

# MODELO DE REGRESIÓN LINEAL APLICADO A LA CANTIDAD DE CONTENEDORES A MOVILIZAR EN PUERTO SANTO TOMÁS DE CASTILLA, GUATEMALA

*Linear regression model applied to the amount of containers to mobilize in puerto Santo Tomás de Castilla, Guatemala*

**Antonio Daniel Asencio Marroquín**

Mtro. en Estadística Aplicada

Correspondencia al autor: antonio.asencio31@gmail.com

Recibido: 01 de junio 2019 | Revisado: 24 de junio 2019 | Aprobado: 20 de septiembre 2019

Asesorado por: Mtro. en Hidrología **Juan Carlos Fuentes Montepeque** jcfuentes24@gmail.com

## Resumen

Los modelos de regresión lineal se utilizan para pronosticar valores, en este caso aplicados a la cantidad de contenedores a movilizar en Puerto Santo Tomás de Castilla. Se ha identificado que en la mayoría de casos estos modelos son elaborados sin realizar un análisis previo de los datos, razón por la cual, en el presente artículo se describe la metodología utilizada para su elaboración, las variables a incluir en los mismos; así como la verificación de los supuestos que deben cumplir los residuos, para su validez. De igual forma se realiza el análisis por medio de series de tiempo, en donde se comprueban las propiedades de tendencia, estacionalidad y variabilidad de los datos, finalizando con la comprobación de la significancia y ajuste de los modelos generados por regresión lineal y series de tiempo. Como resultado de lo descrito se obtienen dos modelos que explican el comportamiento de los datos en el tiempo por medio de análisis regresión lineal; así como dos modelos que proyectan hasta 2030 mediante la simulación con ARIMA.

## Palabras clave

ARIMA, series de tiempo. tendencia, supuestos.

## Abstract

*Linear regression models are used to forecast future values, in this case applied to the number of containers to be mobilized in Port Santo Tomás de Castilla. It has been identified that in the majority of cases these models are elaborated without carrying out a previous analysis of the data, reason for which, in this document the methodology used for its elaboration is described, the variables to be included in them; as well as the verification of the assumption that the residuals must fulfill, for its validity. In the same way, the analysis is carried out by means of time series, where the trend, seasonality and variability properties of the data are checked, ending with the verification of the significance and the adjustment of the models generated by linear regression and time series. As a result, two models have been obtained that explain the behavior of the data over time by linear regression analysis; as well as two models that project data until 2030 through ARIMA modeling.*

## Keywords

ARIMA, Time series, tendency, assumptions.

## Introducción

La estadística aplicada brinda distintas herramientas para la resolución de problemas que ocurren en distintas situaciones. En puerto Santo Tomás de Castilla, se hacen proyecciones del movimiento de carga, las cuales se utilizan para planificar obras de infraestructura necesarias para atender dicho volumen. Es por esto que el análisis de los datos, previo a la elaboración de un modelo de regresión lineal es importante; ya que con ello se determina la validez y ajuste de los datos, siendo de beneficio para dimensionar de mejor manera los proyectos a corto, mediano y largo plazo. Como solución a la problemática encontrada se elaboran modelos para la cantidad de contenedores a movilizar realizando investigación con base en distintos estudios que aplican metodologías similares, la cual también puede ser utilizada para la elaboración de presupuestos y planes operativos basados en proyecciones de carga.

## Desarrollo del estudio

El enfoque del estudio es cuantitativo, ya que los análisis se realizan a partir de una base de datos organizada, logrando la generación de valores futuros mediante el empleo de análisis de regresión y modelos auto regresivos, razón por la cual, el alcance del mismo es descriptivo.

El diseño es no experimental, con corte longitudinal de tendencia, ya que las unidades de análisis y variables del estudio están constituidas por la cantidad de contenedores movilizados por el puerto durante el periodo 2000-2016, en su estado original, y divididas en sub poblaciones dadas por los movilizados en desembarque y embarque.

Para el análisis de regresión lineal se realiza la comprobación de los supuestos que deben cumplir los residuos del modelo, siendo estos: normalidad, linealidad, independencia, homogeneidad y homocedasticidad. Mientras que, para el caso de análisis de series de tiempo, se comprueban las propiedades de los datos: tendencia, estacionalidad y variabili-

dad, para generar posteriormente modelos ARIMA para la proyección de los datos.

A partir de los resultados, se procede a determinar la significancia de los modelos y a su validación, mediante la interpretación y discusión de los mismos, determinando el ajuste que tienen para explicar el comportamiento de los datos.

## Resultados obtenidos

La aplicación de los modelos de regresión lineal, inicia con el análisis previo de los datos, siguiendo la recomendación de Hair, Anderson, Tatham y Black (1999), buscando seleccionar las variables a incluir en los mismos, en búsqueda de obtener el que incluya el menor número posible de variables, principio conocido como parsimonia, verificando la significancia de las mismas mediante la prueba T de Student. Mientras que la validez del modelo se realiza por medio de la prueba F. Los resultados se muestran a continuación.

Tabla 1  
*Análisis de varianza y coeficientes de regresión asociados al modelo*

FUENTE DE VARIACIÓN	F	F TEÓRICO
Modelo importación	270.74	4.60
Modelo exportación	393.64	4.60
COEFICIENTE	T	T TEÓRICO
Gastos de consumo final (variable modelo importación)	16.45	1.96
Exportaciones de bienes y servicios (variable modelo exportación)	19.84	1.96

*Fuente: elaboración propia.*

Para comprobar los supuestos que deben cumplir los residuos de los modelos obtenidos para su validez se realizan las siguientes pruebas, a un nivel de significancia del 0.05. Gómez, Aparicio, y Patiño, (2010); Lind, Marchal, y Wathen, (2012); Díaz, (2009)

1. Linealidad: Significancia r
2. Normalidad: Shapiro – Wilks
3. Independencia: Rachas
4. Homogeneidad: T de Student
5. Homocedasticidad: F de Snedecor

Tabla 2

*Comprobación de supuestos modelos de importación y exportación*

SUPUESTO	ESTADÍSTICO IMPORTACIÓN	ESTADÍSTICO EXPORTACIÓN	VALOR TEÓRICO
1	16.31	21.28	2.14
2	0.8190	0.2662	0.05
3	6	5	4-13
4	0.00014	0.00011	2.14
5	0.59	2.39	2.79

*Fuente: elaboración propia.*

Antunez (2011) recomienda utilizar la prueba de raíz unitaria de Dickey – Fuller para comprobar la estacionalidad de las series, la cual una vez aplicada presenta los siguientes resultados.

Tabla 3

*Resultados prueba de Dickey-Fuller*

MODELO	ESTADÍSTICO DICKEY-FULLER	VALOR CRÍTICO
Importación	0.1421	-2.8766
Exportación	0.4530	-2.8765

*Fuente: elaboración propia.*

El ajuste de los modelos propuestos se determina por el coeficiente de correlación de cada uno de ellos, como se muestra a continuación.

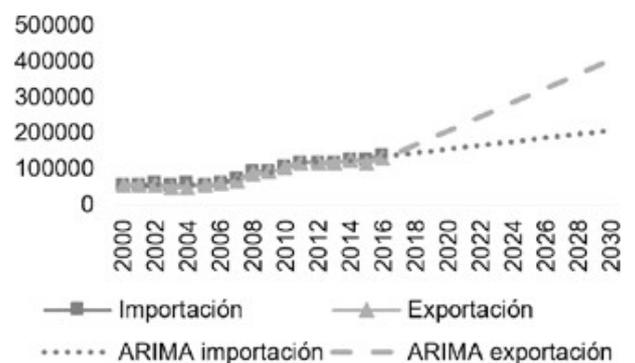
Tabla 4

*Coficiente de correlación de los modelos obtenidos*

ANÁLISIS	MODELO	COEFICIENTE DE CORRELACIÓN
Regresión	Importación	0.9508
Regresión	Exportación	0.9657
Serie de tiempo	ARIMA importación	0.9220
Serie de tiempo	ARIMA exportación	0.9450

*Fuente: elaboración propia.*

Las proyecciones obtenidas mediante la modelación ARIMA se pueden visualizar en la siguiente figura.



*Figura 1. Proyecciones importación y exportación.*

### Discusión de resultados

La colinealidad existente entre las variables elegidas para formar parte de los modelos, hace necesario realizar las pruebas con las que solamente se explica el comportamiento de los datos.

Derivado de lo anterior, se verifica la validez de los modelos parsimónicos, ya que la prueba de contraste global, según lo muestra la tabla 1, indica que se debe rechazar la hipótesis nula ya que el valor de F calculado es mayor que F teórico en ambos casos.

En cuanto al ajuste, éste se considera bueno, según los resultados expresados en la tabla 4. En cuanto a los supuestos, cuyos resultados están en la tabla 2, se verifica el cumplimiento de los mismos, ya que todas las pruebas que se aplican aportan evidencia de ello. En cuanto a las propiedades de las series de tiempo lo más importante de resaltar es que no cuentan con estacionalidad, según el resultado de la prueba unitaria de Dickey-Fuller ya que el valor crítico es menor que el estadístico de prueba, cumpliendo un precepto fundamental para aplicar modelos ARIMA a las mismas, y cuyas proyecciones se visualizan en la figura 1.

### Conclusiones

1. Las variables incluidas en los modelos derivados por el análisis de regresión se determinan mediante la prueba de contraste, buscando la no colinealidad entre las mismas. En las series temporales, se incluyen los componentes de diferenciación para hacer la serie estacional ( $d=1$ ) y el componente de media móvil ( $q=1$ ).
2. Los supuestos se cumplen para los modelos generados por medio de análisis de regresión. Las series temporales carecen de estacionalidad.
3. Se obtiene el modelo  $Y = 0.26 X + 12676.01$  para la importación y para la exportación  $Z = 1.27 W - 7809.25$ , donde X y W corresponden a gastos de consumo final y exportaciones de bienes y servicios, respectivamente. El modelo a utilizar en el caso de series de tiempo es el ARIMA.
4. Se aplica una prueba de contraste global a los modelos obtenidos por regresión lineal determinando con ello la validez en ambos casos.  
En los modelos de series temporales la robustez está dada por los coeficientes de correlación, que para ambos casos son de un buen ajuste.

### Recomendaciones

1. Los modelos de regresión lineal deben utilizarse para describir el comportamiento de los datos a través del tiempo; mientras que las proyecciones se deben realizar con análisis de series de tiempo.
2. Realizar análisis de series de tiempo en los demás puertos del país; así como para los distintos tipos de carga que se movilizan en ellos.

### Referencias

- Antunez, C. (2011). *Análisis de Series de Tiempo*. Perú. Recuperado de <http://www.eumed.net/ce/2011a/chai.htm>
- Hair Jr., Anderson, R., Tatham, R., y Black, W. (1999). *Análisis multivariante*. España: Prentice Hall.
- Gómez, J., Aparicio, J., y Patiño, C. (2010). *Manual de análisis de frecuencias en hidrología*. México: Instituto Mexicano de Tecnología del Agua.
- Díaz, A. (2009). *Diseño estadístico de experimentos*. Colombia: Universidad de Antioquia.
- Lind, D., Marchal, W., y Wathen, S. (2012). *Estadística aplicada a los negocios y la economía*. México: McGraw-Hill.

### Información del autor

Ingeniero Industrial, Antonio Daniel Asencio Marroquín, Facultad de Ingeniería, Universidad de San Carlos de Guatemala, 2010.

Maestro en Estadística Aplicada de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería, Universidad de San Carlos de Guatemala, 2018.

Afiliación laboral; Empresa Portuaria Nacional Santo Tomás de Castilla.