

# Teoría

*Juan Manuel Marin Quintero*

*April 9, 2017*

## **MODELOS GAMLSS**

Los modelos GAMLSS (Modelos Aditivos Generalizados para Ubicación, Escala y Forma) nacen debido a que en la práctica los modelos de regresión lineal estándar en algunas ocasiones no cumplen los supuestos clásicos sobre normalidad y la varianza constante de los errores, además la linealidad de la relación entre la variable respuesta y las explicativas rara vez se mantienen. Nelder y Wedderburn [1972] y Hastie y Tibshirani [1970] introdujeron modelos lineales generalizados (GLM) y modelos aditivos generalizados (GAM) para superar algunas de las limitaciones del modelo lineal estándar.

Mikis Stasinopoulos (2015)

Es un moderno enfoque basado en la distribución de los modelos de regresión (semiparamétricos), donde todos los parámetros de la distribución asumida para la respuesta pueden ser modelados como funciones aditivas de la variable explicativa.

GAMLSS proporciona más de 90 distribuciones continuas, discretas y mixtas para modelar la variable de respuesta. También se pueden usar versiones truncadas, censuradas, logarítmicas y de mezcla finita de estas distribuciones. Se anexa el listado de las Principales Distribuciones Gammls en el siguiente link

<http://www.gamlss.org/wp-content/uploads/2013/01/Distributions-2010-only-The-table.pdf> Mikis Stasinopoulos (2015)

En este caso no se pretende modelar un modelo de regresión, sino ajustar una distribución de las familias GAMLSS a un solo conjunto de datos explicado por una sola variable, mediante un algoritmo de maximización no lineal. No existen variables explicativas, ya que solo se tiene en cuenta el análisis de una sola variable dada por el usuario, solo se desea saber de las distribuciones de la familias GAMLSS cual es la que mejor ajusta ese conjunto de datos explicados por dicha variable.

El usuario ingresa su base de datos en formato csv, especificando el tipo de separación de los datos, si tiene o no encabezado de variables y selecciona una de las variables que tiene la base de datos. A continuación elige una del grupo de familias Reales, Reales Lineales, Reales Positivas, Reales de 0 a 1, Conteos y Binomial. Cabe aclarar que tienen ciertas restricciones por ejemplo si se eligen familia Reales de 0 a 1 tienen que estar los datos en ese intervalo y Los conteos deben tener falso y verdadero. La penalización K es un valor que se da para castigar la cantidad de parámetros de una distribución, entre mayor sea el valor, con mayor facilidad saldrán distribuciones con menor cantidad de parámetros. La aplicación internamente estima cada una de las distribuciones que hay en el grupo de la familia y mediante un criterio de AIC selecciona las mejores cuatro distribuciones. Finalmente se muestra un gráfico con el valor de los parámetros de dichas distribuciones con su correspondiente densidad en los datos.

## **Referencias**

Mikis Stasinopoulos, Vlasios Voudouris, Bob Rigby. 2015. "Flexible Regression and Smoothing the Gamlss Packages in R." *Http://Www.gamlss.org/Wp-Content/Uploads/2015/07/FlexibleRegressionAndSmoothingDraft-1.pdf* 1 (1): 17–18.