Estimación del gasto turístico en Honduras mediante Modelos Multinivel Bayesianos

Eduardo Canales 1\*, Asael Alonzo Matamoros 2

1Universidad Nacional Autónoma de Honduras, ORCID

2Aalto University, [Google scholar](https://scholar.google.com/citations?hl=en&user=_1aaLJYAAAAJ)

\* [edusalo03@hotmail.com](mailto:edusalo03@hotmail.com)

Recibido: Revisado; Aceptado: Publicado

**Resumen:** El turismo es una de las principales actividades de cualquier país en vías de desarrollo. Uno de los aspectos cruciales al estudiar indicadores de turismo es la estimación del gasto promedio por estadía. A la fecha los métodos de estimación utilizados no incorporan estructuras jerárquicas, generando estimaciones globales que fallan al no adecuarse a la simetría de la distribución de los datos. En este trabajo proponemos estimar el gasto turístico desagregado por el área que el turista visitó usando un modelo log-normal multinivel. Las estimaciones de la distribución se realizaron usando métodos Bayesianos con distribuciones a prior jerárquicas y débilmente informativas. Los resultados indican que nuestro modelo presenta un ajuste superior a los modelos globales propuestos por la literatura.

**Palabras clave:** Gasto turístico, Inferencia Bayesiana, Priori jerárquica, Modelos multinivel.

**Abstract:** Tourism is one of the main economic activities of any developing country. One of the crucial aspects when carrying out studies of tourism indicators is the estimation of average expenditure per stay. The methods used for this estimation have yet to incorporate multilevel hierarchies, leading to a failure to consider the asymmetry in the data distribution adequately. In this study, we propose an estimate of tourism expenditure disaggregated by areas visited by tourists using a multilevel log-normal model with hierarchical priors. The results indicate that our model is superior to the global models proposed in the literature.

**Keywords:** Touristic expenditure, Bayesian Inference, Hierarchical priors, multilevel models.

## 1. Introducción

El turismo es una actividad económica que en las últimas décadas se ha desarrollado de manera acelerada a nivel mundial. Según la OMT (2018) “es un sector fundamental de generación de ingresos en las economías emergentes y en desarrollo,” y cada vez son más los países que dan un mayor peso al turismo en la planificación de sus políticas económicas. En Honduras, al igual que en muchos otros países se considera a la actividad turística como prioritaria e importante para dinamizar la economía mediante la atracción de inversión nacional y extranjera, pero sin descuidar el concepto de desarrollo sostenible. Las estadísticas de turismo en Honduras se han centrado únicamente en el análisis a un nivel individual sin tener cuenta estructuras y relaciones más amplias que pueden existir al estudiar un fenómeno físico, sin embargo, muchos problemas de interés en la investigación y la práctica se desarrollan en contextos complejos donde los individuos están agrupados en grandes estructuras jerárquicas como equipos, organizaciones o regiones geográficas.

El Instituto Hondureño de Turismo ([2016](#ref-ITH-et-al-2016)) en su boletín de estadísticas presenta las tendencias del sector turismo en Honduras a nivel global, durante el periodo 2012-2016. Este trabajo usa los datos generados por la encuesta de caracterización del visitante del visitante, 2021 (ECV). Dicha encuesta extrae el gasto por estadía generado por los turistas que ingresan al país, y se realiza al momento que el turista ha realizado una gran parte de su estadía y está cerca de regresar a su país de residencia.

A graph with a line and a red line

Description automatically generated with medium confidence

**Figura 1.** Distribución del gasto turístico, donde aproximadamente el 11% representan valores atípicos o extremos, según los datos recolectados por la encuesta de caracterización del visitante (ECV) del Instituto Hondureño de Turismo. Fuente: Elaboración propia.

El análisis del gasto turístico es fundamental para comprender y evaluar el derrame económico que este conlleva a la economía del país. La estimación adecuada de este indicador mediante modelos estadísticos proporciona una compresión integral y precisa del gasto generado por parte de los turistas que ingresan a realizar actividades económicas en un país. El gasto turístico varía según la zona visitada debido a que se realizan diferentes actividades económicas en cada una de ellas, ver figura 2. Zonas muy populares como la zona insular del país generan alto costos por hospedaje debido a su alta demanda turística. Esto genera una gran variabilidad en el gasto dificultando el proceso de estimación, y métodos clásicos como una media muestral Instituto Hondureño de Turismo ([2016](#ref-ITH-et-al-2016)) o cualquier estimación global del gasto producen valores poco confiables e inexactos. Un fenómeno usual al estudiar estadísticas de turismo son los gastos atípicos generados por los turistas con alto poder adquisitivo, ver figura 1. Estos valores atípicos hacen que la distribución del gasto sea de colas pesadas, invalidando los supuestos de normalidad utilizados en los estimadores de medias muéstrales, y dificultando las comparaciones entre grupos, ver figura 2.

En ese sentido, para obtener estimaciones precisas es necesario utilizar enfoques estadísticos adecuados que permiten considerar la estructura jerárquica presente en el estudio del gasto por estadía generado por los turistas que visitan el país, capturando la variabilidad entre los turistas como individuos y teniendo en cuenta la estructura jerárquica presente según el destino principal que estos visitan. Los modelos multinivel permiten incorporar estructuras jerárquicas en el análisis teniendo en cuenta los efectos fijos como los efectos aleatorios lo cual los hace especialmente útiles para analizar datos longitudinales M. Lindstrom y Bates ([1990](#ref-Lindstrom-et-al-1990)), datos de medidas repetidas o agrupados en diferentes niveles.

Este estudio propone un modelo probabilístico log-normal multinivel estratificado por la zona visitada por los turistas, con medias entre grupos y varianza desconocidas. El objetivo del estudio es establecer la ley de probabilidad que rige el gasto turístico en Honduras, para eso evaluamos si el modelo propuesto ofrece un mejor ajuste y predicciones a los modelos propuestos por la literatura. Para corroborar las hipótesis nos basamos en el trabajo Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022), p. 8), que comparó el ajuste del gasto utilizando cuatro leyes de probabilidad distintas, *Gamma, log-Normal, inversa Gamma, Gamma generalizada, y log-normal asimétrica.* Este trabajo al igual que Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)) descartó el modelo de medias muestrales de Instituto Hondureño de Turismo ([2016](#ref-ITH-et-al-2016)), ya que el supuesto de normalidad de los datos acepta la posibilidad de ocurrir gastos negativos el cual es un supuesto no admisible para los indicadores de turismo.

La literatura realiza las estimaciones con métodos clásicos de máxima verosimilitud, y selecciona el modelo con criterios de información (Chen y Chen, [2008](#ref-BIC2006)) y pruebas de bondad de ajuste (Anderson y Darling, [1952](#ref-anderson1952); Shapiro y Wilk [1965](#ref-SWtest1965); Smirnov [1948](#ref-Smirnov1948)). Este estudio propone los métodos Bayesianos para inferencia que permite al usuario de esta aplicación agregar la información del experto de forma consistente, y una cuantificación adecuada de la incertidumbre para las estimaciones y predicciones, Gelman et al. ([2013](#ref-Gelman-et-al-2013)). Además, se propone seleccionar el modelo que presente una mayor capacidad predictiva, ya que la mayor utilidad de estos indicadores es para la toma de decisiones a futuro a nivel de país.

El modelo propuesto obtuvo estimaciones precisas del gasto turístico en Honduras y las colas pesadas de la distribución log-normal permite tener estimaciones resistentes a gastos atípicos. Adicionalmente, la estructura multinivel del modelo permite determinar la variabilidad del gasto en los diferentes destinos turísticos del país como estructura jerárquica, con el fin de proporcionar información más precisa y relevante para la planificación estratégica para impulsar el crecimiento y desarrollo del sector turismo en Honduras.

El resto del documento se resumen a continuación. El capítulo 2, presenta los modelos multinivel y sus aplicaciones en diferentes áreas de investigación, y se presentan diferentes estudios del gasto turístico en diferentes países. En el capítulo 3 hacemos una descripción estadística del Gasto turístico en Honduras y presentamos el modelo multinivel propuesto, como el modelo reciente propuesto por Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)). En el capítulo 4 presentamos los resultados obtenidos al implementar el modelo propuesto y el modelo de Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)) al gasto registrado por los turistas en Honduras en el año 2021, en un enfoque Bayesiano. Finalmente, en el capítulo 5 y 6 presentamos una discusión de las ventajas de nuestro modelo sobre los demás propuestos en la literatura y conclusiones.

## 2. Métodos

Los modelos multinivel, también conocidos como modelos de efectos mixtos son ampliamente utilizados para el análisis de datos, estos modelos permiten tener en cuenta los efectos fijos como los efectos aleatorios lo cual los hace especialmente útiles para analizar datos longitudinales M. Lindstrom y Bates ([1990](#ref-Lindstrom-et-al-1990)) o de panel Zubizarreta-Barrenetxea ([2020](#ref-Zubizarreta-et-al-2020)), Proença y Soukiazis ([2005](#ref-Proencca-et-al-2005)), datos de medidas repetidas o agrupados en diferentes niveles. En las últimas décadas las metodologías estadísticas para analizar medidas repetidas han tenido un notable desarrollo debido a la facilidad de su implementación gracias a los avances de la computación.

Laird y Ware ([1982](#ref-Nan-et-al-1982)) introdujo los modelos multinivel y su aplicación en el análisis de datos longitudinales que tiene en cuenta los efectos fijos como los efectos aleatorios, este trabajo sentó bases para el desarrollo de los modelos mixtos en diversas áreas de investigación. Adicionalmente, las estimaciones mediante los modelos multinivel se pueden obtener utilizando diferentes métodos; Correa Morales y Salazar Uribe ([2016](#ref-Correa-et-al-2016)) abordan cada una de estas y explica en detalle como pueden ser obtenidas a partir de medidas repetidas o longitudinales. Los primeros métodos numéricos para la estimación en los modelos multinivel son mínimos cuadrados Bates y Watts ([1988](#ref-Bates-et-al-1988)); máxima verosimilitud M. Lindstrom y Bates ([1988](#ref-Lindstrom-et-al-1988)); y máxima verosimilitud restringida M. Lindstrom y Bates ([1990](#ref-Lindstrom-et-al-1990)). En la actualidad los métodos Bayesianos se han vuelto popular para estimar modelos con estructuras multinivel ya que permiten obtener estimaciones confiables de los efectos fijos y aleatorios mediante la introducción de información adicional a través de la distribución a priori, y permiten cuantificar la incertidumbre de los efectos mediante el uso de la distribución a posteriori Bürkner ([2018](#ref-Paul-et-al-2018)).

Estos modelos son una extensión de los modelos de regresión lineal que acoplan varios modelos lineales para cada nivel de análisis, es decir considerar dentro un mismo modelo los distintos niveles de la estructura jerárquica y conocer la variabilidad debida al segundo nivel Fernández ([2012](#ref-Fernande-et-al-2012)), del mismo modo un modelo Bayesiano multinivel se construye a partir del modelo de regresión lineal ordinario e intentará predecir la variable de respuesta () mediante una combinación lineal de un intercepto y una pendiente que cuantifica la influencia de un predictor () , para mas detalles ver Rivas Martínez y Cruz Torres ([2022](#ref-Rivas-et-al-2022)). Las siguientes ecuaciones muestran la estructura clásica de un modelo lineal simple,

|  |  |
| --- | --- |
|  | (1) |

En este modelo las variables de respuesta se distribuyen normalmente alrededor de la media y varianza residual . El modelo anterior se puede extender al siguiente modelo multinivel con niveles o grupos, incorporando un intercepto variable expresado a continuación:

|  |  |
| --- | --- |
|  | (2) |

donde indica que a cada grupo se le da un intercepto único, Rivas Martínez y Cruz Torres ([2022](#ref-Rivas-et-al-2022)) menciona que además de la varianza también se está estimando un componente más que representa la varianza de la distribución de los interceptos variables, esta se considera como la variación del parámetro entre los grupos , siguiendo una metodología similar es posible introducir un termino de pendiente variables que pueda cambiar según el j-ésimo grupo.

(3)

A estas pendientes variables se les asigna una distribución a priori centrada en la gran pendiente , y con varianza . Con lo anterior, la inferencia Bayesiana nos permite realizar todas estas aseveraciones posibles que podemos aplicar en los modelos multinivel, para ello supongamos que tenemos observaciones agrupadas en grupos, la variable de interés es que representa la observación ésima en el grupo , podemos modelar la variable en un modelo multinivel lineal Bayesiano de la siguiente manera;

(4)

con distribuciones a priori:

donde, es la observación i-ésima en el grupo , es la media del grupo para la observación centrada en la media global y escala ; y es la varianza asociada con las observaciones individuales, también conocida como varianza residual o error relativo. Los valores y son hiper-parámetros conocidos y elegidos por el investigador.

Las actividades turísticas se consideran como una de las fuentes más importantes en el crecimiento económico de un país es por ello que la estimación del gasto turístico es un aspecto fundamental para comprender el impacto y poder tomar decisiones adecuadas. En base a ello se han realizado diversos trabajos a lo largo del tiempo, como Wei-Ting, Jui-Kuo, y Fei-Ching ([2013](#ref-Wein-Ting-et-al-2013)) el cual es uno de los primeros trabajos en aplicar un análisis multinivel de los determinantes de gasto turístico de los hogares donde los resultados obtenidos indican que las variables como edad, renta familiar, la propiedad de un vehículo y el uso del internet influyen positivamente en el gasto turístico, Zubizarreta-Barrenetxea ([2020](#ref-Zubizarreta-et-al-2020)) propone un modelo de regresión múltiple para una cuantificación del gasto turístico, asimismo Vena Oya ([2020](#ref-Vena-et-al-2020)) trabajo con determinantes del gasto turístico aplicando varios modelos multinivel.

Uno de los problemas que se generan al momento de realizar estimaciones del gasto turístico es que se obtienen colas pesadas en las distribuciones, esto debido a que hay muchos factores que pueden influir Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)). A lo largo de estos años muchos estudios se han enfocado en abordar este problema de colas pesadas; Craggs y Schofield ([2009](#ref-Ruth-et-al-2009)) mencionan que una solución es segmentar el gasto turístico por categorías en el cual identificaron asociaciones estadísticamente significativas entre los distintos segmentos de gasto, examinando la importancia de una serie de variables socioeconómicas y de comportamiento. Gómez-Déniz, Pérez-Rodríguez, y Boza-Chirino ([2020](#ref-Emilio-et-al-2020)) proponen que otra forma es dividir el gasto total de los turistas basados en el gasto según el país de origen y destino que estos tomen, por otra parte, en estudios más recientes como Nolasco P. ([2022](#ref-Nolasco-et-al-2022)) utilizaron series temporales para pronosticar el gasto medio de turista en España.

El modelo presentado por Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)), cumple con todas las características de nuestra investigación dado que será un punto de comparación con nuestro modelo propuesto, Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)) prueban en su artículo que realizando una parametrización de la distribución log-normal asimétrica de tres parámetros para la modelización del gasto turístico en base usando distintas covariables como el país de origen, destino y el gasto total, se obtienen resultados satisfactorios en los datos del gasto en las partes de la distribución empírica, de igual forma el modelo se adapta bien para captar la asimetría, curtosis y colas pesadas que las tres variables mencionadas tienden a presentar en la práctica, como lo es en nuestro caso, ver figura 2. En tal sentido por ello utilizamos modelos multinivel y así poder obtener estimaciones precisas en la variabilidad del gasto turístico.

A graph of different colored lines

Description automatically generated

**Figura 2.** Distribución del gasto turístico por estadía desagregado por zonas que visitaron en Honduras, según los datos recolectados por la encuesta (ECV). Fuente: Elaboración propia.

Para realizar las estimaciones del gasto turístico de los modelos propuestos, utilizaremos métodos de Markov Chain Monte Carlo (MCMC) como Metrópolis et al. ([1953](#ref-metropolis1953)), Hastings ([1970](#ref-Hasting1970)). En particular emplearemos un Monte Carlo Hamiltoneano Duane et al. ([1987](#ref-Simon-et-al-1987)), Hoffman y Gelman ([2014](#ref-hoffman14)) implementado en el lenguaje de programación probabilista Stan, Development. Team ([2017](#ref-Stan)), y ejecutado en el lenguaje de programación R, Bürkner ([2017](#ref-Bürkner-et-al-2017)). Para cada implementación se corrieron cuatro cadenas con valores iniciales distintas, con un total de iteraciones por cadena, eliminando las primeras iteraciones por cadena (*warm-up*). Presentamos las distribuciones a posteriori del modelo multinivel propuesto y evaluamos la convergencia de las cadenas utilizando el factor de convergencia (), y los tamaños de muestra efectivos () Vehtari et al. ([2020](#ref-Bob2020)). Para las visualizaciones de las cadenas se usó trazas y gráficos de densidades Gabry et al. ([2019](#ref-Gabry2019)), Gabry y Mahr ([2019](#ref-bayesplot2019)). Para, evaluar el ajuste del modelo utilizaremos evaluación de la densidad predictiva (*PPC*)[[1]](#footnote-1) visualizando el ajuste con gráficos de densidades, Säilynoja, Bürkner, y Vehtari ([2021](#ref-Sailynoja-etal-2021:graphical)). Finalmente, comparamos ambos modelos utilizando validación cruzada LOO-CV[[2]](#footnote-2), ver Vehtari, Gelman, y Gabry ([2016](#ref-Vehtari2016)). Dicho método computa la diferencia de las log-predictivas esperadas () removiendo una observación de la muestra y computar la log-predictiva del modelo para dicha observación faltante; este procedimiento es caro computacionalmente, pero puede ser aproximado usando un muestreo por importancia con un suavizado de Pareto Vehtari et al. ([2015](#ref-vehtari2015)), o por submuestreo Magnusson et al. ([2020](#ref-Måns-et-al-2020)).

## 3. Modelización del gasto turístico por estadía en Honduras.

La encuesta de caracterización del visitante (ECV) para el año se aplicó a turistas extranjeros, de los cuales todos ellos declararon un gasto válido para el análisis, alrededor del de los turistas pernoctaron al menos siete noches en el país y en promedio pernoctaron once noches *(estadísticas del número de noches por estadía; min. 1, media 9, mediana 7, max. 90)*. El gasto turístico promedio es de dólares, el gasto mínimo reportado fue de dolares, mientras que el gasto máximo fue de dólares. Además, el de los turistas encuestados reportaron un gasto mayor a dólares, esto es, el doble del gasto promedio registrado para el año .

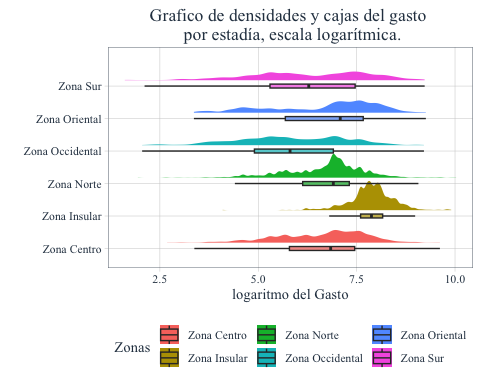
La tabla 1 proporciona un resumen detallado del gasto turístico de los visitantes sin aplicarle la transformación logarítmica a los datos, desglosado por las diferentes zonas que visitaron en Honduras, cada una de ella exhibe características únicas en términos de patrones de gasto, mencionar que las zonas están divididas según la (ECV) de la siguiente manera; *Zona centro, insular, norte, occidental, oriental y sur*. Por ejemplo, la *zona centro* muestra una concentración moderada del gasto alrededor de la mediana indicando una distribución relativamente equilibrada, caso contrario en la zona insular donde se observa una mediana más alta de , sugiriendo una concentración de gastos excepcionalmente altos y de valores atípicos. Además, las *zonas norte, oriental* y *sur* muestran una asimetría notable indicando una distribución desplazada o colas pesadas hacia valores superiores.

**Tabla** **1.** Estadísticas descriptivas de gasto turístico desagregado por zonas, de los visitantes que ingresaron al país año 2021.

| Zona | n | min. | Q1 | Mediana | Q3 | max. | asimetría | kurtósis |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Centro | 1,173 | 14.62 | 327.69 | 934.58 | 1,720.00 | 15,000.00 | 3.46 | 24.18 |
| Insular | 805 | 60.00 | 2,000.00 | 2,635.00 | 3,500.00 | 20,000.00 | 3.52 | 24.35 |
| Norte | 1,338 | 6.98 | 459.06 | 1,000.00 | 1,500.00 | 22,653.31 | 6.24 | 61.67 |
| Occidental | 616 | 7.76 | 134.64 | 332.93 | 1,000.00 | 10,000.00 | 3.54 | 21.94 |
| Oriental | 342 | 28.93 | 295.33 | 1,192.18 | 2,137.50 | 10,500.00 | 1.84 | 8.44 |
| Sur | 437 | 5.00 | 200.00 | 535.52 | 1,739.43 | 10,184.09 | 2.32 | 9.60 |
| Global | 4,711 | 5.00 | 350.00 | 1,050.00 | 2,050.00 | 22,653.00 | 3.58 | 27.27 |

Fuente: Elaboración propia.

Al analizar detalladamente la asimetría en cada zona, logramos identificar patrones en la distribución del gasto turístico; comenzando con la *zona centro* donde observa una asimetría de , indicando que la distribución tiene una cola más larga hacia los valores superiores, lo mismo sucede en las *zonas insular y norte* con valores de y respectivamente, lo que sugiere que hay presencia de valores atípicos, la *zona oriental* también muestra una asimetría más alta de , reflejando una cola más larga hacia la derecha, en la caso de la *zona occidental y sur* indican colas pesada pero no tan pronunciadas. En cuanto a la kurtósis, se notan variaciones en la forma de las distribuciones en el caso de las *zonas centro y oriental* exhiben colas pesadas y picos más notables respectivamente, lo cual implica que hay mayor concentración de valores extremos, la *zona norte* destaca con una kurtósis excepcionalmente alta de , las *zonas occidental y sur* presentan colas más acentuadas a la distribución normal, aunque menos extremas, lo mismo en la *zona insular*.



**Figura 3.** Distribución del gasto turístico logarítmico por estadía desagregado por zonas que visitaron en Honduras, según los datos recolectados por la encuesta (ECV). Fuente: Elaboración propia.

En la figura 2, se presenta las densidades del gasto turístico desagregado por zonas, donde se observa que la mayor parte de la información del gasto se concentra entre los primeros tres mil valores, el gráfico de cajas revela muy poca información, más del de los datos están cercanos a cero, estos gráficos indican que el gasto turístico no sigue una distribución normal. En ese sentido, se propone aplicar una transformación logarítmica a los datos para lidiar con los valores extremos de la muestra, el cual representan aproximadamente el , ver figura 1. En la figura 3, se presentan la transformación logarítmica del gasto desagregado por zonas, donde las densidades indican que los datos siguen una distribución normal, el gráfico de cuantiles mejora significativamente al aplicar dicha transformación en comparación con la figura 2.

En muchas situaciones, los datos empíricos muestran una ligera o marcada asimetría y colas pesadas, que refleja valores extremos, para definir la ley de probabilidad que modela el gasto turístico en Honduras, se propone un modelo lineal jerárquico sin covariables, utilizando un solo nivel con los datos en escala logarítmica. Sea una muestra aleatoria para el gasto turístico, donde es el gasto para la observación en la zona $, tal que y cada observación en escala logarítmica, se distribuye normal jerárquicas con vector de medias entre grupos y varianza global () desconocidas,

con distribuciones a priori:

* es la media entre grupos del gasto turístico para la zona j; que siguen una distribución normal con media cero y varianza diez.
* student-t es la varianza global del logaritmo del gasto, que sigue una distribución *t de Student* definida en los valores positivos, con grados de libertad, centrada en cero y con escala de diez.

En este estudio utilizamos distribuciones a priori débilmente informativas, y se eligen de tal forma que dichas distribuciones provean poca información y mejoren la geometría de la función propuesta tal que el muestreo de la distribución a posteriori sea más estable, Gelman et al. ([2020](#ref-Gelman-et-al-2020:workflow)). Asimismo, realizaremos una comparación con la ley de probabilidad propuesta por Gómez–Déniz, Dávila-Cárdenes, y Boza-Chirino ([2022](#ref-Gómez-et-al-2022)), el cual utilizó un modelo *log-normal asimétrico* para estimar el gasto turístico global. El modelo de Gómez (2022) es equivalente a un modelo lineal global sin covariables y distribución log-normal, el cual se expresa en la siguiente ecuación:

con parámetros de locación (), escala () y de forma () desconocidos, y distribuciones a priori:

* es la media del gasto turístico global; que siguen una distribución normal con media cero y varianza diez.
* student-t es la varianza global del logaritmo del gasto, que sigue una distribución *t de Student* definida en los valores positivos, con grados de libertad, centrada en cero y con escala de diez.
* es la asimetría global del logaritmo del gasto, que sigue una distribución normal estándar.

Donde son los índices de los datos globales, denota el número de observaciones, skew-N denota la densidad de una distribución normal asimétrica. Las distribuciones a priori elegidas en ambos modelos son débilmente informativas, esto con el objetivo de agregar poca información externa al modelo y mejorar el desempeño de los métodos numéricos. Descartamos las otras distribuciones propuestas por Gómez (2022) debido a su mal ajuste en comparación al modelo log-normal asimétrico, y la poca escalabilidad e interpretabilidad que ofrecen esas distribuciones en el contexto de turismo.

|  |
| --- |
| **Figura 4.** Densidades y cadenas de Markov del modelo log normal multinivel. Estos gráficos visualizan la convergencia del MCMC, la no convergencia indica que el método fallo en aproximar la distribución a posteriori, en este caso las cadenas de cada parámetro se mezclaron sin patrones abruptos, indicando convergencia. Por su lado las densidades no presentan múltiples modas indicando convergencia en un máximo local*.* Fuente: Elaboración propia. |

La tabla 3 proporciona un resumen de los resultados obtenidos de las distribuciones a posteriori del modelo log-normal multinivel. En primer lugar, la media representa el estimador puntual de los parámetros al igual que la mediana. Los valores de la desviación (*sd*) y desviación absoluta media (*MAD*) son los estimadores del error de Monte Carlo, valores bien cercanos a cero reflejan que las estimaciones obtenidas tienen un alto grado de confiabilidad y se aprecia que son estables en todas las zonas analizadas, los valores *Q5* y *Q95* señalan los intervalos de credibilidad al , el indicador  se refiere al diagnóstico de convergencia para comparar las estimaciones entre y dentro de la cadenas de cada parámetros estimado, valores muy cercanos a uno implican un buen diagnóstico, y lo interpretamos como que las cadenas del MCMC muestrearon y convergieron eficientemente.

**Tabla 2.** Resumen de la comparación de las predictivas de los modelos log-normal multinivel y Gómez (2022). La tercera columna muestra el ajuste de las log-predictivas esperadas usando LOO-CV, la cuarta columna la estimación del criterio de información de Watanabe, y las primeras dos columnas muestran las diferencias de las log-predictivas y sus errores estándar respectivamente.

| **modelo** |  |  |  | **WAIC** |
| --- | --- | --- | --- | --- |
| log-Normal  multi-nivel | 0.00 | 0.00 | -7,172.74 | 14,345.47 |
| Gómez (2022) | -577.05 | 28.35 | -7,749.78 | 15,499.56 |

Fuente: Elaboración propia.

Los valores y son indicadores para medir la eficiencia de las estimaciones de tamaño de muestra efectivo, la primera se enfoca en la regiones donde la densidad de probabilidad es más alta es decir la parte central de la distribución y la segunda en las colas o extremos de la distribución, estos valores deben ser cercanos a o similares entre si, en nuestro caso los resultados indican que en cada zona refleja la eficiencia con lo que las cadenas de están explorando y muestreando en ambas partes de la distribución.

A group of blue lines

Description automatically generated

**Figura 5.** Posterior Predictive Checks (PPC) del modelo log-normal multinivel. Para cada zona, la línea azul oscuro representa la función de probabilidad de los datos obtenidos, y la línea azul claro, representa el ajuste a posteriori del modelo propuesto. Fuente: Elaboración propia.

Finalmente, los valores obtenidos se observan bien y los indicadores de convergencia muestrean bien nuestro modelo propuesto para la estimación del gasto turístico el cual se puede observar en la figura *4*, donde las cadenas MCMC lograron hacer un *“mixing”* efectivo al juntarse entre sí y parecer estacionaras, las densidades se ven simétricas y sin múltiples modas, siendo estos indicadores de convergencia, poca convergencia indica que el método falló en aproximar la posterior de los parámetros.

Los resultados obtenidos del modelo de Gómez ([2022](#ref-Gómez-et-al-2022)), todos los son cercanos a uno siguiendo la misma explicación de nuestro modelo, asimismo los y también son cercanos a las iteraciones obtenidas, por lo tanto, aceptamos nuevamente la convergencia del modelo. La figura 5 muestra los PPC del modelo log-normal multinivel donde vemos que las predicciones generadas por el modelo son similares a los datos reales para cada una de las zonas, lo cual sugiere que nuestro modelo es capaz de capturar adecuadamente la variabilidad del gasto turístico desagregado por zonas.

La tabla 2muestra los resultados obtenidos por validación cruzada, la primera columna representa la diferencia del log predictivas esperada (), la diferencia es de unidades a favor del modelo multinivel por lo cual nuestro model tiene mayor capacidad predictiva que el modelo global de Gómez (2022), la segunda columna corresponde al error estándar de las diferencias de estimación del valor en este caso la diferencia es poca, aceptando el ajuste del LOO-CV. La tercera columna son los valores del log-predicitiva de cada modelo, mediante el método de validación cruzada *LOO-CV* el cual indica la calidad predictiva, se observa que nuestro modelo presenta mejor resultados. Por último, tenemos el valor del criterio de información de Watanabe-Akaike *WAIC* Watanabe ([2010](#ref-watanabe2010)), dicho criterio elige al modelo con valor menor, el cual confirma que nuestro modelo es el que presenta mejores resultados.

**Tabla 3.** Resumen de las distribuciones a posteriori obtenidas del modelo log-normal multinivel. La media y mediana representan las estimaciones puntuales de los parámetros desconocidos del modelo, los valores sd y MAD representan los errores de Monte-Carlo obtenidos para los estimadores asociados, los valores Q5 y Q95 representan los intervalos de credibilidad del parámetro. Finalmente, los valores y son valores para evaluar la convergencia del MCMC.

| **Zona** | **media** | **mediana** | **sd** | **MAD** | **Q5** | **Q95** |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Centro | 6.62 | 6.62 | 0.03 | 0.03 | 6.57 | 6.67 | 1.00 | 5,429.56 | 2,758.98 |
| Insular | 7.87 | 7.87 | 0.04 | 0.04 | 7.80 | 7.94 | 1.00 | 6,382.14 | 3,077.30 |
| Norte | 6.69 | 6.69 | 0.03 | 0.03 | 6.64 | 6.74 | 1.00 | 5,623.11 | 2,849.54 |
| Occidental | 5.86 | 5.86 | 0.04 | 0.05 | 5.78 | 5.93 | 1.00 | 5,469.62 | 3,097.26 |
| Oriental | 6.69 | 6.68 | 0.06 | 0.06 | 6.59 | 6.78 | 1.00 | 5,847.90 | 2,773.99 |
| Sur | 6.28 | 6.28 | 0.05 | 0.05 | 6.19 | 6.37 | 1.00 | 6,406.71 | 2,982.07 |
| sigma | 1.11 | 1.11 | 0.01 | 0.01 | 1.09 | 1.13 | 1.00 | 5,716.40 | 3,060.18 |

Fuente: Elaboración propia.

## 4. Discusión

En la comparación entre el modelo log-normal multi-nivel propuesto y el modelo de referencia presentado por Gómez ([2022](#ref-Gómez-et-al-2022)) se observa una serie de diferencias sustanciales que respaldan la viabilidad y ventajas de nuestro enfoque. Una característica destacada del modelo propuesto es la eliminación de la necesidad de estimar un parámetro de forma, el cual suele ser bien engorroso de estimar, esto simplifica el proceso de estimación y refuerza la robustez de las predicciones; en cambio el modelo de Gómez ([2022](#ref-Gómez-et-al-2022)) asume una uniformidad del gasto en todo el país sin considerar las variaciones que se presenten en las distintas zonas turísticas es decir para los datos del gasto turístico global como se observa en la figura 1. Nuestro modelo multinivel si aborda explícitamente estas diferencias que surgen con lo cual se obtienen estimaciones más precisas y ajustadas como se observó en los resultados de la validación cruzada. Asimismo, es importante resaltar que según el criterio  del LOO-CV el modelo propuesto puede predecir una nueva información con precisión, proporcionando información más precisa y relevante para la planificación estratégica impulsando el crecimiento y desarrollo del sector turismo en Honduras.

## 5. Conclusiones

En este trabajo, hemos profundizado la relevancia que tiene la distribución log-normal y los modelos multinivel en el análisis del gasto turístico en Honduras. La distribución log-normal es particularmente útil para modelar datos en los cuales se cuenta con colas pesadas como se observó en la figura 2 del gasto turístico, permitiendo capturar de manera efectiva los valores atípicos o extremos y al aplicar transformaciones logarítmicas y estratificar se logró regularizar las distribuciones mejorando la estabilidad y precisión de las estimaciones. En relación con los modelos multinivel hemos podido estudiar la capacidad que estos tienen para identificar la variabilidad y estructura jerárquica de los datos, proporcionando una comprensión enriquecedora y contextualizada de los factores que pueden influir en diversas situaciones de estudio. De esta forma, este trabajo provee una aplicación novedosa para el análisis del gasto turístico en Honduras, que se ajusta a los supuestos teóricos del gasto, a las condiciones de colas pesadas por el exceso de valores atípicos, y a la variabilidad del gasto por zona visitada. Además, el uso de métodos Bayesianos resulta en modelos con alta capacidad predictiva y cuantificar la incertidumbre de las estimaciones de forma automática.

## Agradecimientos

Agradecemos al Instituto Hondureño de Turismo ([2016](#ref-ITH-et-al-2016)) por disponer de la encuesta de caracterización del visitante (2021), sin la cual hubiera sido imposible caracterizar el fenómeno del gasto turístico en Honduras.

## Referencias

Anderson, T., and D. Darling. 1952. “Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes.” *Annals of Mathematical Statistics* 23 (2): 193–212. <https://doi.org/10.1214/aoms/1177729437>.

Bates, D., y D. Watts. 1988. *Nonlinear Regression Analysis y Its Applications*. Wiley.

Bürkner, P. 2017. “Brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software* 80 (1): 1–28. <https://doi.org/10.18637/jss.v080.i01>.

———. 2018. “Advanced Bayesian Multilevel Modeling with the R Package brms.” *The R Journal* 10 (1): 395–411. <https://doi.org/10.32614/RJ-2018-017>.

Chen, J., and Z. Chen. 2008. “Extended Bayesian Information Criteria for Model Selection with Large Model Spaces.” *Biometrika* 95 (3): 759–71. <https://doi.org/10.1093/biomet/asn034>.

Correa Morales, JC., y JC. Salazar Uribe. 2016. “Introducción a Los Modelos Mixtos.” *Escuela de Estadistica*.

Craggs, R., y P. Schofield. 2009. “Expenditure-Based Segmentation and Visitor Profiling at the Quays in Salford, UK.” *Tourism Economics* 15 (1): 243–60. [https://doi.org/10.5367/000000009787536753](%20https://doi.org/10.5367/000000009787536753).

Duane, S., A. Kennedy, B. Pendleton, y D. Roweth. 1987. “Hybrid Monte Carlo.” *Physics Letters B* 195 (2): 216–22. <https://www.sciencedirect.com/science/article/pii/037026938791197X>.

Fernández, Virginia Pérez. 2012. “Los Modelos Multinivel En El análisis de Factores de Riesgo de Sibilancias Recurrentes En Lactantes: Enfoques Frecuentista y Bayesiano.” PhD thesis, Universidad de Murcia.

Gabry, J., y T. Mahr. 2019. “Bayesplot: Plotting for Bayesian Models.” <https://mc-stan.org/bayesplot>.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, y A. Gelman. 2019. “Visualization in Bayesian Workflow.” *J. R. Stat. Soc. A* 182: 389–402. <https://doi.org/10.1111/rssa.12378>.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, y D. Rubin. 2013. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. <https://books.google.nl/books?id=ZXL6AQAAQBAJ>.

Gelman, A., A. Vehtari, D. Simpson, C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P. Bürkner, y M. Modrák. 2020. “Bayesian Workflow.” *arXiv Preprint arXiv:2011.01808*.

Gómez-Déniz, E., J. Pérez-Rodríguez, y J. Boza-Chirino. 2020. “Modelling Tourist Expenditure at Origin y Destination.” *Tourism Economics* 26 (3): 437–60. <https://doi.org/10.1177/1354816619840845>.

Gómez–Déniz, E., N. Dávila-Cárdenes, y J. Boza-Chirino. 2022. “Modelling Expenditure in Tourism Using the Log-Skew Normal Distribution.” *Current Issues in Tourism* 25 (14): 2357–76. [https://doi.org/10.1080/13683500.2021.1960282](%20%20https://doi.org/10.1080/13683500.2021.1960282).

Hastings, W. 1970. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” *Biometrika* 57 (1): 97–109. <http://www.jstor.org/stable/2334940>.

Hoffman, Matthew D., y Andrew Gelman. 2014. “The No-u-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.” *Journal of Machine Learning Research* 15: 1593–623. <http://jmlr.org/papers/v15/hoffman14a.html>.

Instituto Hondureño de Turismo, República de Honduras. 2016. “Boletín de Estadísticas de Turismo 2012 - 2016.” <https://iht.hn/es/boletin-2012-2016>.

Laird, N., y J. Ware. 1982. “Random-Effects Models for Longitudinal Data.” *Biometrics* 38 (4): 963–74. <http://www.jstor.org/stable/2529876>.

Lindstrom, M., y D Bates. 1990. “Nonlinear Mixed Effects Models for Repeated Measures Data.” *Biometrics* 46 (3): 673–87. <http://www.jstor.org/stable/2532087>.

Lindstrom, M, y D. Bates. 1988. “Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data.” *Journal of the American Statistical Association* 83 (404): 1014–22. <http://www.jstor.org/stable/2290128>.

Magnusson, M., M. Andersen, J. Jonasson, y A. Vehtari. 2020. “Leave-One-Out Cross-Validation for Bayesian Model Comparison in Large Data.” <https://arxiv.org/abs/2001.00980>.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. 1953. “Equation of State Calculations by Fast Computing Machines.” *The Journal of Chemical Physics* 21 (6): 1087–92. <https://doi.org/10.1063/1.1699114>.

Nolasco P., Evelyn Y. 2022. “Pronóstico Del Gasto Medio Por Turista En España Mediante El Uso de Series Temporales.”

Proença, S., y E. Soukiazis. 2005. “Demand for Tourism in Portugal: A Panel Data Approach.”

Rivas Martínez, L., and C. Cruz Torres. 2022. “Análisis Multinivel de Factores Que Afectan El Rendimiento Escolar En Español Tercer Grado En Honduras.” *Paradigma: Revista de Investigación Educativa* 29 (48): 93–119. <https://iniees.vrip.upnfm.edu.hn/ojs/index.php/Paradigma/article/view/171>.

Säilynoja, T., P. Bürkner, y A. Vehtari. 2021. “Graphical Test for Discrete Uniformity and Its Applications in Goodness of Fit Evaluation and Multiple Sample Comparison.” *arXiv Preprint arXiv:2103.10522*.

Shapiro, S., y M. Wilk. 1965. “An Analysis of Variance Test for Normality (Complete Samples).” *Biometrika* 52 (3-4): 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>.

Smirnov, N. 1948. “Table for Estimating the Goodness of Fit of Empirical Distributions.” *Annals of Mathematical Statistics* 19 (2): 279–81. <https://doi.org/10.1214/aoms/1177730256>.

Stan, Development. Team. 2017. “Stan: A c++ Library for Probability and Sampling, Version 2.16.0.” <http://mc-stan.org/.>

Vehtari, A., A. Gelman, y J. Gabry. 2016. “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC.” *Statistics and Computing* 27 (5): 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, y P. Bürkner. 2020. “Rank-Normalization, Folding, and Localization: An Improved for Assessing Convergence of MCMC.” *Bayesian Analysis*. International Society for Bayesian Analysis. <https://doi.org/10.1214/20-BA1221>.

Vehtari, A., D. Simpson, A. Gelman, Y. Yao, y J. Gabry. 2015. “Pareto Smoothed Importance Sampling.” <https://arxiv.org/abs/1507.02646>.

Vena Oya, J. 2020. “Determinantes Del Gasto Efectivo En El Turismo Cultural.” <http://hdl.handle.net/10481/62288>.

Watanabe, S. 2010. “Asymptotic Equivalence of Bayes Cross Validation y Widely Applicable Information Criterion in Singular Learning Theory.” *Journal of Machine Learning Research* 11. <http://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf>.

Wei-Ting, H., S. Jui-Kuo, y W. Fei-Ching. 2013. “A Multilevel Analysis on the Determinants of Household Tourism Expenditure.” *Current Issues in Tourism* 16 (6): 612–17. <https://doi.org/10.1080/13683500.2012.725714>.

Zubizarreta-Barrenetxea, L. 2020. “Análisis Estadı́stico Del Gasto Medio Del Turista En España y Los Factores Que Lo Determinan: Una Propuesta de análisis de Datos de Panel.

## 

## 

1. Por sus siglas en ingles, Posterior predictive Checks. [↑](#footnote-ref-1)
2. Por sus siglas en ingles, Leave One Out Cross Validation. [↑](#footnote-ref-2)