

# INTELIGENCIA DE NEGOCIOS ADAPTATIVOS APLICADA A LA EMPRESA DE SERVICIOS PÚBLICOS E.S.P EMSERPUCAR DEL MUNICIPIO DE CARTAGENA DEL CHAIRÁ

Adaptive business intelligence applied to the public utilities  
company E.S.P. EMSERPUCAR of the municipality of  
Cartagena del Chairá

**Jhon William Araque Farfán** 

<sup>1</sup>Estudiante, Matemática Aplicada, Universidad  
Surcolombiana.

Correo electrónico: [cristianandresreyesmora@gmail.com](mailto:cristianandresreyesmora@gmail.com)

**Cristian Andrés Reyes Mora** 

<sup>1</sup>Estudiante, Matemática Aplicada, Universidad  
Surcolombiana.

Correo electrónico: [@xxxxx.com](mailto:@xxxxx.com)

**Ana Victoria Perdomo Fajardo** 

<sup>1</sup>Estudiante, Matemática Aplicada, Universidad  
Surcolombiana.

Correo electrónico: [@xxxx.co](mailto:@xxxx.co)

**Jasmidt Vera Cuenca** 

<sup>2</sup>Docente, Matemática Aplicada, Universidad Surcolombiana

Correo electrónico: [jasmidt.vera@usco.edu.co](mailto:jasmidt.vera@usco.edu.co)

Recibido/Received: 13/01/2021 • Aprobado/Approved: 29/05/2021 • Publicado/Published: 20/06/2021

## Resumen

Este proyecto se basa en la construcción de un modelo artificial a partir del *software Rstudio*, el cual permite al director de una empresa de servicios contar con una idea intuitiva sobre el funcionamiento de su compañía; lo anterior, con el fin de que pueda tomar decisiones acertadas acerca de su negocio. Con este modelo, y siguiendo de ciertos parámetros se podrá explorar el comportamiento futuro de la empresa, gracias al uso de las siguientes herramientas: conceptos estadísticos, técnicas de predicción y optimización para desarrollar el autoaprendizaje. Así mismo, a partir del estudio propuesto se generaron algunas conclusiones, recomendaciones y/o sugerencias, con el fin de orientar a los gerentes de negocio a tomar decisiones que aumenten la eficiencia, la productividad y la competitividad de sus organizaciones. Una de las predicciones que aquí se propone y que está dirigida específicamente a la organización objeto de estudio, es la de que, del 100% de los usuarios que habría para el año 2021, solo el 50% pagaría el servicio de alcantarillado.

**Palabras clave:** Modelo matemático, inteligencia de negocios, regresión lineal múltiple, *Rstudio*, series de tiempo.

## Abstract

This project is based on the construction of an artificial model in *Rstudio*, which allows the director of a service company to have an intuitive idea of its operation in order to make the right decisions about his business. With this model, based on certain parameters, the future behavior of the company can be explored, making use of statistical concepts, prediction and optimization techniques to develop self-learning. In this way, some conclusions, recommendations and/or suggestions were generated to guide business managers to make decisions that increase

## Cómo citar

Araque-Farfán, J. W.; Reyes-Mora, C.A. Perdomo-Fajardo, A.V. y Vera-Cuenca, J. (2021). Inteligencia de negocios adaptativos aplicada a la empresa de servicios públicos e.s.p. emserpucar del municipio de Cartagena del Chairá. *Revista FACCEA*, Universidad de la Amazonia, Vol. 11(1), 55–71.  
DOI: <https://doi.org/10.47847/faccea.v11n1a4>



Este artículo puede compartirse bajo la Licencia Creative Commons  
Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0).

*Revista FACCEA*, Vol. 11 No.(1) Enero-Junio, 2021

efficiency, productivity and competitiveness, where one of the predictions was that of the 100% of users that will exist by the year 2021, only 50% will be paying sewerage.

**Key words:** Mathematical model, business intelligence, multiple linear regression, *Rstudio*, time series.

## Introducción

Zbigniew Michalewicz define el propósito de la inteligencia de negocios adaptativos como una acción con la siguiente tendencia: “resolver problemas de negocios del mundo real que tienen restricciones complejas, (y que) se establecen en entornos que cambian en el tiempo, y donde el número de soluciones posibles es demasiado grande” (Zbigniew, 2007).

El concepto *Business Intelligence* hace referencia al conjunto de estrategias, aplicaciones, datos, tecnologías y arquitecturas que permiten obtener conocimiento útil para la gestión del negocio, a partir de datos procesados (Oliva, 2020). Además, esta es una herramienta de uso libre para cualquier tipo de negocio, la cual se puede extrapolar a cualquier aspecto de la organización con altas posibilidades de obtener buenos resultados y la que a su vez puede permitir la toma de decisiones acertadas.

En tal sentido, por ejemplo, la cadena de restaurantes *Hardee's* y *Friday's* utiliza la inteligencia de negocios para obtener información acerca de los nuevos productos que se pueden añadir a la carta; de los platos que no se deberían ofrecer; de los locales que se deben cerrar y de los lugares donde se debería abrir otros restaurantes. Es decir, a partir de la herramienta mencionada, estas compañías buscan de manera constante “optimizar” sus procesos.

Del mismo modo, una compañía como Coca-Cola utiliza plataformas con características propias de BI para reducir las tareas que solían realizarse durante todo un día, con el fin de que estas sean ejecutadas en tan solo unos pocos minutos.

A su vez, considerando tanto los datos recopilados en su plataforma como las fuentes externas, una compañía como *Netflix* determina qué contenido puede ser del interés de sus usuarios y, en tal sentido, genera recomendaciones para su público. Además, dicha empresa también usa ese conocimiento para comprender a profundidad los intereses de sus usuarios; esto con el fin de determinar los programas de TV y las películas que deben ser adquiridas, así como las que se deben conservar y las que deben ser eliminadas de su lista.

No obstante, las empresas mencionadas anteriormente son solo algunas de las muchas que utilizan los citados procesos, con el objetivo de optimizar, predecir y adaptar los mismos, para alcanzar mayores ingresos.

## Marco Teórico

### **Complejidad**

La palabra complejidad se deriva del latín *complexus*, que refiere a lo que está tejido en conjunto. Por un parte, esta palabra se relaciona comúnmente con confusión e incertidumbre; de hecho, la expresión “es complejo”, implica la dificultad de definir o de explicar algo.

Por otra parte, “dado que el criterio de verdad de la ciencia clásica se expresa con leyes y conceptos simples, la complejidad no tiene que ver más que con apariencias superficiales o ilusorias. Al parecer, los fenómenos se presentan de forma confusa e incierta, pero la misión de la ciencia es descubrir, detrás de esas apariencias, el orden oculto que es la realidad auténtica del Universo, es decir, la complejidad es invisible en la separación disciplinar de lo real” (Morín, 1999).

### **Sistemas complejos**

Un sistema complejo es aquel que está compuesto por muchos elementos y muchas relaciones. Las situaciones complejas, “se caracterizan por la confluencia de múltiples procesos cuyas interrelaciones constituyen la estructura de un sistema que funciona como una totalidad organizada, a la cual hemos denominado sistema complejo” (García, 1986).

“La complejidad de un sistema no está solamente determinada por la heterogeneidad de los elementos (o subsistemas) que lo componen y cuya naturaleza los sitúa normalmente dentro del dominio de diversas ramas de la ciencia y la tecnología. Además de la heterogeneidad la característica determinante de un sistema complejo es la inter-definibilidad y mutua dependencia de las funciones que cumplen dichos elementos dentro del sistema total” (García, 2011-1).

### **Características del estudio de un sistema complejo**

La metodología de trabajo interdisciplinario que se expone en este artículo responde a la necesidad de lograr una síntesis integradora de los elementos de análisis provenientes de las tres fuentes referidas a continuación.

Rolando García afirma que “El objeto de estudio, es decir, el sistema complejo como fuente de una problemática no reducible a la simple yuxtaposición de situaciones o fenómenos que pertenezcan al dominio exclusivo de una disciplina. Además, el marco conceptual desde el cual se aborda el objeto de estudio; es decir, el bagaje teórico desde cuya perspectiva los investigadores identifican, seleccionan y organizan los datos de la realidad que se proponen estudiar. Así mismo, los estudios disciplinarios que corresponden a aquellos aspectos o “recortes” de esa realidad compleja, visualizados desde una disciplina específica. Así el objetivo es llegar a una interpretación sistémica de la problemática original que presenta el objeto de estudio. A partir de allí, será posible lograr un

diagnóstico integrado, que provea las bases para proponer acciones concretas y políticas generales alternativas que permitan influir sobre la evolución del sistema”.

Ilya Prigogine, señala que “en cualquier sistema complejo las partes del sistema están siempre experimentando cambios en pequeña escala, están en constante flujo. El interior de cualquier sistema se halla estremecido de fluctuaciones”

### **Sistemas complejos adaptativos**

Los sistemas complejos adaptativos “adquieren información acerca de su entorno como de la interacción entre el propio sistema y dicho entorno, identificando regularidades, condensándolas en una especie de “esquema” modelo y actuando en el mundo real sobre la base de dicho esquema” (Soler, 2017).

Para Murray Gell-Mann (1995), la adaptación tiene lugar en tres niveles diferentes:

“En primer lugar, tiene lugar cierta adaptación directa (como en un termostato o dispositivo cibernético) como resultado de la operación de un esquema dominante en una época particular. El segundo nivel incluye cambios de esquema, competencia entre esquemas diversos y su promoción o degradación en respuesta a las presiones selectivas en el mundo real. El tercer nivel de adaptación es la supervivencia darwiniana del más apto”

### **Inteligencia de negocios adaptativa**

La inteligencia adaptativa es la capacidad de asumir el cambio con rapidez y de adaptarse a las circunstancias del medio (Berenstein, 2018). En pocas palabras, la Inteligencia de Negocios Adaptativa es la disciplina de combinar predicción, optimización y adaptabilidad de un sistema capaz de responder a las necesidades de su ambiente. Es también, la forma de retroalimentarse un sistema, subsanando sus propias complicaciones, siendo así un sistema inteligente con la capacidad de evolucionar y adaptarse al sistema, generando resultados convincentes.

“En el mundo empresarial y de los negocios comienzan a incorporar distintos análisis y elementos que permitan la optimización en el uso de los recursos disponibles. Pero, lo fundamental en este contexto de la globalización es la relación que comienza a tener lugar entre el desarrollo de la tecnología de la información y la comunicación, junto al procesamiento y clasificación de la información, que inicia su consolidación a partir de la segunda guerra mundial” (Junco & Castellanos, 2013).

“Por lo que, el reto para las empresas es evolucionar, es crecer y esto significa “cambio” . Que tan ágiles son mis procesos para enfrentar los cambios y las necesidades puntuales de la empresa” (Oracle Corporation, pág. 6).

Finalmente, en múltiples entornos empresariales sus representantes se encuentran a diario con diversas problemáticas, las cuales pueden afectar el futuro de la compañía. Como solución a dichas problemáticas emergentes se considera necesario construir modelos matemáticos, que permitan tener una la capacidad de predicción, optimización y adaptación de los altos riesgos de los negocio en el mundo real, de tal manera que el emprendedor pueda conocer lo que podría suceder a futuro con la empresa y, en tal sentido, pueda tomar las decisiones que sean más pertinentes para su beneficio. Lo anterior en tanto: “parece que, independientemente de la decisión que se tome o de su complejidad, primero debemos hacer una predicción de lo que es probable que suceda en el futuro, y luego tomar la mejor decisión en base a esa predicción” (Zbigniew Michalewicz, 2006).

### **Regresión Lineal Múltiple**

“La regresión lineal múltiple permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta (Y) se determina a partir de un conjunto de variables independientes llamadas predictores ( $X_1, X_2, X_3, \dots$ ). Es una extensión de la regresión lineal simple, por lo que es fundamental comprender esta última. Los modelos de regresión múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre ella (esto último se debe que analizar con cautela para no malinterpretar causa-efecto).

Los modelos lineales múltiples siguen la siguiente ecuación:

$$(Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}) + e_i(*))$$

$\beta_0$ : Es la ordenada de origen, el valor de la variable dependiente

$\beta_i$ : Es el efecto promedio que tiene el incremento en una unidad de la variable predictora.

$X_i$ : sobre la variable dependiente

$e_i$ : Es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Es importante tener en cuenta que la magnitud de cada coeficiente parcial de regresión depende de las unidades en las que se mida la variable predictora a la que corresponde, por lo que su magnitud no está asociada con la importancia de cada predictor. Para poder determinar qué impacto tienen en el modelo cada una de las variables, se emplean los coeficientes parciales estandarizados, que se obtienen al estandarizar (sustraer la media y dividir entre la desviación estándar) las variables predictora previo ajuste del modelo” (Rodrigo, 2016).

### **Modelo autorregresivo integrado de media móvil - ARIMA**

“En estadística y econometría, en particular en series temporales, un modelo autorregresivo integrado de promedio móvil o ARIMA (acrónimo del inglés autoregressive integrated moving average) es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no

por variables independientes. Fue desarrollado a finales de los sesenta del siglo XX. Box y Jenkins (1976) lo sistematizaron. El modelo ARIMA necesita identificar los coeficientes y número de regresiones que se utilizarán. Este modelo es muy sensible a la precisión con que se determinen sus coeficientes” (Modelo autorregresivo integrado de media móvil, 2020).

## **Método**

### **Localización y caracterización de la empresa objeto de estudio**

La empresa E.S.P EMSERPUCAR presta servicios domiciliarios a la comunidad de Cartagena del Chairá-Caquetá. Sin embargo, el crecimiento poblacional en la zona urbana ha afectado la calidad en la prestación de dichos servicios. En tal sentido, el objetivo de este estudio es presentar una propuesta que permita fortalecer la prestación de servicios domiciliarios por parte de la empresa E.S.P EMSERPUCAR, mediante la aplicación de análisis de datos, conceptos estadísticos, técnicas de predicción y optimización para desarrollar el autoaprendizaje.

### **Alcance y enfoque**

El tipo de investigación desarrollada con respecto al objeto de estudio es aplicada, dado que se busca la generación de nuevo conocimiento y de nuevas ideas, con el fin de lograr el mejoramiento de los procesos.

De igual manera, la investigación planteada tendría un carácter descriptivo, dado que a partir de esta es posible medir variables o conceptos, con el fin de integrarlos para determinar cómo se manifiesta un proceso y para generar algunos pronósticos a partir de este.

Así mismo, el alcance de la investigación es descriptivo, en tanto con el estudio se busca analizar y describir datos con información relacionada con la prestación de servicios domiciliarios de la empresa objeto de estudio, a partir de un enfoque cuantitativo. Ello en la medida en la que se examinan datos de manera numérica en el campo estadístico, mediante la utilización del software Rstudio, a partir de la aplicación de la técnica de predicción, para lograr un análisis no convencional de los datos de los usuarios que hacen uso del servicio de acueducto, específicamente entre los años 2019 y 2020. Estos datos fueron organizados a partir de un criterio de temporalidad mensual. Lo anterior se realizó a través de series de tiempo, en las cuales se debía indicar el método de predicción a utilizar el que, para este caso, fue el de la regresión lineal múltiple (RLM) y Arima.

Ahora bien, como se ha indicado antes, el objetivo del presente estudio reside en encontrar acciones que incidan en mejorar la capacidad para satisfacer las necesidades en cuanto a los servicios domiciliarios, entre la población del municipio de Cartagena del Chairá: en tal sentido, a partir de este proyecto de investigación se busca dar respuesta a la pregunta sobre ¿Cómo la empresa EMSERPUCAR E.S.P se debe preparar para garantizar la oferta y la demanda en la prestación de servicios domiciliarios, teniendo en cuenta el crecimiento de usuarios para el periodo 2021?

La respuesta a la pregunta planteada se considera como un insumo que puede ayudar al gerente de la mencionada empresa a tomar acciones preventivas y decisiones dentro de un tiempo prudencial, para que en situaciones futuras pueda enfrentarse a fenómenos como el aumento de usuarios, la falta de cobertura del servicio a prestar, y el consumo excesivo del servicio del acueducto, sin que se vea afectado el rendimiento de la empresa.

Como parte del aporte de la investigación, finalmente se presenta un informe detallado al gerente de la mencionada organización, respecto de los resultados obtenidos, en el cual estos son interpretados a partir de la simulación, para así proporcionar sugerencias adecuadas que le sirvan a aquél en la toma de decisiones. Es importante agregar que el uso de herramientas tecnológicas, en este caso del software Rstudio, resulta fundamental para lograr una mejor perspectiva a la hora de analizar y de tomar decisiones a nivel gerencial.

### **Series de tiempo**

Para estudiar el comportamiento a futuro de la variable “Usuarios” se realizó un análisis sustentado en una serie de tiempo.

Las series de tiempo garantizan el conocimiento acerca de si es adecuado o no utilizar el método de regresión lineal múltiple.

Al realizar la serie de tiempo de la variable “Usuario” se encuentra que esta es no estacionaria. La serie de tiempo muestra cambios de varianza, lo que indica la variabilidad del conjunto de datos respecto de la media aritmética.

- Tiene tendencia creciente a través del tiempo.  
La serie de tiempo permite ver que los datos proporcionados entre las variables  $T_t = usuarios$  y  $t = tiempo$  Tienen una relación determinista lineal.

$$T_t = a + bt \quad (1)$$

Por lo anterior, se hace uso del método de la regresión lineal múltiple, para construir un modelo matemático que permita predecir el número de Usuarios que habrá para el año 2021, en el municipio de Cartagena de Chairá.

## **Resultados**

### **Información del modelo a partir de la regresión lineal múltiple (RLM)**

Al realizar la regresión lineal múltiple (RLM) se tomó como variable dependiente “Usuarios” y como variables independientes “Agua”, “Alcantarillado”, “Aseo”, “Si Medidor” y “No Medidor”.

### Análisis estadístico

Se analiza el valor p-value:  $0,0005109 < 0,05$ , con el cual se rechaza la hipótesis nula, por lo que se considera que al menos una de las variables independientes del modelo está significativamente relacionada con la variable "Usuarios".

El Residual Estándar Error (RSE): significa que, en promedio, cualquier predicción del modelo se aleja 66,82 unidades del verdadero valor.

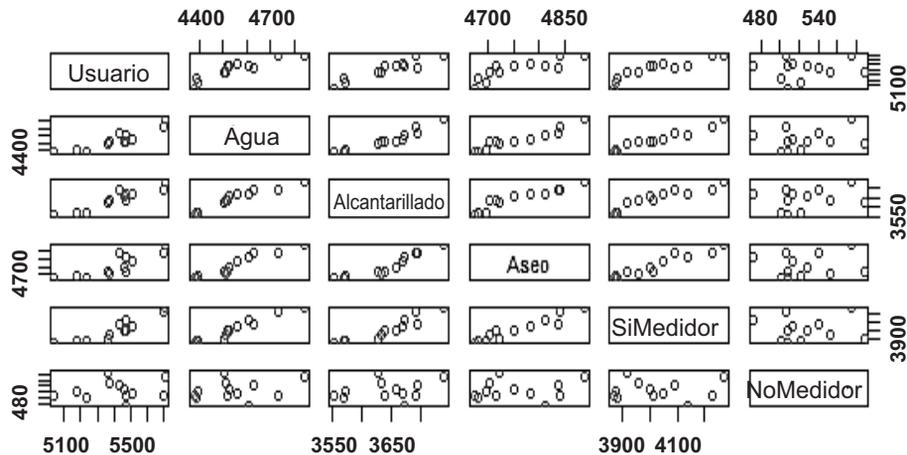
El  $R_2$ : el predictor "Usuarios" empleado en el modelo es capaz de explicar el 92,41% de la variabilidad de las demás variables.

### Gráfica generalizada del Plot

Para estudiar más a fondo el modelo se analiza cada variable de manera independiente, contra la variable "Usuarios".

Figura 1.

Plot generalizado.



Fuente: elaboración propia. (2020).

En la Figura 1 se observa la distribución, la dispersión y la correlación de cada variable, del mismo modo que se evidencia que las variables "Agua" y "Si Medidor", presentan potencialmente un problema para la regresión. Además, la variable "No Medidor" no es considerada para nuestra primera regresión, debido a que no existe una correlación entre la variable predictora "Usuarios" Vs "No Medidor".

### Gráficas de las variables que no tienen una distribución normal

Al depurar la información se encuentra que las variables "Agua" y "Si Medidor" no tienen una

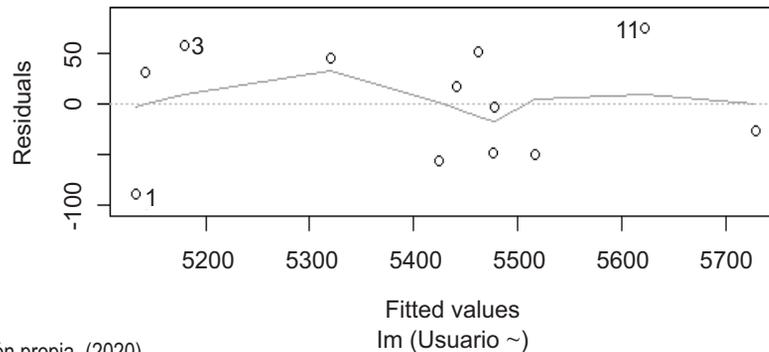
distribución normal, por lo que no son consideradas por el modelo, en tanto no explican la variable del "Usuario".

## Residuos

En la figura 2, en la cual se garantiza si los datos del modelo tienen o no un patrón lineal, se observa que la línea continua crece y decrece para algunos intervalos; sin embargo, sigue una secuencia lineal.

**Figura 2.**

Residuos Vs Ajustados.



Fuente: elaboración propia. (2020).

Los residuos se distribuyen normalmente, puesto que la gran mayoría de los residuos están alineados con la recta discontinua. Sin embargo, el dato 1, 5 y 11 se alejan de la línea discontinua, por lo que podría presentar problemas potenciales para el modelo.

## Ecuación del modelo

A partir de los coeficientes que se encuentran en la información del modelo (Figura 5), se construye la ecuación para la regresión 1 (RLM).

$$Tt = -536,37 + 0,153\text{Agua} + 3,07\text{Alcantarillado} - 2,11\text{Aseo} + 1,02\text{SiMedidor} \quad (2)$$

## Predicción

Se hace la predicción a partir de la ecuación (2), para algunos datos de la tabla.

Al comparar los valores reales con algunos obtenidos se obtiene lo siguiente:

## Error por dato

El %error para el primer dato es de 1,8%; como se muestra en la ecuación 3.1 significa que se tiene un error muy grande por dato, lo cual hace que el modelo no sea confiable.

$$\frac{|v.Aproximado - v.Real|}{v.Real} * 100 \quad (3) \quad \frac{|5135,246 - 5043|}{5043} * 100 = 1,8\% \quad (3.1)$$

### **Ajustes del modelo (adaptabilidad)**

Como las variables "Si Medidor", "Agua" y "No Medidor" no aportan información al modelo, se procede entonces a eliminar dichas variables; esto significa que se adapta el modelo para que sea más efectivo. Se observa que el valor de significancia es:  $p\text{-value}: 9,774e > 0,05$ ; esto indica que la hipótesis nula es verdadera, así, el modelo corre el riesgo de que no funcione con la hipótesis planteada.

### **Ecuación del modelo para la regresión 2**

La siguiente ecuación representa las dos variables que aportan información significativa para nuestro análisis.

$$Usuarios = -6042,51 + 4,03alcantarillado - 0,68aseo \quad (4)$$

### **Predicción**

Se hará la predicción de la variable "Usuario", cuando la variable "alcantarillado" alcance el valor de 3890 y la de "Aseo" el valor de 4992.

Comparando los valores reales con los obtenidos se espera que cuando el número de usuarios que tengan alcantarillado = 3890 y Aseo = 4992, haya aproximadamente 6246 usuarios.

### **Error por dato**

El %error para el primer dato es de 0,16% lo que significa que se tiene un error pequeño por dato; esto hace que el modelo sea más confiable que el anterior.

$$\frac{|5164,637 - 5173|}{5173} * 100 = 0,16\% \quad (5)$$

### **Nuevo ajuste para el modelo**

En este caso, al adaptar nuevamente el modelo se observa que la única variable que aporta información al mismo es la variable "Alcantarillado". Dado que el  $p\text{-value}: 1,234e < 0,05$ ; entonces la hipótesis nula es falsa; esto significa que la variable "Alcantarillado" está significativamente relacionada con la variable "Usuarios".

- Residual Estándar Error (RSE): significa que, en promedio, cualquier predicción del modelo se aleja 74,92 con 10 grados de libertad a las unidades del verdadero valor, esto es, 7.492 aproximadamente.
- El  $R^2$ , el predictor "Usuarios" empleado en el modelo es capaz de explicar el 86,36% de la variabilidad de las demás variables.

### Ecuación del modelo para el nuevo ajuste

$$T_i = - 6305,7037 + 3,21Alcantarillado \quad (6)$$

#### Predicción

A partir de la ecuación 6 se hallarán los valores para la variable “Usuarios”, con base en la variable “Alcantarillado”. Se espera que cuando el número de alcantarillado sea 3.890, entonces haya aproximadamente 6.204 usuarios; para el segundo valor, cuando el número de alcantarillado sea 4.500, se espera que haya aproximadamente 8.166 usuarios.

#### Error por dato

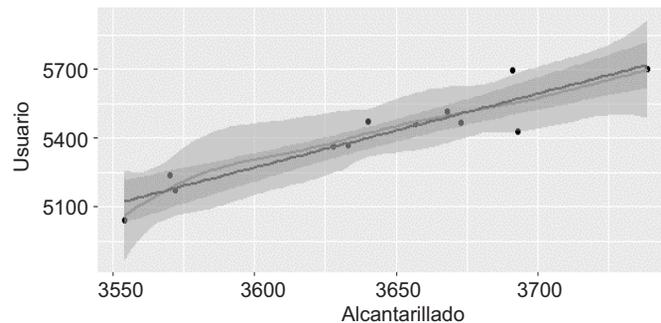
El %error para el segundo dato es del 0,064%, lo que significa que se tiene un error pequeño por dato, lo cual hace que el modelo sea más confiable que las regresiones anteriores.

$$\frac{|5361,559 - 5365|}{5365} * 100 = 0,064\% \quad (7)$$

### Gráfica de regresión lineal simple

#### Figura 3.

Correlación: usuarios-alcantarillado.

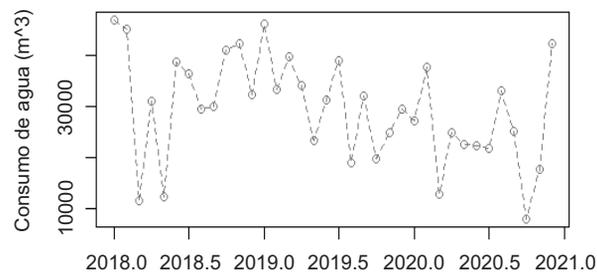


Fuente: elaboración propia. (2020).- Willian Araque

### Predicción para el consumo de acueducto.

#### Figura 4.

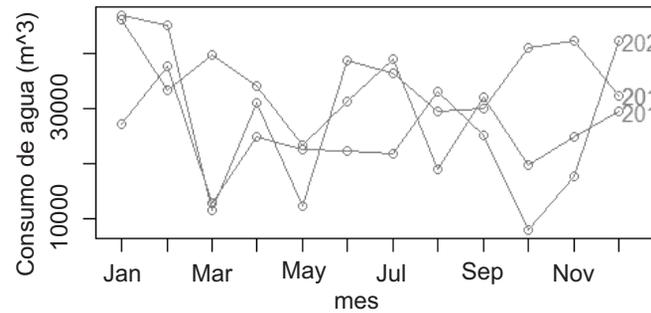
Serie de tiempo.



Fuente: elaboración propia. (2020).

En la serie de tiempo se visualiza el comportamiento de la variable “Consumo”, a través del tiempo que ha tenido la empresa E.S.P EMSERPURCAR desde el 2018 a 2020. Esta serie de tiempo es oscilante (Figura 5).

**Figura 5.**  
Consumo anual.



Fuente: elaboración propia. (2020).

También se observa el comportamiento del consumo durante cada uno de los meses del año, en relación con el servicio de acueducto.

### Prueba de Dickey-Fuller

Se determina que la serie de tiempo del consumo de acueducto es una serie no estacionaria (ver figura 20). Se plantea transformar esta serie no estacionaria en una serie estacionaria, a través de la prueba de Dickey-Fuller; esta prueba además indica si la serie de tiempo es o no es estacionaria.

Aplicando la prueba de Dickey-Fuller (ver figura 21) con una diferenciación, se convierte la serie de tiempo de consumo en una serie estacionaria. Por otra parte, analíticamente se observa que el  $p\text{-value} = 0,01 < 0,05$ , por lo que nuestra serie es estacionaria.

### Prueba de ruido blanco

Indica que, si el modelo satisface las condiciones, entonces el modelo se ajusta bien para predecir.

**Figura 6.**  
Prueba de ruido blanco.

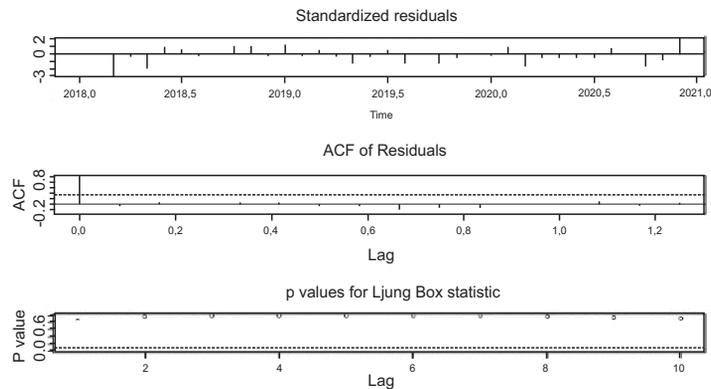


Fuente: elaboración propia. (2020).

Para garantizar las hipótesis de la Figura 7 se puede encontrar respuesta de dos maneras: cualitativa y analíticamente. En tal caso, se llevarán a cabo ambos procedimientos.

**Figura 7.**

Errores estándar, función autocorrelación y la prueba de Ljung Box.



Fuente: elaboración propia. (2020).

En la primera gráfica la media es igual a cero; en la segunda la varianza es constante y en la última gráfica se indica que los datos se encuentran por encima de 0,05, por lo que se concluye que sí hay ruido blanco.

Analíticamente, se observa que el  $p - value = 0,8451 > 0,05$ , por lo que se concluye que la hipótesis del ruido blanco es verdadera. En efecto, el modelo se ajusta bien (ver figura 25).

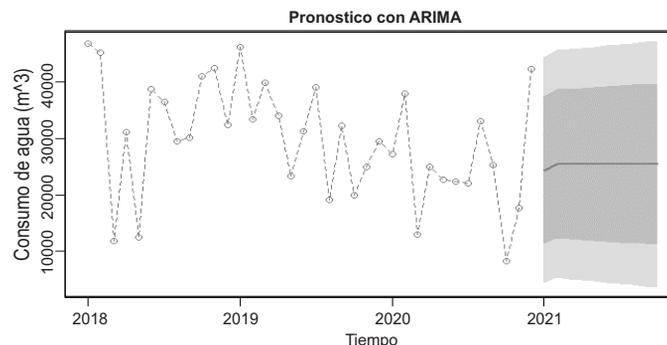
**Pronostico con ARIMA**

A través del método de predicción ARIMA de Rstudio se logró predecir, para los próximos 10 meses del año 2021, el consumo del agua mensual que tendrán los usuarios del municipio de Cartagena del Chairá.

**Gráfico de ARIMA**

**Figura 8.**

Gráfico de predicción con ARIMA.



Fuente: elaboración propia. (2020).

## Discusión

Por una parte, se espera que para los 10 primeros meses del año 2021, el consumo de agua por metro cúbico en el municipio de Cartagena del Chairá sea estable, con un valor aproximado de 25.438. Además, para el mes de octubre el mínimo consumo podría ser de 3.532 y el máximo consumo podría ser de 47.344.

Por otra parte, se encontró correlación entre la variable “Usuarios” y la variable “Alcantarillado”, a partir de las múltiples regresiones lineales aplicadas, con el fin de encontrar el mejor modelo que prediga con gran certeza, con un porcentaje de error por dato del 0,064%.

El análisis del  $R^2$  y la correlación entre las variables permite avanzar en el proceso de la construcción del modelo, como se plasmó en el Flujograma. El modelo de variable “Usuarios” se logró adaptar, las veces que fue necesario, hasta tener un mínimo de error por dato.

A su vez, de acuerdo con lo recabado en la información estadística realizada con anterioridad, se recomienda mejorar la instalación de medidores en el municipio de Cartagena del Chairá; esto, con el objetivo de llevar un control adecuado del consumo de agua por vivienda, e incluso, de los sectores aledaños o de los asentamientos, para así evitar la pérdida de agua y el consumo ilegal del servicio de acueducto. Igualmente, con el fin de que no se sature el servicio de acueducto del municipio de Cartagena del Chairá, por parte de la empresa E.S.P EMSERPUCAR, se debe reducir en un 50% la problemática relacionada con los usuarios que no pagan por dicho servicio.

Finalmente, de modo general se considera que las repercusiones que tendrá la empresa a futuro serán las siguientes:

- Del 100% de los usuarios que habrá para el año 2021, solo el 50% estará pagando alcantarillado.
- Disminución de la presión del servicio del acueducto, lo cual se deberá al incremento de usuarios en asentamientos ubicados en una altura mayor a la ubicación de la planta de tratamiento.
- Incremento de usuarios que harán uso del servicio de acueducto de manera ilegal.
- Crecimiento poblacional en el municipio de Cartagena del Chairá, esto es, por asentamientos nuevos, en los sectores aledaños al municipio.

## Conclusiones

Para la presente investigación se construyó una base de datos a partir de la información mensual proveniente de los usuarios, la cual fue suministrada por la empresa E.S.P EMSERPUCAR del municipio de Cartagena del Chairá, ubicado en el departamento del Caquetá. A su vez, con el uso de las series de tiempo, se construyó un modelo matemático de regresión lineal múltiple (RLM); después del proceso de predecir, optimizar y adaptar, se obtuvo un modelo de regresión lineal

simple (RLS); lo anterior, debido a que las variables de la base de datos no suministraban información suficiente para que el modelo fuese más efectivo. Por su parte, con los datos del consumo mensual del servicio de acueducto del municipio de Cartagena del Chairá, se construyó un modelo artificial utilizando el método de predicción ARIMA, basado en las series de tiempo.

Por otra parte, se considera que el modelo es capaz de predecir el número de usuarios que estarán haciendo uso del servicio de acueducto en el municipio de Cartagena del Chairá para el año 2021, con una confiabilidad del 86,36%. Igualmente, de acuerdo con el segundo modelo se espera que para los 10 primeros meses del año 2021, en el municipio de Cartagena del Chairá el consumo sea estable, lo cual quiere decir que el mismo sería de aproximadamente 25.348 metros cúbicos de agua por mes.

Finalmente, se considera que el uso de herramientas tecnológicas, en este caso del software *Rstudio*, es importante para el ejercicio de toma de decisión en las empresas. En tal sentido, el informe que se elaboró y en el cual se hizo uso del software mencionado brindó la información necesaria para realizar el respectivo análisis estadístico; lo anterior, apoyado además en el uso de conocimientos teóricos de estadística. De esta manera, el informe antes citado se elaboró de manera detallada y sustentada, a partir de información relevante, con el fin de que el gerente de la mencionada empresa tome decisiones en el tiempo adecuado, respecto de los temas relacionados con el crecimiento de usuarios del servicio de acueducto y también respecto del consumo en el municipio de Cartagena del Chairá.

## Referencias bibliográficas

- Morín, E. & Le Moigne, J. (2006). *Inteligencia de la Complejidad Epistemología y Pragmática*. Cerisy: l'aube. <https://pdfcookie.com/documents/inteligencia-de-la-complejidad-x20gg9e0z3l3>
- Alonso, A. M., & Daniel Peña, J. R. (2004). *Introduciendo la incertidumbre del modelo en la serie de tiempo bootstrap*. Instituto de Ciencia Estadística, Academia Sinica, 155-174.
- Andrés M. Alonso. (2004). *Introducción al Análisis de Series Temporales. Cálculo de Tendencia y Estacionalidad*. <http://halweb.uc3m.es/esp/Personal/personas/amalonso/esp/seriestemporales.pdf>
- García, R. (2011). *Interdisciplinariedad y sistemas complejos*. ReLMeCS, 66-101. <file:///C:/Users/Lenovo/Downloads/admin,+Gestor+a+de+la+revista,+v1n1a04.pdf>
- Arias, M. M. (2015, 14 de enero). *No todo es normal. Manejo de datos no normales*. ANESTESIAR. <https://anestesiari.org/2015/no-todo-es-normal-manejo-de-datos-no-normales/>
- Arias, M. M. (2017). *¿Qué significa realmente el valor de p?* Revista Pediatría Atención Primaria, 9(76). [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1139-76322017000500014](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322017000500014)

- Back, T. (2002). Adaptive business intelligence based on evolution strategies: some application examples of self-adaptive software. Elsevier.
- García, R. (2007). Interdisciplinariedad y Sistemas Complejos\*. proglocode.unam. <http://www.proglocode.unam.mx/system/files/interdisciplinariedad%20y%20sistemas%20complejos%20-%20Rolando%20Grac%C3%ADa.pdf>
- b.se-todo.com. (2015). La Ciencia de la Complejidad. En A. & Toffler, La Tercera Ola. Tomado de: [La Ciencia de la Complejidad \(se-todo.com\)](http://www.se-todo.com/la-ciencia-de-la-complejidad)
- Soler B, Y. (2017). Teorías sobre los sistemas complejos. A&D, 52-69. <file:///C:/Users/Lenovo/Downloads/Dialnet-TeoriasSobreLosSistemasComplejos-6403420.pdf>
- Barbatus, C. B. (2008). El Quark y el Jaguar. En M. Gell-Mann, El Quark y el Jaguar. <http://elcajondewatson.blogspot.com/2008/05/el-quark-y-el-jaguar-hacia-un.html>
- Castillo, A. J. (2010). Métodos Estadísticos con R y R Commander. Universidad de Jaén. <https://cran.r-project.org/doc/contrib/Saez-Castillo-RRCmdrv21.pdf>
- Castro Alfaro, A. (2016). Estrategias para la disminución de la carga impositiva en las organizaciones empresariales. Enfoque Disciplinario, 1(1), 21-34. <http://enfoquedisciplinario.org/revista/index.php/enfoque/article/view/6>
- Cervan, D. (2020 29 de mayo). Regresión lineal múltiple con R. (Video). Tomado de [https://www.youtube.com/watch?v=dFiOett-Bgl&ab\\_channel=dheybicervan](https://www.youtube.com/watch?v=dFiOett-Bgl&ab_channel=dheybicervan)
- Corporation, O. (s.f.). ¿Qué es la inteligencia de negocios?. Tomado de: [https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529\\_esa.pdf](https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf)
- Modelo autorregresivo integrado de media móvil. (18 de marzo del 2020). En Wikipedia. [https://es.wikipedia.org/wiki/Modelo autorregresivo integrado de media m%C3%B3vil](https://es.wikipedia.org/wiki/Modelo_autorregresivo_integrado_de_media_m%C3%B3vil)
- Joaquín, A. R. (2016). Introducción a la Regresión Lineal Múltiple: Cienciadedatos.net. [https://www.cienciadedatos.net/documentos/25\\_regresion\\_lineal\\_multiple.html](https://www.cienciadedatos.net/documentos/25_regresion_lineal_multiple.html)
- <https://revistalogos.policia.edu.co:8443/index.php/rict/article/view/11/406>
- D h y b i C e r v a n . ( 1 9 d e M a y o d e 2 0 2 0 ) . D h e y b i C e r v a n . <https://drive.google.com/drive/folders/1hftPFSswzVrJQz5MFifeX4QTtgZeCAAP>
- G., C. S. (2017). Fundamentos Básicos de Estadística. (Sin editorial).
- G., M. Á. (2016). Análisis de integración y modelos de cointegración: Aplicación en software R. En L. Q.

- Romero, & M. Á. Mendoza (Ed.), *ECONOMETRÍA APLICADA UTILIZANDO R* (. 174 - 207 ). Dgapa.
- Grolemund, H. W. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media .
- Berenstein, M. (2018). Inteligencia Adaptativa. *Emprendedores News*. <https://emprendedoresnews.com/tips/inteligencia-aplicada.html#:~:text=La%20inteligencia%20adaptativa%20es%20la,de%20una%20neurona%20a%20otra>.
- Hsinchun Chen, R. H. (2012). *Business intelligence and analytics: Mis Quarterly*.
- Junco, M. J., & Castellanos, G. C. (2013). Business intelligence y la toma de decisiones financieras: una aproximación teórica. *LOGOS CIENCIA & TECNOLOGÍA*, 5(1), (119-138).
- Meyer, P. L. (s.f.). *Probabilidad y aplicaciones estadísticas*. Washington State University, Pearson.
- Rodrigo, J. A. (Julio de 2016). *Introducción a la Regresión Lineal Múltiple*. [https://www.cienciadedatos.net/documentos/25\\_regresion\\_lineal\\_multiple](https://www.cienciadedatos.net/documentos/25_regresion_lineal_multiple)
- Rodrigo, J. A. (Agosto de 2016). *Rpubs by Rstudio*. Obtenido de [https://rpubs.com/Joaquin\\_AR/254575](https://rpubs.com/Joaquin_AR/254575)
- Moro, S., Cortés, P., Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Elsevier. Expert Systems with Applications*, 42(3), (1314-1324). <https://data.library.virginia.edu/diagnostic-plots/>
- Wickham, G. G. (2014). *Hands-On Programming with R write your own functions and simulations*. Matthew Hacker.
- Wickham, H. (2009). *ggplot2 Elegant Graphics for data Analysis*. Springer.
- Zbigniew Michalewicz, M. S. (2006). *Adaptive Business Intelligence*. Springer.
- Zylberberg, A. D. (2006). *Probabilidad y Estadística*. Nueva Librería.