

# Aplicación Shiny para una alternativa de predicción de tiempo en desarrollo de software

**Mateo Restrepo Higuita** *Universidad Nacional de Colombia - sede Medellín, [marestrepohi@unal.edu.co](mailto:marestrepohi@unal.edu.co).*  
**Valentina Yepes Valle** *Universidad Nacional de Colombia - sede Medellín, [vyepesv@unal.edu.co](mailto:vyepesv@unal.edu.co).*

---

**Resumen:** Algunas empresas de desarrollo de software han presentado el problema de estimar el tiempo que tarda un desarrollador para realizar un proyecto dado. Este tiempo se estima inicialmente mediante el juicio de expertos, con metodologías que solo tienen en cuenta el criterio que parte de sus experiencias previas. Sin embargo, se ha presentado que en la práctica dicha estimación resulta ser diferente, demorándose en muchos casos más de lo esperado y ocasionando pérdidas monetarias a estas empresas. En principio buscamos un acercamiento al problema en una empresa particular, planteando desde la organización de la información histórica hasta el análisis estadístico de la misma y teniendo en cuenta variables como: tipo de desarrollo (DT) solicitado por el cliente, el tiempo asignado por juicio del experto (TE), el tiempo real en horas que se tardó el proyecto (TR), entre otras. Se espera generar un modelo estadístico que permita una estimación más precisa del tiempo de duración de los proyectos relacionados con el desarrollo de software y presentar los resultados junto con los detalles técnicos de los datos reales de una empresa particular, este proceso se implementará por medio de una aplicación Shiny para así obtener el tiempo estimado basado en el modelo.

*Keywords:* Shiny, Estimación, Desarrollo de software, Modelo estadístico.

---

## Introducción

En este trabajo se busca hacer un análisis del comportamiento de los datos históricos de una empresa dedicada al desarrollo de software que están almacenados en una base de datos, buscando una solución al problema de estimaciones de tiempo para desarrollo de software en esta empresa. Lo que se busca con este análisis es poder mejorar las estimaciones de tiempo de esta entidad para que así puedan tener una mayor veracidad al dar el veredicto de un tiempo a un determinado cliente que quiere contratar sus servicios, debido a que sus estimaciones no son tan acertadas en todos los casos ya que el tiempo estimado para pactar un trato lo hace un experto basado en su experiencia, el problema se abarca usando los modelos de regresión múltiple en este caso en particular los modelos GAMLSS basado en la implementación de ([Stasinopoulos. and Rigby, 2007](#)) , proponiendo una aplicación Shiny de estos modelos que nos permita obtener el tiempo estimado.

## Resultados

### Elaboración de figuras

De la Figura 1 se logró observar que los datos no tienen una varianza constante, puesto que se aprecia que entre más aumenta el tiempo estimado más dispersión tienen los datos, por lo tanto, no es adecuado usar un modelo lineal simple (lm) estudiado en ([Montgomery., 2002](#)) a causa de que no se cumple el supuesto de varianza constante, por otro lado consideramos un mejor

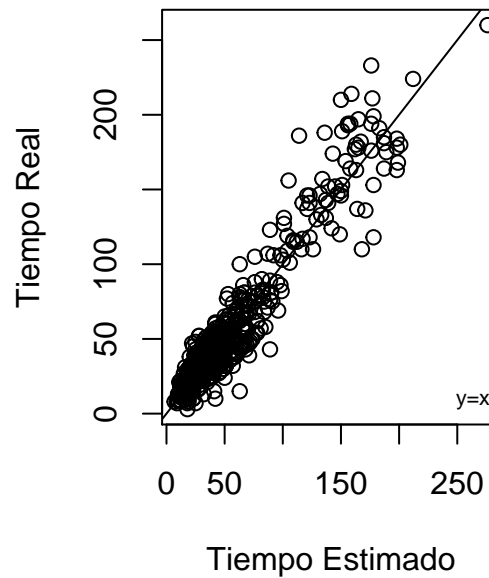


Figure 1: Gráfico dispersión tiempo real vs tiempo estimado

ajuste a estos datos con un modelo GAMLSS, además se comparan con la recta  $Y=X$  para ver el comportamiento de las estimaciones realizadas por el experto y comprar estas con el tiempo real que se tardó el proyecto y tener una idea inicial de que tanto subestima o sobrestima el experto.

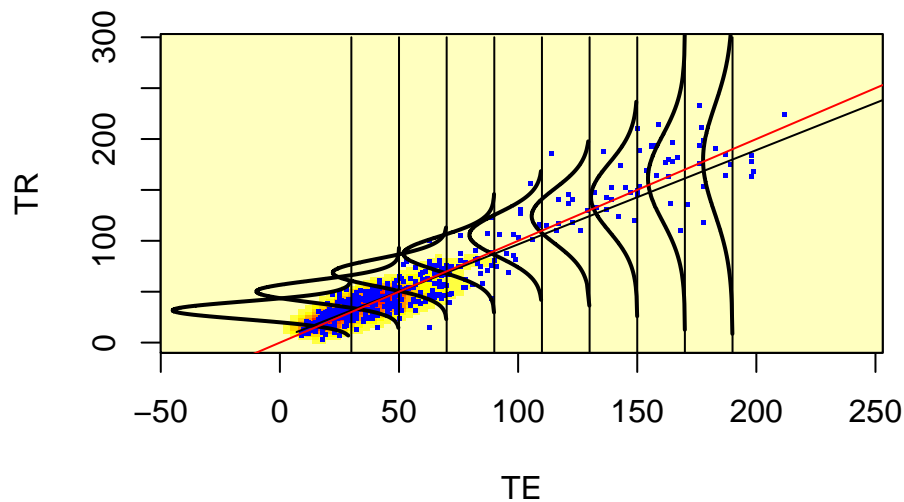


Figure 2: Supuestos del modelo de regresion GAMLSS propuestor por (R.A. and D.M., 2005)( Recta roja:  $y=x$ , recta negra: Modelo GAMLSS)

De la Figura 2 podemos observar que se cumplen los supuestos de regresión lineal ya que cada curva de densidad gaussiana se comporta como una distribución normal, por la estructura de los datos se puede apreciar que entre más aumente el tiempo estimado más variabilidad tienen los datos (mayor varianza).

Además, podemos analizar que de acuerdo con la línea roja el experto en el tema de la empresa en la mayoría de tipos de proyecto está sobrestimando el tiempo que se demora cada proyecto,

en el caso de los desarrolladores con un tiempo estimado entre 0 y 50 el experto da un tiempo acertado para el proyecto de acuerdo con el tiempo real.

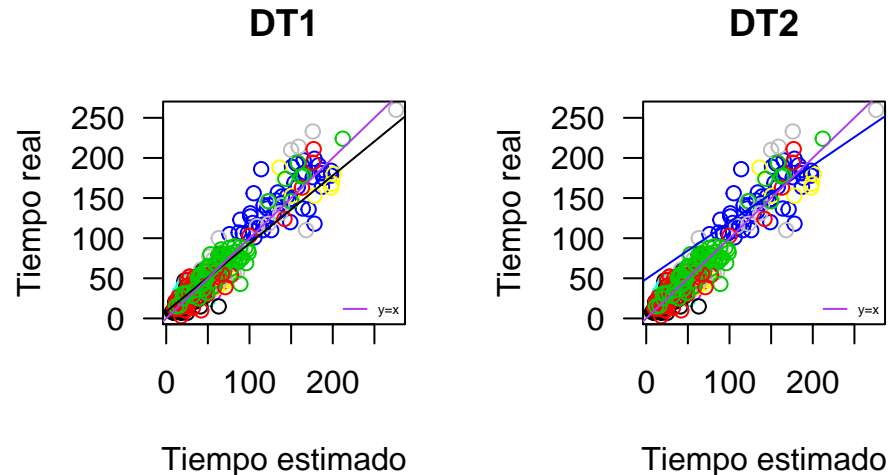


Figure 3: Análisis de cada tipo de proyecto (DT) con modelo GAMLSS.

De la Figura 3 observamos que se toman 2 de los análisis que se le realizó a los 11 tipos de proyectos para ejemplificar, el análisis realizado fue hecho con modelo gamlss y la recta  $Y=X$ .

## Conclusiones

Después de un análisis se pudo observar que una covariable de gran significancia para el modelo de los datos fue el tipo de desarrollo, se pudo apreciar que algunas de las regresiones generadas con el modelo GAMLSS de cada tipo de proyecto (DT) estiman mejor el tiempo teniendo de referencia de la recta  $y=x$ , al realizar estas estimaciones se logró concluir que para proyectos con una duración máxima de 65 horas el experto tenía una buena estimación del tiempo necesario, por el contrario para proyectos con duraciones mayores a 65 horas el experto siempre subestimaba o sobrestimaba el tiempo. Partiendo de esto se plantea darle solución a este problema a partir de un modelo general GAMLSS para los datos, Además al realizar una aplicación Shiny se permite mostrar la información de manera más ágil y rápida, para así facilitar el entendimiento del proceso para todo el público.

## References

- Montgomery., P. 2002. *Introducción al Análisis de Regresión lineal*. 1<sup>a</sup> ed. Ed. Ediciones CECSA. México DF.
- R.A., Rigby and Stasinopoulos D.M. 2005. *Generalized Additive models for location, scale and shape*. Appl. Statist., 54, part 3, pp 507-554.
- Stasinopoulos. and R. Rigby. 2007. *Generalized additive models for location, scale and shape (GAMLSS) in R*. Journal of Statistical Software 23(7), 1–46.