

# ANÁLISIS DE HÁBITOS TABAQUISTAS MEDIANTE SERIES TEMPORALES

J. ROSEL

Universidad Jaime I (Castellón)

E. ELÓSEGUI; T. RIVAS

Universidad de Málaga

## Resumen

El objetivo de esta investigación es hacer un esbozo de los tipos de tabaquismo existentes, utilizando análisis de conglomerados (*cluster analysis*) y el modelo de series temporales Box-Jenkins (ST).

Diez sujetos anotaron diariamente el número de cigarrillos que habían fumado, de los cuales, siete lo hicieron durante 84 días, y los otros tres sujetos durante 56 días (todos comenzaron en la misma fecha).

Se hizo la función de autocorrelación simple (ACF) y parcial (PACF) de los datos de cada sujeto, para posteriormente realizar un análisis de conglomerados de las ACF y de las PACF. Aparecen así tres conglomerados.

Al realizar el análisis individual de ST de cada sujeto, se observa que los cinco sujetos del primer conglomerado siguen un modelo ARIMA(1, 0, 0), los cuatro sujetos del segundo conglomerado se ajustan al modelo ARIMA(0, 0, 0)(1, 0, 0), y el sujeto del tercer conglomerado sigue un modelo ARIMA(3, 0, 0).

La media diaria de los diez sujetos sigue un modelo ARIMA(1, 0, 0)(1, 0, 0).

Se comprueba cómo hay tipos de tabaquistas con diferentes pautas estocásticas en función de lo que fuman en días anteriores.

## Abstract

The object of this research was to classify different types of smoking habits by means of cluster analysis and Box-Jenkins' time series model (TS).

Ten subjects recorded their daily cigarette consumption, seven subjects for 84 consecutive days and three for 56 days. All started recording on the same date.

Autocorrelation function (ACF) and partial ACF (PACF) were made of each subject, followed by a cluster analysis of ACF and PACF. From this, three clusters were apparent. Individual TS analysis of each subject showed that five subjects from the first cluster fit ARIMA(1, 0, 0) model, four subjects from the second cluster correspond to the ARIMA(0, 0, 0)(1, 0, 0) model, and subject from third cluster fit ARIMA(3, 0, 0) model.

The daily average of all ten subjects fit ARIMA(1, 0, 0)(1, 0, 0) model.

This demonstrates that smokers fall into different stochastic smoking patterns as a function of their consumption in previous days.

## Introducción

Son relativamente abundantes los estudios sobre el tabaquismo orientados hacia la modificación de conducta del hábito de fumar (Keutzer, Lichtenstein y Mess, 1968; Bernstein, 1969, 1970; Yates, 1970; Blittner, Goldberg y Merbaum, 1978; Cáceres, 1979; Matesanz, 1981). Igualmente, existe una línea de investigación dirigida hacia el estudio de la conducta de fumar como consecuencia de contingencias ambientales o sociales (Phanishayi, 1951; Ma-

tarazzo y Saslow, 1960; Fernández, 1987); también (aunque los resultados de diferentes investigaciones son a veces contradictorios) hay investigaciones sobre la personalidad del tabaquista (Heath, 1958; Eysenck, 1963, 1965; Simon y Primavera, 1976; Pérez, 1980). En contrapartida, no existen apenas investigaciones tendentes a la descripción del tabaquismo como hábito cuantitativo sometido a regularidades de consumo por parte del individuo fumador (Emurian, Nellis, Brady y Ray, 1982).

En el presente estudio de investigación se tiene

como objetivo comprobar qué modelos de hábitos de tabaquismo siguen los sujetos tabaquistas; para ello se ha realizado un procedimiento de estudio continuo longitudinal, registrándose día a día ininterrumpidamente el número de cigarrillos que fuma en un intervalo diario cada sujeto tabaquista; posteriormente se han analizado los resultados directos de cada sujeto mediante el modelo de series temporales de Box-Jenkins (Arнау, 1981, 1984, 1987; Arredondo, 1983; Box y Jenkins, 1976; Chatfield, 1978; Glass, Wilson y Gottman, 1975; Gottman, 1981; Gregson, 1983; Kendall, 1976; Martínez, 1987; McCain y McCleary, 1979; McCleary y Hay, 1982; Nelson, 1973; Peña, 1986; Uriel, 1985; Vallejo, 1986), para después hacer un análisis de conglomerados de las funciones de autocorrelación simples y parciales de cada sujeto (Anderberg, 1973; Cuadras, 1981; Everitt, 1977; Hartigan, 1975; Martínez, 1984). Por último, se hace el análisis de series temporales de la media diaria de cada conglomerado de sujetos y de la media total de los diez sujetos.

Habida cuenta que el hábito de fumar no es una conducta totalmente sometida al azar, sino probablemente regulada estocásticamente por ritmos de conducta cotidiana, la hipótesis general de esta investigación es la de buscar qué tipos de tabaquistas existen según la secuencia de su consumo diario de cigarrillos.

## Método

### Sujetos

Los individuos participantes en la investigación han sido alumnos tabaquistas de cuarto curso de Psicología de la Universidad de Málaga, definiéndose como tabaquistas (y consiguientemente, indicándolo como requisito para participar en la investigación) a aquel individuo que lleve fumando ininterrumpidamente durante más de dos años una media de más de diez cigarrillos al día.

### Procedimiento

Se pidieron sujetos voluntarios que estuviesen dispuestos a rellenar durante dos o tres meses ininterrumpidamente día a día, en un calendario de respuesta de autoinforme, el número de cigarrillos que habían fumado el día anterior. A los sujetos se les entregaba cada cuatro semanas el correspondiente calendario de autorregistro, siendo recogido al finalizar el mes. Todos los sujetos comenzaron a rellenar los datos en el mismo día. Diez sujetos respondieron correctamente el calendario de autoinforme, de los cuales siete lo rellenaron durante doce semanas ininterrumpidamente (es decir, ochenta y cuatro días), mientras tres sujetos lo hicieron durante ocho semanas (cincuenta y seis días).

De los diez sujetos participantes, seis eran mujeres y el resto hombres. Hubo de rechazarse los calendarios de otros dieciocho sujetos porque tenían lagunas en algunas fechas. El motivo por el que abandonaron tantos sujetos fue la excesiva «pesadez» de tener que rellenar diariamente el número de cigarrillos.

## Resultados

### Análisis de conglomerados

Como paso previo a la identificación del modelo ARIMA de cada sujeto, se comprobaron cuáles eran las autocorrelaciones de los datos de los sujetos. Los datos de cada sujeto fueron analizados mediante el programa P2T del BMDP (Dixon, 1988). En la figura 1 se representan los correlogramas con las funciones de autocorrelación (ACF) y las funciones de autocorrelación parcial (PACF) de los datos de cada uno de los sujetos participantes. Cada participante se identifica con una S seguida de un número (del uno al diez). Se puede observar cómo en las ACF y en las PACF los retardos con mayor número de correlaciones significativas son los correspondientes al 1 y al 2, así como los del 7 y el 14, siendo éstos los que contienen, por tanto, más información para la identificación de parámetros en el modelo ARIMA de Box-Jenkins.

Con el fin de clasificar a los sujetos mediante un procedimiento en el que interviniesen lo mínimo posible los criterios subjetivos de los investigadores, se hizo un análisis de conglomerados (*cluster analysis*) de los sujetos. De los diferentes procedimientos de análisis de conglomerados, se ha optado por utilizar el de la medición de las distancias de asociación de chi-cuadrado, puesto que los datos originales están medidos en frecuencias, mientras el procedimiento seguido para la clasificación de sujetos ha sido el de tipo jerárquico.

Las variables utilizadas como criterios de clasificación han sido las ACF y las PACF de los diferentes sujetos tabaquistas. Puesto que la utilización de los conglomerados que diesen en el análisis iba a ser utilizada en la modelización mediante series temporales, y que al revisar los resultados de las ACF y de las PACF se observa que las correlaciones significativas aparecen con mayor frecuencia en los retardos 1, 2, 7 y 14, y con el fin de evitar la posible multicolinealidad (Everitt, 1974), las variables de análisis de conglomerados para cada sujeto han sido las ACF de los retardos 1, 2, 7 y 14, y las PACF de los retardos 2, 7 y 14. Se ha suprimido la PACF del retardo 1 en el análisis de conglomerados, puesto que es igual a la ACF del retardo 1. El análisis de conglomerados se ha llevado a cabo en el programa P2M del BMDP (Dixon, 1988).

Los resultados obtenidos en el análisis de conglomerados de las ACF y PACF aparecen en forma de diagrama en árbol (*dendograma* de clasificación y

GRAFICO DE AUTOCORRELACIONES DE S1

RET.	CORR.	
1	0.394	I
2	0.268	+ IXXXX+XXXXX
3	0.099	+ IXXXXX+X
4	0.106	+ IXX +
5	0.002	+ I +
6	0.043	+ IX +
7	0.137	+ IXXX +
8	0.025	+ IX +
9	-0.026	+ XI +
10	-0.079	+ XXI +
11	0.071	+ IXX +
12	0.017	+ I +
13	0.229	+ IXXXXXX+
14	0.336	+ IXXXXXX+X

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S1

RET.	CORR.	
1	0.394	I
2	0.133	+ IXXXX+XXXXX
3	-0.054	+ IXXX +
4	0.063	+ XI +
5	-0.065	+ XXI +
6	0.045	+ IX +
7	0.152	+ IXXXX+
8	-0.108	+ XXXI +
9	-0.062	+ XXI +
10	-0.049	+ XI +
11	0.147	+ IXXXX+
12	-0.004	+ I +
13	0.233	+ IXXXX+X
14	0.227	+ IXXXX+X

GRAFICO DE AUTOCORRELACIONES DE S2

RET.	CORR.	
1	0.457	I
2	0.144	+ IXXXX+XXXXX
3	-0.098	+ XXXI +
4	-0.148	+ XXXXI +
5	-0.098	+ XXI +
6	-0.047	+ XI +
7	0.050	+ IX +
8	-0.010	+ I +
9	-0.037	+ XI +
10	-0.078	+ XXI +
11	0.014	+ I +
12	0.027	+ IX +
13	0.140	+ IXXXX +
14	0.127	+ IXXXX +

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S2

RET.	CORR.	
1	0.457	I
2	-0.081	+ IXXXX+XXXXX
3	-0.168	+ XXI +
4	-0.032	+ XXXXI +
5	0.012	+ XI +
6	-0.020	+ I +
7	0.068	+ IX +
8	-0.097	+ XXI +
9	-0.025	+ XI +
10	-0.036	+ IX +
11	0.095	+ XI +
12	-0.028	+ IXX +
13	0.138	+ XXI +
14	0.001	+ IXXXX +

GRAFICO DE AUTOCORRELACIONES DE S3

RET.	CORR.	
1	0.135	I
2	0.030	+ IXXX +
3	-0.114	+ IX +
4	-0.131	+ XXXI +
5	-0.015	+ XXXI +
6	0.130	+ I +
7	0.385	+ IXXX +
8	0.279	+ IXXXX+XXXXX
9	-0.021	+ IXXXXX+X
10	-0.168	+ XI +
11	-0.075	+ XXXXI +
12	-0.118	+ XXI +
13	0.136	+ XXI +
14	0.224	+ IXXX +

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S3

RET.	CORR.	
1	0.135	I
2	0.012	+ IXXX +
3	-0.122	+ I +
4	-0.103	+ XXXI +
5	0.023	+ XXXI +
6	0.130	+ IX +
7	0.348	+ IXXX +
8	0.220	+ IXXXX+XXXXX
9	-0.070	+ IXXXXX+X
10	-0.133	+ XXI +
11	0.051	+ XXXXI +
12	-0.089	+ IX +
13	0.068	+ XXI +
14	0.047	+ IXX +

GRAFICO DE AUTOCORRELACIONES DE S4

RET.	CORR.	
1	0.607	I
2	0.353	+ IXXXX+XXXXX+XXXXX
3	0.290	+ IXXXXXX+XX
4	0.281	+ IXXXXXXX+
5	0.279	+ IXXXXXXX+
6	0.264	+ IXXXXXXX+
7	0.270	+ IXXXXXXX+
8	0.225	+ IXXXXXXX+
9	0.194	+ IXXXXXX +
10	0.217	+ IXXXXXX +
11	0.280	+ IXXXXXXX +
12	0.313	+ IXXXXXXX +
13	0.283	+ IXXXXXXX +
14	0.120	+ IXXXX +

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S4

RET.	CORR.	
1	0.607	I
2	-0.024	+ IXXXX+XXXXX+XXXXX
3	0.134	+ XI +
4	0.091	+ IXXX +
5	0.086	+ IXX +
6	0.055	+ IXX +
7	0.089	+ IXX +
8	-0.017	+ I +
9	0.027	+ IX +
10	0.074	+ IXX +
11	0.125	+ IXXX +
12	0.083	+ IXX +
13	0.023	+ IX +
14	-0.196	+ XXXXI +

GRAFICO DE AUTOCORRELACIONES DE S5

RET.	CORR.	
1	0.073	I
2	-0.170	+ IXX +
3	-0.130	+ XXXXI +
4	-0.129	+ XXXI +
5	-0.067	+ XXI +
6	0.027	+ IX +
7	0.442	+ IXXXXX+XXXXX
8	0.024	+ IX +
9	-0.203	+ XXXXI +
10	-0.157	+ XXXI +
11	-0.158	+ XXXXI +
12	-0.175	+ XXXI +
13	-0.088	+ IXX +
14	0.310	+ IXXXXXX+X

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S5

RET.	CORR.	
1	0.073	I
2	-0.177	+ IXX +
3	-0.106	+ XXXXI +
4	-0.149	+ XXXI +
5	-0.098	+ XXI +
6	-0.032	+ XI +
7	0.411	+ IXXXX+XXXXX
8	-0.062	+ XXI +
9	-0.091	+ XXXI +
10	-0.094	+ XXI +
11	-0.127	+ XXXI +
12	-0.245	+ XXXI +
13	0.043	+ XXXXI +
14	0.042	+ IXX +

Figura 1. Correlogramas de las funciones de autocorrelación (ACF) y de las funciones de autocorrelación parcial (PACF) de los valores de tabaquismo de cada uno de los 10 sujetos.

GRAFICO DE AUTOCORRELACIONES DE S6

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.386						I					
2	0.210						I	I				
3	0.102						I	I	I			
4	0.057						I	I	I	I		
5	0.009						I	I	I	I	I	
6	0.124						I	I	I	I	I	I
7	0.143						I	I	I	I	I	I
8	0.071						I	I	I	I	I	I
9	-0.045						I	I	I	I	I	I
10	-0.046						I	I	I	I	I	I
11	-0.032						I	I	I	I	I	I
12	-0.125						I	I	I	I	I	I
13	-0.060						I	I	I	I	I	I
14	0.075						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S6

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.386						I					
2	0.071						I	I				
3	-0.001						I	I	I			
4	0.007						I	I	I	I		
5	-0.025						I	I	I	I	I	
6	0.145						I	I	I	I	I	I
7	0.069						I	I	I	I	I	I
8	-0.039						I	I	I	I	I	I
9	-0.111						I	I	I	I	I	I
10	-0.012						I	I	I	I	I	I
11	0.023						I	I	I	I	I	I
12	-0.131						I	I	I	I	I	I
13	0.006						I	I	I	I	I	I
14	0.128						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES DE S7

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.382						I					
2	0.137						I	I				
3	-0.008						I	I	I			
4	0.182						I	I	I	I		
5	0.086						I	I	I	I	I	
6	0.136						I	I	I	I	I	I
7	0.160						I	I	I	I	I	I
8	0.198						I	I	I	I	I	I
9	0.155						I	I	I	I	I	I
10	0.184						I	I	I	I	I	I
11	0.079						I	I	I	I	I	I
12	0.005						I	I	I	I	I	I
13	0.021						I	I	I	I	I	I
14	0.119						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S7

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.382						I					
2	-0.010						I	I				
3	-0.067						I	I	I			
4	0.245						I	I	I	I		
5	-0.073						I	I	I	I	I	
6	0.112						I	I	I	I	I	I
7	0.136						I	I	I	I	I	I
8	0.047						I	I	I	I	I	I
9	0.081						I	I	I	I	I	I
10	0.109						I	I	I	I	I	I
11	-0.072						I	I	I	I	I	I
12	-0.055						I	I	I	I	I	I
13	0.036						I	I	I	I	I	I
14	0.033						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES DE S8

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.040						I					
2	-0.268						I	I				
3	-0.177						I	I	I			
4	-0.188						I	I	I	I		
5	-0.130						I	I	I	I	I	
6	0.152						I	I	I	I	I	I
7	0.361						I	I	I	I	I	I
8	0.081						I	I	I	I	I	I
9	-0.286						I	I	I	I	I	I
10	-0.041						I	I	I	I	I	I
11	-0.023						I	I	I	I	I	I
12	-0.194						I	I	I	I	I	I
13	0.147						I	I	I	I	I	I
14	0.231						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S8

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.040						I					
2	-0.270						I	I				
3	-0.165						I	I	I			
4	-0.279						I	I	I	I		
5	-0.278						I	I	I	I	I	
6	-0.063						I	I	I	I	I	I
7	0.195						I	I	I	I	I	I
8	0.063						I	I	I	I	I	I
9	-0.193						I	I	I	I	I	I
10	0.091						I	I	I	I	I	I
11	0.030						I	I	I	I	I	I
12	-0.198						I	I	I	I	I	I
13	0.025						I	I	I	I	I	I
14	0.009						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES DE S9

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.580						I					
2	0.449						I	I				
3	0.478						I	I	I			
4	0.353						I	I	I	I		
5	0.164						I	I	I	I	I	
6	0.219						I	I	I	I	I	I
7	0.054						I	I	I	I	I	I
8	0.014						I	I	I	I	I	I
9	0.081						I	I	I	I	I	I
10	0.045						I	I	I	I	I	I
11	-0.135						I	I	I	I	I	I
12	-0.109						I	I	I	I	I	I
13	-0.082						I	I	I	I	I	I
14	-0.170						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S9

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.580						I					
2	0.170						I	I				
3	0.253						I	I	I			
4	-0.041						I	I	I	I		
5	-0.192						I	I	I	I	I	
6	0.118						I	I	I	I	I	I
7	-0.219						I	I	I	I	I	I
8	0.074						I	I	I	I	I	I
9	0.099						I	I	I	I	I	I
10	-0.005						I	I	I	I	I	I
11	-0.207						I	I	I	I	I	I
12	-0.100						I	I	I	I	I	I
13	0.063						I	I	I	I	I	I
14	-0.061						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES DE S10

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.024						I					
2	-0.067						I	I				
3	-0.150						I	I	I			
4	0.034						I	I	I	I		
5	-0.164						I	I	I	I	I	
6	-0.028						I	I	I	I	I	I
7	0.221						I	I	I	I	I	I
8	-0.010						I	I	I	I	I	I
9	-0.175						I	I	I	I	I	I
10	-0.106						I	I	I	I	I	I
11	0.106						I	I	I	I	I	I
12	-0.117						I	I	I	I	I	I
13	-0.037						I	I	I	I	I	I
14	-0.174						I	I	I	I	I	I

GRAFICO DE AUTOCORRELACIONES PARCIALES DE S10

RET.	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.024						I					
2	-0.068						I	I				
3	-0.148						I	I	I			
4	0.037						I	I	I	I		
5	-0.190						I	I	I	I	I	
6	-0.039						I	I	I	I	I	I
7	0.220						I	I	I	I	I	I
8	-0.093						I	I	I	I	I	I
9	-0.157						I	I	I	I	I	I
10	-0.059						I	I	I	I	I	I
11	0.055						I	I	I	I	I	I
12	-0.126						I	I	I	I	I	I
13	-0.043						I	I	I	I	I	I
14	0.122						I	I	I	I	I	I

Figura 1. (Continuación.)

distancia), indicando que hay tres conglomerados de sujetos (Fig. 2):

- un primer conglomerado (C1) estaría formado por los sujetos S1, S6, S7, S2 y S4;
- otro conglomerado (C2) sería el formado por S5, S8, S10 y S3;
- el tercer conglomerado (C3) lo formaría un único sujeto, el S9.

DISTANCIA AMALG.	(SUJETOS)									
	1	6	7	2	4	5	8	0	3	9
	*	*	*	*	*	*	*	*	*	*
2.970	I	I	I	I	I	+-	I	I	I	
3.230	I	I	+-	I	I	I	I	I	I	
3.421	I	---	+	I	I	I	I	I	I	
3.829	I	I	I	I	I	+-	I	I	I	
4.427	---	+	I	I	I	I	I	I	I	
6.590	I	I	---	+	I	I	I	I	I	
7.507	+-	---	+	I	I	I	I	I	I	
8.334	---	+	---	+	---	+	---	+	---	I
11.442	+-	---	+	---	+	---	+	---	+	---

Figura 2. Diagrama en árbol (dendograma) del análisis de conglomerados de los 10 sujetos, utilizando como variables de clasificación: ACF(1), ACF(2), ACF(7), ACF(14), PACF(2), PACF(7), PACF(14).

### Análisis para cada sujeto mediante series temporales ARIMA

Una vez realizado el análisis de conglomerados de las ACF y de las PACF de cada sujeto, se llevó a cabo la identificación ARIMA (p, d, q) (P, D, Q)<sub>s</sub>, de

cada sujeto mediante el método de *backcasting*, dando los resultados que aparecen en la tabla 1.

Como puede comprobarse, los sujetos del conglomerado C1, compuesto por los sujetos S1, S6, S7, S2 y S4, se ajustan a un modelo ARIMA(1, 0, 0), a lo cual correspondería una fórmula:

$$(1 - \phi_1 B) Y_t = a_t + \theta_0$$

Es decir, lo que fuma cada uno de estos sujetos del primer conglomerado es función de lo que han fumado el día anterior, habida cuenta que:

$$Y_t = \phi_1 Y_{t-1} + a_t + \theta_0$$

Los sujetos del segundo conglomerado, como se observa en la tabla 1, siguen un modelo de serie temporal ARIMA(0, 0, 0) (1, 0, 0)<sub>7</sub>, cuya formulación sería:

$$(1 - \Phi_7 B^7) Y_t = a_t + \theta_0$$

Cuya interpretación cualitativa sería que el número de cigarrillos fumados en un determinado día por cualquiera de los sujetos de este segundo conglomerado es función del número fumado en el séptimo día anterior, teniendo en cuenta que si se despeja  $Y_t$ :

$$Y_t = \Phi_7 Y_{t-7} + a_t + \theta_0$$

Igualmente, al hacer el análisis mediante serie temporal del sujeto S9, componente único del tercer conglomerado, se comprueba que sigue un modelo ARIMA(3, 0, 0), aunque el componente AR(2) no es significativo, por lo que su ecuación de ajuste quedaría:

$$(1 - \phi_1 B - \phi_3 B^3) Y_t = a_t + \theta_0$$

TABLA 1

En esta tabla se incluyen los valores del coeficiente  $\phi_1$ , su correspondiente error típico como estimador, el coeficiente  $\phi_3$ , y su error típico, el coeficiente  $\Phi_7$ , su error típico, el valor de la constante  $\theta_0$ , el estimador de los residuales de Ljung-Box (LBQ para 20 retardos), y su estimación sobre si estos residuales son ruido blanco (\*), o no son ruido blanco (N.S.) a un nivel de significación del 5 por 100

Sujeto	$\phi_1$	S.E.	$\phi_3$	S.E.	$\Phi_7$	S.E.	$\theta_0$	LBQ20	Estim. Res.
S1	0,400	0,102	—	—	—	—	8,74	22	*
S2	0,461	0,098	—	—	—	—	6,00	16	*
S3	—	—	—	—	0,475	0,093	7,00	21	*
S4	0,636	0,084	—	—	—	—	7,37	12	*
S5	—	—	—	—	0,561	0,088	5,50	22	*
S6	0,390	0,102	—	—	—	—	21,33	13	*
S7	0,383	0,102	—	—	—	—	11,10	15	*
S8	—	—	—	—	0,416	0,137	10,70	18	*
S9	0,476	0,114	0,353	0,113	—	—	2,92	18	*
S10	—	—	—	—	0,287	0,134	16,63	14	*

Nota: El estadístico LBQ de Ljung y Box sigue una distribución chi-cuadrado con un número de grados de libertad igual al de retardos tomados para el cálculo del LBQ (en nuestro caso, 20) menos el número de parámetros del modelo (p + q + P + Q); si se incluyese la constante de la serie ( $\theta_0$ ), habría que aumentar en una unidad el número de parámetros. Si el estadístico LBQ calculado sobrepasa el valor crítico de la chi-cuadrado, los residuales no son ruido blanco (Ljung y Box, 1978).

En las series que se están analizando, para la estimación del LBQ no se ha restado de los grados de libertad la unidad correspondiente a  $\theta_0$  para el cálculo de la chi-cuadrado crítica, porque esta investigación va dirigida hacia las propiedades del perfil de los tabaquistas; en el caso de que se investigase con el objetivo de realizar un pronóstico diario concreto de la tasa de tabaquismo de cada sujeto, habría sido preciso tener en cuenta el grado de libertad correspondiente al estadístico  $\theta_0$  (Makridakis, Wheelwright y McGee, 1983).

Significando que la tasa de cigarrillos fumados un día cualquiera por S9 es función de la tasa fumada el día anterior y de la del tercer día previo al del pronóstico, puesto que:

$$Y_t = \phi_1 Y_{t-1} + \phi_3 Y_{t-3} + a_t + \theta_0$$

Nótese cómo en los anteriores modelos se ha incluido  $\theta_0$  (el término constante) y aunque esta información no la proporciona la salida del programa BMDP, se ha calculado manualmente. En el texto de Box y Jenkins (1976, pág. 194) se indica la relación existente entre  $\mu_y$  (la media de la serie  $Y_t$ ) y  $\theta_0$ , una vez que  $Y_t$  ya está diferenciada, cuando no tiene componentes estacionales; generalizando esa fórmula, queda:

$$\mu_y = \frac{\theta_0}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)(1 - \Phi_{1S} - \Phi_{2S} - \dots - \Phi_{pS})}$$

### Análisis de cada conglomerado mediante series temporales

Una vez realizados los análisis ARIMA de cada sujeto, se ha pasado a hacer los análisis ARIMA de los valores medios de la tasa diaria de cigarrillos que fuman los sujetos integrantes de cada conglomerado (Anderson, 1978; Bowerman y O'Connell, 1987; Mills, 1990; Peña, 1986; Vandaele, 1983).

Los resultados mediante análisis ARIMA para los valores medios diarios de los sujetos del conglomerado C1 (véase tabla 2) han sido no significativos.

Se observa cómo el modelo ARIMA de los valores medios del conglomerado C1 no se corresponden con los modelos ARIMA de cada sujeto por separado; ello puede ser debido a que la variabilidad de la tasa de tabaquismo entre días sea mayor que la presentada entre los sujetos del conglomerado; así, cada sujeto tiende a presentar máximos y mínimos en distintos días de la semana, anulándose parcial o totalmente los efectos ARIMA(1, 0, 0) de cada sujeto por separado.

Se han analizado detenidamente mediante el modelo ARIMA los datos del conglomerado C1, no encontrándose ningún modelo adecuado, es decir, con parámetros estadísticos significativos y que dejen

unos residuales de *ruido blanco* (Box y Jenkins, 1976), por lo que el modelo de tabaquismo de la media de este conglomerado sería (habida cuenta que la media de esta serie es 19,75, y su desviación típica es 1,98):

$$Y_t = \theta_0 + a_t$$

con los valores:

$$Y_t = 19,75 + a_t$$

[siendo  $a_t \sim \text{NID}(0, 3,92)$ ]

También se ha hecho el análisis mediante series temporales de la media diaria del número de cigarrillos de los sujetos del conglomerado C2. Los resultados que se obtienen (véase tabla 2) indican que la media de los sujetos de C2 también sigue un modelo ARIMA(0, 0, 0) (1, 0, 0)<sub>7</sub>, por lo que hay una coincidencia entre el análisis individualizado de cada sujeto de C2 y el de la media de ese mismo conglomerado. Ello seguramente es debido a que hay una tendencia (en los sujetos y, por tanto, en la media de su conglomerado) a descender en la tasa de cigarrillos fumados los domingos, por lo que el ciclo semanal se repetiría regularmente (véase Fig. 3).

Obviamente, el tercer conglomerado lo forma un solo sujeto (S9), por lo que los resultados de la media diaria de este conglomerado son los mismos que los del sujeto S9.

Por último, se ha efectuado el análisis mediante series temporales ARIMA de la media diaria total de los diez sujetos (véase tabla 2). Se observa que el modelo de tabaquismo diario que siguen en conjunto los diez sujetos se ajusta a un ARIMA (1, 0, 0) (1, 0, 0)<sub>7</sub>, por lo que la ecuación de este proceso sería:

$$(1 - \phi_1 B)(1 - \Phi_7 B^7) Y_t = a_t + \theta_0$$

Lo que significa que la cantidad de cigarrillos fumados en un día cualquiera depende de la cantidad fumada el día anterior y de la que se fumó el séptimo día inmediatamente anterior. Puede afirmarse que el modelo ARIMA(1, 0, 0) (1, 0, 0)<sub>7</sub>, correspondiente a la media de la población tabaquista, es una síntesis de los modelos ARIMA de los conglomerados más numerosos (C1 y C2).

TABLA 2

Valores de los coeficientes de los modelos ARIMA de las medias diarias de los sujetos de los conglomerados C1, C2 y C3 y de la media total de los 10 sujetos, el error típico de los coeficientes, junto con la estimación de sus residuales (véase tabla 1)

Sujeto	$\phi_1$	S.E.	$\phi_3$	S.E.	$\Phi_7$	S.E.	$\theta_0$	LBQ20	Estim. Res.
C1	0,342	0,104	—	—	—	—	12,99	42	N.S.
C2	—	—	—	—	0,616	0,091	6,04	21	*
C3	0,476	0,114	0,353	0,113	—	—	2,92	18	*
Media	0,293	0,108	—	—	0,395	0,101	7,72	17	*

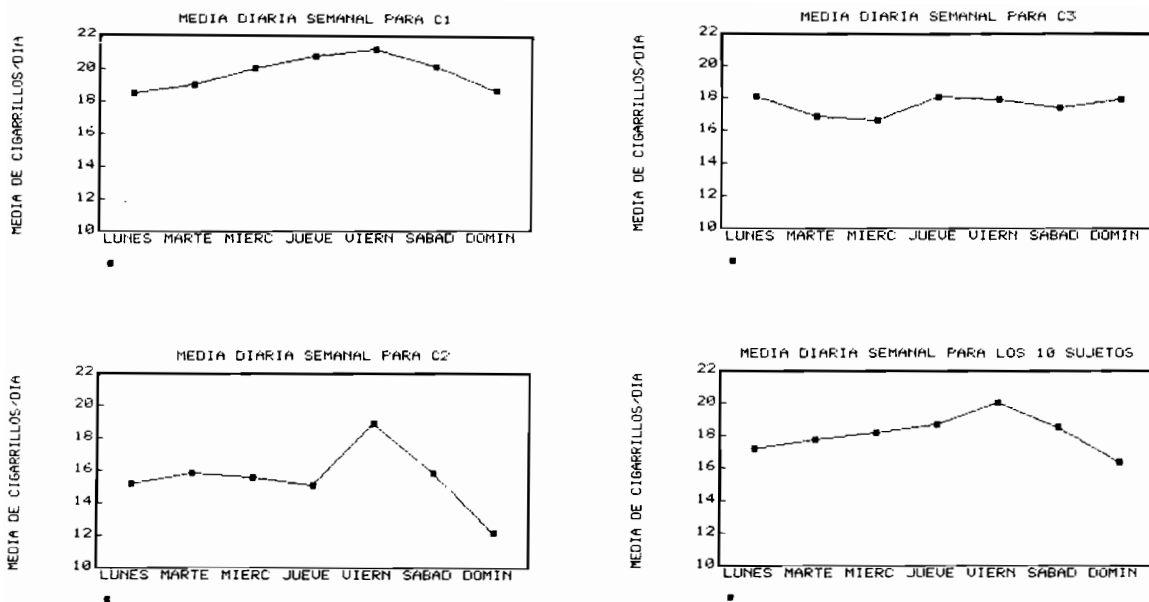


Figura 3. Representación gráfica de las medias para cada día de la semana de los conglomerados C1, C2 y C3, y de la media total de los 10 sujetos.

En la figura 3 se han representado las medias para cada día de la semana de los diez sujetos, de los conglomerados, y de la media total de los diez sujetos; en ellas se observan las diferencias relativas entre los perfiles de los distintos tipos de tabaquistas.

## Discusión

Ha de destacarse que de los diez sujetos que formaron la muestra, cinco de ellos (los cuales han formado un conglomerado al analizar las variables más significativas del ACF y del PACF) siguen una pauta diaria de consumo que se corresponde a un modelo ARIMA(1, 0, 0), lo cual significa que el mejor predictor de la cantidad de cigarrillos fumada por cada uno de estos sujetos para un día cualquiera es la tasa de cigarrillos fumada por ese mismo sujeto el día anterior [ $Y_t = f(Y_{t-1})$ ].

Una interpretación inadecuada de los modelos ARIMA(1, 0, 0) puede hacer pensar en este caso que el tabaquismo es un hábito en el cual sólo influye la tasa de cigarrillos consumida el día antes. El modelo implícitamente asume que la conducta de un día depende de la del anterior, pero ésta, a su vez, es función de la del día de antes, y así sucesivamente. Es decir, existiría una cadena de dependencias cuantitativas retardadas, en la cual sólo se precisa conocer lo acontecido el día previo. En nuestro caso, la unidad de medición ha sido el día, pero siempre ha de tenerse en cuenta el intervalo de medición utilizado. Lo indicado para este modelo es válido también para los otros modelos Box-Jenkins.

En un segundo conglomerado, integrado por cuatro sujetos, se ha comprobado que los datos de cada sujeto se ajustan a un consumo diario de cigarrillos según un modelo ARIMA(0, 0, 0) (1, 0, 0)<sub>7</sub>, por lo que se puede pronosticar, con un margen de error insignificante, la cantidad de cigarrillos fumada por cada uno de estos sujetos en un determinado día, a partir de la cantidad de cigarrillos consumida el séptimo día inmediato anterior, puesto que estos sujetos siguen unas pautas cíclicas tendentes a repetir la tasa de tabaquismo cada siete días. Se insiste en que estos sujetos no fuman dependiendo de lo que han fumado el día inmediatamente anterior, sino de lo que fumaron en el mismo día de la semana anterior [ $Y_t = f(Y_{t-7})$ ], tendiendo a fumar el lunes una cantidad, el martes otra, etc., produciéndose el máximo consumo los viernes, con gran diferencia entre los sujetos del segundo conglomerado (véase Fig. 3), lo cual indica el carácter cíclico del hábito de fumar en función de la fatiga, los hábitos o de demandas sociales que se establecen según cadencia semanal.

Un solo sujeto integra un tercer conglomerado, el cual sigue un proceso ARIMA(3, 0, 0), en el que el componente AR(2) es no significativo, por lo cual se ha de indicar que este sujeto tiende a fumar en un día cualquiera según la cantidad de cigarrillos fumados el día previo y según la tasa de cigarrillos consumidos el tercer día anterior al del pronóstico [ $Y_t = f(Y_{t-1}, Y_{t-3})$ ].

Al hacer la media de la tasa diaria de tabaquismo de los sujetos pertenecientes a cada conglomerado, se comprueba cómo el modelo ARIMA del conglomerado C1 no coincide con el de cada uno de los individuos por separado, lo cual indica que hay dife-

rencias individuales en lo que a hábitos diarios de tabaquismo se refiere, o bien que la muestra no es lo suficientemente numerosa. La media del segundo conglomerado contiene el mismo tipo de parámetros que el de cada individuo por separado, debido, probablemente, a que estos sujetos tienden a fumar cada semana repitiendo el mismo patrón.

La media total de los diez sujetos se adecua a un modelo ARIMA(1, 0, 0) (1, 0, 0)<sub>7</sub>, lo cual denota que el perfil del tabaquista promedio es una síntesis entre los dos perfiles hallados en los dos conglomerados más numerosos en cuanto al número de sujetos (C1 y C2), pudiéndose pronosticar la media de cigarrillos que fuma un determinado día ese paradigmático sujeto intermedio de la muestra de población, a partir de la cantidad fumada el día anterior y de la del mismo día de la semana inmediatamente anterior a la del pronóstico [ $Y_t = f(Y_{t-1}, Y_{t-7})$ ].

Estos resultados indican también que el hábito diario de fumar no es un fenómeno completamente aleatorio, sino que sigue unas pautas generales de consumo individual, y esas pautas pueden ser descritas y clasificadas mediante procedimientos estocásticos. De hecho, en todos los individuos de nuestra muestra son significativos los coeficientes de los componentes determinísticos del proceso estadístico descriptivo (mediante modelos ARIMA) y la varianza explicada de la tasa diaria de tabaquismo es significativamente mayor que la no explicada, puesto que simultáneamente los residuales (o errores de pronóstico) son *ruido blanco*.

Desde una perspectiva metodológica, ha de resaltarse que hay una adecuación entre un modelo de clasificación de poblaciones, como es el *análisis de conglomerados*, y un modelo de descripción de una variable temporal, el modelo ARIMA univariado de series temporales de Box-Jenkins. Desde un planteamiento sustantivo más general, probablemente las conductas referidas a hábitos cotidianos (como la ingesta, el índice de actividad física o laboral diaria, etc.) están sometidas a procesos estocásticos que pueden ser analizados y descritos mediante modelos estadísticos adecuados.

## Referencias

Anderberg, M. R. (1973). *Cluster Analysis Applications*. New York: Academic Press.

Anderson, T. W. (1978). Repeated Measurements on Autoregressive Processes. *Journal of the American Statistical Association*, 73, 362, 371-378.

Arnau, J. (1981). Uso de los modelos de series temporales como técnica de análisis de los diseños conductuales. *Anuario de Psicología*, 25, 17-35.

Arnau, J. (1984). *Diseños experimentales en psicología y en educación*. Vol. II. México: Trillas.

Arnau, J. (1987). *Técnicas de análisis de series temporales*. Barcelona: Documento de Trabajo. Departamento de Metodología. Facultad de Psicología. Universidad de Barcelona.

Arredondo, J. M. (1983). Análisis de la temperatura en ratas de laboratorio según los modelos de series temporales. *Informes de Psicología*, 415-434.

Bernstein, D. A. (1969). The modification of smoking behavior: An evaluative review. *Psychology Bulletin*, 71, 418-444.

Bernstein, D. A. (1970). The modification of smoking behavior: A search for effective variables. *Behavior Research Therapy*, 8, 133-146.

Blittner, M., Goldberg, J. y Merbaum, M. (1978). Cognitive self-control factors in reduction of smoking behavior. *Behavior Therapy*, 9, 553-561.

Bowerman, B. L. y O'Connell, R. T. (1987). *Time Series Forecasting*. Boston: Duxbury Press.

Box, G. E. P. y Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Cáceres, J. (1979). Modificación del comportamiento y hábito de fumar. *Revista de Psicología General y Aplicada*, 34, 225-243.

Cuadras, C. M. (1981). *Métodos de análisis multivariante*. Barcelona: Eunibar.

Chatfield, C. (1978). *The Analysis of Time Series: Theory and Practice*. London: Chapman and Hall.

Dixon, W. J. (ed.) (1988). *BMDP Statistical Software Manual (Vol. 1)*. Berkeley, CA: University of California Press.

Emurian, H. H., Nellis, M. J., Brady, J. V. y Ray, R. L. (1982). Event time-series relationship between cigarette smoking and coffee drinking. *Addictive Behaviors*, 7(4), 441-444.

Everitt, B. (1974). *Cluster Analysis*. London: Heinemann Educational Books.

Eysenck, H. J. (1963). Smoking, personality and psychosomatic disorders. *Journal of Psychosomatic Research*, 7, 107-130.

Eysenck, H. J. (1965). *Smoking, Health and Personality*. London: Weidenfeld and Nicolson.

Fernández, J. (1987). Causas y aspectos psicológicos del consumo de tabaco: una revisión. *Estudios sobre consumo*, 10, 119-130.

Glass, G. V., Wilson, V. I. y Gottman, J. M. (1975). *Design and Analysis of Time Series Experiments*. Boulder, Co.: Colorado Ass. University Press.

Gottman, J. M. (1981). *Time Series Analysis*. Cambridge: Cambridge University Press.

Gregson, R. A. M. (1983). *Time Series in Psychology*. Hillsdale, N.J.: L. Erlbaum.

Hartigan, J. A. (1975). *Clustering Algorithms*. New York: Wiley and Sons.

Heath, C. W. (1958). Differences between smokers and non-smokers. *Archive International of Medicine*, 101, 377-388.

Kendall, M. G. (1976). *Time Series*. London: Charles Griffin and Co.

Keutzer, C. S., Lichtenstein, E. y Mees, H. L. (1968). Modification of smoking behavior: A review. *Psychological Bulletin*, 70, 520-533.

Ljung, G. M. y Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 2, 297-303.

Makridakis, S., Wheelwright, S. C. y McGee, V. E. (1983). *Forecasting: Methods and Applications*. New York: John Wiley and Sons.

Martínez, E. (1984). Aspectos teóricos del análisis de cluster y aplicación a la caracterización del electorado potencial de un partido. En J. J. Sánchez (Ed.): *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas.

Martínez, M. R. (1987). El análisis de los datos con diseños de sujeto único. En J. Mayor y F. J. Labrador (Eds.): *Manual de modificación de conducta*. Madrid: Alhambra.

Matarazzo, J. D. y Saslow, G. (1960). Psychological and related characteristics of smokers and non-smokers. *Psychological Bulletin*, 57 (6), 493-513.



- Matesanz, A. (1981). La terapia de conducta y el hábito de fumar: problemas metodológicos y situación actual. *Análisis y Modificación de Conducta*, 15 (7), 247-286.
- McCain, L. J. y McCleary, R. (1979). The Statistical Analysis of the Simple Interrupted Time Series Quasi-Experiment. En T. D. Cook y D. T. Campbell, *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin, Co.
- McCleary, R. y Hay, R. A. (1982). *Applied Time Series Analysis for the Social Sciences*. Beverly Hills, Ca.: Sage.
- Nelson, C. R. (1973). *Applied Time Series Analysis*. San Francisco: Holden-Day.
- Mills, T. C. (1990). *Time Series Techniques for Economists*. Cambridge: Cambridge University Press.
- Peña, D. (1986). *Estadística. Modelos y métodos (Vol. II): Modelos lineales y series temporales*. Madrid: Alianza.
- Pérez, J. (1980). Personalidad y hábito de fumar: una revisión. *Revista de Psicología General y Aplicada*, 35 (4), 661-672.
- Phanishayi, R. A. (1951). Causes of smoking. *Journal of Educational Psychology*, 9, 29-37.
- Simon, W. E. y Primavera, L. H. (1976). The personality of the cigarette smoker: some empirical data. *International Journal of the Addictions*, 11, 81-94.
- Uriel, E. (1985). *Análisis de series temporales: Modelos ARIMA*. Madrid: Paraninfo.
- Yates, A. J. (1970). *Behavior Therapy*. New York: John Wiley.
- Vallejo, G. (1986). Aplicación de análisis de series temporales en diseños con  $N = 1$ : Consideraciones generales. *Revista Española de Terapia del Comportamiento*, 4 (1), 1-29.
- Vandaele, W. (1983). *Applied Time Series and Box-Jenkins Models*. New York: Academic Press.