

# Una aproximación no lineal a la distribución del gasto total de turistas cruceristas en Uruguay

G. Brida, B. Lanzilotta, L. Moreno, F. Santiñaque

25 de octubre de 2017

## 1 Objetivo

## 2 Antecedentes

## 3 Metodologías

- CART
- Random Forest
- Modelos de Regresión en dos pasos con Lasso

## 4 Base de Datos

## 5 Resultados

- Modelización del Gasto Total con CART
- Modelización del Gasto Total con Random Forest
- Modelización del Gasto Total con Regresión a dos pasos con LASSO

## 6 Conclusiones

Modelizar el gasto total de cruceristas en Uruguay en función de características socio-económicas, de satisfacción y contexto recogida en la encuesta del MINTUR de temporada 2014-2015.

Para cumplir dicho objetivo se aplican las siguientes técnicas:

- CART
- Random Forest
- Modelo de Regresión a dos pasos con Lasso

Se ponen en competencia dichos modelos para conocer cual de ellos brinda el mejor ajuste.

Los trabajos académicos que analizan los determinantes del gasto turístico a nivel individual (microeconómico) ha sido relativamente menor que el que estudia la demanda macroeconómica agregada.

Algunos antecedentes recientes de trabajos sobre demanda macroeconómica: (Dogru et al. 2017), (Peng et al. 2014 y 2015), (Lin et al. 2015), (Dwyer et al. 2012), (Divisekera 2013), (Song y Li 2008) y (Lim y otros 2006), entre otros

A nivel microeconómico: nuevas metodologías innovadoras y originales para estudiar los factores determinantes del gasto turístico se pueden ver en (Bellani 2016), (Marrocu et al. 2015), (Olya Mehran 2017), (Melstrom 2017), (Abbruzzo et al. 2014)

# Metodología: CART

Una aproximación no lineal a la distribución del gasto total de turistas cruceristas en Uruguay

G. Brida,  
B. Lanzilotta,  
L. Moreno, F. Santiaque

Objetivo

Antecedentes

Metodologías

CART

Random Forest

Modelos de Regresión en dos pasos con Lasso

Base de Datos

Resultados

Modelización del Gasto Total con CART

Modelización del Gasto

Las metodologías conocidas como Árboles de clasificación y Regresión (CART), son útiles para realizar predicciones para nuevas observaciones, cuya estimación se basará en el promedio o moda de la región en la que caiga (ver (Breiman et al. 1984), (James et al. 2013) )

Su implementación se basa en estratificar o segmentar el espacio de variables predictoras (covariables) según ciertos criterios: criterio de partición, criterio de parada y criterio de asignación de etiqueta.

Es una técnica complementaria a los modelos de regresión clásicos, con la ventaja de no necesitar de ningún supuesto a priori (método *machine learning*).

# Metodología: Random Forest

Una aproximación no lineal a la distribución del gasto total de turistas cruceñistas en Uruguay

G. Brida,  
B. Lanzilotta,  
L. Moreno, F. Santiaque

Objetivo

Antecedentes

Metodologías

CART

Random Forest

Modelos de Regresión en dos pasos con Lasso

Base de Datos

Resultados

Modelización del Gasto Total con CART

Modelización del Gasto

Métodos basados en combinaciones de árboles:

- Bagging: Combinación de  $n$  árboles basados en  $n$  muestras bootstrap.
- Random forest: Idem que Bagging, pero en cada split sólo se considera un subconjunto de variables. Esto reduce la correlación entre árboles.
- Boosting: La generación de cada árbol es secuencial, con una reponderación de los datos en cada paso.

Es una metodología que se basa en la combinación de muchos árboles.

Mejoran capacidad predictiva de los CART, obteniendo estimadores con menor varianza.

Pierden interpretabilidad, aunque disponen de métodos para clasificar la importancia de las covariables.

# Metodología: Regresión en dos pasos con Lasso

Una aproximación no lineal a la distribución del gasto total de turistas crucevistas en Uruguay

G. Brida,  
B. Lanzilotta, L.  
Moreno, F.  
Santiaque

Objetivo

Antecedentes

Metodologías

CART

Random  
Forest

Modelos de  
Regresión  
en dos  
pasos con  
Lasso

Base de  
Datos

Resultados

Modelización  
del Gasto  
Total con  
CART

Modelización  
del Gasto

Hay muchos enfoques posibles para el análisis de datos que contienen una gran cantidad de ceros. La metodología apropiada depende de la fuente generadora de ceros en los datos. En el caso que las observaciones  $y_i$  sean ceros genuinos los modelos que suelen utilizarse son del tipo *Hurdle Models* o *Two-Part Models*.

Si además se dispone de muchas variables explicativas, es necesario realizar selección de variables para obtener un modelo parsimonioso y con alto poder de predicción.

La metodología denominada *LASSO* es una técnica de penalización de los coeficientes de las variables, que tiene la ventaja de que permite hacer regresión y selección de variables a la vez.

$$\max_{\beta} Ln(f(y|\beta)) + ||\beta||_1$$

Se consideran datos individuales de la encuesta realizada a cruceristas por parte del Ministerio de Turismo correspondiente a la temporada 2014-2015.

Las variables que se incluyen en el modelo son (total 59 variables):

- Variables socio-económicas: Residencia, Edad, Ocupación.
- Variables de viaje: Mes visita, Número de visitas anteriores, Puerto, Sitios visitados.
- Tipología de grupos: Total personas del grupo (Solo, en pareja o más de dos personas) y Tipo de grupo (Mixto, solo mujeres ó solo hombres).
- Variables de satisfacción: variables de agrado y desagrado.

La aplicación fue realizada utilizando la variable Gasto Total categorizada según "No Gasta", "Gasto reducido" (hasta 20 dólares), "Gasto moderado" (entre 20 y 50 dólares) y "Gasto alto" (más de 50 dólares).

Para comparar el error de clasificación de cada modelo, se separa la muestra original en una muestra de entrenamiento y una muestra test.



# Modelización del Gasto Total con CART

Una aproximación no lineal a la distribución del gasto total de turistas cruceristas en Uruguay

G. Brida,  
B. Lanzilotta,  
L. Moreno, F. Santiaque

Objetivo

Antecedentes

Metodologías

CART

Random Forest

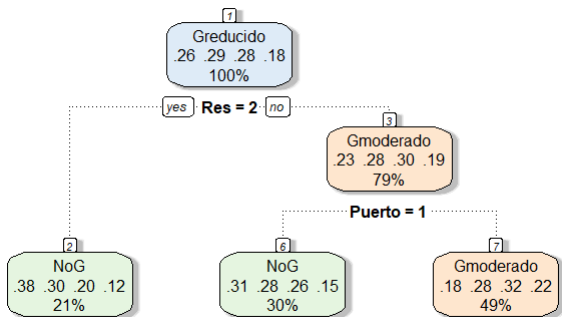
Modelos de Regresión en dos pasos con Lasso

Base de Datos

Resultados

Modelización del Gasto Total con CART

Modelización del Gasto



# Modelización del Gasto Total con Random Forest

Una aproximación no lineal a la distribución del gasto del gasto total de turistas cruceristas en Uruguay

G. Brida,  
B. Lanzilotta,  
L. Moreno, F. Santiaque

Objetivo

Antecedentes

Metodologías

CART

Random Forest

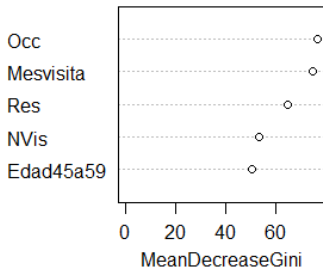
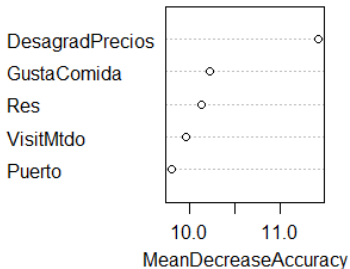
Modelos de Regresión en dos pasos con Lasso

Base de Datos

Resultados

Modelización del Gasto Total con CART

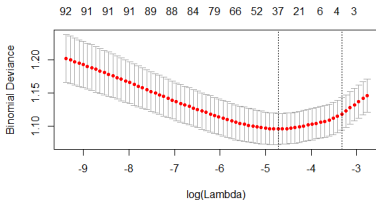
Modelización del Gasto



# Modelización del Gasto Total con Regresión a dos pasos con LASSO

PRIMER PASO: Modelo *Logit* (Gasta o no gasta) con penalización Lasso

A través de validación cruzada, se estudia la bondad del ajuste del modelo en función de la variación del parámetro de penalización  $\lambda$ .



Variables	Est. Coef.
Intercept	1.1907
Desagrada Precios	-0.1405
Puerto (Punta del Este)	-0.1965
Residencia (Arg)	-0.2444
Edad30a44	0.0592

# Modelización del Gasto Total con Regresión a dos pasos con LASSO

Una aproximación no lineal a la distribución del gasto total de turistas cruceristas en Uruguay

G. Brida,  
B. Lanzilotta, L.  
Moreno, F.  
Santíaque

Objetivo

Antecedentes

Metodologías

CART

Random Forest

Modelos de Regresión en dos pasos con Lasso

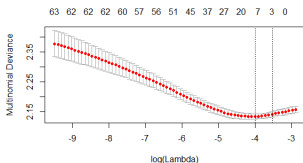
Base de Datos

Resultados

Modelización del Gasto Total con CART

Modelización del Gasto

## SEGUNDO PASO: Modelo Multinomial con penalización Lasso



Est. coef para prob. de "Gasto reducido"

Variabes	Est. Coef
Intercept	0.1873
Gusta Gente	-0.1855
Gusta Comida	-0.0399
Residencia (Arg.)	0.1219
Desagrada Precios	0.0096
Mes visita (Enero)	-0.0300
Mes visita (Marzo)	0.0679
Grupos > 2 personas	0.2497
Edad (<18 años)	0.0873

Est. coef para prob. de "Gasto alto"

Variabes	Est. Coef
Intercept	-0.3345
Ocupacion (Empresario, Gerente, etc)	0.0797
Loyalty (> 3 visitas)	0.1079
Edad (45 a 59)	0.0088

Se aplicaron tres metodologías alternativas para modelar el gasto total de pasajeros de cruceros en Uruguay en la temporada 2014-2015 con respecto a un conjunto de covariables.

Las tres metodologías son innovadoras.

Del análisis de competencia de los modelos surge que Random Forest evidencia una mayor capacidad predictiva con respecto a los otros dos modelos aplicados.

Los resultados de las tres metodologías coinciden en resaltar el papel predominante de la residencia y el puerto de desembarco, como las variables más asociadas con el gasto total.

Estos resultados son similares a los encontrados en [1] en donde se destaca como variables determinantes del gasto la residencia, el hecho de no viajar solo, el puerto de desembarque y la satisfacción.

## Bibliografía



A. BELLANI , *El turismo de cruceros en Uruguay: determinantes socioeconómicos y comportamentales del gasto en los puertos de desembarco*, Master thesis, Universidad de la República, 2016.



G. JAMES, D. WITTEN, T. HASTIE, R. TIBSHIRANI , *An Introduction to Statistical Learning*, Springer, 2013.



Brida, J.G., Bukstein, D. Tealde, E. (2013) *Exploring cruise ship passengers' spending patterns in two Uruguayan ports of call*. Current Issues in Tourism, DOI: 10.1080/13683500.2013.861391



Brida, J.G., Bukstein, D., Garrido, N., Tealde, E Zapata-Aguirre, S. (2010). *Cruise passengers expenditure in the Caribbean port of call Cartagena de Indias: A cross-section data analysis*. Estudios y Perspectivas de Turismo, 19, 607-634.



Brida, J.G., Fasone, V., Scuderi, R. Zapata-Aguirre, S. (2014). "ClustOfVar and the segmentation of cruise passengers from mixed data: some managerial implications". Knowledge-Based Systems 70: 128–136.



Brida, J.G., Garrido, N. Such Devesa, M.J. (2012). *Cruise passengers satisfaction: Cartagena de Indias*. Benchmarking: An International Journal, 19(1), 55-69.



Brida, J.G., Pulina, M., Riaño, E. Zapata-Aguirre, S. (2012). *Cruise passenger's experience embarking in a Caribbean Home Port*. Ocean and Coastal Management, 55, 135-145.



Brida, J.G. Scuderi, R. (2013). *Determinants of tourist expenditure: A review of microeconomic models*. *Tourism Management Perspectives*, 6, 28-40.



Brida, J.G. Zapata-Aguirre, S. (2010). *Cruise tourism: economic, socio-cultural and environmental impacts*. *International Journal Leisure and Tourism Marketing*.



Dwyer, L. Forsyth, P. (1998). "Economic significance of cruise tourism". *Annals of Tourism Research*, 25 (2), 393–415.



Risso, W. A. (2012). *El gasto de los cruceristas en Uruguay 2008 – 2010*. *Revista de Turismo y Patrimonio*, 10 (3), 393 – 406