

Revista de Psicología del Deporte
2010. Vol. 19, núm. 2, pp. 319-333
ISSN: 1132-239X

Universitat de les Illes Balears
Universitat Autònoma de Barcelona

CLAVES EN LA APLICACIÓN DEL ALGORITMO *CHAID*. UN ESTUDIO DEL OCIO FÍSICO DEPORTIVO UNIVERSITARIO

Eva Sanz Arazuri y Ana Ponce de León Elizondo*

KEYS TO APPLYING THE CHAID ALGORITHM: A STUDY OF UNIVERSITY PHYSICAL-SPORT LEISURE ACTIVITIES

KEYWORDS: CHAID, Hierarchical segmentation, Satisfaction, Physical-sport activity, Leisure.

ABSTRACT: The purpose of this article is to explain the usefulness of and the procedure involved in hierarchical segmentation based on the CHAID algorithm as a multivariate analysis technique. A study carried out on leisure physical-sport behaviour of a university population served to facilitate an understanding of this method, the exhibition of its use and the interpretation of this technique. The study aimed to define the profiles of university students according to different degree of satisfaction with their physical-sport practices, as well as verify the existence of predictors of that satisfaction when all of them are related to one other.

The article demonstrates the valuable capacity of this hierarchical segmentation technique in predicting and explaining certain behaviours, as well as determining the cause-effect relationship of these behaviours.

Correspondencia: Eva Sanz Arazuri. Departamento de Ciencias de la Educación. Universidad de La Rioja. C/ Luis de Ulloa s/n (Edificio Vives). 26004 Logroño, La Rioja. E-mail: eva.sanz@unirioja.es

¹ Investigación financiada dentro del marco del Plan Riojano de I+D+I de la Consejería de Educación, Cultura y Deporte del Gobierno de La Rioja (España) en los años 2005-2007 (Ref. 2005/17).

* Universidad de La Rioja (Logroño).

— Fecha de recepción: 8 de Julio de 2009. Fecha de aceptación: 9 de Febrero de 2010.

La investigación en el ámbito de las ciencias sociales y del comportamiento está requiriendo la utilización de técnicas propias de análisis multivariado que, utilizando la segmentación jerárquica basada en el algoritmo CHAID, permitan buscar y describir grupos poblacionales que respondan a determinados rasgos comunes, así como, explicar y pronosticar las causas de determinados fenómenos, comportamientos y situaciones con el fin de conducir programas de intervención.

A lo largo de estos últimos años se viene detectando un uso predominante de estrategias de segmentación sobre todo en estudios de marketing y empresariales, algunos de ellos son los llevados a cabo por Baron y Phillips (1994), Céspedes y Sánchez, (1996), Chi y Chen (2003), Díaz de Rada (2005), Galguera y Méndez (2004); también en el ámbito del turismo, como los de Arentze y Timmermans (2004), Chen (2003), Chung, Oh, Kim y Han (2004), Van Middelkoop, Borgers y Timmermans (2003) y el de Zhang y Fujiwara (2004), del mismo modo, en las ciencias de la salud, los estudios de Ho, Jee, Lee y Park (2004), Liu, Gao y Wong (2003), Saltini, Mazzi, Del Piccolo y Zimmermann (2004), Spratt, Keller, Szpalski, Vandeputte y Gunzburg (2004), Uter, Pfahlberg, Kalina, Kölmer y Gefeller (2004) y el de Welte, Barnes, Wiczorek y Tidwell (2004). Por el contrario, en el ámbito de la educación en general, la utilización de esta técnica sólo la encontramos en los estudios de Borden (1994), Burley, England y Beran (1996) y Thomas y Galambos (2004); si nos ceñimos al ámbito del ocio, en general, las posibilidades de encontrar estudios que utilicen esta técnica se reduce a la de Wilson y Hattingh (1992), mientras que García Ferrando (2001) y Sanz (2005) la utilizan para estudiar el ocio físico-deportivo, en particular.

La aplicación de una técnica en la investigación científica supone conocer los objetivos a los que sirve, el contexto de utilización, el planteamiento metodológico, los criterios que rigen su aplicación así como las normas básicas a tener en cuenta en el análisis e interpretación de sus resultados.

En este artículo se presenta, por una parte, la conceptualización y fundamentación teórica del análisis multivariable, así como de los métodos de dependencia y, más concretamente, de la técnica de segmentación jerárquica, enmarcando y definiendo el contexto de su utilización y profundizando en las principales investigaciones que, de una manera u otra, emplean las técnicas de segmentación jerárquica a partir del algoritmo CHAID. Por otra parte, se describe el planteamiento metodológico de la técnica de segmentación jerárquica a partir del algoritmo CHAID, a través del cual se explican los principales pasos a seguir para su aplicación con el *Answer-tree versión 3.0* y se pormenoriza cada detalle del programa estadístico con el fin de mostrar los resultados y profundizar en cómo llevar a cabo el análisis y la interpretación de dichos resultados. Para ello, nos basaremos en uno de los 20 modelos de segmentación jerárquica analizados en una investigación realizada sobre el comportamiento físico-deportivo de una población universitaria en la que se han obtenido abundantes elementos de reflexión a partir de esta técnica multivariada criterial.

Por último, se remarca la capacidad con la que cuenta este método a la hora de elevar conclusiones, determinando, en la ejemplificación, aquellos factores que influyen en el grado de satisfacción que los universitarios presentan ante la práctica físico-deportiva.

En la investigación aplicada en las ciencias sociales y del comportamiento son muchas las ocasiones que el investigador se enfrenta a fenómenos complejos que requieren para su solución del análisis multidimensional de los datos dado que son más de una variable las que intervienen directamente en el fenómeno objeto de estudio. Ello ha ocasionado que, cada vez más, los investigadores incorporen en sus estudios las llamadas técnicas de análisis multivariable que permiten analizar simultáneamente conjuntos de datos y variables, en función de unos objetivos (Bisquerra, 1989), permitiendo una mayor aproximación a la realidad social.

Si bien todas las técnicas del análisis multivariable sirven a objetivos descriptivos, tan sólo algunas de ellas van a permitir, además, pronosticar y explicar determinados comportamientos llegando a determinarse las relaciones causa-efecto de dichos comportamientos.

Atendiendo a estos objetivos, autores como Santesmases (2001), Iglesias y Sulé (2003), Sheth (1971), Kinnear y Taylor (1971), Evrard y Le Maire (1971), Dillon y Goldstein (1984) y Prieto (1985) clasificaron las múltiples técnicas que comprende el análisis multivariable en dos tipos: las técnicas multivariadas de interdependencia y las de dependencia.

Las técnicas multivariadas de interdependencia, también denominadas técnicas descriptivas (Evrard y Le Maire, 1971) o reductivas (Prieto, 1985), identifican grupos o categorías homogéneas con respecto a ciertas variables definidas sin partir de una variable dependiente o variable criterio, lo que supone analizar las relaciones entre variables considerándolas todas ellas como variables independientes. La funcionalidad de estas técnicas, en general, es describir el conjunto de objetos o sujetos con características comunes.

Sin embargo, las técnicas multivariadas de dependencia, también conocidas como técnicas explicativas (Evrard y Le Maire, 1971) o predictivas (Prieto, 1985), se diferencian de las descriptivas en que utilizan una variable criterio o dependiente para crear grupos o categorías a través de la relación de variables independientes consideradas como variables predictoras. La funcionalidad, en general, de estas técnicas es explicar el comportamiento de la/s variable/s dependiente/s en función de los cambios producidos en las variables independientes. Con ello se determina la relación causa/efecto entre las variables predictoras (independientes) y las variables criterio (dependientes).

En las investigaciones de las ciencias sociales y del comportamiento a menudo se pretende explicar el comportamiento de variables categóricas a partir de la influencia de factores también categóricos. El análisis conjunto y la segmentación jerárquica permiten este tipo de análisis.

Concretamente, la segmentación jerárquica se abre a todo tipo de variables categóricas, tanto nominales como ordinales, y tiene como finalidad diferenciar grupos de elementos homogéneos a partir de los valores adoptados en la variable dependiente en las distintas categorías de las variables independientes.

En definitiva, la segmentación jerárquica es un método de dependencia donde todos los subgrupos creados son excluyentes y exhaustivos, ya que cada sujeto u objeto analizado pertenece a un único subconjunto y la suma de los sujetos u objetos de cada uno de los segmentos se corresponde con toda la muestra analizada.

Dentro de las técnicas multivariadas de dependencia la segmentación jerárquica va más allá que otros métodos porque no sólo

revela las interacciones muestrales, sino que también prioriza la importancia de las variables predictoras (Chi y Chen, 2003). La ventaja de esta técnica con respecto a otras está en la capacidad para conocer aquellas variables independientes que presentan una fuerte interacción entre ellas, lo que supone que, la segmentación jerárquica, además de explicar el efecto principal de las variables predictoras sobre la variable criterio, también detecta la influencia de cada variable independiente sobre la variable dependiente en función del comportamiento de las demás variables independientes, si bien no cuantifica la fuerza de la relación existente entre la variable criterio y las variables predictoras.

A lo largo del siglo XX se han utilizado distintos algoritmos en la segmentación jerárquica, caso de AID (Sonquist y Morgan, 1964), THAID, CHAID (Kass, 1980), CART (Breiman Friedman, Olshen y Stone, 1984) y C4.5 (Quinlan, 1993), de ellos el más utilizado es el AID (*Automatic Interaction Detection*) o Detección Automática de Interacciones y el CHAID (*Chi-square Automatic Interaction Detection*) o Detección Automática de Interacciones a partir de Chi-cuadrado.

El algoritmo CHAID, creado por Kass (1980) y adaptado por Magidson (1994), permite trabajar con una variable dependiente categórica (nominal u ordinal).

Por otro lado, el análisis AID estudia la varianza secuencialmente mediante divisiones exclusivamente dicotómicas de la variable dependiente. Sin embargo, el CHAID permite crear en cada partición desde dos subconjuntos hasta el mismo número de grupos que de categorías de la variable predictora, consiguiendo reducir en mayor medida la varianza residual y, al mismo tiempo, mejorar la selección de otras variables explicativas. Asimismo, cuando el número de categorías de las variables

predictoras son muy diferentes entre sí, el algoritmo AID tiende a seleccionar como mejor factor explicativo a aquella variable con mayor número de categorías, produciendo un sesgo en el análisis que es rectificado por el CHAID.

Por todo ello, la aplicación del algoritmo CHAID puede resultar más conveniente que la del AID en la investigación de las ciencias sociales y del comportamiento.

El CHAID analiza todos los valores de cada variable predictora potencial a través del Chi-cuadrado, el cual refleja cuan similares o asociadas están las variables. A partir de aquí, selecciona el predictor más significativo para formar la primera partición en el árbol de decisión, de tal forma que cada nodo está conformado por aquellas categorías similares de la variable seleccionada. El proceso continúa sucesivamente hasta que el árbol queda completado.

La segmentación jerárquica empleando el algoritmo CHAID reúne unas características muy particulares que la convierten en una de las técnicas multivariadas de dependencia más apropiadas en el ámbito de las ciencias sociales y del comportamiento (Fowdar, Bandar y Crockett, 2004 y Román y Lévy, 2003).

El objetivo de la investigación que nos ocupa, enmarcada en el contexto de la comunidad universitaria de la Universidad de La Rioja (España), tiene por objeto conocer el grado de satisfacción de la colectividad universitaria de La Rioja con respecto a su práctica de actividades físicas o deportivas, al mismo tiempo que definir los perfiles de los universitarios en función de los diferentes grados de satisfacción resultantes, llegando a constatar la existencia de determinadas variables predictoras de dicha satisfacción físico-deportiva, al relacionar globalmente todas.

Para ello se establecen ciento dos variables, estructuradas en dos bloques: la dependiente, aquella cuyo comportamiento queremos explicar por medio de las independientes, que en este caso concreto mide el grado de satisfacción de cada sujeto con su propia práctica físico-deportiva, y, el bloque configurado por la ciento una variables independientes dispuestas en tres grupos: las de identificación

(quinze variables) que describen las características de la muestra analizada, las de tiempo libre (veintinueve variables) que se corresponden con la disponibilidad y ocupación del tiempo libre y las de práctica físico-deportiva (cincuenta y siete variables) correspondientes a los comportamientos, actitudes y motivaciones en torno la ejercitación motriz de tiempo libre (Tabla 1).

VARIABLES INDEPENDIENTES				
FOCO	DIMENSIONES	VARIABLES INDEPENDIENTES	Nº var.	
CARACTERÍSTICAS IDENTIFICATIVAS - 15-	Personales	Sexo	1	
		Edad	1	
		Estado civil	1	
	Laborales o Educativas	Colectivo universitario	1	
		Departamento, Servicio o Titulación	1	
		Antigüedad en la UR	1	
		Curso más alto matriculado	1	
		Tipo de dedicación	1	
		Turno laboral/clases	1	
	Ambientales	Lugar de procedencia	1	
		Lugar de residencia	3	
		Medio de transporte empleado	1	
		Tiempo de desplazamiento	1	
	TIEMPO LIBRE Y OCIO -29-	Disponibilidad de tiempo libre	Comportamental	Cantidad de TL 4
				Distribución de TL 6
Actitudinal			TL percibido 4	
Ocupación del tiempo libre		Comportamental	Ocupación del TL 14	
		Actitudinal	Satisfacción con el TL 1	
PRÁCTICA FÍSICO DEPORTIVA -57-		Historial de práctica físico-deportiva	Comportamental	Situación de práctica 2
	Años de práctica 3			
	Actividades físico- deportivas	Comportamental	Lugar de práctica 2	
			Contexto de práctica 4	
			Con quién practica 4	
	Motivos en torno a la práctica físico-deportiva	Motivaciones	Motivos de inicio 6	
			Motivos de mantenimiento 8	
			Motivos de abandono 3	
			Motivos para no practicar nunca 6	
			Motivos de práctica en la universidad 8	
			Motivos de práctica fuera de la universidad 9	

Tabla 1. Variables independientes de identificación, tiempo libre y práctica físico-deportiva.

Método

La técnica de segmentación jerárquica empleando el algoritmo CHAID, tal y como la contemplan Baron y Phillips (1994), Díaz de Rada (1999), Kass (1980) y Magidson (1994), permite identificar grupos homogéneos de sujetos de acuerdo con los valores mostrados por la variable criterio en función de la significación estadística de las variables explicativas y, a su vez, establecer una relación jerárquica de las variables explicativas en función de su nivel de significación a la hora de explicar la variable criterio.

Antes de definir los métodos causales vinculados en los que aplicar el algoritmo CHAID, se debe tener en cuenta que las variables predictoras necesariamente tienen que estar relacionadas significativamente con la variable dependiente a explicar. Esta condición viene a reforzar la afirmación de Bisquerra (1989) la cual defiende que toda segmentación jerárquica debe fundamentarse y apoyarse en análisis estadísticos previos.

Así pues, los análisis relacionales bivariados previos comprobarán si la variable criterio se correlaciona significativamente con cada una de las variables predictoras hipotéticas (condición necesaria para que se dé una relación causa-efecto entre la variable predictora y el criterio); esta exploración relacional permitirá clarificar, concretar y definir el análisis de segmentación jerárquica posterior.

El análisis relacional bivariado, fijado en nuestro estudio para un nivel de confianza del 95 por ciento, pone de relieve las relaciones existentes entre la variable dependiente y las independientes lo que lleva a obtener una primera aproximación hacia los posibles descriptores de algunos perfiles relacionados con la satisfacción con la práctica físico-deportiva de los universitarios.

Dado que la *satisfacción con la práctica físico-deportiva* es una variable ordinal, para el

análisis de su asociación con otras variables ordinales se utilizará la *D* de Somers, medida asimétrica que elimina la influencia de los pares empatados en la variable dependiente y, según Díaz de Rada (1999), nos indica *la reducción proporcional en el error cometido al predecir el ordenamiento de los casos de una variable mediante el conocimiento de la ordenación de los casos en otra variable* (p. 192). Sus valores oscilan entre -1 y $+1$, donde el valor 0 muestra la no existencia de relación y el valor absoluto $|1|$ la máxima relación. El signo señala el sentido de la asociación de manera que el signo negativo muestra una relación inversa entre variables y el positivo una relación directa.

Al tratar la asociación de la variable dependiente con variables nominales se recurre al estadístico *V* de Cramer; su rango va de 0 a 1 e indica la existencia y fuerza de la relación.

De los análisis inferenciales de la variable dependiente, *la satisfacción con la práctica físico-deportiva*, con las variables independientes de identificación, tan sólo se encuentra relación significativa con el *sexo*, mientras que con las independientes de tiempo libre se localiza, con un nivel bajo (0,229), el cruce con la *satisfacción con la ocupación del tiempo libre* y con nivel moderado la asociación con la *práctica de actividades físico-deportivas* (0,453).

Al relacionar la variable dependiente con las cincuenta y siete variables de práctica físico-deportiva se ha obtenido un total de cinco asociaciones estadísticamente significativas (Figura 1).

La satisfacción con su situación de práctica físico-deportiva se asocia con: *años de práctica físico-deportiva* (0,351), *comportamiento físico-deportivo actual y pasado* (0,290), *comportamiento físico-deportivo encubierto* (0,284), con la *práctica de deportes federados* (0,235) y con la *práctica en compañía de familiares* (0,232).

No obstante, el hecho de que una variable esté relacionada significativamente con otra no

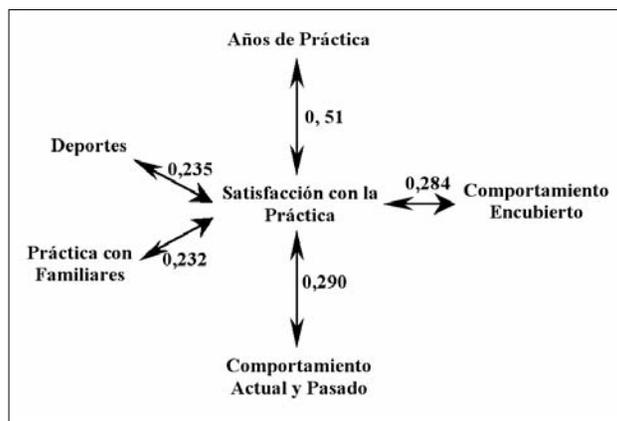


Figura 1 Análisis relacional de la satisfacción con su situación de práctica físico-deportiva con las variables de práctica físico-deportiva.

necesariamente expresa una conexión de tipo causal entre ambas, no siempre una de ellas, según Sánchez Carrión (1992), es la causa del comportamiento de la otra. En efecto, es un hecho que el grado de satisfacción con la práctica físico-deportiva es diferente en las mujeres que en los hombres, sin embargo, no por ello podemos decir que el sexo genere mayor o menor satisfacción con la práctica físico-deportiva. Cabe pensar que los verdaderos agentes de este contraste sean variables psicológicas y sociopersonales, tal y como nos advierten Pavón y Moreno (2008), al detectar que estas diferencias en el grado de satisfacción ante la práctica físico-deportiva entre hombres y mujeres se encuentran asociadas a distintas motivaciones. Mientras los hombres se sienten más satisfechos cuando juegan mejor que en otras ocasiones o mejor que los demás, las mujeres se sienten más satisfechas por motivos de salud.

Por ello, en toda segmentación jerárquica no deben existir vinculaciones significativas entre las variables predictoras con el fin de

evitar la introducción de “aparentemente ciertas variables predictoras” que puedan distorsionar los resultados. Así, entre aquellas variables predictoras relacionadas significativamente con la variable criterio que al mismo tiempo se asocien significativamente entre ellas, sólo se toman las que se consideran causantes de la variable dependiente.

Este criterio refleja la necesidad de realizar también, previamente al método de segmentación jerárquica, un análisis relacional bivariado entre todas las variables asociadas a la variable criterio a estudiar. Desde esta perspectiva se ponen de relieve las asociaciones existentes para cada combinación posible de pares de variables independientes, posibles predictoras.

Así pues al relacionar entre sí a estas ocho variables independientes se detecta un total de seis asociaciones significativas tal y como muestra la Figura 2.

Una vez conocidas las asociaciones bivariadas significativas se intenta mostrar los factores que provocan, en mayor medida, un

alto grado de satisfacción. Para ello se lleva a cabo el análisis de segmentación jerárquica que, como todo análisis multivariado de dependencia, debe comenzar con la formulación de una hipótesis causal fundamentada en teorías, estudios previos y en los resultados del análisis relacional bivariado.

Estas hipótesis causales formuladas como punto de partida se conocen como modelos causales y están compuestos por una variable dependiente relevante para el estudio (variable criterio) y aquellas variables independientes asociadas significativamente con la variable a explicar en el análisis bivariado y posibles predictoras de la variable criterio.

En nuestro caso concreto, el método causal a emplear estará constituido por la variable dependiente *grado de satisfacción con su situación de práctica actual* y las variables predictoras *satisfacción con la ocupación de tiempo*

libre, comportamiento físico-deportivo actual y pasado, práctica de deporte y práctica en compañía de familiares, dado que de las ocho variables que mostraban una asociación significativa en el análisis bivariado, cuatro se han desestimado por estar a su vez asociadas significativamente con la variable *comportamiento físico-deportivo actual y pasado*, se trata del *sexo* y la *práctica de actividades físicas*, variable esta última que, a su vez, se le asocian otras dos que también se desechan, se trata del *comportamiento físico-deportivo encubierto* y los *años de práctica* o la adherencia físico-deportiva.

La hipótesis que encabeza este modelo teórico de dependencia en la investigación que nos ocupa, viene expresada en términos como:

El grado de *satisfacción con la situación de práctica físico-deportiva-actual* de los universitarios riojanos se ve influido positivamente por

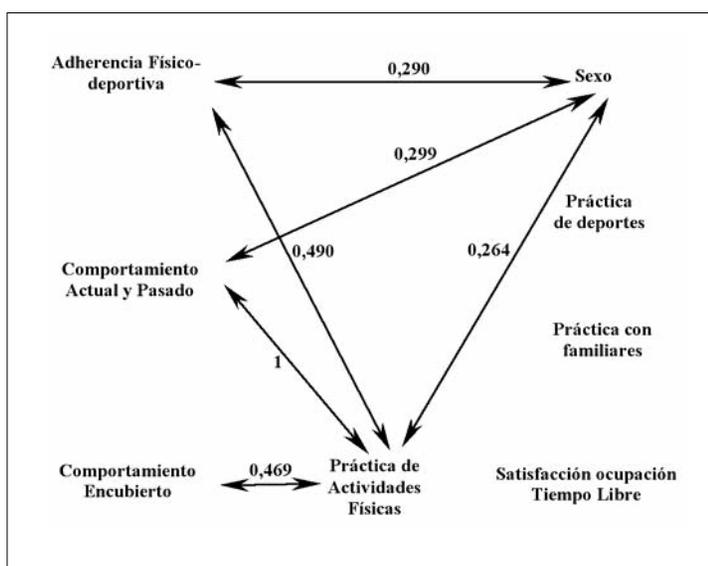


Figura 2. Análisis relacional de las variables independientes entre sí

el *comportamiento físico-deportivo actual y pasado*, así como por la *satisfacción con la ocupación del tiempo libre*. Al mismo tiempo, entre quienes se ejercitan físicamente, la participación en deportes o *en compañía de familiares* ejerce una fuerza directa.

Estos modelos causales suelen verse representados esquemáticamente mediante diagramas de paso o diagramas causales tal y como nos presenta la Figura 3.

El modelo teórico de dependencia definido teóricamente será analizado estadísticamente con el programa informático *Answertree 3.1*, distribuido bajo la licencia de SPSS, como uno de los programas más adecuados y utilizados para operar con el algoritmo CHAID.

A partir de los criterios establecidos en el análisis de segmentación, a través del algoritmo CHAID, selecciona las variables predictoras más significativas y su orden de prelación, creando distintos grupos poblacionales que explican las causas de los comportamientos de la variable dependiente.

Resultados

Los resultados de la segmentación jerárquica se presentan formando una estructura de árbol, donde el tronco se constituye por los valores obtenidos por la variable criterio para el conjunto de la población, mientras que las ramas se conforman a partir de los grupos formados por las variables predictoras.

Esta estructura arbolada (Figura 3) ofrece una visión general de los resultados obtenidos y facilita la identificación rápida de los factores predictores significativos para explicar la variable criterio así como su orden de prelación.

Así, el árbol resultante de la aplicación del CHAID al *grado de satisfacción con la situación de práctica físico-deportiva* de la comunidad

universitaria de La Rioja muestra cuatro variables significativas para explicar el *grado de satisfacción con la situación de práctica físico-deportiva*, éstas son: el *comportamiento físico-deportivo actual y pasado*, la *satisfacción con la ocupación del tiempo libre*, la *práctica de deportes federados con competiciones oficiales* y la *práctica físico-deportiva en compañía de familiares*.

De esta forma se comprueba la hipótesis para el *comportamiento físico-deportivo actual y pasado*, la *práctica de deportes* y la *satisfacción con la ocupación del tiempo libre*; también se acepta la influencia de la *práctica físico-deportiva en compañía de familiares*, pero en sentido negativo (no en el positivo que se planteaba en la hipótesis inicial).

El sistema arbolado refleja en su rama superior la variable más influyente y, a partir de ella, aparecen las demás en orden de prelación descendente. De las relaciones de las variables predictoras con la variable criterio surge una clasificación de la población estudiada según la variable dependiente.

En el caso que nos ocupa las asociaciones de las cuatro variables predictoras con el *grado de satisfacción con su situación particular de práctica físico-deportiva* generan un total de quince nodos, nueve de ellos son nodos finales que definen los perfiles de los universitarios en función de los diferentes grados de satisfacción con su situación de práctica físico-deportiva. (Figura 3).

En esta línea, la variable más significativa para explicar el *grado de satisfacción con su situación particular de práctica físico-deportiva* de los universitarios es su *comportamiento físico-deportivo actual y pasado* que muestra una relación causal positiva. De esta forma, aquellos que actualmente practican actividades físico-deportivas en su tiempo libre manifiestan un mayor grado de satisfacción que los físicamente inactivos.

Así mismo los que nunca han realizado ejercicio físico están menos satisfechos con la situación de práctica.

La segunda variable que hemos de tener en cuenta para interpretar la variable criterio para el grupo de los practicantes actuales es la práctica de deportes (entendidos como actividades físicas de competición institucionalizadas), dividiendo al grupo de practicantes en otros dos nodos, el de no deportistas y el de deportistas, este último de carácter final (nodo 5).

Sin embargo, para quienes no se ejercitan motrizmente en su tiempo libre el grado de satisfacción con su propia ocupación del tiempo libre constituye la segunda causa más determinante de la satisfacción físico-deportiva. Influye positivamente y genera 4 nodos finales, dos entre quienes nunca han practicado (nodos números 1 y 2) y otros dos entre los que han abandonado el ejercicio físico-deportivo (nodos 3 y 4).

Esta última variable incide positivamente en un tercer nivel entre los físicamente activos pero no deportistas creando tres grupos: los que se encuentra entre muy insatisfechos e indiferentes, los satisfechos y los muy satisfechos. Estos dos últimos conforman los nodos finales 8 y 9.

El grupo de los no deportistas entre muy insatisfechos o indiferentes con su propia ocupación de su tiempo libre vuelve a subdividirse en función de si practican actividades físico-deportivas acompañados de sus familiares o no. Se forman así dos nuevos nodos finales (nodos números 6 y 7).

Al comparar los nueve nodos resultantes se observa que el primero lo forman aquellos universitario riojanos que, no habiendo practicado nunca actividades físico-deportivas de tiempo libre, se muestran menos satisfechos con la ocupación de su tiempo libre. Representan al 14,25% de la comunidad universitaria riojana y son los que

presentan un menor grado de satisfacción (2,70 sobre 5), ya que se encuentran muy por debajo de la percepción del total de la muestra (3,83 sobre 5).

Quienes nunca han practicado actividades físico-deportivas, pero se declaran muy satisfechos con la ocupación de su tiempo libre (grupo 2), muestran un grado de satisfacción con su situación físico-deportiva (3,19 sobre 5) muy similar a la de los practicantes pasados menos satisfechos con su ocupación actual del tiempo libre (grupo 3: 3,25 sobre 5) o a la de los universitarios riojanos practicantes actuales no satisfechos con su ocupación del tiempo libre y que se ejercitan acompañados de sus familiares (grupo 6: 3,43 sobre 5). Todos ellos muestran un grado de satisfacción inferior a la media de la muestra (3,83 sobre 5).

Los universitarios riojanos practicantes pasados, que se encuentran muy satisfechos con su ocupación de tiempo libre (grupo 4), son los únicos inactivos actuales más satisfechos que el conjunto de la población universitaria (4,05 frente a 3,83).

Entre los practicantes actuales, los deportistas (grupo 5) se muestran los más satisfechos con su situación de práctica físico-deportiva (4,84 sobre 5).

Los que practican actividades físicas no deportivas también están más satisfechos con su situación de práctica físico-deportiva que el conjunto de la población universitaria riojana; su grado de satisfacción es mayor a medida que se encuentren más complacidos con su ocupación del tiempo libre (grupos 8 y 9).

En cambio, los que dedican su tiempo libre a actividades físicas no deportivas y no se muestran contentos con su ocupación real de tiempo libre, se sienten motrizmente más satisfechos que el conjunto de la UR si no se ejercitan en compañía de la familia (grupo 7).

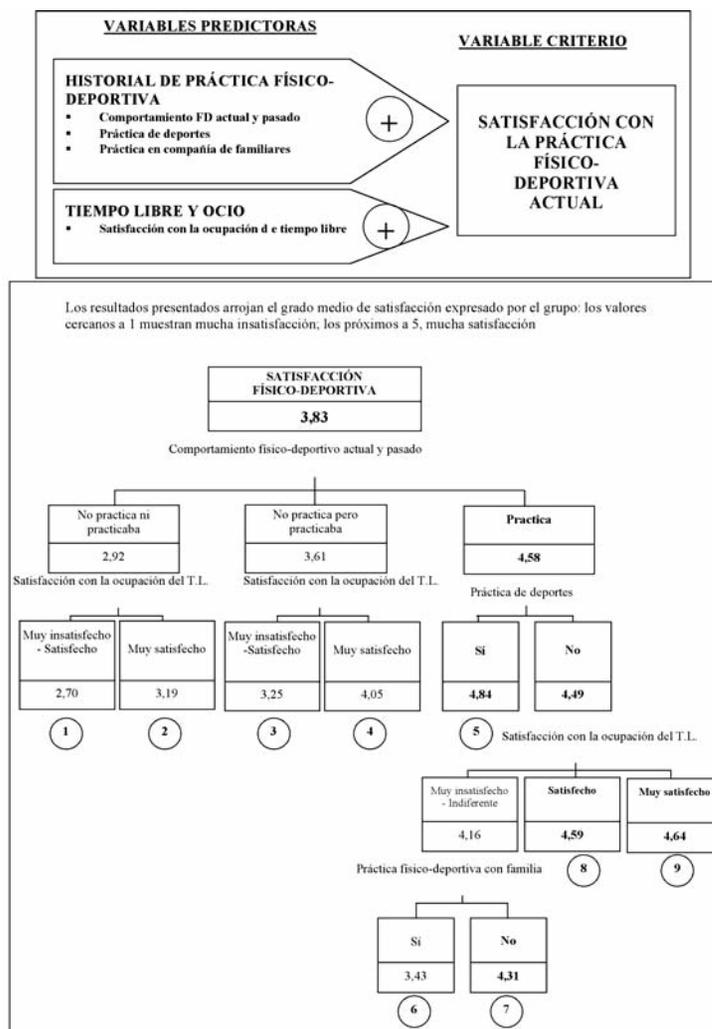


Figura 3. Modelo teórico CHAID y resultados de la segmentación jerárquica para el grado de satisfacción con la situación de práctica físico-deportiva.

Discusión

En el estudio concreto aquí presentado, la técnica de segmentación jerárquica empleando el algoritmo CHAID nos ha

permitido conocer las variables que mejor describen el grado de satisfacción con su situación de práctica físico-deportiva en los universitarios de la Universidad de La Rioja. Éstas son el realizar o no y/o haber realizado

o no actividad físico-deportiva, el grado de satisfacción con la ocupación del tiempo libre, practicar o no deportes y practicar o no en compañía de familiares. Así mismo nos ha permitido determinar los perfiles de los universitarios riojanos en función de diferentes grados de satisfacción con su situación de práctica físico-deportiva.

A partir de las aportaciones ofrecidas por los resultados se elevan las principales conclusiones y se sugiere una serie de iniciativas en torno a la práctica físico-deportiva de tiempo libre en la población universitaria que nos conduzcan a la obtención de un mejor aprovechamiento del tiempo libre y de una mayor calidad de vida.

Los físicamente inactivos que no se encuentran muy satisfechos con su ocupación del tiempo libre tampoco se sienten muy contentos con su situación físico-deportiva, lo que les convierte en posibles practicantes potenciales ya que se muestran susceptibles al cambio, a quienes presentan mayor predisposición a la ejercitación motriz. A este respecto, las entidades que proporcionan servicios físico-deportivos deberían tenerlos en cuenta intentando producir el ingreso de estos en la práctica motriz; para ello, han de planificar su oferta atendiendo a las actitudes,

aspiraciones, motivaciones y expectativas de estos universitarios posibles practicantes potenciales.

Si a los practicantes de actividades no deportivas que no están contentos con la forma de ocupar su tiempo libre se les proporcionara la posibilidad de ejercitarse motrizmente separados de su familia, se conseguiría aumentar el grado de satisfacción con su comportamiento físico reduciendo la posibilidad de su abandono. Por lo que condría ofertar paralelamente unas actividades dirigidas a los hijos y otras a los padres; de esta manera los padres aprovechan el tiempo de práctica de sus hijos para su propia ejercitación motriz y se evaden por un momento de sus obligaciones familiares.

Estos casos sugieren la necesidad de una educación físico-deportiva de ocio continua, a través de programas de intervención, como el planteado por Pérez y Delgado (2003), encauzados, por un lado a producir modificaciones en las actitudes hacia la práctica físico-deportiva orientada a la salud y, por otro, a ayudar, desde la propia práctica, a apreciar sus ejercitaciones motrices de tiempo libre como experiencias valiosas, que se conviertan en vivencias de un verdadero ocio, divertido, satisfactorio, creativo y autotélico.

CLAVES EN LA APLICACIÓN DEL ALGORITMO CHAID. UN ESTUDIO DEL OCIO FÍSICO DEPORTIVO UNIVERSITARIO

PALABRAS CLAVE: CHAID, Segmentación jerárquica, Satisfacción, Actividad físico-deportiva, Ocio.

RESUMEN: El propósito de este artículo es argumentar y explicar la utilidad y el procedimiento de la segmentación jerárquica basada en el algoritmo CHAID como técnica de análisis multivariado. Para facilitar la comprensión de este proceso, la exposición del uso, así como, la interpretación de esta técnica nos hemos servido de una investigación realizada sobre el comportamiento físico-deportivo de tiempo libre de una población universitaria. Este estudio busca definir los perfiles de los universitarios en función de los distintos grados de satisfacción con su práctica físico-deportiva, así como, constatar la existencia de determinadas variables predictoras de dicha satisfacción al relacionar globalmente todas.

En el artículo queda demostrada la valiosa capacidad que posee esta técnica de segmentación jerárquica para pronosticar y explicar determinados comportamientos, así como para determinar la relación causa-efecto de dichos comportamientos.

CHAVES NA APLICAÇÃO DO ALGORITMO CHAID. ESTUDO LAZER FISICO-DESPORTIVO UNIVERSITÁRIO

PALAVRAS-CHAVE: CHAID, Segmentação hierárquica, Satisfação, Actividade físico-desportiva, Lazer.

RESUMO: O propósito deste artigo é argumentar e explicar a utilidade e o procedimento da segmentação hierárquica baseada no algoritmo CHAID como técnica de análise multivariada. Para tal, foi levada a cabo uma investigação sobre o comportamento físico-desportivo de tempo livre de uma população universitária visando facilitar a compreensão deste método, a demonstração do seu uso, bem como a interpretação desta técnica. Este estudo procura definir os perfis dos universitários em função dos diferentes graus de satisfação com a sua prática físico-desportiva, assim como constatar a existência de determinadas variáveis predictoras dessa mesma satisfação quando relacionadas globalmente.

Neste artigo é demonstrada a valiosa capacidade que possui esta técnica de segmentação hierárquica para prever e explicar certos comportamentos, assim como para determinar a sua relação de causa-efeito.

Referencias

- Arentze, T. A. y Timmermans, J.P. (2004). A learning-based transportation oriented simulation system, *Transportation Research Part B: Methodological*, 38 (7), 613-633.
- Barón, S. y Phillips, D. (1994). Attitude Survey Data Reduction Using CHAID: An Example in Shopping Centre Market Research. En Hooley, G. J. y Hussey M. K., *Quantitative Methods in Marketing* (75-88). Londres: Academic Press.
- Bisquerra, R. (1989). *Introducción conceptual al análisis multivariable*. Vol I y II. Barcelona: PPU
- Borden, V. M. H. (1994). Segmenting Student Markets with a Student Satisfaction and Priorities Survey. *AIR 1994 Annual Forum Paper*, 4-27.
- Breiman, L., Friedman, J. H., Olshen, R. A. y Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth.
- Burley, H., England, D. y Beran, P. (1996). A CHAID analysis of a diagnostic writing sample as a placement tool for Freshman Composition. *Annual Meeting of the Southwest Educational Research Association*, 4-33.
- Céspedes, J. y Sánchez, M. (1996). Tendencias y desarrollos recientes en métodos de investigación y análisis de datos en dirección de empresas. *Revista Europea de Dirección y Economía de la Empresa*, 5 (3), 23-40.
- Chen, J. S. (2003). Market segmentation by tourists' sentiments. *Annals of Tourism Research*, 30 (1), 178-193.
- Chi, C. F. y Chen, C. L. (2003). Reanalyzing occupational fatality injuries in Taiwan with a model free approach. *Safety Science*. 41 (8), 681-700.
- Chung, K. Y., Oh, S. Y., Kim, S. S. y Han, S. Y. (2004). Three representative market segmentation methodologies for hotel guest room customers. *Tourism Management*. 25 (4), 429-441.
- Díaz de Rada, V. (1999). *Técnicas de análisis de datos para investigadores sociales. Aplicaciones prácticas con SPSS para Windows*. Madrid: RA-MA.
- Díaz de Rada, V. (2005). Measure and control of non-response in a mail survey. *European Journal of Marketing*, 39 (1-2), 16-32.

- Dillon W.R. y Goldstein M. (1984). *Multivariate analysis. Methods and applications*. Nueva York: John Wiley.
- Evrard y Le Maire, P. (1976). *Information et decision en marketing*. Paris: Dalloz.
- Fowdar, J., Bandar, Z. y Crockett, K. (2004). Inducing fuzzy decision trees in non-deterministic domains using CHAID. En Barr, V. y Markou, Z. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, 838-843. California: AAAI Press.
- Galguera, L. y Méndez, M.P. (2004). Una aproximación empírica al marketing de relaciones. Las tarjetas de fidelización de las compañías aéreas. *Esic-Market*, 119, 223-256.
- García Ferrando M. (2001). *Los Españoles y el Deporte: Prácticas y comportamientos en la última década del siglo XX*. Madrid: Consejo Superior de Deportes.
- Ho, S. H., Jee, S. H., Lee, J. E. y Park, J. S. (2004). Analysis on risk factors for cervical cancer using induction technique. *Expert Systems with Applications*, 27 (1), 97-105.
- Iglesias Antelo, S y Sulé Alonso, M. A. (2003). Introducción al análisis multivariable. En Lévy Magin, J. P. y Varela Mallou, J. (Eds.) *Análisis Multivariable para las Ciencias Sociales*, (pp. 1-41). Madrid: Pearson Educación.
- Kass, G. (1980) An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29 (119-127).
- Kinnear, P. R. y Taylor J. R. (1971). Multivariate methods in marketing research. A further attempt at classification. *Journal of Marketing*, 35 (4), 56-59.
- Liu, J. Y., Gao, S. y Wong, K. S. (2003). Optimal flow velocity of transcranial Doppler in diagnosing and grading middle cerebral artery stenosis in comparison with magnetic resonance angiography. *Chinese Journal of Neurology*, 36 (6), 458-460.
- Magidson, J. (1994). The CHAID Approach to Segmentation Modeling: Chi-square Automatic Interaction Detection. En Bagozzi, R. P., (Ed.) *Advanced Methods of Marketing Research* (pp. 118-159). Oxford: Blackwell.
- Mayntz, R. (1976). *Introducción a los métodos de sociología empírica*. Madrid: Alianza Universidad.
- Pavón Lores A. y Moreno Murcia, J. A. (2008). Actitud de los universitarios ante la práctica físico-deportiva: diferencias por géneros. *Revista de Psicología del Deporte*, 17, 7-23.
- Pérez I. J. y Delgado, M. (2003). Modificación de las actitudes del alumnado de secundaria hacia la práctica de actividad física orientada a la salud tras un programa de intervención. *Revista de Psicología del Deporte*, 12, 165-179.
- Prieto G. (1985). Análisis multivariable. En De la Orden Hoz, A. (Ed.) *Investigación educativa*, (pp. 26-28). Madrid: Anaya.
- Quinlan, J. R. (1993). *Programs for Machine Learning*. California: Morgan Kaufmann Publishers.
- Román González, M^a. V. y Lévy Mangin, J. P. (2003). Clasificación y segmentación jerárquica. En Lévy Magin, J. P. y Varela Mallou, J. (Eds) *Análisis multivariable para las Ciencias Sociales*, (pp. 567-630). Madrid: Pearson Educación.
- Saltini, A., Mazzi, M. A., Del Piccolo, L. y Zimmermann, C. (2004). Decisional strategies for the attribution of emotional distress in primary care. *Psychological Medicine*, 34 (4) 729-739.
- Sánchez Carrión, J. J. (1992). *Análisis de tablas de contingencia*. Madrid: Centro de Investigaciones Sociológicas y Siglo XXI de España Editores S. A.
- Santesmases Mestre, M. (2001). *DYANE Versión 2. Diseño y análisis de encuestas en investigación social y de mercados*. Madrid: Pirámide.

- Sanz Arazuri, E. (2005). *La práctica físico-deportiva de tiempo libre en universitarios: análisis y propuestas*. Logroño: Universidad de La Rioja.
- Sheth, J. N. (1971). The multivariate revolution in marketing research. *Journal of Marketing*, 35 (1), 13-19.
- Sonquist, J. N. y Morgan, J. A. (1964). *The detection of interaction effects*. Monográfico nº 35. The University of Michigan: Survey Research Center, Institute for Social Research.
- Spratt, K. F., Keller, T. S., Szpalski, M., Vandeputte, K. y Gunzburg, R. (2004). A predictive model for outcome after conservative decompression surgery for lumbar spinal stenosis. *European Spine Journal*, 13 (1), 14-21.
- Thomas, E. H. y Galambos, N. (2004). What satisfies students? Mining student-opinion data with regresión and decisión tree analysis. *Research in Higher Education*, 45 (3), 251-269.
- Uter, W., Pfahlberg, A., Kalina, B., Kölmer, K. F. y Gefeller, O. (2004). Inter-relation between variables determinig constitutional UV sensitivity in Caucasian children, *Photodermatology Photoimmunology and Photomedicine*, 20 (1), 9-13.
- Van Middelkoop, M., Borgers, A. y Timmermans, H. (2003). Inducing heuristic principles of tourist choice of travel mode: A rule-based approach. *Journal of Travel Research*, 42 (1), 75-83.
- Welte, J. W., Barnes, G. M., Wiczorek, W. F. y Tidwell, M. C. (2004). Gambling participation and pathology in the United States. A sociodemographic analysis using classification trees. *Addictive Behaviors*, 29 (5), 983-989.
- Wilson, G. D. H. y Hattingh P. S. (1992). Environmental Preferences for Recreation within Deprived Areas: the case of black townships in South Africa. *Geoforum*, 23 (4), 477-486.
- Zhang, J. y Fujiwara, A. (2004). Development of methodology for analyzing travel patterns in context of developing countries, *Proceedings of Conference on Traffic and Transportation Studies, ICTTS*, 4, 222-234.

Agradecimientos

Agradecemos a la Consejería de Educación, Cultura y Deporte del Gobierno de La Rioja el soporte financiero concedido al estudio dentro del Plan Riojano de I+D+I en el año 2005 (Ref. 2005/17).