

CARACTERIZACIÓN DEL GASTO DE TURISTAS DE CRUCEROS EN URUGUAY MEDIANTE TÉCNICAS DE MINERÍA DE DATOS

Silvia Altmark¹; Ramón Álvarez¹; Florencia Santiñaque¹

RESUMEN

El turismo de cruceros en Uruguay ha crecido, determinando un importante aporte de divisas cada temporada (abril a octubre de cada año). Por ello son necesarias herramientas para proyectar y caracterizar las variables económicas involucradas en esta actividad, en particular, el gasto.

Se comparan los resultados con la tipología surgida de aplicar métodos de clusters jerárquicos y no jerárquicos. Se usan datos de las temporadas 2010-2011 y 2011-2012, del Ministerio de Turismo y Deporte, obtenidos de una muestra de pasajeros a través de una encuesta cara a cara con diseño muestral complejo.

Surgen 3 tipologías de cruceristas, una aplicando el algoritmo de *k-means* sobre los gastos por rubro; otra aplicando el método *PAM* sobre los gastos en proporciones y otra usando algoritmos de cluster para datos binarios al considerar presencia o ausencia de gastos por rubro. Las diferentes tipologías se analizan y asocian con las características socio-demográficas.

Palabras clave: Clusters, cruceros, minería de datos, gasto.

Introducción

El turismo es una actividad muy relevante para la economía del Uruguay, en términos de divisas, valor agregado (PIB) y empleo. De acuerdo a las estimaciones del Ministerio de Turismo, surgidas del Anuario, en 2015, el aporte del sector a la economía uruguaya es del orden del 14% de las exportaciones totales y el 58% de las exportaciones de servicios del 7,7% del PIB, en lo que de alrededor de 4% en términos de generación del PIB y 8 % con respecto a la contribución a los puestos de trabajo del país.

También se pueden señalar otros impactos de la actividad turística que dinamizan la economía: creación de infraestructuras y servicios, mejora de recursos humanos, aplicación de nuevas tecnologías, surgimiento de nuevas oportunidades de negocios, recuperación y/o preservación del patrimonio, puesta en valor de recursos. Sin embargo, el turismo también puede afectar negativamente un destino, cuando no se toman en consideración los impactos medioambientales y socioculturales de ciertas actividades turísticas.

De acuerdo a los datos relevados por el Ministerio de Turismo a partir de sus encuestas, el turismo de cruceros presenta una evolución creciente en Uruguay, partiendo de 75 arribos, 56.167 pasajeros desembarcados y U\$S 18.899.464 de gasto en la temporada 2004-2005, desde cuando se dispone de datos. Entre esa temporada y la temporada 2011-2012, la cantidad de personas desembarcadas aumentó más del 500% y el gasto en U\$S, el 11%.

Es importante señalar que la temporada de cruceros se extiende generalmente entre octubre de un año y abril del siguiente.

¹ IESTA – DMMCC – Facultad de Ciencias Económicas y de Administración - UdelaR

Tabla 1 Gastos de cruceristas por Temporada

DATOS DE CRUCERISTAS POR TEMPORADA			
Temporada	Gasto en U\$S	Personas	Gasto p/ pers. en u\$s
2006-2007	11.235.466,00	149.062,00	75
2007-2008	16.818.273,00	256.593,00	66
2008-2009	14.384.413,00	247.120,00	58
2009-2010	17.830.909,00	292.048,00	61
2010-2011	13.291.304,00	278.627,00	48
2011-2012	20.884.091,00	353.727,00	59

Fuente: Elaborada en base a los datos de los Anuarios del Ministerio de Turismo

Además del interés económico del producto turístico cruceros, se señala que es muy habitual que los cruceristas regresen al destino visitado, ya no en esta modalidad, sino como turistas, es decir, alojándose y permaneciendo algunos días en él, con el consiguiente impacto positivo en la exportación de servicios turísticos.

Aunque el turismo de cruceros en general ha mostrado un importante crecimiento, no ha sido muy estudiado; se pueden citar antecedentes de 1985 para Miami (Mescon y Vozikis), un estudio general del año 1998 realizado Chapapría, otro de 2002 para Barbados, de Cuellar-Río y Kido-Cruz, entre otros, señalándose el aporte de investigación de la Organización Mundial del Turismo del año 2008

Por otra parte, son muy pocos los antecedentes de estudios acerca del gasto de los cruceristas en Uruguay: “El gasto de los cruceristas en Uruguay 2008-2010”, de Risso , “Patrones de gasto de cruceristas en dos puertos uruguayos”, de Brida et al.

Por esta razón, el presente trabajo, además de lo metodológico en sí mismo, tiene interés también desde la perspectiva del diseño de políticas públicas orientadas a la promoción del turismo de cruceros, a partir de más y mejor información de los cruceristas. En particular, conocer los componentes del gasto y sus determinantes, puede contribuir al mejor diseño de campañas promocionales, así como a definir, junto con los operadores de este producto turístico, cambios en la oferta, de acuerdo al perfil de gasto encontrado.

El artículo está organizado de la siguiente forma: en la segunda sección se explican las técnicas aplicar y los datos que se utilizan, la tercera sección presenta los resultados, la cuarta sección presenta las conclusiones y finalmente se incluyen Bibliografía y Anexos.

Metodología

En esta sección se presenta brevemente el herramental estadístico que será usado sobre un juego de datos.

Técnicas de minería de datos

Existen muchas definiciones de minería de datos (*datamining*) y para eso tomamos la que propone Jiawei Han (2012): “La minería de datos es el proceso de descubrir patrones interesantes y conocimiento a partir de grandes cantidades de datos. Las fuentes de datos pueden incluir bases de datos, los datos almacenes, la web, otros repositorios de información o de datos que se transmiten en el sistema dinámicamente”. Por lo tanto, a partir de esta

definición, lo que interesa es proponer las diferentes técnicas con las que se pueden descubrir patrones en las grandes masas de datos. La tarea que hay que efectuar en la minería de datos es el análisis automático de grandes cantidades de datos donde, para extraer patrones interesantes desconocidos, se pueden agrupar registros de datos, identificar registros poco usuales y, lo más importante, dependencias entre registros para un mismo atributo o para atributos entre sí.

Para eso surgen desde el campo de la estadística lo que se conoce como aprendizaje estadístico (statistical learning), conceptos que autores como Tibshirani y Hastie (2013) han instaurado, permitiendo clasificar los problemas de aprendizaje estadístico como

- *Supervisado*, donde en general se dispone de información correspondiente a varios atributos para un conjunto de datos, en los cuales se conoce además un atributo en particular que se toma como variable de respuesta y lo que interesa es poder desarrollar modelos que permitan encontrar relaciones entre la variable de respuesta y sus predictores con fines esencialmente predictivos. Dentro de esta clase de técnicas podemos encontrar la regresión, el análisis discriminante y los métodos CART (Classification and Regression Trees)
- *No supervisado*, donde se dispone de muchos atributos, pero donde no existe una variable de respuesta que puede usarse para supervisar el análisis. Dentro de estos encontramos todos los métodos de clustering o análisis de conglomerados

Análisis de Conglomerados

Para el análisis de conglomerados o análisis de clusters se consideran varios escenarios

- Las variables de gasto en escala original (salvo que el método de aglomeración exija trabajar con datos estandarizados según la distancia usada). Para este tipo de variables en este caso se usan los métodos de **k-means** (k-medias) que es un tipo de método de clusters no jerárquicos
- Las variables de gasto en escala de proporciones, es decir que los gastos por componente están relativizados contra el gasto total de cada observación. Para este tipo de variables, en este caso, se usan los métodos de k-means (k-medias) que es un tipo de método de clusters no jerárquicos y otro método de cluster jerárquico como es el **PAM** (Partition Around Medoids)
- Las variables de gasto se transforman en variables binarias (ausencia o presencia de gasto en cada rubro). Este tipo de análisis deja de lado el considerar el gradiente del gasto y, a su vez, por la naturaleza de las variables considerada, exige el uso de algoritmos específicos, así como de determinadas distancias.

Dentro de los métodos no jerárquicos se encuentra el de **k-means**, que es uno de los más utilizados, en función de su simplicidad y velocidad de convergencia, que es de orden ($n \cdot p$), donde n es la cantidad de observaciones y p el número de variables .

A partir de un conjunto de n observaciones (x_1, x_2, \dots, x_n), que se puede considerar un vector p , el método de **k-means** busca encontrar una partición de los n individuos en k subconjuntos con $k \leq n$, de manera de minimizar la suma de cuadrados intraclase (SCIC):

$$\arg \min_{S} \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

considerando μ como el centroide de los puntos en el grupo S_i .

Al inicio, todos los centros de los conglomerados están en la media de las celdas de Voronoi, que se puede interpretar como el conjunto de puntos de los datos que están más cerca del centro de ese grupo que de cualquier otro grupo.

El algoritmo funciona de la siguiente manera:

- Se eligen en forma aleatoria los centros iniciales. Queda entonces la siguiente secuencia $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_k$ de k centros
- Se asigna cada observación al cluster con la media más próxima, es decir que la partición queda determinada con las medias iniciales
- Se calculan los S_i de la siguiente manera

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\| \leq \|x_p - m_j^{(t)}\| \forall 1 \leq j \leq k\}$$

donde cada uno de los x_p queda asignado a uno de los S_i .

- El algoritmo se actualiza calculando las nuevas medias del grupo
- El algoritmo se detiene luego que al reasignar alguna observación a otro grupo no hay cambios menores a una tolerancia prefijada en la (SCIC)

Los algoritmos habitualmente usados en los paquetes estadísticos están basados en los que plantearon MacQueen (1967), Lloyd (1957) y Forgy (1965). El algoritmo de Hartigan-Wong es el que se usa habitualmente y el que está implementado en R, al trabajar con centros iniciales que se eligen en forma aleatoria, se recomienda ejecutar el algoritmo varias veces ($n=10$) de manera de ver la estabilidad de los resultados.

El algoritmo (PAM), que se basa en la búsqueda de k objetos representativos o medoides entre las observaciones del conjunto de datos, de manera que representen adecuadamente la estructura de los datos. Un **medoide** se podría definir como el objeto perteneciente a un cluster o conglomerado, cuyo promedio de disimilaridad a todos los objetos en el conglomerado es mínima, es decir, que se puede considerar como el punto más céntrico de la agrupación considerada.

La implementación más frecuente de **k-medoides** es el agrupamiento alrededor de medoides (Partition Around Medoids) que tiene el siguiente algoritmo:

- Inicialización: Selección al azar de k de los n puntos de datos como los candidatos a medoides en la fase de construcción.
- Se asigna cada observación al cluster con el medoide más próximo, dependiendo de la distancia elegida (euclidiana, Manhattan o Minkowski). Luego se encuentra un mínimo local para la función objetivo, es decir, una solución de tal manera que el cambio de observación con un medoide haga que la función objetivo decrezca (esto se denomina la fase de intercambio).
- Se repiten los pasos anteriores hasta que los medoides queden estables (es decir que haya cambios en los medoides)

En este caso, a diferencia de lo que ocurre con los k-means, los centros son los medoides, los que pueden resultar más robustos.

Del mismo modo que para el método anterior, se evalúa cuál es la mejor estructura de grupos en base a la silueta promedio que se obtiene con cada configuración (en el párrafo que sigue se define el concepto de silueta), mostrando que la estructura adecuada podría ser con 5, donde aparece el primer máximo local.

El método **Silueta** en realidad no es método de clustering en sí mismo, sino un método de interpretación y validación del número de conglomerados o clusters hallado. Esta técnica permite obtener una representación gráfica de qué tan bien está cada observación dentro de un determinado grupo.

Tiene la ventaja que puede ser utilizada para datos que hayan sido clasificados a través de cualquier método, como por ejemplo **k-medias** o **k-medoides**. Para cada observación i , sea $a(i)$ la disimilaridad promedio de i con todas las demás observaciones dentro del mismo grupo. Se puede interpretar que (i) está bien emparejado con los restantes integrantes del grupo cuanto menor sea $a(i)$. Para los demás clusters C se define $d(i,C)$ como la disimilaridad promedio de i con los datos de C . Se repite el procedimiento para cada grupo del cual el i no es miembro y se determina $b(i) := \min\{C\} d(i,C)$, que representa la mínima disimilaridad promedio de i con cualquier otro grupo, lo que representa la disimilaridad entre i y los clusters vecinos. Se define entonces el estadístico $S(i)$ (silueta) como:

$$S(i) = \frac{(b(i) - a(i))}{\max[(b(i) - a(i))]}$$

De la definición anterior, se puede ver que $-1 \leq S(i) \leq 1$, en donde si $S(i)$ está cerca de 1 significa que los individuos están correctamente clasificados, valores cercanos a 0, que los grupos no están bien determinados y valores negativos de $S(i)$, que esa observación debería pertenecer a otro grupo.

Los métodos jerárquicos se caracterizan por generar una serie de particiones encajadas y requieren la definición de una distancia. Inicialmente, cada objeto se le asigna a su propio grupo, y entonces los algoritmos proceden iterativamente; en cada etapa unen los dos grupos más similares, continuando hasta que sólo quede un grupo. En cada etapa las distancias definidas entre las agrupaciones se recalculan por la fórmula de disimilitud de Lance-Williams, actualizándose de acuerdo con el método de agrupación particular que se utilice.

Dentro de los métodos jerárquicos se considera el de **WARD**, que consiste en descomponer la variación total (T) en variación en los grupos W (**within**) y variación entre los grupos B (**between**) y, al estar frente a una partición dada, el método unirá aquellos grupos que produzcan el efecto de hacer mínima la variación **within** en la nueva partición.

$$T = W + B$$

Descripción de datos

Se utilizan los datos de la temporada 2010-2011 del Ministerio de Turismo y Deporte. Los datos sobre escala de buques, cantidad de cruceros, cantidad de personas arribadas y desembarcadas, para Punta del Este, son proporcionados por la Dirección Nacional de Hidrografía y la Dirección de Transporte Fluvial y Marítimo y, para Montevideo, por Operaciones Portuarias de la Administración Nacional de Puertos (ANP). Los datos restantes

surgen de encuestas realizadas por el Área de Investigación y Estadísticas del Ministerio de Turismo, con un diseño muestral en base a dos muestras independientes, una para Montevideo y otra para Punta del Este.

En la temporada 2010-2011 llegaron a Uruguay 171 cruceros, 76 a Montevideo y 95 a Punta del Este, desembarcando en total 278.627 personas, de las cuales 99.851 lo hicieron en Montevideo y 178.776 en Punta del Este. La nacionalidad mayoritaria de quienes desembarcaron es brasileña (97.272), seguida de argentina (95.547); norteamericanos y europeos siguen en importancia, con 36.608 y 30.394 personas respectivamente.

El gasto total de las personas desembarcadas fue de U\$S 13.291.304, discriminados en U\$S 5.232.921 en Montevideo y U\$S 8.058.383 en Punta del Este. Los brasileños gastaron U\$S 6.907.191, los argentinos U\$S 2.899.730, los norteamericanos U\$S 1.514.731 y los europeos U\$S 1.116.230. Si se atiende el gasto per cápita (en U\$S) en la temporada de referencia, en promedio fue de 48, mientras que los brasileños gastaron 71, los argentinos 30, los norteamericanos 41 y los europeos 37.

Una primera observación, de suma importancia, es entender la estructura de los datos con los que se trabajará.

En la base de datos obtenida de la página web del Ministerio, las unidades de análisis, que son, en este caso, grupos de cruceristas, provienen de una muestra probabilística bietápica estratificada por conglomerados, donde las unidades primarias de muestreo (UPM), son los cruceros, que se clasifican en 2 estratos (cruceros que desembarcan en Montevideo (estrato 1) y en Punta del Este (estrato 2)). Éstos fueron seleccionados mediante muestreo π -ps (con probabilidad proporcional), en este caso al número de cruceristas. En una segunda etapa se selecciona una cantidad fija de 40 grupos de cruceristas, unidades secundarias de muestreo (USM), que tienen, por lo tanto, un número variable de personas. De esta manera, al ser una cantidad fija de grupos de cruceristas, se obtiene un diseño auto-ponderado. Los expansores que tienen los datos, fueron calculados y calibrados por el Área de Investigación y Estadística del Ministerio de Turismo y Deporte, teniendo, para la temporada estudiada, un total de 3176 grupos (filas en la tabla de datos). Entonces, dado que el interés del presente trabajo es el gasto de los cruceristas, se descartaron los grupos que no presentaban gasto o que tenían un monto de gasto imputado, reduciéndose el estudio a 2311 casos (filas). Esto significa que 2311 grupos de cruceristas realizaron algún tipo de gasto.

Los rubros que se utilizan para desagregación del gasto de los cruceristas son: Shopping (Compras), Alimentación, Tours, Transporte y Otro gasto, siendo los de mayor importancia Shopping y Alimentación.

Resultados

Teniendo en cuenta los datos correspondientes a 2010-2011, para su análisis es importante, en primer lugar, ver cómo es la distribución del gasto total.

A su vez, para trabajar adecuadamente con los componentes del gasto y, teniendo en cuenta que para cada grupo el gasto total varía mucho, se propone una transformación en cada componente en proporciones, relativizando cada rubro en función del gasto total.

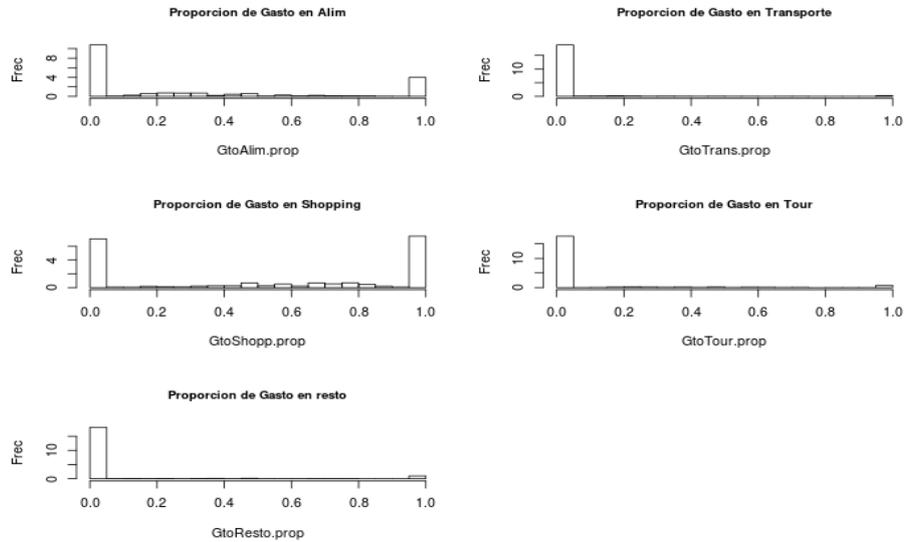


Figura 1 Distribución de Gasto Total y componentes. Fuente: Elaboración propia

Considerando la gran asimetría que tiene cada componente y que, además, tiene un rango de variación muy diferente, se propone, para comparar, visualizar las funciones de distribución de cada uno de ellos.

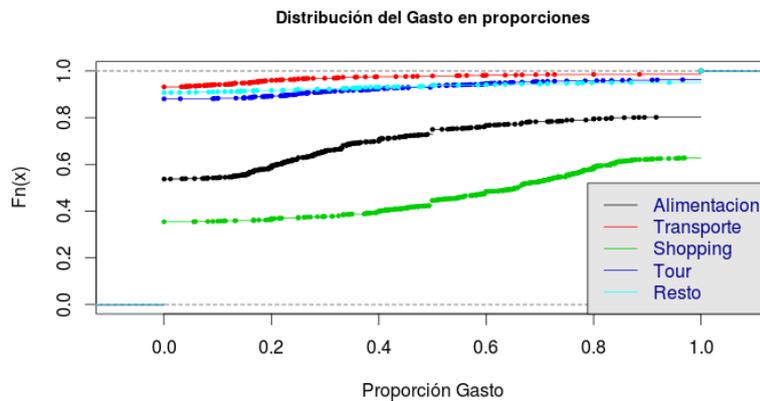


Figura 2 Distribución de Gasto Total y componentes en proporciones. Fuente: Elaboración propia

Puede verse según Figura 2 que aparece una clara distinción entre tipo de gastos, al haber dos de ellos que, si bien comienzan en 0, la proporción de cruceristas que no gastan va desde un 20 % y 40 % (grupo 1 de gastos) versus el resto, que tiene casi 85 % de cruceristas como mínimo (grupo 2 de gastos). De acuerdo a lo que se observa en el cuadro anterior, se decide considerar solamente los gastos en Tour, Alimentación y Shopping, los cuales se toman como elementos centrales en la caracterización del gasto. A fin de ver más claramente ese comportamiento, se presentan en la Figura 2 las relaciones entre esos tres componentes, a través de los gráficos que siguen.

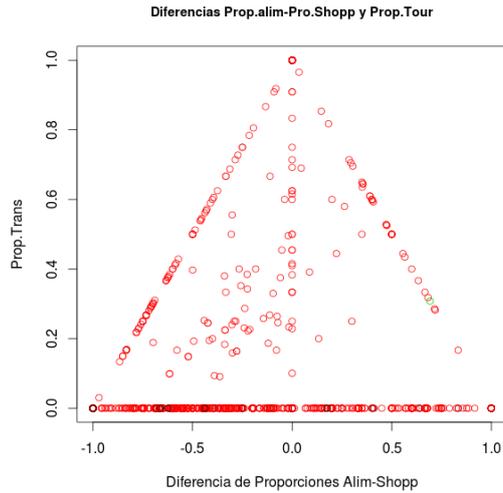


Figura 3: Relación entre proporciones de Gasto, Shopping y Tour. Fuente: Elaboración propia

Debe tenerse en cuenta que hay una cantidad importante de cruceristas que concentran el gasto en valor fijo (en 0) o cantidades mayores a 0 que se repiten, con lo cual al establecerlo en proporciones se produce gran “granularidad”, que es un efecto no deseado al ser variables continuas. A su vez, visualmente resulta difícil captar el patrón, sobre todos en las aristas de triángulo que aparece en la Figura 3.

Para superar esta dificultad se modifican los valores de las proporciones de gasto, incluyendo una perturbación aleatoria con distribución uniforme, lo que permite ver exactamente cómo es la masa de cruceristas que están en los vértices del triángulo (donde vale 0)

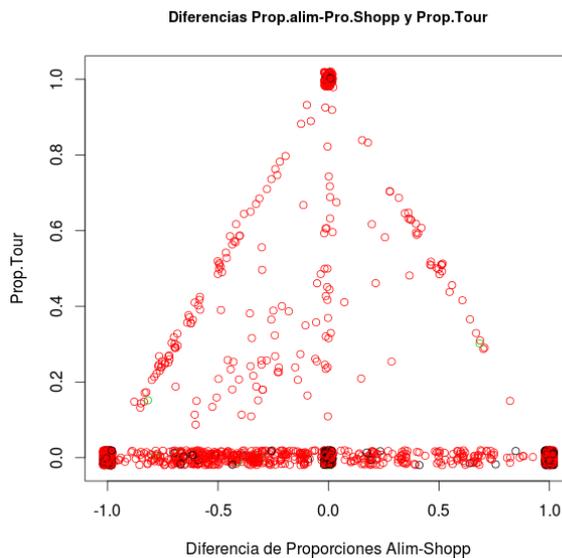


Figura 4: Relación entre proporciones de Gasto, Shopping y Tour (perturbación aleatoria). Fuente: Elaboración propia

La Figura 4 muestra que hay una cantidad importante de cruceristas que no gastan en Tour y que no gastan en Alimentación: es el vértice del triángulo que tiene coordenadas (-1,0); lo mismo sucede para los cruceristas que sólo gastan en Alimentación, al tener coordenadas (1,0).

Los cruceristas que aparecen proyectados en el punto (0,0) son aquellos cruceristas que no gastan en Tour y que puede ser que gasten sólo en Alimentación y en Shopping, repartiendo en mitades el gasto, o que sean cruceristas que sólo gastan los rubros del estrato 2 de gasto (Transporte u otro gasto). El punto con coordenadas (0,1), que resulta interesante y que aparece con una superposición de puntos importantes, es el de cruceristas que gastan sólo en Tour. El resto de los cruceristas son los que reparten el gasto en los tres componentes.

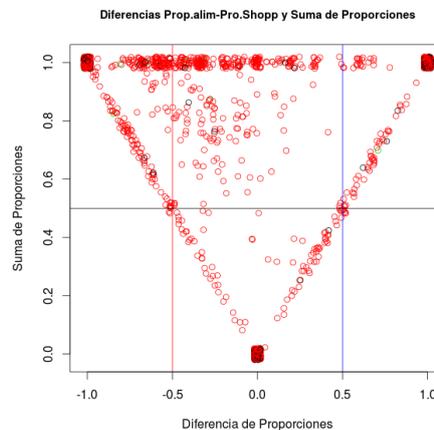


Figura 5: Relación entre suma y diferencias de proporciones de Gasto, Shopping y Tour. Fuente: Elaboración propia

Análisis de conglomerados con los gastos en escala original

Para el análisis de las variables en la escala original (o estandarizada), teniendo en cuenta los resultados primarios que se muestran en el estudio univariado de cada rubro y en los gráficos de dispersión se decide trabajar formando tipologías de gasto con 3 de los 5 componentes (Tour, Alimentación y Shopping) y filtrando los datos de acuerdo al gasto total, usando como umbral el valor de 500, ya que los cruceristas que gastan más de ese valor pueden considerarse como valores extremadamente atípicos (outliers), que tendrían una gran influencia en la formación de clusters, producto de la asimetría que provocan en las variables (en particular al generar medias distorsionadas). El umbral de 500 provoca una pérdida de 4% de las observaciones, una forma de ver como la variabilidad intracluster disminuye. Tal como aparece en la figura 10 la cantidad de grupos o clusters parece ser 6 donde la VIC tiene una inflexión o descenso en el ritmo de crecimiento. Por otra parte 6 grupos parece una cantidad razonable.

Tabla 2 Promedios de cada tipo de Gasto según los grupos obtenidos por método k-means

	Gasto Total	Gasto Tour	Gasto Alim	Gasto Shopping	N
1	-0,577	-,0179	-0,267	-0,454	1417
2	2,495	-0,212	-0,114	2,908	134
3	0,940	3,810	0,0823	-0,079	102
4	0,675	-0,142	-0,158	0,815	384
5	0,982	-0,274	2,368	-0,269	188
Total	0	0	0	0	2225

Fuente: Elaboración propia

Análisis de conglomerados con los gastos en escala de proporciones

En este caso se va a construir una segunda tipología través del método de k, usando las cinco variables.

Tabla 3 Promedios de cada tipo de Gasto según los grupos obtenidos por método k-means (gasto en proporciones)

	Gasto Tour	Gasto Alim	Gasto Shopping	Gasto Transporte	Gasto Resto	N
1	0,654	0,074	0,066	0,187	0,019	212
2	0,035	0,341	0,573	0,029	0,021	378
3	0,004	0,962	0,009	0,017	0,008	529
4	0,016	0,047	0,005	0,006	0,926	144
5	0,008	0,15	0,971	0,004	0,002	188
Total	0	0	0		0	2225

Fuente: Elaboración propia

Para la construcción de la tipología de cruceristas en base a las proporciones de gastos cambiando de algoritmo, PAM en este caso, que nos lleva a considerar 5 o 6 grupos, en base a la silueta media como índice para mostrar homogeneidad de cada grupo.

Se usa una muestra de aprendizaje: se seleccionan 20% mediante muestreo aleatorio simple, para poder visualizar cómo es la silueta en cada grupo. Se ensayaron varias muestras donde siempre se ve que hay un grupo mayoritario muy homogéneo mientras que hay un grupo muy heterogéneo que, con valores negativos del índice en silueta, son un indicio que puede parecerse ese conglomerado más a los cruceristas de los demás clusters.

Análisis de conglomerados con los gastos en escala dicotómica

Para este análisis se cambia la perspectiva del análisis, tal como dijimos antes, al considerar las variables de gasto como variables binarias de ausencia/presencia, con la cuales construir la tipología de cruceristas, teniendo que recurrir a algoritmos adecuados, en este caso el de Ward, usando para eso distancia de tipo binaria.

Analizando el dendrograma que surge al aplicar el algoritmo de Ward se propone un corte en el árbol de agregación de manera de tener cinco grupos.

Tabla 4: Promedios de cada tipo de Gasto según los grupos obtenidos por método jerárquico para variables binarias

	Gasto Tour	Gasto Alim	Gasto Shopping	Gasto Transporte	Gasto Resto	N
1	0,00	0,00	1,00	0,00	0,00	860
2	0,014	0,873	0,870	0,00	0,287	539
3	0,00	1,00	0,00	0,00	0,00	458
4	0,114	0,261	0,195	1,00	0,01	210

5	1,00	0,348	0,500	0,00	0,00	244
Total	0,119	0,462	0,645	0,09	0,068	2311

Fuente: Elaboración propia

Una forma de ver más en detalle cuáles son las diferencias entre grupos es a través de la Tabla 4, que cruza los grupos con una variable que cuenta la cantidad de tipos de gasto de cada crucerista clasificado.

Tabla 5 Número de Gastos por clusters jerárquicos para gastos en variables binarias

Nº de rubros donde hay Gasto	1	2	3	4	5	Total
1	860	33	458	114	87	1552
2	0	447	0	74	107	628
3	0	56	0	17	50	123
4	0	3	0	5	0	8
Total	860	539	458	210	244	2311

Fuente: Elaboración propia

Discusión

Caracterización de los clusters para variables cuantitativas

Teniendo en cuenta la tabla 2, cuando se considera el nivel de gasto total y los 3 gastos que se consideraron más relevantes, como el gasto en Tour, el gasto en Alimentación y el gasto en Shopping, se puede decir que hay un cluster mayoritario (el 1 que representa el 64, 5% de los grupos de cruceristas) que se caracteriza por tener gastos promedios negativos (recordar que están estandarizados los valores), es decir gastos por debajo de la media; el cluster 2 aparece como el segmento de cruceristas que más gastan (casi 3 desvíos del gasto medio, con tamaño casi 10 veces menor que el cluster 1). En particular gastan mucho en el ítem Shopping.

Otro segmento diferenciado es el cluster 3, que aparece como el que concentra cruceristas que gastan muy por encima del promedio en el rubro Tour, siendo éste un grupo también pequeño (102 grupos de cruceristas).

El cluster 4 aparece nuevamente como otro segmento de grupos de cruceristas más grande, que también gasta en promedio por encima de la media global para el rubro Shopping pero acompañado de gastos totales menores.

Por último el cluster 5, es el segmento de cruceristas con un gasto total por encima de la media, que se caracteriza por gastar mucho en el rubro Alimentación

Cuando cambiamos la forma de segmentar los grupos de cruceristas, considerando los 5 rubros de gasto (como proporciones), pero dejando de lado el nivel (gasto total), los resultados que se desprenden de la tabla 3 muestran lo siguiente.

En primer lugar los tamaños de los clusters son más homogéneos, mostrando un único cluster (el 3) de tamaño 529, que se caracteriza por tener una proporción de gasto media de 96% . Aparece luego un segundo cluster el 5 que aparece como el segmento de cruceristas que gastan en Shopping, casi en exclusividad; del mismo modo, el cluster 4 aparece como el

segmento de cruceristas que gasto en el rubro Resto. Los restantes clusters (1 y 2) son segmentos que tienen un gasto más repartido, donde predomina el gasto en Tour, mientras que en el 2 se reparten Alimentación y Shopping.

Estas 2 clasificaciones o grupos de segmentos pueden analizarse de manera más detallada buscando en qué se diferencian para los atributos propios de los grupos de cruceristas y que no se usaron para la formación de clusters que son

- Total de personas del grupo de cruceristas, porcentaje (%) de hombres, porcentaje (%) de adultos, número de visitas y gasto total

Para la construcción de los indicadores % de hombre se relativiza el total de hombres contra el total de personas, % de adultos es total de personas entre 15 y 64 años, contra total de personas

En función de esto podemos decir para la segmentación de viajeros en base a gasto total, gasto en Tour, gasto en Alimentación y gasto en Shopping:

- Cluster 3 y 5 se diferencian de los demás por ser segmentos con proporción de grupos de cruceristas unitarios muy por debajo de la media, siendo, a su vez, el cluster que tiene, en proporción, mayor cantidad de personas por grupo de cruceristas
- El Cluster 5 se diferencia de los demás por ser el que tiene % de hombres 0 más bajo que el promedio (casi la mitad), mientras que el cluster 3 se caracteriza por tener mayor % relativo de grupos con 3 hombres
- El cluster 2 es el segmento que tiene menor % de 0 adultos (casi la mitad) y el cluster 3 es el que tiene reparto más uniforme de las categorías en cuanto a % de adultos, siendo el segmento donde la totalidad de cruceristas son todos adultos es menor (casi 2/3 partes del promedio)
- Un aspecto interesante es que los 5 clusters en términos del gasto total, que se usó para su construcción, identifican segmentos muy marcados como el 2, que concentra el 100% de su gasto en el tercer quintil del gasto total, mientras que los clusters 3,4 y 5 no tienen grupos de cruceristas con gasto total en el 1^{er} quintil y el tercer quintil. A su vez el cluster 4 se diferencia del 3 y 5 por no tener cruceristas con el mayor tramo de gasto (quinto quintil)
- El cluster 3 se caracteriza por tener una composición mayoritaria de cruceristas que llegan por primera vez

La segmentación de viajeros en base a gasto en proporciones (que ignora el nivel de gasto total), muestra las siguientes características

- El cluster 4 se diferencia de los restantes por ser el que tiene mayor proporción de grupos de cruceristas individuales y el que tiene menor proporción de grupos de 3 viajeros
- el cluster 2 es el que tiene menor proporción de grupos con 0 % de hombres
- el cluster 4 es el que tiene menor proporción de cruceristas con 0% de adultos, mientras que en este segmento aparece la mayor proporción de cruceristas donde hay solamente adultos
- Para el gasto total es donde se ve la mayor diferenciación, con el cluster 3 y 4, donde predominan cruceristas que gastan en el 1^{er} quintil. El cluster 2 concentra cruceristas que gastan en el 2do y 3er quintil, mientras que el cluster 5 es el que tiene una

distribución de gasto similar al promedio marginal en quintiles.

- El cluster 1 se diferencia por ser el segmento donde los cruceristas que viene por primera vez predominan, con un 10% más que para la distribución marginal

Caracterización de los clusters para variables binarias

- Se pueden distinguir 5 clusters, dos de ellos que gastan solamente en un rubro: el 1 con 860 elementos, con un gasto exclusivo en Shopping y el cluster 3, cuyo gasto es en Alimentación.
- Los otros 3 clusters destinan su gasto a varios rubros: el cluster 2, de mayor gasto relativo en Alimentación y Shopping, pero que también gasta en Transporte y en menor medida en Tours y Otros rubros (Resto); el cluster 4, de mayor gasto relativo en Otros rubros, seguido de Alimentación, Shopping y Tours; el cluster 5, gasta mayoritariamente en Tours, seguido de Shopping y Alimentación.

Para caracterizar esos segmentos de cruceristas se analizan algunos aspectos de los viajeros, tal como se muestra en las tablas que aparecen en los Anexos y se hizo para los clusters sobre variables cuantitativas en la sección anterior.

En total, los cruceristas estudiados viajan en grupos de 2 personas, integrados por: más mujeres que hombres, adultos que vienen por primera vez y con un gasto total con distribución bastante uniforme, con un peso mayor de los grupos que gastan entre 100 y 200 dólares.

Según la cantidad de personas, el cluster 5 se distingue del resto en que se integra generalmente por 5 personas; muy pocos de este tipo de cruceristas viaja solo.

El cluster 1 es altamente femenino (lo cual es predecible, dado que es el segmento que gasta exclusivamente en Shopping), en el cluster 4 predominan los adultos y el cluster 5 es más joven.

En cuanto al número de visitas, en general los cruceristas estudiados vienen por primera vez, destacándose el cluster 5.

Al analizar el rango de gasto, se destaca el cluster 3 con un menor gasto que el resto, lo cual es coherente con la identificación del segmento como el que gasta exclusivamente en Alimentación.

Conclusiones

Como conclusiones preliminares podemos ver que hay tipologías bien marcadas de gastos en términos de los niveles, así como de las estrategias que los cruceristas usan para dividir su gasto potencial.

Se trata de un estudio primario, donde las diferentes estrategias de análisis usadas son

- considerar la totalidad de los gastos (en proporciones)
- los 3 rubros más importantes
- dejar de lado el gradiente de gasto y analizarlos como atributos de ausencia o presencia (para combatir la gran asimetría de cada rubro)

Todas estas estrategias dan respuestas diferentes; sin embargo, las 3 tienen como elemento común que parten de la base de que la segmentación de los cruceristas puede hacerse independientemente de otras características, como son las socio-demográficas; es decir que

los clusters sólo usan la relaciones de gastos internas para los 3 o 4 rubros y nada más. Para eso es necesario considerar una lógica diferente, que permita pensar que el gasto responde a otros factores, para los que se proponen las siguientes alternativas

Construir perfiles de gasto usando toda la información que se dispone complementaria al tipo y nivel de gasto, en particular las socio-demográficas. Para eso se puede probar aplicar análisis factorial de correspondencias sobre las variables binarias de gastos, además del bloque de variables socio-demográficas. Sobre los factores creados a posteriori pueden crearse la tipología o perfiles de gasto, usando cualquiera de los métodos de clustering presentados. Por otra parte, es importante ver cómo funcionan los métodos de cluster difusos.

A diferencia de los métodos presentados, donde se obtiene una partición en k subconjuntos, es decir

$$N = \bigcup_1^n x_i = \bigcup_1^k \bigcup_{j=1}^{j=n_k} c_j$$

los clusters son determinados mediante algún algoritmo, en el agrupamiento **difuso (fuzzy)**, cada observación se 'extiende' a través de los distintos grupos. En este método se puede establecer: m parámetros de incertidumbre (fuzziness parameter), v_k es el centro del cluster c y u_{ik} el grado de pertenencia del individuo i al grupo k . Si se tiene n el número de observaciones, k el número de clusters, r es el parámetro de ajuste del modelo y $d(i,j)$ la disimilaridad entre las observaciones i y j , el método busca minimizar alguna función objetivo

$$\sum_{v=1}^{v=n} \frac{\sum_1^i \sum_1^j u_{i,v}^r u_{j,v}^r d(i,j)}{2 \sum_1^n u_{j,v}^r}$$

Luego para poder evaluar, ya no el perfil de gastos de los cruceristas, sino para poder hacer predicción, se propone otra metodología: Modelos de respuesta discreta, donde la variable de respuesta es

- (a) Si gasta en algún rubro Cuando gasta en alguno, la propensión a gastar en ese componente (Respuesta Binaria)
- (b) lo mismo que en el ítem (b) pero adecuando la variable de respuesta en niveles , para poder evaluar gradiente de gasto, con el uso de variables de respuesta politómicas ordinales
- (c) Modelos de respuesta de variable continua truncada, al considerar cada componente a través de su proporción del gasto total

Bibliografía

Andriotis, K. & Agiomirgianakis, G. (2010) "Cruise visitors' experience in a Mediterranean port of call". *International Journal of Tourism Research* 12(4): 390-404

Archer, B. (1995) "The impact of international Tourism on the economy of Bermuda". *Journal of Travel Research* 34(2): 27-30

- Ben-Akiva, M. & Lerman S. (1993) "Discrete choice analysis: theory and application to travel demand." The MIT Press, Cambridge
- Braun, B. M.; Dander, J. A. & White, K. R. (2002) "The impact of the cruise industry on a region's economy: a case study of Port Canaveral, Florida". *Tourism Economics* 8(3): 281–288
- Bresson, G. & Logossah, K. (2011) "Crowding-out effects of cruise Tourism on stay-over Tourism within the Caribbean. A non parametric panel data evidence". *Tourism Economics* 17(1): 127-158
- Brida, J. G.; Riaño, E. & Zapata Aguirre, S. (2011a) "Residents' attitudes and perception towards cruise Tourism development: a case study of Cartagena de Indias (Colombia)". *Tourism and Hospitality Research* 11(3): 187 – 202
- Brida, J. G.; Pulina, M.; Riaño, E. & Zapata Aguirre, S. (2011b) "Cruise passenger's experience embarking in a Caribbean Homeport". *Ocean & Coastal Management* 55: 135-145 DOI: 10.1016/j.ocecoaman.2011.10.003
- Brida, J. G.; Pulina, M.; Riaño, E. & Zapata Aguirre, S. (2011c) "A market segmentation analysis of cruise passengers visiting the port of Cartagena de Indias". *Tourism Geographies* 18(2): 431-447
- Brida, J. G.; Bukstein, D.; Garrido, N. & Tealde, E. (2011d) "Cruise passengers' expenditure in the Caribbean port of call of Cartagena de Indias: A Cross-Section data analysis". *Tourism Economics* 18(2): 425-434
- Brida, J. G. & Risso, W. A. (2010) "Cruise passenger expenditure analysis and probability of repeat visit to Costa Rica: a Cross-Section data analysis". *Tourism Analysis* 15(4): 425-434
- Brida, J. G. & Zapata Aguirre, S. (2010) "Economic impacts of cruise Tourism: the case of Costa Rica". *Anatolia, an International Journal of Tourism and Hospitality Research* 21(2): 322-338
- CLIA (2010) "Cruise Lines International Association - CLIA Cruise Market Overview: Executive summary". Available at <<http://clia-facts-cdn-remembers.s3.amazonaws.com/785dfa9a2c5360c08856c30b0e46e705.pdf>>, ingresado en 5 de Noviembre de 2011
- Cessford G. R. & Dingwall P. R. (1994) "Tourism on New Zealand's sub-Antarctic islands". *Annals of Tourism Research* 21(2): 318–332
- Chase, G. L. & McKee, D. L. (2003) "The economic impact of cruise Tourism on Jamaica". *Journal of Tourism Studies* 14(2): 16-22
- Chessel, D. and Dufour, A.B. and Thioulouse, J. (2004): The ade4 package-I- One-table methods. *R News*. 4: 5-10.
- Diedrich, A. (2010) "Cruise ship Tourism in Belize: The implications of developing cruise ship Tourism in an ecoTourism destination". *Ocean & Coastal Management* 53(5-6): 234-244

- Douglas, N. & Douglas, N. (2004) "Cruise ship passenger spending patterns in Pacific island ports". *International Journal of Tourism Research* 6(4): 251-261
- Dray, S. and Dufour, A.B. and Chessel, D. (2007): The ade4 package-II: Two-table and K-table methods. *R News*. 7(2): 47-52.
- Duman T. & Mattila, A. S. (2005) "The role of affective factors on perceived cruise vacation value." *Tourism Management* 26(3): 311–323
- Dwyer, L. & Forsyth, P. (1996) "Economic impacts of cruise Tourism in Australia". *Journal of Tourism Studies* 7(2): 36–43
- Dwyer, L. & Forsyth, P. (1998) "Economic significance of cruise Tourism". *Annals of Tourism Research* 25(2): 393–415
- Dwyer, L.; Douglas N. & Livaic, Z. (2004) "Estimating the economic contribution of a cruise ship visit". *Tourism in marine environments* 1(1): 5-16
- John Fox and Sanford Weisberg (2011). *An {R} Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage. URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>
- Gabe, T.; Lynch, C. & McConnon, J. (2006) "Likelihood of cruise ship passenger return to a visited port: the case of Bar Harbor, Maine". *Journal of Travel Research* 44: 281-287
- Gibson, P. & Bentley, M. (2006) "A study of impacts—cruise Tourism and the South West of England". *Journal of Travel & Tourism Marketing* 20(3/4): 63-77
- Hall, J. A. & Braithwaite, R. (1990) "Caribbean cruise Tourism: a business of transnational partnerships". *Tourism Management* 11(4): 339-47
- Hamilton, N. (2015). ggtern: An Extension to 'ggplot2', for the Creation of Ternary Diagrams. R package version 1.0.6.0. <http://CRAN.R-project.org/package=ggtern>
- Henthorne, T. L. (2000) "An analysis of expenditures by cruise ship passengers in Jamaica". *Journal of Travel Research* 38(3): 246–250
- Johnson, D. (2006) "Providing ecoTourism excursions for cruise passengers". *Journal of Sustainable Tourism* 14(1): 43-54
- Kester, J. G. C. (2003) "Cruise Tourism". *Tourism Economics* 9(3): 337–350
- Kim, S.; Prideaux, B. & Chon, K. (2010) "A comparison of results of three statistical methods to understand the determinants of festival participants' expenditures". *International Journal of Hospitality Management* 29(2): 297-307
- Lekakou, M. B.; Pallis, A. & Vaggelas, G. (2009) "Which homeport in Europe: the cruise industry's selection criteria". *TOURISMOS* 4(4): 215-240
- Lois, P.; Wang, J.; Wall, A. & Ruxton, T. (2001) "Fundamental considerations of competition

at sea and the application of cost-benefit analysis". *Tourism Today* 1: 89-102

Lumley, T. (2004) Analysis of complex survey samples. *Journal of Statistical Software* 9(1): 1-19

Lumley, T. (2014) "survey: analysis of complex survey samples". R package version 3.30.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2015). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.2.

Marti, B. (1992) "Passenger perceptions of cruise itineraries". *Marine Policy* 24(4): 360-370

Meyer, D., Achim Zeileis, and Kurt Hornik (2015). *vcd: Visualizing Categorical Data*. R package version 1.4-1.

Meyer, D., Achim Zeileis, and Kurt Hornik (2006). The *Strucplot* Framework: Visualizing Multi-Way Contingency Tables with *vcd*. *Journal of Statistical Software*, 17(3), 1-48. URL <http://www.jstatsoft.org/v17/i03/>

Ministerio de Turismo y Deporte, (2010) Anuario estadístico de turismo de cruceros". Disponible en www.turismo.gub.uy.

Miriela, C. G. L. & Lennie, P. (2010) "Cruise Tourists returning to Curaçao for a land-based vacation: A logit model". Working Paper N° 10/1, Research Department of the Central Bank of the Netherlands Antilles, Bank of the Netherlands Antilles, Willemstad

Morrison, A. M.; Yang, C. H.; O Leary, J. T. & Nadkarni, N. (2003) "Comparative profiles of travellers on cruises and land-based resort vacations". *Journal of Tourism Studies* 14(1): 99-111

Moscardo, G.; Morrison, A.; Cai, L.; Nadkarni, N. & Leary, O. (1996) "Tourist perspectives on cruising: multidimensional scaling analyses of cruising and other holiday types". *Journal of Tourism Studies* 7(2): 54-63

Petrick, J. F. & Sirakaya, E. (2004) "Segmenting cruisers by loyalty". *Annals of Tourism Research* 31(2): 472-475

Petrick, J. F. (2005) "Segmenting cruise passengers with price sensitivity". *Tourism Management* 26:753-762

Polydoropolou, A. & Litinas, N. (2007) "Demand models for Greek passenger shipping". *Research in Transportation Economics* 21: 297-322

Qu, H. & Ping, E.W.Y. (1999) "A service performance model of Hong Kong cruise travelers' motivation factors and satisfaction". *Tourism Management* 20(2): 237-244

Risso, W. (2012) "El gasto de los cruceristas en Uruguay 2008-2010", Risso W. A, *Revista Pasos*, vol 10 N°3

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Seidl, A.; Guillano, F. & Pratt, L. (2007) “Cruising for colones: cruise Tourism economics in Costa Rica”. *Tourism Economics* 13(1): 67-85

Seidl, A.; Guillano, F. & Pratt, L. (2006) “Cruise Tourism and community economic development in Central America and the Caribbean: The case of Costa Rica”. *Pasos* 4(2): 213-224

Silvestre, A.; Santos, C. M. & Ramalho, C. (2008) “Satisfaction and behavioral intentions of cruise passengers visiting the Azores”. *Tourism Economics* 14(1): 169–184

Teye, V. B. & Leclerc, D. (1998) “Product and service delivery satisfaction among North American cruise passengers.” *Tourism Management* 19(2): 153-160

Thurau, B.; Carver, A.; Mangun, J.; Basman, C. & Bauer, G. (2007) “A market segmentation analysis of cruise ship Tourists visiting the Panama Canal watershed: opportunities for ecotourism development”. *Journal of EcoTourism* 6(1): 1-18

Tibshirani, T. ; Gareth, J.; Hastie, T. y Witten, D. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer. Recuperado a partir de <http://link.springer.com/content/pdf/10.1007/978-3-319-00327-6.pdf>

Tobin, J. (1958) “Estimation of relationships for limited dependent variables”. *Econometrica* 26: 24–36

Vina, L. D. L. & Ford, J. (1998) “Economic impact of proposed cruise ship business”. *Annals of Tourism Research* 25(4): 205–208

Wilkinson, P. (1999) “Caribbean cruise Tourism: delusion or illusion?” *Tourism Geographies* 3:261–282

World Tourism Organization (2008) “Cruise Tourism- current situation and trends”. World Tourism Organization, Madrid

Anexos

Para los grupos creados considerando el nivel de gasto tenemos

Tabla 6: Porcentaje de cruceristas por clusters jerárquicos según número de personas (Gastos en niveles)

Número de Personas	Cluster por k-means con Gasto total, G de Tour, G en alimen y G en Shopping					
	1	2	3	4	5	Total Resultado
1	8,78%	6,41%	0,52%	5,05%	2,26%	6,87%
2	54,84%	35,17%	30,35%	46,44%	37,60%	49,10%
3	13,78%	20,78%	15,23%	19,01%	12,50%	15,07%
4	14,00%	25,23%	29,33%	14,72%	24,34%	16,68%
5	2,55%	3,36%	16,64%	2,99%	3,71%	3,59%

Más de cinco	6,07%	9,04%	7,93%	11,80%	19,58%	8,69%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia

Tabla 7: Porcentaje de cruceristas por clusters jerárquicos según porcentaje de hombres (Gastos en niveles)

Cluster por k-means con Gasto total, G de Tour, G en alimen y G en Shopping						
% Hombres	1	2	3	4	5	Total Resultado
0%	20,05%	18,98%	13,28%	15,06%	9,10%	17,64%
(0- 50]%	68,01%	66,74%	68,70%	76,83%	80,37%	70,75%
[50-100%)	6,89%	9,67%	16,80%	5,93%	7,63%	7,52%
100 %	5,05%	4,62%	1,21%	2,18%	2,90%	4,09%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 8: Porcentaje de cruceristas por clusters jerárquicos según % de adultos (Gastos en niveles)

Cluster por k-means con Gasto total, G de Tour, G en alimen y G en Shopping						
% adultos	1	2	3	4	5	Total Resultado
0%	14,76%	6,34%	18,39%	8,66%	11,18%	13,04%
(0- 50]%	9,12%	11,57%	18,31%	10,78%	15,53%	10,71%
[50-100%)	9,18%	12,07%	15,53%	16,26%	20,25%	12,05%
100 %	66,94%	70,02%	47,78%	64,31%	53,03%	64,19%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 9: Porcentaje de cruceristas por clusters jerárquicos según quintiles de gasto (Gastos en niveles)

Cluster por k-means con Gasto total, G de Tour, G en alimen y G en Shopping						
Gto Total quintiles	1	2	3	4	5	Total Resultado
(0,30]	34,13%					20,80%
(100,200]	6,36%		29,96%	72,55%	46,41%	22,90%
(200,5e+03]		100,00%	44,28%	27,45%	29,70%	16,20%
(30,60]	33,38%					20,34%
(60,100]	26,12%		25,76%		23,88%	19,75%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 10: Porcentaje de cruceristas por clusters jerárquicos según número de visitas (Gastos en niveles)

Cluster por k-means con Gasto total, G de Tour, G en alimen y G en Shopping						
Visitas	1	2	3	4	5	Total Resultado
1-Primera vez	64,11%	68,85%	84,61%	65,56%	72,04%	66,59%
2-Dos veces	14,17%	14,01%	7,19%	16,99%	8,15%	13,66%
3-Tres veces	7,42%	8,14%	3,87%	7,65%	7,44%	7,30%
4-Cuatro veces	3,68%	2,14%		1,50%	1,63%	2,80%
5-Cinco veces	0,97%			2,09%	0,47%	1,01%
6-Más de cinco veces	9,65%	6,86%	4,32%	6,21%	10,27%	8,64%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 11: Porcentaje de cruceristas por clusters jerárquicos según número de personas (Gastos en proporciones)

Cluster por k-means con Gastos en proporciones						
Número de Personas	1	2	3	4	5	Total Resultado
1	2,64%	4,25%	7,44%	11,05%	7,97%	6,87%
2	46,30%	51,29%	48,10%	52,00%	49,01%	49,10%
3	17,74%	15,27%	13,48%	9,88%	16,12%	15,07%
4	17,77%	17,16%	18,90%	12,81%	15,49%	16,68%
5	7,88%	4,15%	3,82%	3,42%	2,21%	3,59%
Más de cinco	7,66%	7,87%	8,25%	10,85%	9,21%	8,69%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 12: Porcentaje de cruceristas por clusters jerárquicos según porcentaje de hombres (Gastos en proporciones)

Cluster por k-means con Gastos en proporciones						
% de Hombres	1	2	3	4	5	Total Resultado
0%	19,29%	11,83%	16,01%	13,74%	21,27%	17,64%
(0- 50]%	67,98%	76,39%	69,99%	71,66%	69,38%	70,75%

[50-100%)	10,14%	9,08%	8,47%	5,72%	5,96%	7,52%
100 %	2,60%	2,70%	5,53%	8,88%	3,40%	4,09%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 13: Porcentaje de cruceristas por clusters jerárquicos según porcentaje de adultos (Gastos en proporciones)

Cluster por k-means con Gastos en proporciones						
% de adultos	1	2	3	4	5	Total Resultado
0%	14,73%	9,81%	12,93%	6,52%	15,08%	13,04%
(0- 50]%	13,99%	13,63%	10,18%	9,33%	9,25%	10,71%
[50-100%)	13,27%	10,26%	13,07%	13,02%	11,74%	12,05%
100 %	58,01%	66,31%	63,82%	71,13%	63,93%	64,19%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 14: Porcentaje de cruceristas por clusters jerárquicos según quintiles de gasto Total (Gastos en proporciones)

Cluster por k-means con Gastos en proporciones						
Gto Total quintiles	1	2	3	4	5	Total Resultado
(0,30]	19,36%	4,61%	34,50%	30,45%	18,07%	20,80%
(100,200]	20,73%	40,11%	13,70%	18,19%	22,58%	22,90%
(200,5e+03]	14,73%	27,67%	3,62%	17,46%	19,21%	16,20%
(30,60]	18,84%	9,64%	27,82%	20,05%	20,67%	20,34%
(60,100]	26,34%	17,98%	20,36%	13,85%	19,47%	19,75%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 15: Porcentaje de cruceristas por clusters jerárquicos según número de visitas (Gastos en proporciones)

Cluster por k-means con Gastos en proporciones						
Visitas	1	2	3	4	5	Total Resultado
1-Primera vez	77,13%	65,55%	62,59%	74,85%	65,64%	66,59%
2-Dos veces	10,37%	14,23%	12,71%	11,61%	15,11%	13,66%
3-Tres veces	4,25%	7,36%	9,54%	0,74%	7,68%	7,30%
4-Cuatro veces	1,59%	3,53%	2,67%	1,39%	3,07%	2,80%
5-Cinco veces	0,19%	1,70%	0,75%	2,11%	0,89%	1,01%
6-Más de cinco veces	6,47%	7,62%	11,74%	9,31%	7,61%	8,64%

Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
------------------------	----------------	----------------	----------------	----------------	----------------	----------------

Fuente: Elaboración propia .

Para cluster jerárquicos sobre variables binarias

Tabla 16: Porcentaje de cruceristas por clusters jerárquicos según número de personas (Gastos variables binarias)

Cluster por algoritmo Ward con Gastos en variables binarias						
Número de Personas	1	2	3	4	5	Total Resultado
1	8,36%	3,96%	7,60%	10,36%	2,44%	6,71%
2	48,56%	47,92%	47,85%	49,77%	45,83%	48,07%
3	15,71%	17,62%	12,29%	10,28%	16,29%	15,00%
4	14,20%	20,75%	17,59%	15,13%	15,74%	16,69%
5	2,07%	3,00%	3,39%	5,09%	9,09%	3,61%
Más de cinco	11,09%	6,74%	11,29%	9,37%	10,61%	9,92%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 17: Porcentaje de cruceristas por clusters jerárquicos según porcentaje de hombres (Gastos variables binarias)

Cluster por algoritmo Ward con Gastos en variables binarias						
% de Hombres	1	2	3	4	5	Total Resultado
0%	21,94%	13,28%	16,73%	14,60%	15,35%	17,44%
(0- 50]%	67,69%	75,61%	70,23%	70,61%	71,84%	70,78%
[50-100%)	6,86%	8,01%	7,83%	6,18%	11,31%	7,76%
100 %	3,51%	3,10%	5,21%	8,61%	1,50%	4,01%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 18: Porcentaje de cruceristas por clusters jerárquicos según porcentaje de adultos (Gastos variables binarias)

Cluster por algoritmo Ward con Gastos en variables binarias						
% de adultos	1	2	3	4	5	Total Resultado
0%	14,64%	9,45%	13,80%	6,41%	14,97%	12,56%
(0- 50]%	9,46%	12,51%	9,25%	9,55%	14,13%	10,64%
[50-100%)	12,61%	12,40%	12,87%	11,58%	15,45%	12,84%
100 %	63,29%	65,64%	64,08%	72,46%	55,45%	63,96%

Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
------------------------	----------------	----------------	----------------	----------------	----------------	----------------

Fuente: Elaboración propia.

Tabla 19: Porcentaje de cruceristas por clusters jerárquicos según quintiles de Gasto total (Gastos variables binarias)

Cluster por algoritmo Ward con Gastos en variables binarias						
Gto Total Quintiles	1	2	3	4	5	Total Resultado
(0,30]	18,87%	9,51%	37,44%	24,72%	7,05%	19,85%
(100,200]	19,68%	35,86%	10,20%	20,61%	22,84%	21,86%
(200,5e+03]	21,19%	24,25%	5,96%	18,92%	35,37%	20,04%
(30,60]	21,29%	11,46%	28,34%	19,07%	13,22%	19,41%
(60,100]	18,98%	18,92%	18,07%	16,68%	21,52%	18,85%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.

Tabla 20: Porcentaje de cruceristas por clusters jerárquicos según número de visitas (Gastos variables binarias)

Cluster por algoritmo Ward con Gastos en variables binarias						
Nro de Visitas	1	2	3	4	5	Total Resultado
1-Primera vez	65,60%	66,93%	58,94%	71,23%	77,12%	66,28%
2-Dos veces	15,35%	13,62%	13,16%	10,71%	10,67%	13,55%
3-Tres veces	6,98%	7,79%	10,09%	3,29%	5,25%	7,30%
4-Cuatro veces	3,16%	3,43%	5,28%	2,87%	0,71%	3,37%
5-Cinco veces	0,90%	1,22%	0,85%	1,45%	0,48%	0,97%
6-Más de cinco veces	8,01%	7,01%	11,69%	10,45%	5,77%	8,53%
Total Resultado	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración propia.