Resultados

4.1.- Diseños de reversión

a) *Enuresis Nocturna*. (A-B-A₁). El análisis estadístico realizado ha sido un test Kruskal-Wallis con la distribución χ^2 . De estos análisis se deduce que no hay diferencias significativas entre las distintas fases del diseño, por tanto el tratamiento no parece ser efectivo. En el análisis visual, a través de la curva de regresión, no parece encontrarse cambios significativos entre las fases de línea base y la de tratamiento, por lo que podemos concluir lo mismo, que no es

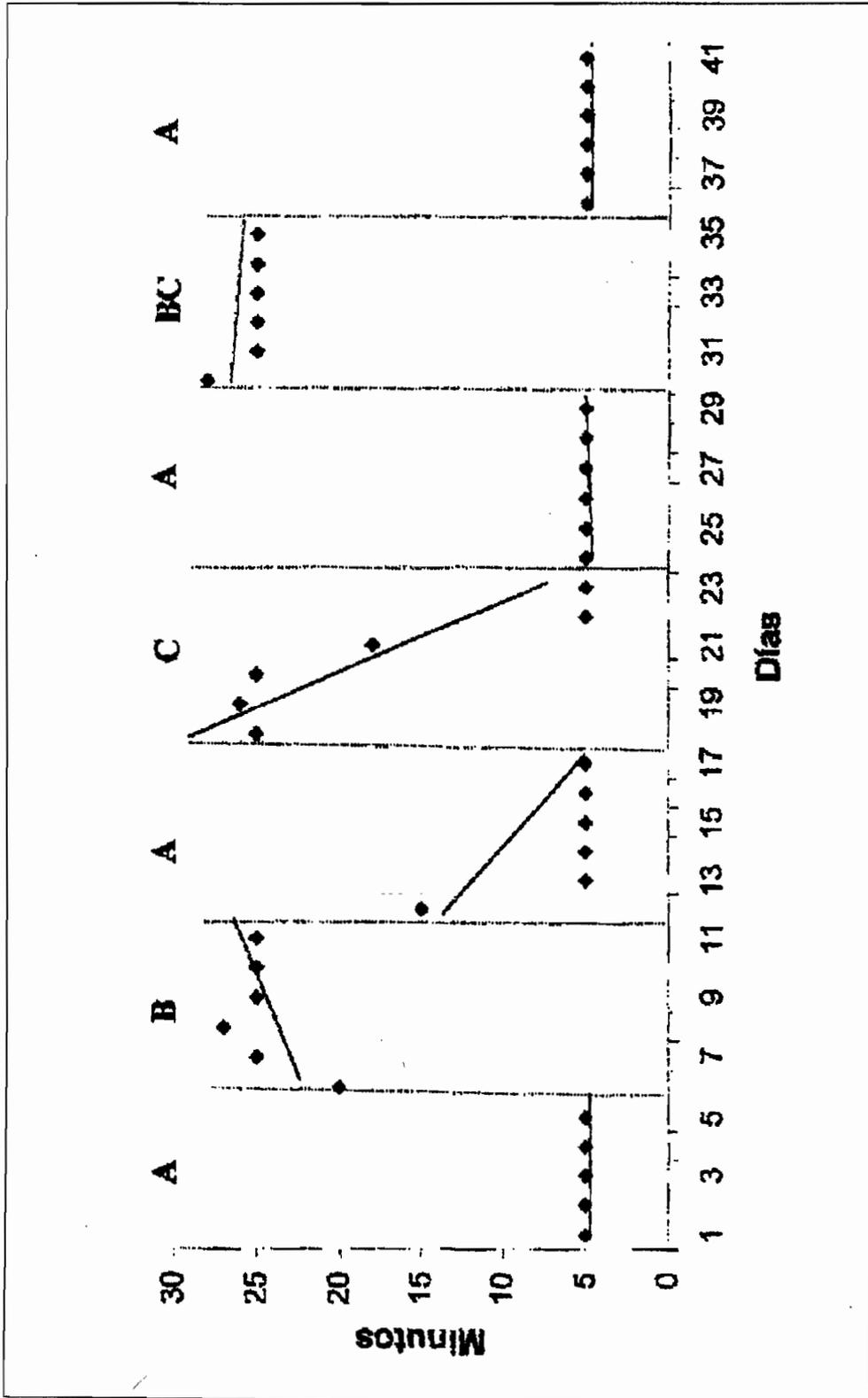


Figura 12.- Hábitos de estudio (III) (A-B-A-C-A-B-C-A)

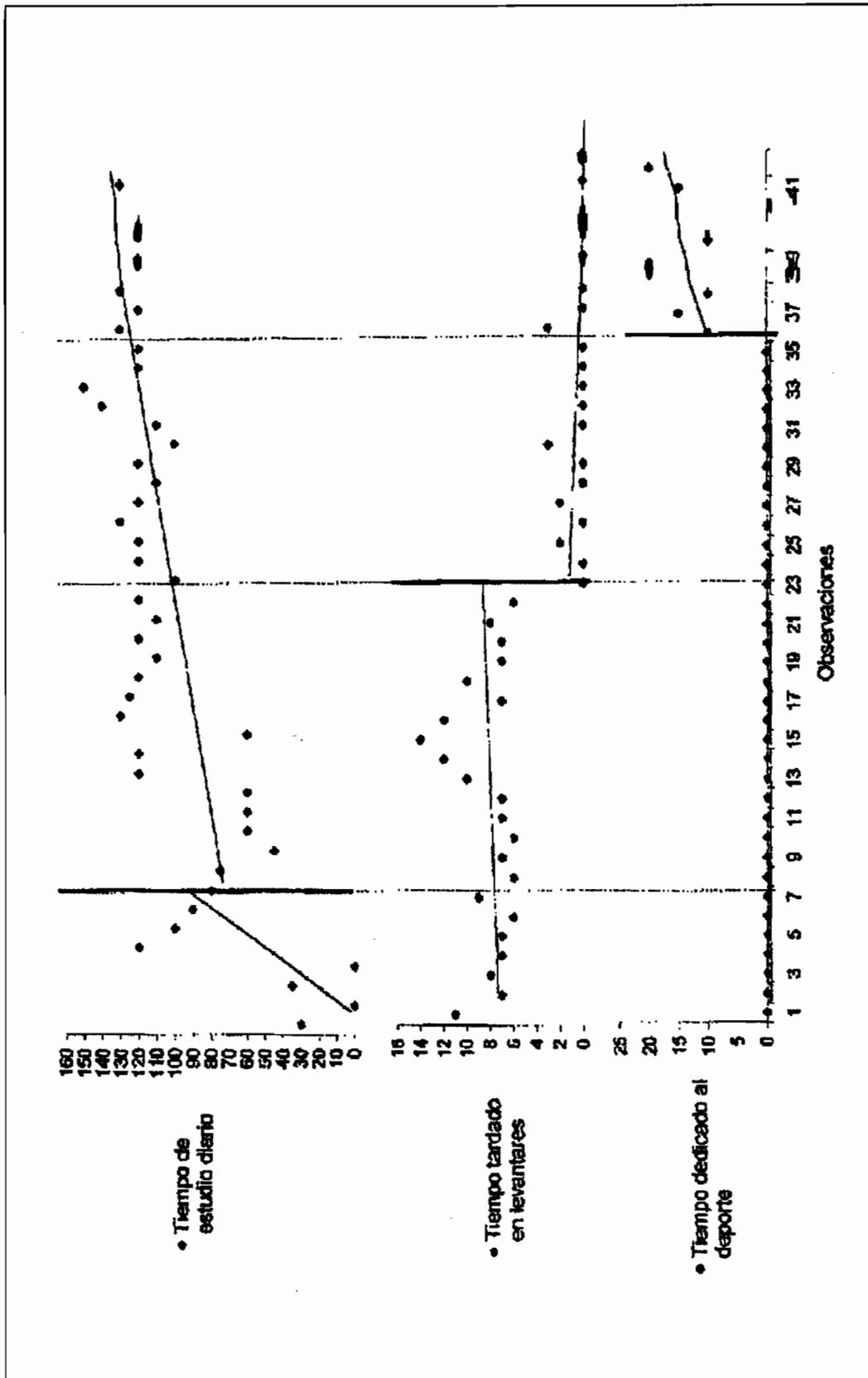


Figura 13.- Modificación de tres hábitos de conducta ($A_1-B_1-B_1-B_1; A_2-B_2-B_2; A_3-A_3-A_3-B_3$)

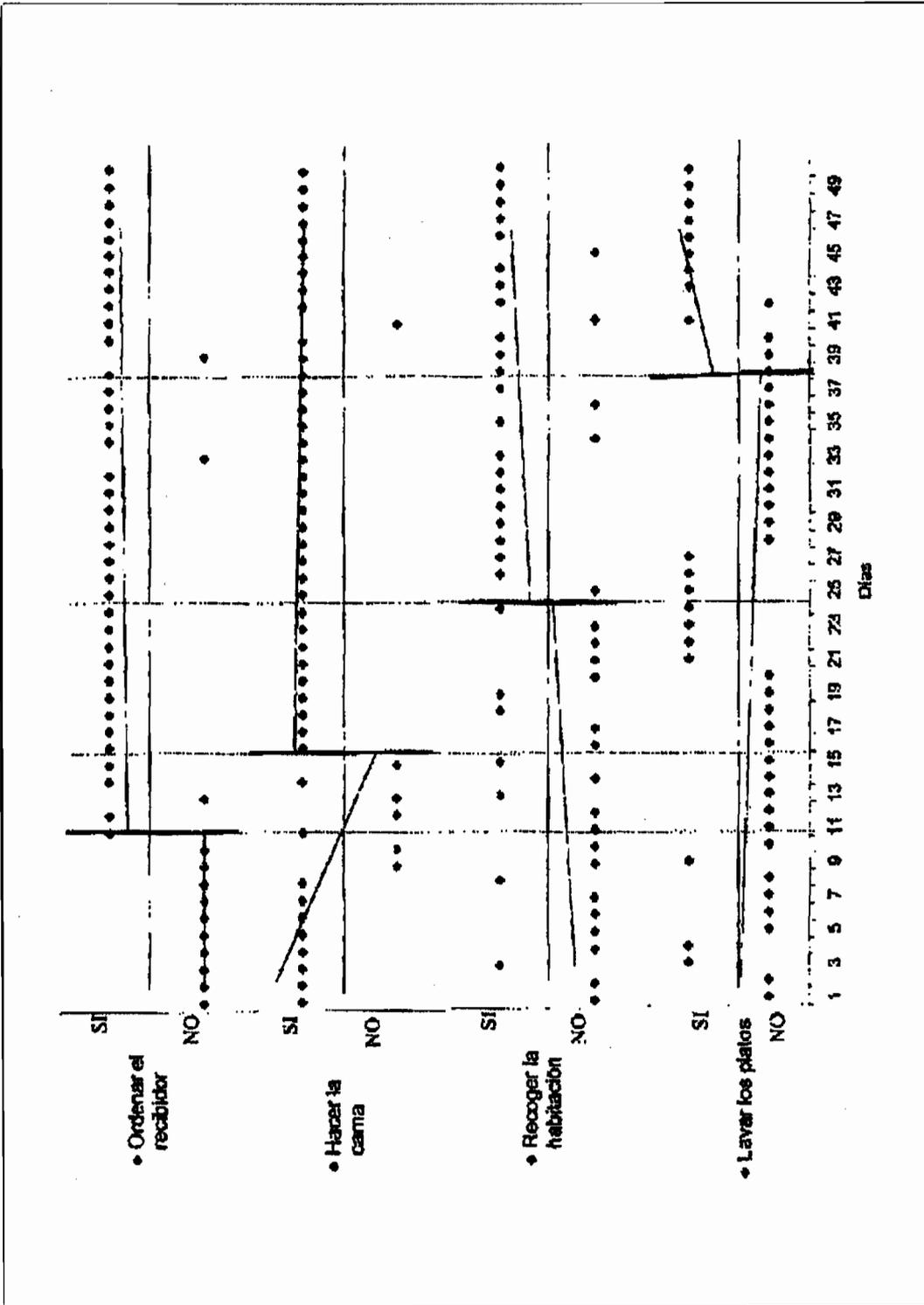


Figura 14.- Modificación de cuatro hábitos de conducta (A₁-B₁-B₁; A₂-B₂-B₂; A₃-B₃-B₃; A₄-B₄-B₄)

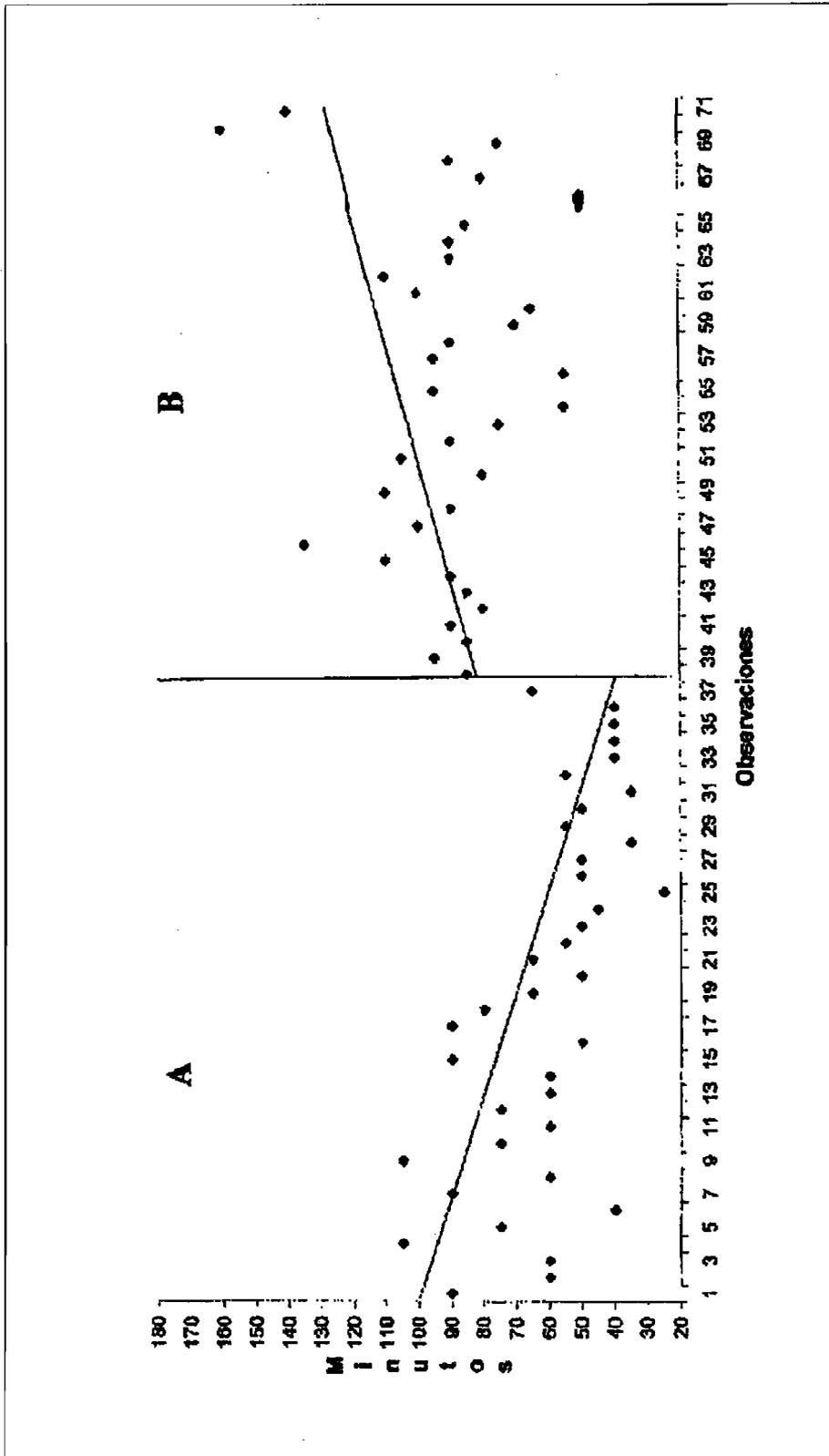


Figura 15.- Hábitos de estudio IV (A-B)

eficaz el tratamiento. En este caso llegamos a las mismas conclusiones con el análisis estadístico que con el visual.

b) *Hiperactividad*. (A-B-A₁-B₁). A partir del Kruskal-Wallis y de χ^2 podemos decir que hay diferencias significativas entre todas las fases consecutivas. Los efectos del tratamiento se dan en el sentido que se esperaba. El tratamiento afecta en igual medida a todos los momentos dentro de cada fase. (Análisis de Medidas Repetidas). En el análisis visual, si construimos una recta de regresión con las dos Líneas Base (A-A₁) y otra con los dos Tratamientos (B-B₁), vemos un cambio de nivel y no de tendencia en la conducta «Levantarse del asiento» (↔) el cambio de nivel es descendente, como corresponde a una conducta a extinguir, y en «Levantar la mano para preguntar» (D) el cambio de nivel es mayor y ascendente, como corresponde a una conducta que se está reforzando. Por lo que podemos concluir que hay concordancia entre unas y otras conclusiones.

c) *Tartamudez* (A-B-B₁). En cuanto al análisis estadístico, hemos realizado un test de Friedman, un ANOVA para medidas repetidas, un ANOVA para medidas independientes, y un test Kruskal-Wallis de los que se concluye que existen diferencias significativas entre las fases, lo que refleja una efectividad del tratamiento. En el análisis visual a través de la recta de regresión, se aprecian cambios de nivel entre las distintas fases, esto es, nos confirma la efectividad del tratamiento de igual forma.

d) *Hora de ir a dormir*. (A-B-A₁-B₁). El análisis estadístico, a través de la prueba de «t» tanto para medidas repetidas como independientes, muestra que hay diferencias significativas entre las distintas fases, además, la varianza explicada entre las fases, aumenta según se trate de Carlos (menor), Ana (mediana) y David (mayor). De igual forma, en el análisis visual, se muestran diferencias de nivel considerables entre las rectas de regresión que describen las distintas fases, pero no cambios de tendencia, por lo que el tratamiento parece ser efectivo con los tres niños.

4.2.- Diseños con cambio de criterio

a) *Obesidad*. (A-B₁-B₂-B₃). Como análisis estadístico se ha realizado el test de Friedman, un ANOVA (medidas repetidas) y el test de Kruskal-Wallis, encontrándose diferencias significativas en las distintas fases, tanto en las comidas como en las cenas. A través de las líneas de regresión (análisis visual), encontramos cambios de nivel en todas las fases y de tendencia entre la fase de línea base y la de tratamiento. (para comida y cena).

b) *Dependencia del tabaco*. (A-B₁-B₂-B₃-B₄-B₅-B₆-B₇-B₈-B₉). En este estudio el análisis estadístico no es necesario, puesto que el criterio siempre se consigue. El tratamiento es efectivo. En el análisis visual, a través de las líneas de regresión, se observan cambios de nivel en todas las fases, este cambio viene dado por el criterio.

c) *Hábitos de estudio I*. (A-B₁-B₂-B₃-B₄-B₅). Se ha realizado como análisis estadístico un ANOVA para medidas repetidas, encontrándose diferencias significativas entre la fase de línea base y las de tratamiento, excepto en la primera fase de tratamiento. Se ha realizado también un ANOVA de medidas repetidas, para ver la homogeneidad dentro de cada fase, y encontramos que esta homogeneidad existe en todas ellas. la interpretación de las rectas de regresión para el análisis visual, muestra cambios de nivel y de tendencia en todas las fases.

d) *Hábitos de estudio II*. (A-B₁-B₂-B₃-B₄-B₅-B₆-B₇). El ANOVA para medidas repetidas concluye que no hay diferencias dentro de cada fase, el ANOVA de medidas independientes muestra que no hay diferencias entre las fases. El análisis visual, a través de las rectas de regresión, no muestra cambios de nivel ni de tendencia en ninguna fase, sino tan sólo el cumplimiento del criterio.

4.3.- Diseños con interacción

a) *Hábitos de estudio III.* (A-B-A-C-A-BC-A). Se encuentran diferencias significativas entre la fase de línea base y las fases de tratamiento en el análisis estadístico (ANOVA para medidas independientes). En el análisis visual concluimos lo mismo, puesto que a través de las líneas de regresión podemos observar cambios de nivel y de tendencia en todas las fases.

4.4.- Diseños con línea base múltiple

a) *Modificación de tres hábitos de conducta:* Tiempo dedicado al estudio diario, Tiempo tardado en levantarse y Tiempo dedicado al deporte ($A_1-B_1-B_1-B_1$; $A_2-A_2-B_2-B_2$; $A_3-A_3-A_3-B_3$). Para el análisis estadístico, se ha realizado un ANOVA (para cada fase independientemente), encontrándose siempre diferencias significativas entre la fase de línea base y la de tratamiento para todas las conductas. De igual forma, con el análisis visual, encontramos cambios de nivel entre la recta de regresión de la fase de línea base y la del tratamiento en todas las conductas. Se observan cambios de tendencia también en la conducta número 3. También podemos decir, que no hay interferencias entre los distintos tratamientos, ya que cuando empieza un tratamiento esto no produce cambios en la evolución de las demás conductas.

b) *Modificación de cuatro hábitos de conducta:* Ordenar el recibidor, Hacer la cama, Recoger la habitación y Lavar los platos, ($A_1-B_1-B_1-B_1$; $A_2-A_2-B_2-B_2$; $A_3-A_3-A_3-B_3$; $A_4-A_4-A_4-B_4$). Se ha realizado un test Kruskal-Wallis (para cada tarea independientemente), encontrándose diferencias entre la fase de línea base y la de tratamiento en todas las tareas. Con el análisis visual, observamos, a través de las líneas de regresión, cambios de nivel y de tendencia entre la fase de línea base y la de tratamiento en todas las tareas; también se observa que no hay interferencia de un tratamiento a otro.

4.5.- Diseño simple de no reversion

a) *Hábitos de estudio IV.* (A-B). Para el análisis estadístico hemos utilizado como prueba no paramétrica la T de Wilcoxon, ésta indica una clara diferencia entre las dos fases, siendo los parámetros de la fase A inferiores a los de la fase B. Por otra parte, si suponemos que las variables minutos y fase no están relacionadas y hacemos la U de Mann-Whitney, obtenemos los mismos resultados, esto es, hay eficacia del tratamiento. Realizadas las pruebas paramétricas tenemos que con la prueba de «F» las varianzas son significativamente diferentes por lo que elegimos los resultados de «t» para varianzas distintas y se concluye que las dos fases son significativamente distintas, por lo que afirmamos que hay efectividad del tratamiento. En el ANOVA, obtenemos el mismo resultado, las dos fases son significativamente distintas, por lo que hay efectividad del tratamiento. Con el análisis de series temporales (ARIMA), llegamos a la misma conclusión de los anteriores análisis, que existe efectividad del tratamiento al haber diferencias significativas entre la fase de Línea Base y la del Tratamiento. En el análisis visual, se muestran las puntuaciones obtenidas y las líneas de regresión que permiten observar que se producen cambios de nivel y de tendencia significativos, por lo que podemos concluir que hay efectividad del tratamiento. La tendencia en la fase A es descendente y en la B ascendente, con lo que el tratamiento produce los efectos deseados, con el refuerzo de la conducta estudiada. Por tanto podemos concluir, que existe concordancia con cualquiera de los análisis realizados para comprobar la eficacia del tratamiento.

5. Conclusiones

Puesto que nuestro objetivo en esta investigación no era analizar la eficacia del tratamiento, sino la concordancia entre distintos métodos de análisis de datos, nos limitaremos a la interpretación de dichos resultados.

Podemos observar que a lo largo de los 11 casos estudiados se han utilizado diversas técnicas estadísticas de análisis de datos dentro del mismo experimento, con la finalidad de someter a prueba la concordancia entre los distintos análisis que han sido defendidos o criticados por diferentes autores, como pruebas paramétricas frente a no paramétricas, o medidas independientes frente a repetidas. De ello se puede concluir que en todos los casos se ha dado esta concordancia, tanto en los casos en que los tratamientos resultaban eficaces como en los que no lo eran.

Tras el estudio de los cambios de nivel y de tendencia a partir de la línea de regresión para el análisis visual, los resultados obtenidos resultan concordantes con los obtenidos mediante las diferentes técnicas estadísticas.

En el último caso (diseño A-B) ha sido posible la utilización de un análisis de Series Temporales (ARIMA), llevado a cabo por ser actualmente defendido por numerosos autores para estos diseños comprobándose nuevamente, la concordancia con los métodos de análisis anteriores; por lo que en este caso concreto no nos ha aportado mayor información, que las restantes técnicas.

Como conclusión general, y siempre dentro de la línea de planificar nuevos estudios para una mejor comprobación, defendemos la validez del análisis visual, especialmente para la Psicología Aplicada, siempre y cuando se lleve a cabo el estudio de los cambios de nivel y de tendencia (García Jiménez, 1992) mediante la línea de regresión.

El análisis estadístico presentará una indudable ayuda a la interpretación de los datos, sobre todo en los casos dudosos, en que el nivel de significación de las pruebas estadísticas puede ser el criterio de decisión sobre si es o no efectivo el tratamiento y sobre todo creemos que deben ser estudiadas nuevas técnicas que cada vez proporcionen resultados más consistentes.

En cuanto al análisis de Series Temporales (ARIMA), consideramos que aún no podemos concluir nada contundentemente, ya que solamente ha sido aplicado a un diseño, pero nos parece que es ya un indicativo importante el concluir en este estudio que su interpretación se corresponde con la obtenida mediante los restantes análisis, tanto visuales como estadísticos.

Referencias

- Box, G.E.P. y Tiao, G.M. (1965). A change in level of a nonstationary time series. *Biometrika*, 2, 181-192.
- Box, G.E.P. y Jenkins, G.M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- DeProspero, A. y Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. *Journal of Applied Behavior Analysis*, 12, 573
- Eddington, E.S. (1967): Statistical inference from N=1 experiments. *Journal of Psychology*, 65, 195-199.
- Eddington, E.S. (1969): Aproximate randomization tests. *Journal of Psychology*, 72, 143-149.
- Eddington, E.S. (1975): Randomization tests for one-subject operant experiments. *Journal of Psychology*, 90, 57-68.
- Eddington, E.S. (1980): Random assignment and statistical test for one-subject experiments. *Behavioral Assessment*, 2, 19-28.
- Fisher, R.A. (1935): *The design of experiments*. Londres: Oliver y Boyd.
- Gentile, J.R.; Roden, A.H. y Klein, R.D. (1972). An Analysis of variance model for the intra-subject replication design. *Journal of Applied Behavior Analysis*, 5, 193-198.
- Glass, G.V.; Willson, V.L. y Gottman, J.M. (1975). *Design and analysis of time-series experiments*. Boulder, Associated University Press.
- G^a-Jiménez, M.V. (1992). *El método experimental en la investigación psicológica*. Barcelona: PPU, S.A.(Ed.) Col: LCT-21.

- G^a-Jiménez, M.V.; González, M.S.; Pamos, A. y Sebastián, E.M. (1994a). Data analysis in time series experimental designs with multiple criterial, intervence and treatments. III Network Meeting of the Biometric Society Caribbean, Central American, Colomba & Venezuela. Caracas. Venezuela.
- G^a-Jiménez, M.V.; González, M.S.; Pamos, A. y Sebastián, E.M. (1994b). Data analysis in time series experimental designs with reversion of the treatment. XVII^m IBC Ontario. Canada.
- G^a-Jiménez, M.V.; Sebastián, E.M. and Domenech, A. (1995a). Congruencia entre diversos modos de analizar los datos en un experimento de caso único. IV Symposium de Metodología de las C.C del Comportamiento. Murcia. España.
- G^a-Jiménez, M.V.; Domenech, A. y Sebastián, E.M. (1995b). Diseños y análisis de datos en experimentos de series temporales. V Conferencia Española de Biometría. Valencia. España.
- G^a-Jiménez, M.V.; Domenech, A. y Sebastián, E.M. (1995c). Análisis visual y estadístico en diseños de series temporales. V Conferencia Española de Biometría. Valencia. España.
- Gottman, J.M. y Glass, G.V. (1978) Analysis of interrupted time-series experiments. En T.R. Kratochwill (Ed.) Single subject research: Strategies evaluating chance. Nueva York: Academic Press, 197-235.
- Hartmann, D.P. (1974). Forcing square pegs into round holes: Some comments on An analysis-of-variance model for the intra subject replication design. Journal of Applied Behavior Analysis, 7, 635-638.
- Hartmann, D.P.; Gottman, J.M.; Jones, R.R.; Gardner, V.; Kazdin, A.E. y Vaught, R.S. (1980): Interrupted time-series analysis and its application to bahavioral data. Journal of Applied Behavior Analysis, 13, 543-559.
- Jones, R.R.; Vaught, R.S. y Weinrott, M. (1977). Time series analysis in operant research. Journal of Applied Behavior Analysis. 10, 151-166.
- Jones et al. (1978). Effects of serial dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis. 11, 277-283
- Kazdin, A.E. (1976). Statistical analysis for single-case experimental designs. En M. HERSEN y D.H. BARLOW. Single-case experimental designs: Strategies for studying behavior change. New York, Pergemon Press.
- Keselman, M.J. y Leventhal, L. (1974). Concerning the statistical procedors enumerated by gentile et al. : Another perspective. Journal of Applied Behavior Analysis, 7, 643-645.
- Levin, J.R.; Marascuilo, L.A. y Hubert, L.J. (1978). N=Nonparametric randomization tests. En T.R. Kratochwill (Ed.). Single subject research: Strategies for evaluating change. Nueva York: Academic Press, 167-196.
- Michael, J. (1974). Statistical inference for individual organism research. Mixed belssing or curse?. Journal of Applied Behavior Analysis. 7.647-653.
- Revusky, R.H. (1967): Some statistical treatments compatible woth individual organism methodology. Journal of the Experimental Analysis of Behavior. 10, 319-330.
- Scheffe, H. (1959). The analysis of variance. New York, John Wiley.
- Shine, L.C. y Bower, S.M. (1971). A one-way analysis of variance for single subject designs. Educational and Psychological Measurements, 31, 105-113.
- Sidman, M. (1973). Tácticas de investigación científica. Barcelona. Fontanella (orig. 1960).