

ECONOMÍA DEL OCIO Y EL TURISMO EN ESPAÑA: APROXIMACIÓN MEDIANTE TÉCNICAS ESTADÍSTICAS MULTIVARIANTES

Patricio Sánchez Fernández (*)

M. Esther López Vizcaino (**)

(*) Universidade de Vigo

(**) Instituto Galego de Estatística

1. INTRODUCCIÓN.

La presente comunicación tiene como objetivo explorar la situación de las actividades de ocio y turismo a nivel provincial en España. Conscientes de la terciarización que se está produciendo en economías desarrolladas como la española, estos sectores en todos sus componentes adquieren un papel destacado en las mismas. A través de este trabajo pretendemos adentrarnos en el conjunto de actividades como las anteriormente mencionadas realizando esta tarea con una perspectiva amplia y genérica sin ánimo de exclusividad.

Para ello se lleva a cabo un análisis multivariante de una serie de variables relacionadas con estas actividades para las distintas provincias. En concreto, los datos son sometidos a un análisis factorial, cluster y discriminante que nos permite obtener una serie de características sobre la realidad del ocio en España, así como realizar una comparación entre las distintas provincias.

Somos conscientes de la complejidad del tema que pretendemos abordar, pero lo hacemos con el convencimiento de que la importancia de actividades de ocio y turismo, como parte integrante del sector servicios, demanda una mayor atención a las mismas. En cuanto a las técnicas estadísticas multivariantes, su empleo se viene mostrando muy adecuado en el seno de las ciencias sociales, reflejando una gran operatividad y versatilidad en temas poco estudiados como es el que se está tratando.

2. VARIABLES Y MÉTODOS UTILIZADOS.

La pregunta de qué variables utilizar en un estudio sobre las actividades de ocio y turismo en España parte de la premisa de que estas actividades en dimensión espacial tienen una estructura típica multivariante, ya que no se puede medir directamente mediante una variable sino que es resultado de un conjunto de características variadas.

Es por ello que se consideró que la mejor forma de proceder a su estudio era mediante técnicas estadísticas multivariante sobradamente contrastadas para llegar a unos resultados que proporcionasen una primera visión de la realidad de las distintas provincias. Así se partió de un conjunto de información muy amplia que engloba diversas parcelas del ocio y el turismo. Éste fue sometido a análisis, teniendo en cuenta criterios habituales en la literatura especializada, llegando finalmente a retener 22 variables que recogen con cierto detalle los aspectos más importantes del conjunto de actividades objeto de estudio.

En concreto, las variables seleccionadas, todas ellas referidas al año 1998, se presentan a continuación:

Cuadro 1: Variables seleccionadas y unidad de medida

Variable	Unidad
Bibliotecas	número
Películas distintas exhibidas	número
Prensa escrita	Nº de ejemplares vendidos
Automóviles	número
Actividades de bares y restauración	número
Salas escénicas	número
Emisoras de radio	número
Plazas en establecimientos hoteleros	número
Establecimientos hoteleros	número
Sesiones música clásica	número
Gasto medio por espectador en el cine	Importe
Producción editorial	Nº de títulos
Espectadores de cine	número
Teléfonos	número
Actividades comerciales de grandes superficies	número
Actividades comerciales de no alimentación	número
Espectadores de espectáculos escénicos	número
Pernoctaciones en establecimientos hoteleros	número
Actividades comerciales minoristas	número
Audiencia radio	Nº de oyentes
Licencias de caza y pesca	número
Recaudación en espectáculos escénicos	importe

Con respecto a estas variables retenidas, hay que destacar que son todas variables relativizadas por el número de habitantes de la provincia correspondientes al año 1998. Esta relativización es importante porque pone de manifiesto la estructura provincial respecto de un concepto determinado, lo cual constituye un indicador de cada campo. Por otro lado, hay que considerar las unidades de las variables, puesto que si las variables emplean escalas diferentes los resultados pueden ser equívocos o de difícil interpretación. De ahí que en este caso se procediese a la estandarización de las variables relativizadas.

Asimismo hay que dejar constancia que, a pesar de la flexibilidad existente a la hora de seleccionar las variables (recogiendo éstas tanto del lado de la oferta como de la demanda), sí se fue riguroso en cuanto a su relación directa con las actividades de ocio y el turismo y no la mera influencia en las mismas. Así en particular, cabe señalar que la variable renta y, de modo fundamental, la renta disponible incide en actividades como las aquí recogidas pero no puede ser considerada exclusivamente como un variable de ocio ya que su ámbito se extiende más allá de este sector.

Algo semejante ocurre con las variables climáticas que también inciden en los distintos aspectos estudiados. Sin embargo, igual que en el caso anterior, no se puede considerar propiamente una variable de este campo y, además, su influencia se produce

en dos sentidos. En primer lugar, la existencia de un clima más favorable llevaría a pensar que, dadas las mayores posibilidades en cuanto a la oferta turística existente, actividades como las aquí recogidas tendrían una menor importancia en términos relativos. En sentido opuesto, esta circunstancia es motivo de una mayor presencia de visitantes que, generando una importante población turista flotante, repercutirían en el conjunto de actividades económicas de estas zonas y, por tanto, en las de ocio.

En cuanto a los métodos empleados, resulta un hecho evidente que el número de variables finalmente seleccionadas no permite interpretar con claridad y precisión la situación relativa de cada provincia. Además, por otra parte se comprueba que la información seleccionada es fuertemente redundante. Es por ello que ante esta situación se intentó buscar una técnica que permitiese reducir la dimensión del conjunto de variables observables llegando a un número más reducido que, por su importancia conceptual, las puedan sustituir.

Con este objetivo se utilizaron algunas técnicas multivariantes que constituyen uno de los ejemplos más claros de utilización de las técnicas estadísticas en el campo de los estudios económicos. Se emplearon tres técnicas multivariantes: análisis **factorial**, análisis **cluster** y, finalmente, análisis **discriminante**.

La finalidad del **análisis factorial** es expresar, cuando esto es posible, un conjunto de variables en función de unas pocas características subyacentes, pero no observables llamadas factores. Por lo tanto, este análisis no es sólo una técnica de reducción de la dimensionalidad sino que se basa en un modelo explícito de relaciones entre las variables – en este caso el modelo tradicional: el lineal - que interpreta los factores como características que las variables miden en común. Además establece que son los niveles o valores que las provincias poseen de estos factores los que influyen en mayor o menor medida sobre los observados en las variables.

De esta manera se puede motivar el modelo factorial del siguiente modo: supóngase que las variables originales pueden ser agrupadas por sus correlaciones lineales, es decir, que las variables correspondientes a un determinado grupo están muy correlacionadas entre ellas mismas, pero tienen relativamente pequeñas correlaciones con variables de otro grupo diferente, entonces se puede pensar que cada grupo de variables representa una construcción subyacente, el factor, que es responsable de las correlaciones observadas.

Por otro lado, la pretensión última del **análisis cluster** o de conglomerados es el establecimiento de grupos entre observaciones de manera que las agrupaciones resultantes presenten la mayor homogeneidad posible dentro de cada grupo, así como la mayor heterogeneidad entre grupos diferentes, medidas ambas de acuerdo con criterios establecidos en función de los objetivos del estudio. Los elementos a agrupar vendrán caracterizados por los valores que en ellos toman ciertas variables, que pueden ser muy distintos de unos elementos a otros.

En este trabajo se utilizó un método jerárquico de clasificación. Como algoritmo de clasificación se utilizó el método de *Ward* o de mínima variación intragrupo que se basa en el principio de minimizar en cada etapa de unión la heterogeneidad dentro de cada cluster. En primer lugar calcula la media de todas las variables en cada cluster y luego calcula la distancia euclídea al cuadrado entre cada individuo y la media del grupo. A continuación se suman las distancias de todos los casos y, en cada fase, los clusters

que se forman son aquellos que resultan con el menor incremento en la suma total de las distancias al cuadrado intra-cluster.

Por su parte, el **análisis discriminante** es una técnica de clasificación y asignación de un individuo a un grupo conocidas sus características. Este método parte de una serie de grupos definidos “a priori”, con una serie de observaciones para cada individuo referidas a un conjunto de variables. Con esta información se analizan cuáles son las variables que contribuyen más a discriminar entre los grupos considerados. Para ello lo que se hace es “reducir” las variables que mejor discriminan a unas pocas nuevas variables, que se denominan “variables canónicas”. Las variables canónicas son combinación lineal de las originales y vienen expresadas por una “función discriminante”. Esta es una ecuación lineal con una variable dependiente que representa la pertenencia a un grupo al tiempo que combinaciones lineales de variables independientes sirven de base para clasificar a los individuos entre los grupos. Así en este estudio en particular se utilizó este análisis para validar los resultados obtenidos con el análisis cluster.

Finalmente hay que señalar que en la selección de variables se utilizó un método de inclusión de variables paso a paso, que ofrece la posibilidad de “arrepentirse” de decisiones tomadas con precipitación, bien sea eliminando del conjunto seleccionado una variable introducida en el conjunto de discriminación en un paso anterior del algoritmo, bien sea introduciendo en dicho conjunto una variable eliminada con anterioridad. Para determinar qué variables entran y salen en cada paso de este tipo de algoritmos se utilizan diversos criterios de entrada y salida. Uno de los más utilizados es el de la lambda de Wilks que será el utilizado en este trabajo. Este criterio permite medir la potencia discriminante ganada/perdida al introducir/sacar una variable del conjunto de discriminación.

3. RESULTADOS.

3.1. Análisis factorial.

Antes de proceder al análisis e interpretación de resultados hubo que estudiar la adecuación de la muestra a los métodos que se pretende utilizar, mediante procedimientos tales como el examen de la matriz de correlaciones, el Test de Esfericidad de Bartlett, el Indicador de Adecuación Simple *Kaiser-Meyer-Olkin* o la anti-imagen de la matriz de correlación. Las pruebas fueron favorables lo que constituye un síntoma de que el análisis factorial puede considerarse aceptable y permitió seguir con su aplicación.

En la misma, empleando el criterio de elegir los factores cuyo autovalor sea mayor que 1, se obtuvieron tres factores, que explican un total del 70,34% de la varianza total de las variables originales. En función de esto el objetivo de síntesis que se pretende parece conseguido (cuadro 2).

Cuadro 2: Autovalores y varianza explicada por cada componente

COMPON.	AUTOVALOR	% VARIANZA	% ACUMULADO
1	7,12	32,38	32,38
2	5,74	26,12	58,5
3	2,61	11,84	70,34

Otro instrumento que ayuda a evaluar la bondad de ajuste del análisis factorial es la comunalidad de cada variable, es decir, el porcentaje de variación total de una variable explicada por los factores que, por lo tanto, indica hasta que punto los factores ayudan a explicar la variable. En este caso, las comunalidades fueron todas superiores a 0,6, exceptuando el número de bibliotecas y las actividades comerciales de grandes superficies.

Una vez evaluada la calidad de este análisis, se hace necesario realizar la interpretación de los resultados. Para esto es necesario obtener los coeficientes que relacionan las variables originales con cada uno de los factores, una vez realizada la rotación *quartimax*. Estas relaciones constituyen la clave de la interpretación de los mismos (Cuadro 3):

Cuadro 3.- Matriz de componentes rotada

Rotated Component Matrix ^a			
	Component		
	1	2	3
Bibliotecas	.657	.117	-.203
Producción editorial	-.155	-.169	.810
Películas distintas exhibidas	.900	4.82E-02	-.113
Espectadores de cine	-.136	-8.9E-02	.795
Prensa escrita	1.75E-02	.287	.740
Teléfonos	.411	.421	.676
Automóviles	-9.3E-02	.728	.459
Act.comerciales de grandes superficies	.452	6.32E-02	-.459
Act. Bares y restauración	.518	.620	2.95E-02
Act comerciales no alimentación	.114	.814	.305
Salas escénicas	.824	9.61E-02	2.93E-02
Espectadores de espectáculos escénicos	.869	9.20E-02	.147
Emisoras de radio	.735	.192	-.342
Pernotaciones en establecimientos hoteleros	8.59E-02	.888	-4.6E-02
Plazas en establecimientos hoteleros	-.115	.896	5.15E-02
Act. Comerciales minoristas	.482	.628	.127
Establecimientos hoteleros	.349	.843	-2.1E-02
Audiencia radio	.874	8.04E-03	-.148
Sesiones musica clasica	.816	3.59E-02	.249
Licencias de caza y pesca	.708	-.199	-.321
Gasto medio por espectador en el cine	-.186	.232	.786
Recaudacion en espectáculos escénicos	-.127	.293	.739

Extraction Method: Principal Component Analysis.
Rotation Method: Quartimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Teniendo en cuenta las variables que tienen más peso en cada factor, la interpretación de los resultados es la siguiente:

Primer Factor (32,38% de varianza explicada): aparece definido por las correlaciones positivas con las sesiones de música clásica, la audiencia de radio, las emisoras de radio, los espectáculos escénicos, las salas escénicas, las actividades de comercio en grandes superficies, películas distintas exhibidas en el cine, licencias de caza y pesca y las bibliotecas

Segundo Factor (26,12% de varianza explicada): este factor aparece definido por las correlaciones positivas con los automóviles, actividades de restauración y bares, actividades comerciales de no alimentación, actividades comerciales minoristas y pernoctaciones, plazas y establecimientos hoteleros.

Tercer Factor (11,84% de varianza explicada): relacionado positivamente con los espectadores de cine, la producción editorial, teléfonos, prensa escrita, gasto medio por espectador en el cine, y la recaudación de los espectáculos escénicos.

A partir de los factores obtenidos se construyó un indicador sintético, transformando ese indicador y cada uno de los factores en índices. En el anexo I se presentan los resultados de los factores y del indicador sintético para cada una de las provincias españolas, ordenadas de forma decreciente teniendo en cuenta el indicador sintético.

3.2. Análisis cluster.

Una vez resumida la información proporcionada por las 22 variables en tres factores mediante el análisis factorial se aplicó el análisis cluster basándose en los valores que tomaban los factores en cada una de las provincias, ya que los factores resumen y simplifican la mayor parte de la información inicial además de eliminar redundancia y, por lo tanto, pueden sustituir a los datos originales.

Los grupos obtenidos después de aplicar el análisis cluster jerárquico se presentan a continuación con las provincias que los forman y, posteriormente, para su mejor visualización se recogen en un mapa:

□Cluster 1: Burgos, Salamanca, Cantabria, Castellón, Valencia, Valladolid, Vizcaya, Navarra, Rioja, Álava, Zaragoza y Guipúzcoa.

□Cluster 2: Barcelona y Madrid.

□Cluster 3: Soria, Teruel, Ávila, Segovia, Palencia y Huesca.

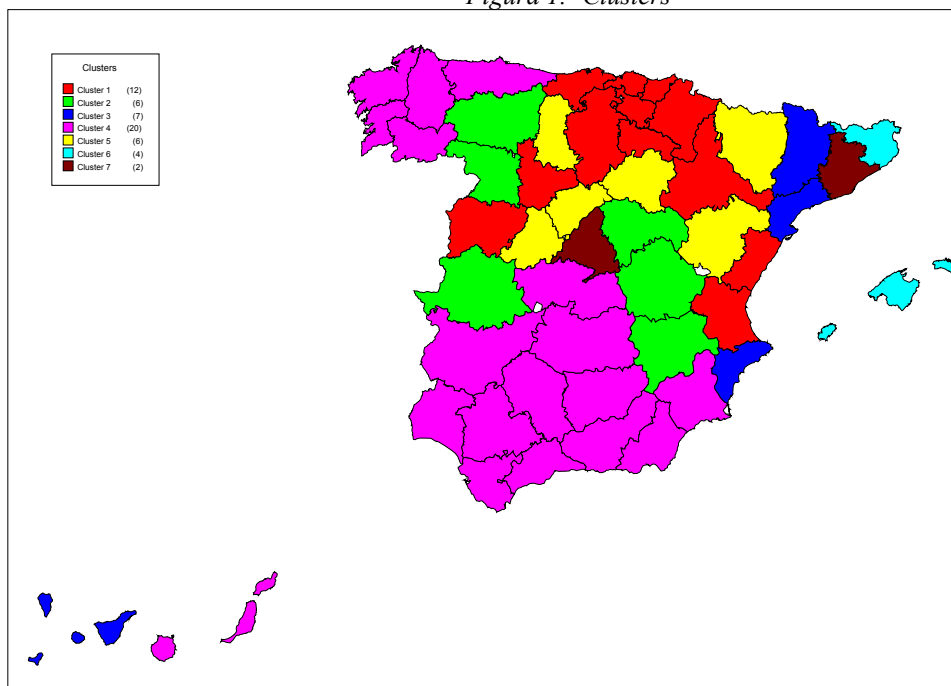
□Cluster 4: Islas Baleares y Gerona

□Cluster 5: Alicante, Tarragona, Lérida y Santa Cruz de Tenerife.

□Cluster 6: Cáceres, Cuenca, Zamora, Albacete, León, y Guadalajara.

□Cluster 7: Ourense, Toledo, Lugo, Badajoz, Ciudad Real, Huelva, Jaén, Granada, Murcia, Cádiz, Córdoba, Sevilla, Asturias, A Coruña, Almería, Pontevedra, Málaga y Las Palmas.

Figura 1.- Clusters



3.3. Análisis discriminante.

Teniendo en cuenta que el principal objetivo de aplicar el análisis discriminante es validar los resultados del análisis cluster, se presentan a continuación los resultados más destacados:

□ Se obtuvieron tres funciones discriminantes que explican la totalidad de la varianza intergrupos. En concreto, la primera función discriminante explica el 44,1% de la varianza intergrupos, mientras que la segunda el 34.7% y se deja, por tanto, la explicación del resto de varianza intergrupos para la tercera función discriminante

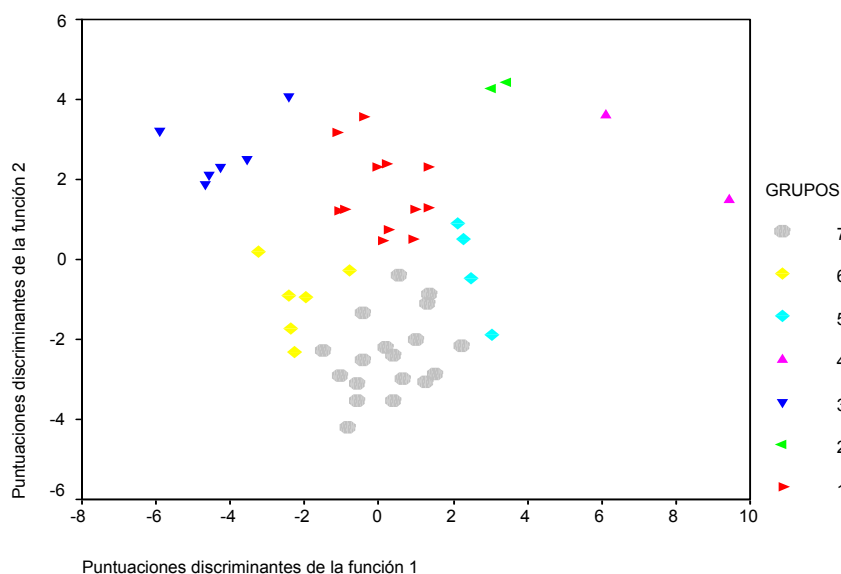
□ También hay que señalar que los factores obtenidos en el análisis factorial anterior que más contribuyen a la discriminación entre grupos son, por orden de importancia, el primer factor, el tercero y, finalmente, el segundo.

□ De las probabilidades de clasificación de cada provincia en un grupo se observa que sólo hay una provincia mal clasificada, con lo cual el análisis discriminante confirma el 98% de los casos. Concretamente, la provincia mal clasificada es Ciudad Real que aparece en el cluster 4 y según el análisis discriminante debería aparecer en el cluster 2.

Para ver como se distribuyen las provincias en los grupos es de gran utilidad representar las puntuaciones discriminantes mediante gráficos. La figura 2 presenta el diagrama de dispersión de las puntuaciones discriminantes de todas las provincias para las dos funciones más discriminantes. En este gráfico se observa como las puntuaciones discriminantes de la primera y segunda función permiten separar y distinguir los correspondientes clusters, que en la figura 2 se representan cada uno mediante un símbolo. También permite observar que los clusters 1 y 2 son los que presentan unas

diferencias más marcadas con respecto a los demás y justifica el hecho de que con dos provincias cada uno formen clusters independientes.

Figura 2.- Puntuaciones discriminantes



4. RECAPITULACIÓN Y CONCLUSIONES.

Las limitaciones de espacio necesariamente obligan a la concisión a la hora de extraer conclusiones, debiendo dejar solamente apuntadas las que consideramos más destacables. Con respecto al **análisis factorial** los resultados ofrecen una interesante simplificación de las 22 variables utilizadas. Claramente el segundo factor se decanta por las actividades relacionadas con la *movilidad turística* que, no en vano, coincide con una de las ramas del ocio y turismo más estudiadas y desarrolladas. Por lo que se refiere a los otros dos factores, un análisis de las variables contenidas en cada uno de ellos, permite vislumbrar que el primero, que explica casi la tercera parte de la varianza, incluye alguna de las actividades más tradicionales como la caza y la pesca, la radio y espectáculos relacionados con el teatro. Quizás está captando, entre otras, la idea de la dotación que puede venir influida por causas históricas o administrativas como, por ejemplo, la dotación mínima característica de la capital de provincia. Es, entonces, un factor que parece estar eminentemente relacionado con la *infraestructura básica* en las provincias españolas. En función de esto, no es de extrañar que este sea el factor que explique mayor proporción de varianza. Por otro lado, las variables contenidas en el tercer factor pretenden recoger el dinamismo en las actividades de ocio y turismo, más concretamente, en el ámbito cultural. Por ello, se puede afirmar que parece un factor eminentemente relacionado con el *dinamismo cultural* de las provincias españolas.

Por otra parte es importante destacar la importancia de relativizar las variables puesto que, tal y como se comprobó en trabajos previos al aquí llevado a cabo, la no relativización de las mismas conlleva que el primer factor se convierta de modo exclusivo en un factor tamaño.

A continuación la información suministrada por el análisis factorial es complementada con los datos del análisis cluster a través de los siete conglomerados que se obtienen. En cuanto al número de provincias contenidas en cada uno de ellos, se puede realizar una clasificación en tres grupos: dos son de tamaño pequeño, tres medios y otros dos son grandes cada uno de los cuales con características diferenciadores que responden claramente a sus señas de identidad.

Además se puede afirmar que quizás lo más importante de estos resultados es que no son motivo de extrañeza o, como mínimo, no ofrecen sorpresas llamativas. La consistencia de los distintos grupos con los resultados del análisis factorial es evidente y, de hecho, se podría establecer una tipificación en función de la importancia relativa en cada uno de los tres factores y, en la mayoría de los casos, respondiendo a una lógica territorial.

Todo esto se ve reflejado en la última parte del estudio, el análisis discriminante. Este no vino a hacer otra cosa sino confirmar el paso previo, de tal modo en la práctica totalidad de los casos la asignación de provincias a sus cluster fue realizada correctamente. Si bien ya fue comentado que todos ellos respondían a cierta lógica, es de gran ayuda que esta lógica o, en cierta medida, intuición se vea refrendada estadísticamente.

Sin embargo, existe una única excepción a la correcta asignación de provincias en cada cluster que viene dada por el caso de Ciudad Real. El análisis cluster la localizó en el séptimo cluster que es el cluster más numeroso y que está formado por provincias del sur junto con Galicia y Asturias. El análisis discriminante corrobora en este caso la correcta asignación de las provincias septentrionales pero, sin embargo, señala que para la provincia manchega su correcta situación debería ser en el cluster 6.

BIBLIOGRAFÍA

- BISQUERRA, R. (1989): *Introducción conceptual al análisis multivariante*. PPU CONSEJERÍA DE HACIENDA. Instituto de Estadística de la Comunidad de Madrid (2000); *Tipología Socioeconómica de las Regiones Europeas. Comparativa Estadística*. Madrid
- GARCÍA FERRANDO, M. (1997); *Los españoles y el deporte 1980-1995. (Un estudio sociológico sobre comportamientos, actitudes y valores)*; Estudios de Economía. Consejo Superior de Deportes & Tirant le blanch, Madrid.
- HARMAN, H.H. (1980): *Análisis Factorial moderno*. Editorial Saltés
- IGLESIAS, C. ET AL (1999): "Dimensionalidade da capacidade económica nas comarcas galegas". *Revista Galega de Economía*, vol.9, num.2 (2000), pp. 67-90. Santiago de Compostela.
- JOHNSON, R.A & D.W. WICHERN (1998); *Applied mutivariate statistical analysis*. Prentice Hall

MARTÍN GUZMÁN, M.P. (1988): "Métodos estadísticos en el análisis regional". *Estudios regionales*, número 22, pp. 149-170

PEREIRA, E. (2000); *Notas sobre a metodologia seguida na construção de índices con base nos factores obtidos e no indicador de síntese*; Memoria no publicada, Instituto Nacional de Estatística. Oporto.

SÁNCHEZ CARRIÓN, J.J. (1984): *Introducción a las técnicas de análisis multivariante aplicadas a las ciencias sociales*. Centro de Investigaciones Sociológicas.

TRIBE, J. (1999); *Economía del ocio y del turismo*; Editorial Síntesis.

VALLS, J.F. (2000); *Gestión de empresas de turismo y ocio*; Gestión 2000.

ANEXO I: ÍNDICES

Provincia	Índice 1	Índice 2	Índice 3	Indicador sintético
Huesca	491.75	240.62	60.72	242.27
Gerona	166.62	368.00	114.23	236.51
Baleares (Islas)	31.95	465.84	82.88	217.09
Soria	533.14	111.57	49.16	194.32
Teruel	460.34	136.13	38.07	176.24
Segovia	392.67	102.49	70.09	164.51
Navarra (Comunidad Foral de)	253.04	88.81	134.47	161.69
Rioja (La)	284.84	94.62	115.92	160.85
Ávila	421.59	89.43	49.48	153.88
Barcelona	69.94	70.13	206.68	149.65
Lérida	163.52	231.11	61.27	145.48
Palencia	378.72	51.53	77.03	144.54
Tarragona	148.86	190.87	81.88	138.64
Guipúzcoa	157.09	112.12	125.51	137.92
Madrid (Comunidad de)	61.85	40.90	209.48	137.39
Álava	194.31	76.12	122.28	131.84
Cantabria	170.96	140.13	92.41	131.37
Zaragoza	173.16	62.80	131.07	126.24
Castellón	194.83	141.13	73.00	125.79
Salamanca	225.27	92.83	85.21	122.98
Burgos	226.94	88.40	84.61	121.27
Alicante	84.66	174.99	67.82	104.73
Guadalajara	284.13	57.54	50.22	102.42
Vizcaya	88.00	61.00	127.18	98.74
Santa Cruz de Tenerife	61.66	223.44	31.76	94.00

Provincia	Índice 1	Índice 2	Índice 3	Indicador sintético
Zamora	246.79	94.14	27.31	91.49
Valencia	95.08	65.84	105.92	88.71
Asturias (Principado de)	117.95	99.17	72.84	86.88
Valladolid	121.58	53.16	99.45	86.86
León	160.68	69.24	69.14	84.50
Palmas (Las)	92.42	134.39	56.88	83.34
Lugo	153.71	112.99	37.80	79.55
Cuenca	226.73	99.42	11.58	77.56
Coruña (A)	71.46	105.77	71.88	75.69
Albacete	191.21	57.83	46.16	73.44
Toledo	107.11	120.40	28.13	62.93
Málaga	25.44	141.03	47.73	60.95
Cáceres	196.76	84.80	7.59	60.52
Ourense	125.79	113.75	17.30	58.42
Ciudad Real	146.32	69.25	24.19	50.83
Murcia (Región de)	44.28	86.01	54.61	48.62
Almería	49.15	139.47	19.46	48.44
Granada	65.57	87.93	39.46	45.47
Pontevedra	26.80	122.31	32.54	43.78
Badajoz	113.63	75.06	16.48	38.80
Huelva	81.24	82.52	8.79	27.54
Córdoba	57.69	41.02	32.01	19.38
Sevilla	1.73	44.48	48.79	15.93
Cádiz	14.96	61.13	29.54	13.73
Jaén	53.52	54.78	4.12	5.39