

Análisis Multivariado de la Varianza para la Conformación de Indicadores Sintéticos en las Cadenas Hoteleras de Pinar del Río

Multivariate Analysis of the Variance for the Conformation of Synthetic Indicators in the Hotel Chains of Pinar del Río

Resumen:

En la actualidad el proceso de medición de indicadores turísticos de Pinar del Río no cuenta con un indicador sintético que ofrezca un valor como medida de agregación del comportamiento de los indicadores de turismo, al no emplearse en su obtención ningún procedimiento multivariado para este fin. Lo anterior provoca que el proceso de toma de decisiones se vea afectado. En este sentido el presente trabajo consiste en elaborar un conjunto de indicadores sintéticos para las distintas cadenas hoteleras mediante el empleo de técnicas de Análisis Multivariante de la Varianza, que permitan la obtención de una medida global para establecer un ranking que sustente el proceso de toma de decisiones en las distintas cadenas hoteleras de Pinar del Río.

Palabras clave: bootstrap, indicador, MANOVA, turismo

Abstract:

At present, the process of measuring tourist indicators in Pinar del Río does not have a synthetic indicator that offers a value as a measure of aggregation of the behavior of tourism indicators, since no multivariate procedure is used for this purpose. This causes the decision-making process to be affected. In this sense, the present work consists of developing a set of synthetic indicators for the different hotel chains through the use of Multivariate Variance Analysis techniques, which allow obtaining

a global measure to establish a ranking that supports the process of making decisions in the different hotel chains of Pinar del Río.

Keywords: bootstrap, indicator, MANOVA, tourism

Introducción

El análisis multivariante es una disciplina difícil de definir, aunque por lo general agrupa diversas técnicas estadísticas que, si bien muchas de ellas fueron ideadas por autores que se puede denominar clásicos, deben su auge y puesta en práctica a la difusión del software estadístico y a la creciente demanda que de ellas exige el desarrollo de otras disciplinas (Fernández, 2008).

Es por ello que las investigaciones han utilizado, de manera creciente en los últimos años, el análisis de varianza con varias variables dependientes como técnica del análisis multivariante. Un acercamiento típico ha sido, realizar el análisis de varianza univariado para cada una de las variables dependiente. Sin embargo, esto presenta como dificultad la inflación del error de tipo I. (Rosales, 1990). El análisis multivariado de varianzas (MANOVA) dispone de técnicas globales de significación (Lambda de Wilks, Traza de Hotteling, -Lawley, Raíz máxima de Roy).

MANOVA es una generalización del análisis de la varianza univariante para el caso de más de una variable dependiente (Álvarez, 2017). Se trata de contrastar la significación de uno o más factores (variables independientes) para el conjunto de variables dependientes. Es un método estadístico para explorar simultáneamente la relación entre varias variables categóricas y dos o más variables dependientes medibles o métricas (Horta, 2006).

En el presente trabajo se planteó como objetivo elaborar un conjunto de indicadores sintéticos, mediante el empleo de técnicas de Análisis Multivariante de la Varianza, para las distintas cadenas hoteleras de Pinar del Río.

La aplicación del procedimiento MANOVA se hace difícil si no se dispone de un programa estadístico adecuado, por lo cual se utiliza en esta investigación el

lenguaje estadístico R 3.5.3 y el software R Studio 1.1.463, como soporte para el procesamiento de los datos en cuestión.

El lenguaje R como herramienta de análisis estadístico es uno de los paquetes informáticos cuyo uso ha crecido de manera más acelerada en los últimos años, teniendo como ventajas que cuenta con una amplia comunidad de expertos que continuamente introducen innovaciones (López & González, 2018).

Metodología

Se utilizaron métodos empíricos de investigación, basados en la observación científica y el análisis documental técnica para valorar la bibliografía relacionada con el problema de investigación y caracterizar la situación actual de la medición de los indicadores turísticos en Pinar del Río. La técnica de la entrevista, para determinar las cadenas hoteleras que fueron incluidas en la investigación y obtener información acerca de los indicadores turísticos. Entre los métodos estadísticos matemáticos, se utilizaron técnicas de análisis multivariado como MANOVA. También se empleó Bootstrap como herramienta de la minería de datos que permitió realizar transformaciones en las variables que no contemplaban normalidad. Se emplearon los softwares R 3.5.3 y R Studio 1.1.463 para el procesamiento de los datos y la información disponible.

Paralelamente, se empleó el método de medición para la descripción y análisis del comportamiento de los indicadores en cada una de las dimensiones, determinando la validez de éstos como representativos del concepto que se pretende calcular.

También fueron utilizados métodos teóricos para reseñar el desarrollo de los actuales procesos de gestión del turismo en Pinar del Río, a partir del empleo de indicadores, y como método lógico, la modelación, para la construcción de las funciones que garantizan la confección del nuevo procedimiento de agregación, se emplearon operaciones de análisis y síntesis, a través del estudio de los procedimientos de agregación para la construcción de indicadores sintéticos.

Resultados y discusión

La aplicación de entrevista semiestructuradas con los actores del Ministerio del Turismo en Pinar del Río (MINTUR), se determinaron las cadenas hoteleras o entidades a tratar. Las seleccionadas para establecer los indicadores sintéticos fueron Cadena Hotelera Cubanacan, Cadena Hotelera Islazul y Campismo Popular. De estas entidades se tomaron dos indicadores, uno referente a la eficacia y otro referente a la eficiencia: ingresos entre la cantidad de turistas extranjeros y el costo por peso para la aplicación de la herramienta MANOVA.

En el análisis descriptivo de los datos mediante gráficos de cajas (figura 1), se observa la existencia de diferencias entre las entidades analizadas, siendo Campismo Popular la menos eficiente pero la más eficaz, mientras que Cubanacan mantiene bajos valores de costo por peso y de ingresos por turistas. Islazul mantiene evidencia valores de costo por peso similares a Cubanacan, aunque la supera en los ingresos/turista.

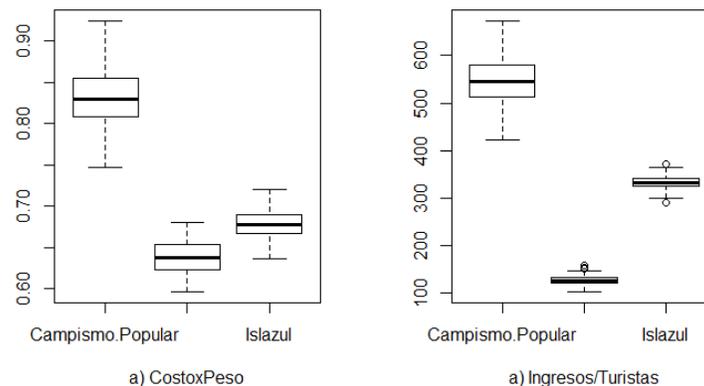


Figura 1. Gráficas de cajas para los indicadores de eficiencia y eficacia

La correlación entre las variables dependientes analizadas, se obtuvo un coeficiente de 0,89, significando, de manera global, que a mayor ingreso/turista el costo por peso aumenta. Sin embargo, no se considera un resultado favorable, ya que sería lógico una relación inversa entre estos indicadores.

Sin el empleo de pruebas no paramétricas referente a la normalidad de los datos, se puede observar en la figura 2 que este supuesto se viola, implicando la no

igualdad de varianzas. Como se puede observar, el conjunto de variables dependiente no mantiene normalidad, por definición de la normalidad invariante no se mantendrá la normalidad multivariada del conjunto de variables dependientes, lo cual, el modelo de análisis multivariado de la varianza perdería validez (Sanz, 2011).

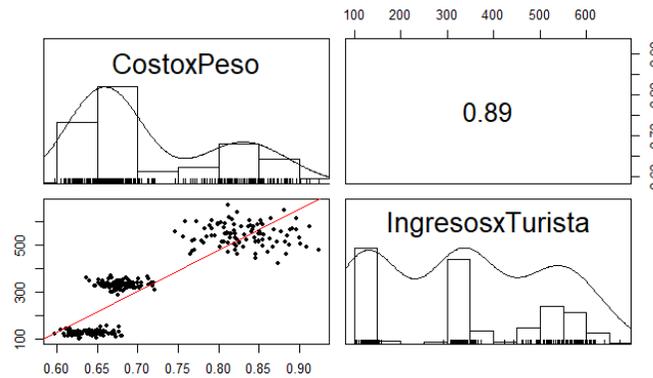


Figura 2. Gráficas de dispersión e histograma

Comprobando las sospechas de la falta de normalidad, se realizaron las pruebas de normalidad múltiple de Mardia (tabla 1), las mismas concretan la no existencia de la normalidad multivariante (MVN) (Cerron, 2016) teniendo como consecuencia un bajo coeficiente de determinación. Un análisis preliminar arrojó un coeficiente de determinación de 4,24% para la prueba ANOVA de la variable dependiente costo por peso y 0,63% para la variable dependiente ingresos por turistas.

Tabla 1. Test de Mardia de normalidad multivariada

Test	Statistic	P value
Mardia Skewness	51.87	0.00
Mardia Kurtosis	-3.22	0.00

En este momento, se hace necesario encontrar una transformación aceptable como respuesta a este problema. Existe un método que permite obtener, de una forma rápida una transformación que proporciona ciertos beneficios.

Bootstrap, el cual está basado en la idea de tratar a la muestra como una especie de "universo estadístico", muestreando repetidamente y utilizando las muestras

para estimar medias, varianzas, sesgos e intervalos de confianza para los parámetros de interés (Montoya, Vergara, Mora, & Gómez, 2015).

Después de descrito los elementos anteriores, se puede realizar el MANOVA. Partiendo de las bases conceptuales, Álvarez (2017) explica que al aplicar la técnica multivariante MANOVA se contrasta una sola hipótesis: que la media de los g grupos son iguales en las p variables dependientes lo que se traduce en

$$\left\{ \begin{array}{l} H_0: \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{pmatrix} = \begin{pmatrix} \mu_{31} \\ \mu_{32} \\ \mu_{33} \end{pmatrix} \\ \text{Al menos dos difieren} \end{array} \right.$$

Que equivale a $H_0: v_1 = v_2 = v_3$ donde v_k son la media de cada tratamiento.

En este caso los g son las distintas entidades y las p variables dependientes son los indicadores. En la siguiente tabla salida de R se muestra la aplicación del MANOVA con sus respectivas pruebas de significación.

Tabla 2. MANOVA

Test	Statistic	P value
Pillai	342.38	0.00
Wilks	1150.55	0.00
Hotelling-Lawley	3287.59	0.00
Roy	6513.44	0.00

En las distintas pruebas de significación se puede concluir que existe diferencias significativas entre las distintas entidades en los distintos indicadores.

Si se analiza cada variable dependiente a través de un ANOVA, se puede verificar en cuál variable dependiente existe diferencias significativas (tabla 3).

Tabla 3. ANOVA

Factor	Costo por peso (valor p)	Ingreso por turista (valor p)
Entidades	0.00	0.00
Coficiente R^2	0.9041	0.97

Se hace imprescindible el análisis de los residuos para comprobar si cumplen el supuesto de normalidad necesario para la aplicación de estas herramientas

estadísticas. En la figura 3 y 4 se observa el cumplimiento del supuesto de normalidad (Normal Q-Q Plot) no ocurriendo así con el supuesto de homocedasticidad (vs. Fitted values).

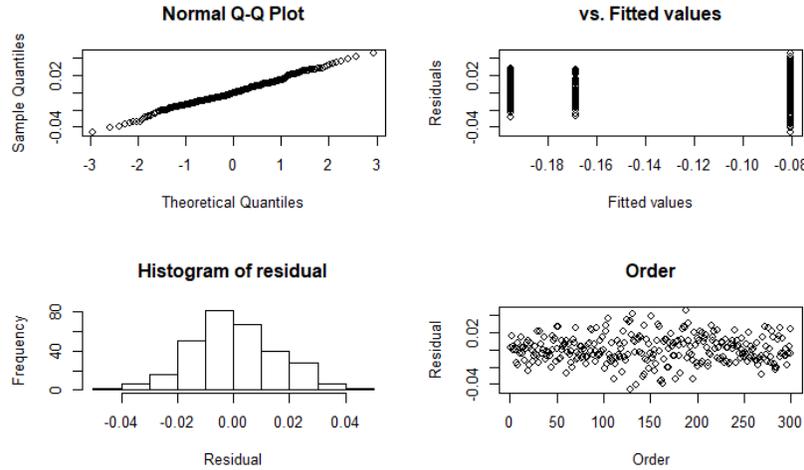


Figura 3. Supuestos teóricos indicador costo por peso

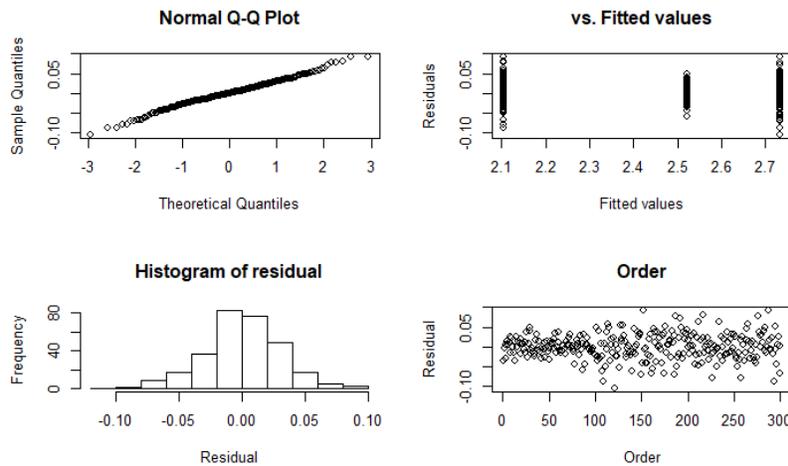


Figura 4. Supuestos teóricos indicador ingresos por turista

La causa de violación del supuesto de igualdad de varianza se debe a la alta diferencia que presenta la entidad Campismo Popular con respecto a las demás entidades. Lo anterior se evidencia mediante la prueba de comparaciones múltiples de Tukey (figura 5).

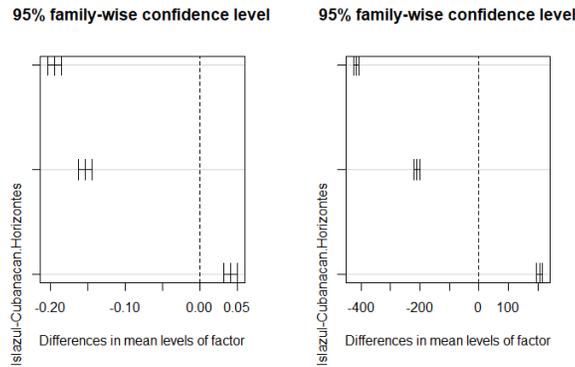


Figura 5. Prueba gráfica para la igualdad de varianzas

Repitiendo el procedimiento para el análisis de la varianza, pero sin incluir Campismo Popular, se logra la igualdad de varianzas para las entidades Islazul y Cubanacan, corroborando lo explicado anteriormente. A partir de esto, se puede obtener resultados más limpios al aplicar la herramienta del análisis multivariado de datos. En la figura 6 se muestra el cumplimiento de este supuesto.

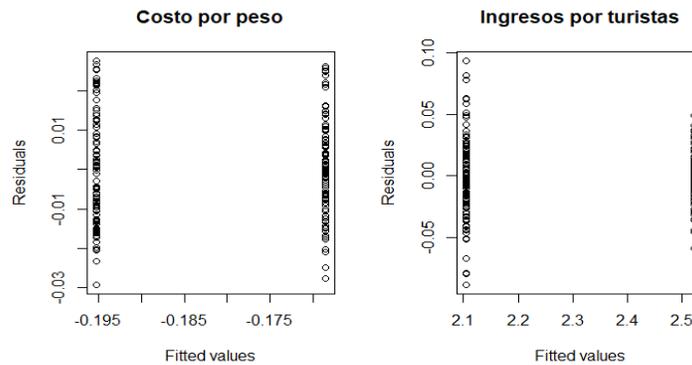


Figura 6. Supuestos teóricos para la homocedasticidad

Efectuando el ANOVA univariado sin incluir Campismo, los resultados arrojados por el software R, se introdujeron en la siguiente tabla, donde se muestran diferencias significativas entre estas entidades para ambos indicadores ($p \leq 0,01$), con coeficientes de determinación de 0,53 y 0,985 respectivamente.

Tabla 4. ANOVA sin campismo

Factor	Costo por peso (valor p)	Ingreso por turista (valor p)
Entidades	0.00	0.00
Coefficiente R^2	0.53	0.985

Para la conformación del indicador sintético (tabla 5), fue necesario asignar a cada subindicador el mismo peso que a los demás, en este caso los coeficientes de determinación R^2 (ver tablas 3 y 4), agregando la información mediante una suma (Delgado & Palomeque, 2017). La ponderación y agregación suele hacerse en niveles sucesivos de manera que previamente se ponderan y agregan una serie de variables para construir los subindicadores relativos a una determinada dimensión y, posteriormente, se agregan éstos para construir el indicador sintético (Nardo et al., 2005). Así, el indicador para una unidad i se define como $IS = \sum_{j=1}^m w_j I_j$ donde w_j es el peso asignado al indicador j .

Tabla 5.

Entidades	Costo por peso Coeficiente R^2	Ingreso por turista Coeficiente R^2	Costo por peso 1-CV	Ingreso por turista 1-CV	Suma ponderada	Ranking
Cubanacán	0.5313	0.985	0.6012	0.0794	0.3976	2
Campismo	0.9856	0.9094	0.3353	0.0208	0.3493	3
Islazul	0.5313	0.985	0.6676	0.3734	0.7224	1

Como indicador normalizado se utilizó el complemento del coeficiente de variación (CV) el cual mide el grado de homogeneidad de los valores de la variable. El CV es una medida del grado de heterogeneidad; se utiliza fundamentalmente para comparar períodos o etapas y permite hacer comparaciones entre conjuntos de datos heterogéneos.

Una vez determinados los pesos w_j , los valores del indicador sintético se obtienen mediante una suma ponderada de los valores normalizados de los indicadores del sistema (Parada, Fiallo, & Blasco, 2015).

En la tabla anterior se aprecia que la mejor cadena hotelera es Islazul, mientras que Campismo Popular evidencia una situación menos adecuada en cuanto eficiencia y eficacia, la cual está más alejada en cuanto a puntuación del resto de las entidades, sobre todo por los problemas de eficiencia que presenta.

Conclusiones

El análisis multivariado de la varianza es una herramienta estadística que permite identificar diferencias significativas entre distintos factores (entidades) en función de varias variables dependientes (indicadores), posibilitando determinar un coeficiente de determinación que mide el grado de variabilidad explicada por los factores, tomando estos coeficientes como peso para conformar el indicador sintético. Los indicadores sintéticos confeccionados dieron la posibilidad de determinar las puntuaciones para establecer un ranking entre las entidades. La entidad Campismo Popular es la menos eficiente en cuanto a la gestión turística.

Recomendaciones

Se sugiere al Ministerio del Turismo de Pinar del Río:

1. Tener en cuenta los indicadores sintéticos determinados para el análisis de la eficiencia y eficacia de la gestión turística.
2. Socializar los resultados de la investigación con los actores explícitos del análisis en la provincia.
3. Investigar los indicadores implícitos que influyen de forma negativa en la eficiencia de la actividad turística.

Bibliografía

- Álvarez, M. M. R. (2017). *Curso de análisis de investigaciones con programas informáticos* Universidad de Jaén.
- Cerron, J. P. (2016). Comparaciones de prueba de normalidad multivariada. *Anales Científicos*, 6.
- Delgado, A. T., & Palomeque, F. L. (2017). TheISOST index: A tool for studying sustainable tourism. *Journal of Destination Marketing & Management*.
- Fernández, J. M. (2008). *Análisis Multivariante*. Universidad de Extremadura: Colección manualesuex. Espacio Europeo de Educación Superior.
- Horta, D. S. (2006). Métodos estadísticos multivariados.
- López, R. F., & González, L. R. D. (2018). Algoritmo regresivo de integración numérica. *Revista Caribeña de Ciencias Sociales*, 10.
- Montoya, J. R., Vergara, I. O., Mora, J. R., & Gómez, S. G. (2015). Bootstrap and Jackknife resampling in reliability: Case Exponential and Weibull. *Revista Facultad de Ingeniería*, 25.

- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide* OECD Statistics Working Paper
- Parada, S. E., Fiallo, J., & Blasco, O. B. (2015). Construcción de indicadores sintéticos basados en juicio experto: Aplicación a una medida integral de la excelencia académica *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 16.
- Rosales, J. C. (1990). Interpretación del MANOVA: Análisis de la importancia de las variables dependientes. *Investigaciones*, 10.
- Sanz, J. O. (2011). *Métodos estadísticos y econométricos en la empresa y para finanzas*. España: Universidad Pablo de Olavide.