



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado

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

Antonio Daniel Asencio Marroquín

Asesorado por el Mtro. Juan Carlos Fuentes Montepeque

Guatemala, mayo de 2018

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

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

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

ANTONIO DANIEL ASENCIO MARROQUÍN

ASESORADO POR EL MTRO. JUAN CARLOS FUENTES MONTEPEQUE

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, MAYO DE 2018

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. Angel Roberto Sic García
VOCAL II	Ing. Pablo Christian de León Rodríguez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Oscar Humberto Galicia Nuñez
VOCAL V	Br. Carlos Enrique Gómez Donis
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADORA	Dra. Mayra Virginia Castillo Montes
EXAMINADOR	Mtro. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. Pedro Dionisio Remis Salguero
SECRETARIA	Inga. Lesbia Magalí Herrera López.

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

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

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 3 de junio de 2017.

Antonio Daniel Asencio Marroquín

ACTO QUE DEDICO A:

- Dios** Por ser mi guía a lo largo de los años, por la vida, salud y fuerzas para levantarme y lograr cumplir este gran sueño.
- Mi madre** María Magdalena Marroquín Gómez. Gracias mamá por tu esfuerzo y dedicación para sacarme adelante, y a pesar de todo, haber formado al hombre que soy.
- Mi esposa** Sanddy Maricela Valenzuela Arreaga, por ser parte de mi vida e instarme a culminar este trabajo de graduación; como también, por su apoyo y amor incondicional.
- Mi hija** Fátima Daniela, por ser la estrella que alumbra mi vida, mi razón de ser y mi mejor maestra, te amo hija.
- Mis abuelos** Óscar y Carlota, por su sabiduría y consejos de lucha por la vida, ya que siempre han estado a mi lado para motivarme a cumplir todos mis anhelos. Gracias por preocuparse por mí.
- Mi familia** Por apoyarme siempre.

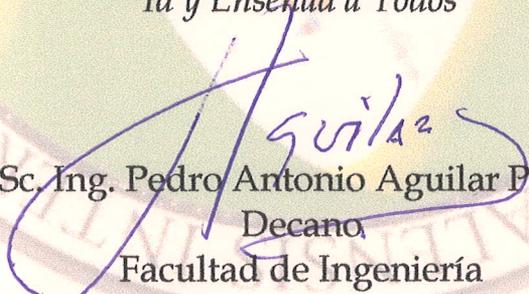
AGRADECIMIENTOS A:

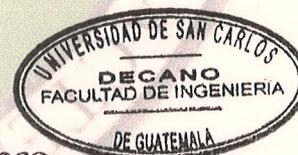
Universidad de San Carlos de Guatemala	Por haberme albergado todos estos años en tan prestigiosa casa de estudios.
Facultad de Ingeniería	Por haber participado durante toda mi formación académica superior.
Empresa Portuaria Nacional Santo Tomás de Castilla	Por haber facilitado los datos para la realización del presente trabajo.
Ing. Juan Carlos Fuentes Montepeque	Por su valiosa colaboración en la asesoría, revisión y corrección del presente trabajo.
Doctora Mayra Castillo y Maestro Edwin Bracamonte	Por todo el apoyo brindado durante el transcurso de la maestría.
Mis amigos	Por su valiosa amistad y apoyo.

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Postgrado, al Trabajo de Graduación de la Maestría en Artes en Estadística Aplicada titulado: "MODELO DE REGRESIÓN LINEAL APLICADO A LA CANTIDAD DE CONTENEDORES A MOVILIZAR EN PUERTO SANTO TOMÁS DE CASTILLA, GUATEMALA" presentado por el Ingeniero Industrial Antonio Daniel Asencio Marroquín, procede a la autorización para la impresión del mismo.

IMPRÍMASE.

"Id y Enseñad a Todos"


MSc. Ing. Pedro Antonio Aguilar Polanco
Decano
Facultad de Ingeniería
Universidad de San Carlos de Guatemala



Guatemala, mayo de 2018.

Cc archivo/L.Z.L.A.

Ref.APT-2018-012

El Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen y dar el visto bueno del revisor y la aprobación del área de Lingüística al Trabajo de Graduación titulado **"MODELO DE REGRESIÓN LINEAL APLICADO A LA CANTIDAD DE CONTENEDORES A MOVILIZAR EN PUERTO SANTO TOMÁS DE CASTILLA, GUATEMALA"** presentado por el Ingeniero Industrial **Antonio Daniel Asencio Marroquín**, correspondiente al programa de Maestría en Artes en Estadística Aplicada; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

MSc. Ing. Murphy Olympto Paiz Recinos
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala



Guatemala, mayo de 2018.

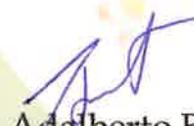
Cc: archivo/L.Z.L.A.

Ref.APT-2018-012

Como Coordinador de la Maestría en Artes en Estadística Aplicada del Trabajo de Graduación titulado **"MODELO DE REGRESIÓN LINEAL APLICADO A LA CANTIDAD DE CONTENEDORES A MOVILIZAR EN PUERTO SANTO TOMÁS DE CASTILLA, GUATEMALA"** presentado por el Ingeniero Industrial **Antonio Daniel Asencio Marroquín**, apruebo y recomiendo la autorización del mismo.

Atentamente,

"Id y Enseñad a Todos"


MSc. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador(a) de Maestría
Escuela de Estudios de Postgrado
Facultad de Ingeniería
Universidad de San Carlos de Guatemala



Guatemala, mayo de 2018.

Cc: archivo/LZLA.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
GLOSARIO.....	IX
RESUMEN.....	XI
PLANTEAMIENTO DEL PROBLEMA.....	XIII
OBJETIVOS.....	XIX
RESUMEN DEL MARCO METODOLÓGICO.....	XXI
INTRODUCCIÓN.....	XXV
1. MARCO REFERENCIAL.....	1
2. MARCO TEÓRICO.....	7
2.1. Conceptos básicos.....	7
2.1.1. Puerto.....	7
2.1.2. Contenedor.....	8
2.1.3. El papel del puerto en el comercio internacional..	9
2.2. Fundamentación estadística.....	10
2.2.1. Análisis previo de los datos.....	10
2.2.2. Modelos de regresión.....	19
2.2.2.1. Análisis de correlación.....	19
2.2.2.2. Análisis de regresión.....	20
2.2.2.3. Series de tiempo.....	23
2.2.2.4. Modelos autoregresivos.....	24
3. PRESENTACIÓN DE RESULTADOS.....	27
3.1. Selección de las variables.....	27

3.1.1.	Variables a incluir en el modelo de regresión lineal.....	27
3.1.2.	Variables a incluir en el modelo generado por series de tiempo.....	33
3.2.	Comprobación de los supuestos y propiedades del modelo.....	33
3.2.1.	Supuestos modelo de regresión lineal.....	33
3.2.1.1.	Linealidad.....	33
3.2.1.2.	Normalidad.....	34
3.2.1.3.	Independencia.....	35
3.2.1.3.	Homogeneidad.....	37
3.2.1.3.	Homocedasticidad.....	37
3.2.2.	Propiedades de las series de tiempo.....	38
3.2.2.1.	Tendencia y aleatoriedad.....	38
3.2.2.2.	Estacionalidad.....	40
3.3.	Construcción del modelo.....	40
3.3.1.	Validación del modelo de regresión lineal.....	40
3.3.2.	Validación del modelo generado por series de tiempo.....	42
3.4.	Robustez del modelo.....	44
3.4.1.	Ajuste modelo de regresión lineal.....	44
3.4.2.	Ajuste del modelo generado por series de tiempo.....	45
4.	DISCUSIÓN DE RESULTADOS.....	47
4.1.	Selección de las variables.....	47
4.1.1.	Variables a incluir en el modelo de regresión lineal.....	47

4.1.2.	Variables a incluir en el modelo generado por series de tiempo.....	49
4.2.	Comprobación de los supuestos y propiedades del modelo.....	53
4.2.1.	Supuestos modelo de regresión lineal.....	53
4.2.1.1.	Linealidad.....	54
4.2.1.2.	Normalidad.....	54
4.2.1.3.	Independencia.....	55
4.2.1.3.	Homogeneidad.....	56
4.2.1.3.	Homocedasticidad.....	56
4.2.2.	Propiedades de las series de tiempo.....	57
4.2.2.1.	Tendencia y variabilidad.....	57
4.2.2.2.	Estacionalidad.....	57
4.3.	Construcción del modelo.....	58
4.3.1.	Validación modelo de regresión lineal.....	58
4.3.2.	Validación del modelo generado por series de tiempo.....	58
4.4.	Robustez del modelo.....	60
4.4.1.	Ajuste modelo de regresión lineal.....	60
4.4.2.	Ajuste del modelo generado por series de tiempo.....	60
	CONCLUSIONES.....	63
	RECOMENDACIONES.....	65
	BIBLIOGRAFÍA Y REFERENCIAS	67
	APÉNDICES.....	71
	ANEXOS.....	89

ÍNDICE DE ILUSTRACIONES

FIGURAS

1. Unidades de contenedores manipuladas en los puertos 2008-2015.....	XIV
2. Diagrama de dispersión toneladas-recaladas 2012 a 2015....	19
3. Recta de regresión para el modelo toneladas-recaladas.....	23
4. Q-Q plot residuos modelo importación de contenedores.....	34
5. Q-Q plot residuos modelo exportación de contenedores.....	35
6. Correlograma residuos y predichos, modelo importación contenedores.....	36
7. Correlograma residuos y predichos, modelo exportación contenedores.....	36
8. Serie de tiempo importación contenedores.....	39
9. Serie de tiempo exportación contenedores.....	39
10. FAS y FAC residuos modelo importación.....	43
11. FAS y FAC residuos modelo exportación.....	44
12. Diagrama de dispersión importación-gastos de consumo final.....	45
13. Diagrama de dispersión exportación-exportaciones de bienes y servicios.....	45
14. Contenedores importación y proyección ARIMA 2017.....	46
15. Contenedores exportación y proyección ARIMA 2017.....	46
16. Diferencia contenedores importación (d=1).....	50
17. Diferencia contenedores exportación (d=1).....	50
18. FAC importaciones (d=1).....	51

19.	FACP importaciones (d=1).....	51
20.	FAC exportaciones (d=1).....	52
21.	FACP exportaciones (d=1).....	52

TABLAS

I.	Histórico unidades de contenedores manipuladas.....	XIV
II.	Variables a considerar en el modelo de importación de contenedores.....	27
III.	Variables a considerar en el modelo de exportación de contenedores.....	28
IV.	Análisis de regresión lineal contenedores importación.....	28
V.	Coefficientes de regresión y estadísticos asociados modelo de importación.....	29
VI.	Análisis de regresión lineal contenedores importación (modelo parsimonioso).....	29
VII.	Coefficientes de regresión y estadísticos asociados modelo de importación (modelo parsimonioso).....	30
VIII.	Análisis de varianza.....	30
IX.	Análisis de regresión lineal contenedores exportación.....	30
X.	Coefficientes de regresión y estadísticos asociados modelo de exportación.....	31
XI.	Análisis de regresión lineal contenedores exportación (modelo parsimonioso).....	31
XII.	Coefficientes de regresión y estadísticos asociados modelo de exportación (modelo parsimonioso).....	32
XIII.	Análisis de varianza.....	32
XIV.	Prueba de significancia para el coeficiente r.....	33
XV.	Prueba de Shapiro-Wilks para normalidad.....	34

XVI.	Prueba de rachas para aleatoriedad.....	35
XVII.	Prueba T de Student para homogeneidad.....	37
XVIII.	Prueba F de Snedecor para homocedasticidad.....	38
XIX.	Prueba de raíz unitaria de Dickey-Fuller.....	40
XX.	Contraste global del modelo, prueba de F.....	41
XXI.	Prueba de bondad de ajuste Kolmogórov-Smirnov.....	43

GLOSARIO

ARIMA	Modelo autoregresivo de media móvil.
Carga general	Se presenta en estado sólido, líquido o gaseoso y puede ser tratada como unidad, embarcándose y desembarcándose por medio de grúa.
Contenedor	Caja construida especialmente para movilizar carga y que tiene un carácter reutilizable. En ella se pueden consolidar mercancías para ser trasladadas de punto a punto como una unidad.
EIEWS	Paquete de software estadístico
FAC	Función de autocorrelación.
FACP	Función de autocorrelación parcial.
FAP	Función de autocorrelación parcial serial.
FAS	Función de autocorrelación serial.
Granel líquido	Líquidos embarcados y desembarcados por medio de tuberías.

Granel sólido	Producto sólido movilizado sin empaque o envase, embarcado y desembarcado por medio de grúa.
INFOSTAT	Paquete de software estadístico.
JICA	Agencia de Cooperación Internacional del Japón.
Línea naviera	Empresa encargada de transportar carga desde su punto de origen hasta su destino final.
Microsoft Excel	Aplicación de hojas de cálculo que forma parte del paquete de oficina Microsoft Office.
PIB	Producto interno bruto.
SPSS	Paquete de software estadístico.

RESUMEN

El presente trabajo de graduación tiene como objetivo brindar la metodología necesaria para la elaboración de modelos de regresión lineal que permitan pronosticar valores a futuro para la cantidad de contenedores de importación y exportación en el puerto Santo Tomás de Castilla. La necesidad del mismo surge derivado del poco o nulo análisis que se les hace a los datos previo a la elaboración de modelos, de los cuales, en la mayoría de los casos, son modelos de regresión lineal. Por lo anterior se realizó un modelamiento de regresión lineal para la cantidad de contenedores de importación y exportación, seleccionando las variables a incluir en ellos; para posteriormente verificar el cumplimiento de los supuestos de normalidad, linealidad, independencia, homogeneidad y homocedasticidad; principios fundamentales para que un modelo de regresión sea válido. De igual manera se comprobó la significancia del mismo y el ajuste que tiene para explicar el comportamiento de los datos por medio de los modelos obtenidos.

De igual forma se hizo un análisis por medio de series de tiempo, en donde se comprobaron las propiedades de la tendencia, variabilidad y estacionalidad de los datos y se estableció el modelo a utilizar para pronosticar los mismos; eligiendo en ambos casos un modelo autoregresivo de media móvil (ARIMA). Para darles validez a los modelos obtenidos, se verificó el cumplimiento de la normalidad e independencia de los residuos, obteniendo el ajuste a los datos históricos por medio del coeficiente de correlación.

PLANTEAMIENTO DEL PROBLEMA

Contexto general

El comercio exterior de Guatemala está constituido por el volumen de importaciones y exportaciones que ingresan o salen al país, ya sea por la vía terrestre, marítima o aérea. Según el informe estadístico El Sistema Portuario Nacional en apoyo al comercio exterior de Guatemala, elaborado por la Comisión Portuaria Nacional en el año 2016, el 81 % del volumen de dicho comercio tiene como punto de entrada o salida los puertos marítimos nacionales, movilizándose el otro 19 % por los puertos aéreos y terrestres. Este comercio está constituido por distintos tipos de productos, los cuales se agrupan en cuatro clasificaciones generales, tanto para importación como para exportación, siendo estos: carga general, carga en contenedores, graneles sólidos y graneles líquidos. De estos tipos de carga, la que se transporta en contenedores es la que mayor tonelaje moviliza en su totalidad, llegando a representar un 36 % del tonelaje total manipulado en los puertos. La importancia de este tipo de carga es que son unidades homogéneas donde se pueden introducir una gran cantidad de productos y su manipulación en puerto es muy sencilla, ya que existe maquinaria especializada para ello.

Debido a la importancia que tiene para el comercio por vía marítima, es necesario contar con proyecciones para tener una idea de su comportamiento a futuro y de esta manera, realizar planes de reordenamiento de las áreas físicas dentro de los puertos, ampliaciones de las instalaciones actuales, entre otras; las cuales representan una alta inversión de capital para las empresas portuarias, razón por la que, las cantidades proyectadas deben ser lo más acertadas posibles.

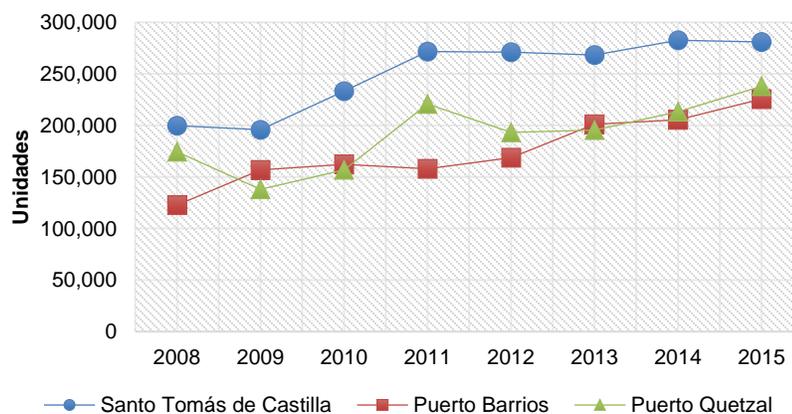
Para visualizar de mejor manera lo comentado en los párrafos anteriores, se presenta en la tabla I la cantidad de contenedores manipulados en los puertos guatemaltecos durante el periodo 2008-2015.

Tabla I. **Histórico unidades de contenedores manipuladas**

Puerto/año	Santo Tomás de Castilla	Puerto Barrios	Puerto Quetzal	Total sistema portuario
2008	199,706	122,685	174,374	496,765
2009	195,815	156,906	138,000	490,721
2010	233,262	162,292	156,786	552,340
2011	271,688	157,997	220,702	650,387
2012	271,121	168,768	193,223	633,112
2013	268,325	201,121	195,487	664,933
2014	282,593	205,535	213,382	701,510
2015	280,973	225,589	237,964	744,526

Fuente: elaboración propia, con información de la Comisión Portuaria Nacional, 2008-2015.

Figura 1. **Unidades de contenedores manipuladas en los puertos 2008 - 2015**



Fuente: elaboración propia.

Según los datos de la figura 1 se observa que el puerto en el que se manipula una mayor cantidad de contenedores es Santo Tomás de Castilla, llegando a tener el 38 % de unidades movilizadas en todos los puertos de Guatemala, colocándolo como el de mayor importancia para este tipo de carga.

Descripción del problema

Durante el 2015 se realizaron proyecciones del movimiento de carga en los puertos, a partir de información histórica mensual, la cual reportan los diferentes puertos y es consolidada en una base de datos para todo el Sistema Portuario Nacional. Estas proyecciones se realizan mediante un análisis de tendencias proyectando los valores en el tiempo. Sin embargo, no se puede precisar si el método con el que se elaboraron las proyecciones cumple con los análisis de datos necesarios para elaborarlas, debido a que se desconocen las características y las tendencias de las series de datos, o en determinado caso las mismas se realizan de manera empírica.

Lo anterior puede tener como consecuencia el riesgo de realizar proyecciones poco precisas, incurriendo en errores a la hora de elaborar estudios de reordenamiento de espacios físicos dentro de los puertos, planes de expansión y proyectos de inversión, entre otros, lo cual repercutiría en una asignación de recursos, que sobrepase o no, sea suficiente a la hora de realizar planificaciones a futuro, por ejemplo, en los presupuestos de ingresos, los cuales en gran medida dependen de la carga.

Derivado de lo anterior, hay que definir una metodología para proyectar el movimiento de carga, basada en el análisis de datos mediante herramientas estadísticas y compararlas con los métodos actuales, como también, cuál utilizar, optimizando los procesos e inversiones.

Es por ello que debe ser una prioridad para las empresas portuarias analizar la serie de datos de la carga y realizar una proyección utilizando las herramientas que brinda la estadística aplicada, para compararlas con los datos que genera el método actual, y así definir la metodología que se utilizará para proyectar los datos.

Formulación del problema

Se basó en la búsqueda de respuesta a la pregunta principal:

¿Cuál es el modelo estadístico de mejor ajuste para realizar proyecciones de la cantidad de contenedores en puerto Santo Tomás de Castilla?

Esta interrogante principal se relaciona con las siguientes preguntas auxiliares:

1. ¿Qué variables se deben incluir en un modelo que permita realizar proyecciones de la cantidad de contenedores en Puerto Santo Tomás de Castilla?
2. ¿Cuáles son los supuestos y las propiedades que debe tener la serie de datos histórica para realizar modelos por medio de análisis de regresión y modelos autoregresivos, que permitan proyectar la cantidad de contenedores en puerto Santo Tomás de Castilla?
3. ¿Qué análisis se debe aplicar para construir el modelo que permita realizar proyecciones de la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla?

4. ¿Cuál es la robustez del modelo que se ajuste mejor a las características de las series de contenedores en puerto Santo Tomás de Castilla?

Las respuestas a estas interrogantes proporcionaron la solución a la problemática que se presenta en puerto Santo Tomás de Castilla.

Delimitación del problema

Los contenedores son la carga con mayor peso en los puertos guatemaltecos, y puerto Santo Tomás de Castilla es el que mayor número de unidades manipula, se utilizaron los datos mensuales de unidades de contenedores desde el año 2000 hasta 2016 de este puerto para realizar los análisis correspondientes, los cuales permitieron elaborar modelos para proyectar a futuro los datos de desembarque/embarque de contenedores.

OBJETIVOS

General

Elaborar un modelo para proyectar la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla a partir del año 2017, mediante el análisis de regresión y de series temporales, determinando el que mejor se ajuste a las características de los datos.

Específicos

1. Definir la o las variables a incluir en el modelo de proyección de la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla, mediante análisis de regresión y modelos autoregresivos, para incluirlas en el mismo.
2. Comprobar los supuestos y propiedades que debe tener la serie de datos histórica para establecer si se pueden realizar modelos mediante análisis de regresión y modelos autoregresivos, los cuales permitan realizar proyecciones de la cantidad de contenedores en puerto Santo Tomás de Castilla.
3. Construir un modelo de proyección mediante el análisis de regresión y modelos autoregresivos, para proyectar la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla.

4. Determinar la robustez del modelo mediante la evaluación que mejor se ajuste a las características de la serie real de contenedores movilizados en puerto Santo Tomás de Castilla.

RESUMEN DEL MARCO METODOLÓGICO

Características del estudio

El presente estudio tiene un enfoque cuantitativo debido a que se construyó una base de datos organizada referente a las unidades de contenedores movilizados en Puerto Santo Tomás de Castilla, la cual fue analizada, logrando la generación de valores futuros mediante el empleo de análisis de regresión y modelos autoregresivos.

El alcance es descriptivo, ya que se especificaron las características y tendencias de las series de datos para realizar estimaciones de valores a corto y mediano plazo.

El diseño adoptado fue no experimental de corte longitudinal de tendencia, ya que se utilizó la cantidad de unidades de contenedores que se movilizaron en el puerto durante el periodo 2000-2016 en su estado original, sin realizar ninguna modificación a los mismos.

Unidades de análisis

La población en estudio fueron las unidades de contenedores que se movilizan por el puerto Santo Tomás de Castilla, las cuales se encuentran divididas en subpoblaciones dadas por los movilizados en la operación de desembarque, así como los movilizados en embarque, durante el periodo 2000-2016, estudiándolas en su totalidad.

Variables

Las variables en estudio se describen a continuación:

Nombre variable	Definición teórica	Definición operativa
Contenedor desembarque	Unidad descargada de un buque	Número entero referente a unidades descargadas de los buques mensualmente
Contenedor embarque	Unidad cargada a un buque	Número entero referente a unidades cargadas a los buques mensualmente
Mes	Identificación del mes	Nombre del mes
Año	Identificación del año	2000 a 2016

Fases del estudio

A continuación, se describe el proceso que permitió realizar el estudio.

- Fase 1. Revisión de literatura

Se hizo una lectura de distinta bibliografía para ampliar los aspectos teóricos y técnicos que contiene el informe final, con el objetivo de estar bien documentado al respecto y abarcar la mayor parte del tema.

- Fase 2. Gestión de la información

Se realizó la solicitud de datos estadísticos a puerto Santo Tomás de Castilla, principalmente los referentes a las unidades de contenedores movilizados durante el periodo 2000-2016.

- Fase 3. Diagnóstico de la situación actual

Se realizó un análisis de la situación inicial en cuanto al movimiento de unidades de contenedores durante el periodo 2000-2016, para conocer cómo se ha comportado esta operación en el puerto Santo Tomás de Castilla.

- Fase 4. Elaboración de base de datos

Se construyó una base de datos en Excel para elaborar los análisis correspondientes, y así realizar el estudio propuesto.

- Fase 5. Análisis de información

Ya con la base de datos elaborada se realizó el análisis de la misma en dos vertientes, siguiendo la línea de modelos de regresión y, por aparte, con series de tiempo; obteniendo un modelo con cada análisis para contrastar los mismos, y elegir el de mejor ajuste a los datos reales.

Para realizar el análisis mediante la línea de modelos de regresión se comprobaron los supuestos de linealidad, homogeneidad, homocedasticidad, normalidad e independencia. Los mismos se verificaron de manera gráfica y teórica, obteniendo un modelo y realizando su validación.

El análisis mediante la línea de series de tiempo se realizó de manera similar, comprobando las propiedades de longitud, datos faltantes y tendencia; utilizando luego modelos autoregresivos para la obtención del mismo, procediendo a validarlo.

En principio, el análisis de la información se realizó utilizando Excel 2010; sin embargo, también se hizo uso de los programas estadísticos INFOSTAT, SPSS y EVIEWS 9.

- Fase 6. Interpretación y discusión de resultados

Luego de obtener los resultados de los análisis realizados a las series de datos, se interpretaron los mismos incluyéndolo en el presente informe.

Conjuntamente con el asesor del trabajo se realizó la discusión de resultados, determinando el aporte de los análisis realizados, los modelos que se obtuvieron, haciendo la selección del que mejor se ajustó a las características de los datos.

- Fase 7. Conclusiones y recomendaciones

Al tener finalizado el trabajo se realizaron las conclusiones, en concordancia con los objetivos planteados e igualmente se elaboraron las recomendaciones pertinentes.

INTRODUCCIÓN

La estadística aplicada brinda distintas herramientas que pueden ser utilizadas para la resolución de problemas que se presentan en varias situaciones. Es por ello que se hizo uso de estas, para dar solución a uno de dichos problemas, específicamente el de realizar proyecciones del movimiento de contenedores en puerto Santo Tomás de Castilla, Guatemala.

Actualmente se desconoce si se practican los análisis estadísticos necesarios a las series de datos correspondientes a la cantidad de contenedores que se movilizan en el puerto, los cuales deriven en algún modelo de proyección que permita obtener datos a corto, mediano y largo plazo. Dichas proyecciones ayudarían a que se evite cometer cierto tipo de errores, que a la larga parecen comunes, pero que pueden evitarse; lo cual será de beneficio a las empresas portuarias, minimizando la asignación de recursos monetarios, materiales y humanos a obras de infraestructura que no cumplan con la finalidad con las que fueron concebidas, ya que pueden ser insuficientes o sobredimensionadas para satisfacer la demanda que se planificó.

Como solución a la problemática encontrada se hizo un análisis estadístico de las series de datos disponibles, elaborando un modelo con las variables incluidas en el mismo, para proyectar la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla a partir del 2017, realizando una investigación con base en distintos estudios que han aplicado análisis similares; determinando el que mejor se ajuste a las características del puerto, mediante la comparación con datos reales.

La inversión en nueva infraestructura, la correcta asignación de los recursos, la elaboración de presupuestos, planes operativos y estratégicos son algunos temas en los que se puede hacer uso de la metodología utilizada para la elaboración del modelo de proyección que se construya en el presente trabajo. Derivado de que en los temas mencionados con anterioridad se debe hacer uso de proyecciones, ya que de las mismas dependen las decisiones que puedan tomar los directivos del puerto y cuyo objetivo principal sea el crecimiento del mismo, tanto en sus instalaciones físicas como en el servicio que se preste a la carga y a los buques.

El informe final tiene la siguiente estructura:

En el capítulo 1 se describen los antecedentes, que básicamente referencian los trabajos de investigación realizados anteriormente respecto al tema, obteniendo de ellos técnicas y procedimientos en los que se basaron los autores para la obtención de modelos de proyección en el ámbito portuario.

El capítulo 2, marco teórico, se divide en una sección de conceptos básicos, específicamente en el tema portuario, y una segunda parte donde está la fundamentación estadística que se utilizó para realizar los modelos de proyección de la cantidad de contenedores en puerto Santo Tomás de Castilla.

La presentación de los resultados obtenidos mediante los análisis estadísticos necesarios, se dan a conocer en el capítulo 3, descritos de manera numérica y gráfica.

Por último, se presenta la discusión de los resultados en el capítulo 4, en este se brinda solución al problema que generó el estudio, donde la realización de proyecciones se realiza de forma empírica, ya que con el análisis realizado, permite tener una metodología para realizar las mismas de manera fundamentada, la cual puede ser replicada para los distintos tipos de carga que se movilizan, no solo en puerto Santo Tomás de Castilla, sino también en los puertos del Sistema Portuario Nacional guatemalteco.

1. MARCO REFERENCIAL

La realización de proyecciones para el movimiento de carga en los puertos nacionales debe ser considerada como una labor crucial, ya que de estas se desprenderán datos muy importantes, los cuales permitirán realizar planes a futuro en cuanto a expansiones del recinto portuario, adquisición de nueva maquinaria, construcción de bodegas o patios, etc. Es por ello que, en el puerto Santo Tomás de Castilla no se puede rezagar el tema, primordialmente en lo que se refiere a la carga de contenedores, ya que la misma es el principal tipo de carga que se moviliza en el puerto, razón por la cual debe contar con series históricas de datos que permitan generar un modelo para proyectar lo que pueda ocurrir en el futuro.

De acuerdo con Rodríguez (2015), los estudios de planificación en los puertos españoles se basan en modelos empíricos para el desarrollo de nuevas terminales; así como para la elaboración de sus planes a futuro. El autor utiliza dos técnicas basadas en la minería de datos para realizar modelos de proyección en puertos, siendo estas las redes bayesianas y las neuronales artificiales, haciendo una comparación entre los modelos obtenidos por ambos métodos; llegando a la conclusión que ambas técnicas pueden ser utilizadas para dicho fin. Tomando en consideración lo desarrollado por el Sistema Portuario de España y la falta de modelos de proyección en el mismo, podría ser muy difícil que en países como Guatemala se tengan modelos de proyección desarrollados, que permitan planificar los aspectos que ayuden a administrar de mejor forma un puerto.

Gavarrete y Fernández (2001) recalcan la importancia de contar con datos históricos para realizar proyecciones en puerto, sobre todo los centroamericanos, ya que con estos datos se contarán con tasas de variación y crecimiento, las cuales incorporan a su modelo para la proyección a futuro y utilizándolas a la vez como un parámetro de comparación con los datos que genere el modelo.

Lo anterior da una idea de cómo se puede lograr obtener datos que puedan responder al movimiento histórico que se ha presentado durante un periodo determinado; sin embargo, según el Ministerio de Océanos y Pesca de la República de Corea (2015), el plan maestro portuario de Guatemala utilizó en el modelo, para realizar las proyecciones de movimiento de carga en puertos, variables macroeconómicas, tales como el PIB y el crecimiento poblacional del país en estudio, en su modelo de proyección.

La Agencia de Cooperación Internacional del Japón JICA (1988) utilizó una metodología similar en el estudio sobre el proyecto de desarrollo del puerto Santo Tomás de Castilla en la República de Guatemala, en donde se menciona la utilización de dos técnicas para pronosticar el volumen de carga, mediante el método micro y el método macro. Haciendo uso de la relación entre los índices macroeconómicos (tales como el PIB) y el movimiento de carga para el método macro, ya que el mismo se basa en que la actividad económica del país tiene una relación con el movimiento de carga. Por su parte, el método micro pronostica productos individualmente, tomando en consideración la demanda y el suministro en su modelo.

Royal Haskoning (2008) utilizó una metodología de proyección similar al método micro y macro (descrito anteriormente) en el plan maestro para el complejo portuario Limón-Moín, Costa Rica. La diferencia es que utilizan el método macro, en donde consideran las variables macroeconómicas como el PIB para el cálculo de tráfico de importación, considerando relevante y constante la relación entre ambas variables. Sin embargo, para el cálculo de proyecciones de exportación realiza el análisis micro; proyectando los volúmenes de carga para los principales productos de exportación, realizando una sumatoria de los mismos al final para obtener los volúmenes de carga proyectados.

En la misma línea realiza las previsiones de tráfico la Autoridad Portuaria Nacional de Perú (2012), en el plan nacional de desarrollo portuario, utilizando una estimación macro para el cálculo del volumen total de carga, donde las variables del modelo son el movimiento histórico de mercancías, la correlación de los principales índices macroeconómicos del país y del mundo; así como la incidencia de los proyectos de inversión y los tratados de libre comercio. Por su parte, en la estimación micro se estima el volumen por producto tomando en cuenta su movimiento histórico y la correlación de índices y cuentas per cápita, las cuales pueden ser algunas variables a utilizar en el presente estudio.

De igual manera, Puerto de Valparaíso Chile (2015), utiliza una metodología micro para estimar las proyecciones de sus exportaciones, es decir, realizan una proyección para varios sectores analizados, consolidando cada uno para obtener un volumen total. Mientras que, para tener datos proyectados de las importaciones se toma en cuenta que las mismas son dependientes en buena medida del PIB nacional. Esto resulta importante, ya que se tomó la idea de realizar un modelo para el desembarque y otro distinto para el embarque.

Iztapa Management S. A. (2003), no da mayor detalle de las variables que utiliza en el modelo de proyección para la carga en contenedores en Puerto Quetzal, Guatemala; sin embargo, resulta interesante mencionar que, para realizarlo utilizaron tres escenarios distintos, en los cuales se analizan el tipo de servicio que satisface el comercio nacional y el comercio de otros países, un supuesto de varias líneas navieras utilizando al puerto como un puerto de trasiego regional y la utilización del puerto como un punto de transbordo que permita servir no solo a la región sino al mercado asiático y la costa occidental del sur de América. Lo rescatable de este estudio fue el análisis que se le puede dar a las proyecciones, por lo que se consideró la idea de realizar tres tipos de escenarios: uno optimista, neutro y pesimista.

En el Plan Nacional de Desarrollo Portuario, elaborado por la Subsecretaría de Transportes (2013), se recopila el método de proyección utilizado para cada uno de los puertos de Chile, en donde cada uno depende, principalmente del tipo de producto que se manipula. En este sentido llamó la atención el método de proyección utilizado para el puerto de Austral, en donde se elaboró un modelo de correlación entre el volumen de la carga y la actividad económica de la región, medida en este caso como el valor total facturado, combinándolo con un análisis de las series de tiempo de las cargas en contenedores, gráneles y carga fraccionada. Tomando como supuestos determinadas tasas de variación para distintos periodos de tiempo.

Considerando que las referencias anteriores utilizan metodologías similares para elaborar modelos de proyección en los puertos, estas fueron utilizadas como base para construir un modelo de proyección de contenedores en el puerto Santo Tomás de Castilla, ya que a pesar de que las metodologías no son explicadas a detalle, brindaron una idea de las variables que puedan formar parte del mismo.

En este sentido se utilizaron las variables de la cantidad de contenedores y el producto interno bruto del país, realizando un análisis de correlación en busca de obtener un modelo mediante esta técnica.

Derivado de lo anterior, se consideran muy valiosos los esfuerzos que se han venido realizando para generar modelos de proyección de las cargas en los puertos, por lo que estos sirvieron de guía para elaborar un modelo acorde a las características de puerto Santo Tomás de Castilla, el cual permitió proyectar a futuro las cifras de contenedores a movilizar.

El carácter innovador de este trabajo de investigación radica en la obtención de otro modelo mediante la utilización de modelos autoregresivos, ya que según reflejan los antecedentes, las metodologías utilizadas no mencionan el uso de este análisis estadístico para proyectar datos.

2. MARCO TEÓRICO

2.1. Conceptos básicos

2.1.1. Puerto

Un puerto marítimo, para cualquier país que tenga el beneficio de tenerlo, constituye la puerta de entrada y salida de la mayor parte del comercio exterior del mismo, razón por la que se le debe dar la importancia del caso.

Según Cendrero y Truyols (2008), un puerto es el lugar donde se realiza intercambio comercial por medio del tráfico marítimo y terrestre. El mismo está conformado por instalaciones, organizaciones e infraestructura las cuales en conjunto permiten que el mismo tenga funcionamiento.

Por su parte Marí, De Souza, Martín y Rodrigo (2003), puerto es como un lugar situado a la orilla del mar, río o lago y en donde se encuentran carreteras y vías férreas con los buques para el intercambio de mercancías.

Por último, la Conferencia de las Naciones Unidas para el Comercio y Desarrollo (2009), conceptualiza el puerto como:

Entidad geográfica y económica, con un nombre concreto, situado en la orilla del mar, de un río o de un lago, al servicio de los buques, y donde se lleva a cabo el tránsito de mercancías y de pasajeros desde el agua al transporte terrestre y donde se encuentran las instalaciones en la tierra y en el agua para prestar los servicios complementarios requeridos por los buques, las mercancías y para el desarrollo del comercio internacional, la industria y, de forma más genérica, la economía de los países en el marco de la zona de influencia del puerto. (p. 7).

La definición anteriormente descrita da una explicación bastante amplia de lo que es un puerto, por lo tanto, es la que se utiliza en el presente trabajo.

2.1.2. Contenedor

Este trabajo de investigación tiene como base realizar proyecciones del movimiento de unidades de contenedores, es por esto que resulta importante definirlo, así como describir los tipos y tamaños de unidades que se manipulan en el puerto.

El contenedor nace a partir de la necesidad de unificar la carga, lo que consiste en agrupar mercancías en unidades que sean capaces de facilitar su manipulación, transporte y movimiento.

Marí et al. (2003), contenedor lo definen como una caja de dimensiones establecidas en la cual se pueden cargar distintos tipos de mercaderías, y el cual puede ser transportado vía marítima, terrestre, aérea y férrea.

Por su parte, Castellanos (2015), identifica a un contenedor como un recipiente para transporte de mercancías, el cual debe garantizar su uso varias veces, por lo que el mismo debe ser de calidad duradera y resistente.

Las dimensiones de los mismos son de 8 pies de ancho y 8 pies con 6 pulgadas de alto, encontrando variaciones en el largo, el cual puede variar entre 20, 40 y 45 pies para el mercado guatemalteco.

2.1.3. El papel del puerto en el comercio internacional

Tal y como se ha mencionado anteriormente, el puerto es solamente un eslabón en toda la cadena logística y de transporte. Sin embargo, como bien lo menciona Sánchez (2004), el mismo es parte fundamental en la determinación del grado de competitividad, integración y desarrollo económico, no solo del país, sino que también de la región a la que este pertenezca. En este sentido, para mejorar la productividad de la economía del país, es necesario que la red de servicios de infraestructura funcione de manera eficiente y a un menor costo, por lo que estos deben ser temas importantes para cualquier país, más cuando el mismo está abierto a la integración y al comercio internacional.

El comercio exterior de un país está constituido por el valor de las mercancías y el volumen de las mismas, tanto en importaciones como exportaciones que ingresan o salen del país, ya sea por la vía terrestre, marítima o aérea. Es por ello que se puede afirmar con seguridad, que el desarrollo económico de determinado país está íntimamente ligado con el desarrollo comercial, siendo esta la razón que ha llevado a las autoridades de distintos países a promover su comercio exterior, el cual sigue teniendo expectativas de crecimiento a futuro.

Los datos mencionados anteriormente dan una idea clara y precisa de la importancia de un puerto para los países, razón por la cual hacerlos eficientes es una tarea crucial para aumentar la competitividad y dinamizar la economía de su entorno.

2.2. Fundamentación estadística

Muchas veces realizar análisis de datos tiene como objetivo principal medir la relación existente entre dos o más variables en el transcurso de determinado periodo de tiempo. Para ello existe una técnica estadística llamada análisis de correlación. Este muchas veces deriva en la construcción de una ecuación que permite estimar el valor futuro de una de las variables respecto de la otra, a esto se le llamará análisis de regresión.

De la misma manera que el análisis de regresión se utiliza para estimar el valor de determinada variable, el análisis de series de tiempo consiste en el uso de modelos para predecir el comportamiento de la misma a futuro, tomando como base un conjunto de datos ordenados en un periodo establecido.

En el presente capítulo se describe en qué consisten estas técnicas que sirvieron de base para desarrollar los modelos que permitan realizar proyecciones de la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla, el cual es tema central de este trabajo de investigación.

2.2.1. Análisis previo de los datos

Un error común que se comete al realizar análisis, ya sea de correlación o de series temporales, es no realizar un examen previo de los datos, desconociendo que al realizarlo es muy probable obtener predicciones más acertadas.

Según Hair, Anderson, Tatham y Black (1999) existen métodos diversos para el análisis de los datos previo a aplicar alguna técnica para estimar el comportamiento de los mismos, en el caso de algún modelo de regresión es muy importante la comprobación de los supuestos, dentro de los cuales se pueden mencionar la normalidad, homocedasticidad y linealidad.

A estos tres supuestos se les pueden agregar la homogeneidad y la independencia, a los cuales hacen referencia en su trabajo los autores Gómez, Aparicio y Patiño (2010).

En consecuencia, estos son los cinco supuestos que se analizaron en el presente trabajo, en su apartado de análisis de regresión. A continuación, se da una breve descripción de cada uno de ellos.

- Normalidad

Hair et al. (1999) expresan en su obra que la normalidad de los datos es uno de los supuestos fundamentales que se deben comprobar a la hora de realizar análisis de los mismos, ya que constituye el punto de referencia de la mayor parte de las pruebas estadísticas.

La importancia de este supuesto se fundamenta en que si los datos no siguen una distribución cercana a la normal toda prueba estadística paramétrica que se le haga a los mismos pierde su validez.

Para verificar el cumplimiento de este supuesto, Guisande (2006), recomienda la utilización de la prueba de Shapiro-Wilks para una muestra menor a 30 datos. Esta prueba permite obtener un p-valor, el cual debe ser contrastado con el nivel de significancia que se esté trabajando, partiendo de las siguientes hipótesis:

Ho: Los datos se ajustan a una distribución normal.

Ha: Los datos no se ajustan a una distribución normal.

Una vez planteadas las hipótesis se realiza la prueba para obtener el valor crítico de p, para tomar la decisión de la siguiente manera: si p-valor es menor que el nivel de significancia, se rechaza la hipótesis nula; por lo tanto los datos no se ajustan a una distribución normal. Mientras que si el valor de p es mayor que la significancia se acepta la hipótesis nula, concluyendo que los datos se ajustan a la distribución normal, cumpliéndose el supuesto.

Esta prueba se complementa con el gráfico Q-Q plot, en donde se busca verificar de manera visual el ajuste de los datos a una recta probabilística normal, interpretándose un ajuste perfecto cuando los puntos de la distribución formen una recta con pendiente de 45° , es decir un ajuste perfecto entre las frecuencias observadas y frecuencias esperadas.

Para una serie con más de 50 datos se recomienda utilizar la prueba Kolmogórov–Smirnov, la cual se basa en una comparación de la distribución acumulada teórica y muestral, buscando que las mismas no sean significativamente distintas, por lo que de ser así se acepta la hipótesis nula, es decir, la hipótesis de normalidad según lo indica Marques (2001).

Para ello se debe comparar el p-valor con el nivel de significancia al que se esté trabajando la prueba, aceptando la hipótesis nula si p-valor es mayor que la significancia establecida.

- Homocedasticidad

De acuerdo con Hair et al. (1999), la homocedasticidad de los datos se refiere, principalmente a que los valores de la varianza entre los datos presenten valores similares a lo largo del periodo estudiado, basándose en la dispersión de esta. Es decir, una dispersión constante de la varianza refleja homocedasticidad mientras que una dispersión variable expresa heterocedasticidad.

Este supuesto, según Guàrdia, Freixa, Però, y Turbany (2008) puede ser verificado mediante la prueba F de Snedecor, la cual busca en esencia la igualdad de las varianzas. Las hipótesis planteadas para esta prueba son las siguientes:

Ho: No existe diferencia significativa entre las varianzas de los grupos; por lo tanto existe homocedasticidad entre las mismas.

Ha: Existe diferencia significativa entre las varianzas de los grupos; por lo tanto no se cumple la homocedasticidad entre las mismas.

El estadístico de contraste F de Snedecor se define por medio del cociente de dos varianzas, de acuerdo a la siguiente formula:

$$F \text{ Snedecor} = \frac{\sigma_1^2}{\sigma_2^2} \quad (\text{Ec.1})$$

Ya con el estadístico de contraste definido se toma la decisión de la manera siguiente: si F de Snedecor es menor que el estadístico F de la tabla, se acepta la hipótesis nula, concluyendo que existe homocedasticidad de varianzas; mientras que, por el contrario, si el estadístico de contraste es mayor que el de la tabla se rechaza la hipótesis nula y se concluye la no existencia de homocedasticidad.

- Linealidad

El supuesto de linealidad, acorde a lo que expresan Hair et al. (1999) está implícito en cualquier análisis de regresión, ya que el mismo depende directamente de la correlación que exista entre las variables analizadas; sin embargo, en esta relación no están representados los efectos no lineales. Derivado de lo anterior, es necesario investigar todas estas relaciones para evaluar cualquier impacto que pueda afectar la correlación. Lo que se busca es determinar si los datos se ajustan a una recta de regresión, ya que de lo contrario el modelo obtenido por medio de este análisis no representa la estructura de los mismos.

Para verificar el cumplimiento de este supuesto, se debe evaluar que la correlación entre las variables sea significativa y por ende, demuestre que existe relación entre las mismas, de no ser así, la correlación se deberá a la casualidad, según lo expresado por Lind, Marchal y Wathen (2012).

Derivado de lo anterior se debe realizar una prueba de significancia al coeficiente de correlación, la cual parte de las siguientes hipótesis:

Ho: La correlación entre las variables es igual a cero.

Ha: La correlación entre las variables es diferente de cero.

El análisis que se utiliza para establecer la significancia se realiza por medio de la prueba T de Student, aplicándola al coeficiente de correlación, y para la cual se calcula el estadístico de contraste con la siguiente fórmula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (\text{Ec.2})$$

La regla de decisión en este caso se debe tomar basada en el valor calculado de t, ya que, si este es menor que t teórica se acepta la hipótesis nula, y por lo tanto no existe correlación entre las variables; mientras que, si el valor calculado es mayor que el de la tabla, se rechaza la hipótesis nula, concluyendo que existe significancia en el coeficiente de correlación y, por lo tanto, el supuesto de linealidad se cumple.

- Homogeneidad

Gómez, Aparicio y Patiño (2010) especifican que este supuesto se presenta cuando existen cambios sensibles en la media o la varianza de los datos, provocando que se pierda la homogeneidad y se presente la inconsistencia de los mismos.

Este supuesto descansa sobre la hipótesis de que el valor promedio o esperado de los errores debe ser igual o aproximadamente cero.

Para validar el supuesto hay que utilizar la prueba estadística T de Student, a partir de las siguientes hipótesis:

Ho: Los datos siguen una distribución homogénea.

Ha: Los datos no siguen una distribución homogénea.

Es por ello que, para realizar la prueba hay que calcular el estadístico t de prueba, de acuerdo a la siguiente fórmula:

$$t = \frac{\mu_1 - \mu_2}{\left[\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{1/2}} \quad (\text{Ec.3})$$

Para ello se debe comparar el estadístico t calculado con el valor de la distribución T de Student, utilizando la regla de decisión de que si t calculado es menor que el t de la distribución, se acepta la hipótesis nula; existiendo por lo tanto, la homogeneidad; mientras que si el estadístico de prueba es mayor que el de la distribución, se rechaza la hipótesis nula, ya que la diferencia entre las media evidencia inconsistencia y, por lo tanto se concluye la no existencia de homogeneidad.

- Independencia

Por último, se busca que se cumpla el supuesto de independencia de los datos, es decir, que cada una de las observaciones no se vea reflejada en otra observación distinta, según lo comentan Gómez et al. (2010).

González, P. (2009) explica que existe una prueba gráfica de dispersión entre los valores de los residuos y de los predichos (valores esperados) y en la cual se espera no se muestre ningún patrón, de no ser así; existiría la posibilidad de que se estuviera omitiendo alguna variable dependiente dentro del modelo de regresión considerado.

Por su parte, Díaz (2006) expresa que la independencia de los datos se debe en gran parte a la aleatorización de los mismos, es por ello que la dependencia de los residuos se caracteriza por tener secuencias largas de valores positivos o negativos; o bien la misma se puede dar por la correlación en el tiempo. Derivado de lo anterior, y si existe sospecha de correlación o de tener una serie dependiente, lo más recomendable es aplicar una prueba de rachas, cuyo resultado permita determinar si se cuenta con una serie dependiente o independiente.

En lo que corresponde al apartado de series de tiempo, se analizaron las siguientes propiedades para el conjunto de datos.

- Estacionalidad

Está definida como un factor que se repite o es similar en intervalos de tiempo similares, que por lo general, no puede presentarse en periodos mayores a 12 meses, y para verificarla se hace uso de la prueba de raíz unitaria de Dickey-Fuller, según lo indica Antúnez (2011) partiendo de las siguientes hipótesis:

Ho: La serie de tiempo no es estacional

Ha: La serie de tiempo es estacional

Concluyendo la estacionalidad de la serie, una vez el estadístico de Dickey-Fuller sea mayor que el valor crítico o teórico, en ambos casos utilizando su valor absoluto.

- Tendencia

La serie de datos puede tener una tendencia lineal o no lineal, y representa el comportamiento predominante de la serie de datos. Según González (2009) es el cambio de la media durante un largo periodo de tiempo.

- Aleatoriedad

Es la componente de una serie que es impredecible, ya que según González (2009) son movimientos erráticos que no tienen un comportamiento específico, obedeciendo a diversas razones.

- Ajuste

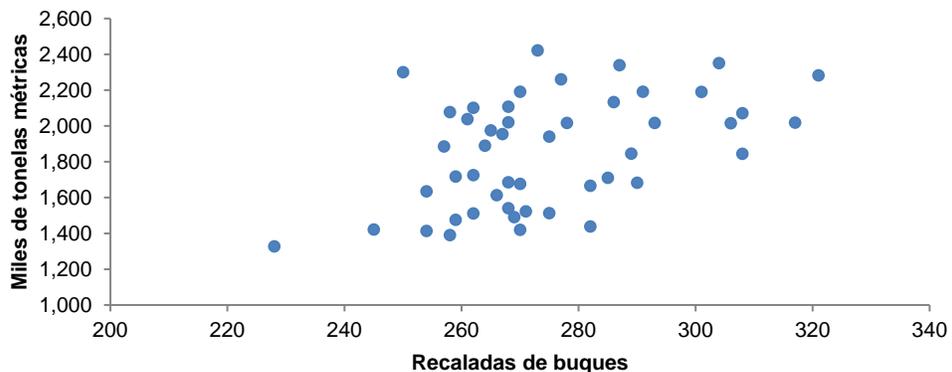
Según Gómez et al. (2010), los datos de la muestra deben ajustarse a alguna función de distribución de probabilidad. Esto consiste en la estimación de los parámetros de la función de distribución de probabilidad que le corresponda.

2.2.2. Modelos de regresión

2.2.2.1. Análisis de correlación

Según Lind et al. (2012), al estudiar la relación existente entre dos variables se debe iniciar con un análisis visual, mediante un diagrama de dispersión, el cual permitirá observar gráficamente el comportamiento y la relación entre ambas, tal y como se ejemplifica en la figura 2.

Figura 2. Diagrama de dispersión toneladas - recaladas 2012 a 2015



Fuente: elaboración propia, con información proporcionada por Comisión Portuaria Nacional.

Luego de realizar el diagrama de dispersión, se efectúa el cálculo matemático del coeficiente de correlación según la Ec.1 y el cual según Navidi (2006): “es una medida numérica de la fuerza de la relación lineal entre dos variables” (p. 476).

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_x S_y} \quad (\text{Ec.4})$$

Donde:

X_i y Y_i corresponden a cada uno de los valores de las observaciones, \bar{X} y \bar{Y} son los valores correspondientes al promedio de cada serie de datos, S_x y S_y es la desviación estándar de cada serie de datos y n es el número de observaciones.

El coeficiente de correlación posee las siguientes características: se denota por "r", tiene signo (+ ó -) mediante el cual se determina su dirección y su valor debe estar comprendido entre -1 y 1. Mientras más cercano a estos valores esté, mayor es la fuerza de asociación de las variables, por lo que cuando su valor se acerque a 0, la fuerza de asociación disminuye.

La interpretación de este coeficiente es de suma importancia en los análisis que se realicen, ya que, si bien indica el grado de asociación entre variables, el mismo no indica que un cambio en una variable obedece directamente a un cambio en la otra, por lo que se debe tener suma precaución en qué tipo de variables son las que se relacionan y así determinar si la correlación entre las mismas se debe o no a la casualidad, mediante la prueba de significancia del coeficiente r.

2.2.2.2. Análisis de regresión

En la sección anterior se definieron los análisis que se deben realizar para determinar el grado de asociación entre dos variables, tanto a nivel gráfico como a nivel matemático. En esta sección se realizará una descripción de lo que es un análisis de regresión, el cual Lind et al. (2012) definen como la elaboración de una ecuación para expresar la relación que existe entre dos variables; así como la utilización de la misma para estimar valores futuros.

Según Walpole, Myers, Myers y Ye (2012), en todo modelo de regresión existe la variable dependiente, explicada o de respuesta, que es el valor que se busca estimar, y está denotada por la letra “y”, mientras que la variable o variables independientes, explicativas o regresoras son los valores que ayudarán a explicar el comportamiento de “y”, por lo que están denotadas por letra “x”.

La ecuación general de un modelo de regresión lineal simple, según Webster (2000) es de la siguiente forma:

$$Y = b_0 + b_1X \quad (\text{Ec.5})$$

En donde b_0 es el intercepto y b_1 es la pendiente de la recta. La relación existente entre las variables Y y X puede ser determinística o estocástica. La diferencia de las mismas se da debido a que una relación determinística es una relación exacta, en donde no existe el error; mientras que una relación estocástica está determinada por la aleatoriedad, por lo que seguramente a la hora de realizar una predicción habrá un error. Derivado de lo anterior se expresará el modelo estocástico de la siguiente manera:

$$Y = \beta_0 + \beta_1X + \varepsilon \quad (\text{Ec.6})$$

La fórmula anterior expresa el valor real o verdadero de Y cuando se hace una regresión de la misma en la base de X, mientras que $\beta_0 + \beta_1X$ es la parte determinística de la relación y la letra griega ε representa el factor aleatorio que muestra la variable dependiente, y por tanto, es el término del error que interviene en la función. La mayoría de los valores de este modelo permanecerán desconocidos, por lo que la única forma de estimarlos es mediante la función muestral, tal y como se muestra a continuación:

$$Y = b_0 + b_1X + e \quad (\text{Ec.7})$$

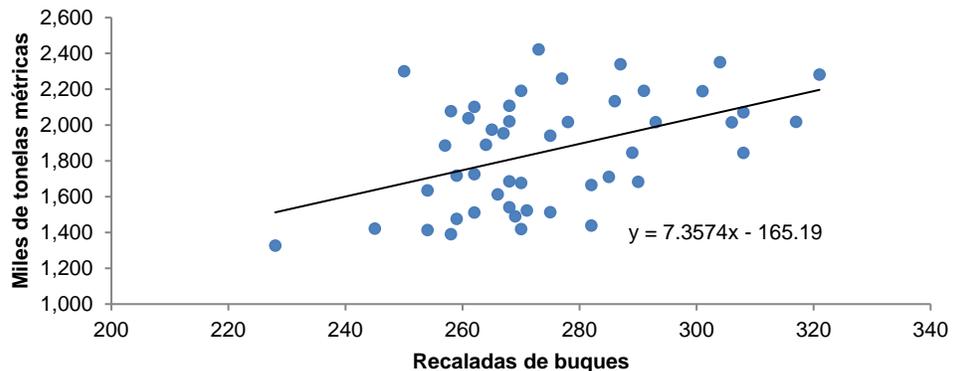
Donde b_0 y b_1 son variables estimadoras de β_0 y β_1 ; y e constituye el término aleatorio. Si en determinado caso se conociera el valor exacto del término aleatorio se podría conocer el valor de Y ; sin embargo, como el mismo solo puede estimarse, el modelo de regresión toma la siguiente forma:

$$\hat{Y} = b_0 + b_1X \quad (\text{Ec.8})$$

Donde \hat{Y} es el valor estimado de la serie con base en el modelo de regresión y; b_0 y b_1 continúan siendo el intercepto y la pendiente de la recta de regresión que se ha estimado.

De acuerdo con Webster (2000) se dice que el propósito principal del análisis de regresión es encontrar la recta que mejor ajuste a los datos muestrales. Para encontrarla se utiliza el método matemático denominado de mínimos cuadrados ordinario, en donde el resultado obtenido será una recta que se ajuste a todos los valores de los datos mejor que cualquier otra. Regresando a la figura 2, se puede observar el diagrama de dispersión de los datos, por lo que la figura 3 ilustra los mismos con la adición de la recta de regresión.

Figura 3. **Recta de regresión para el modelo toneladas - recaladas**



Fuente: elaboración propia, con información proporcionada por Comisión Portuaria Nacional.

2.2.2.3. **Series de tiempo**

Una serie de tiempo, de acuerdo con Lind et al. (2012), es un conjunto de datos registrados de forma cronológica, pudiendo ser los mismos medidos de forma anual, mensual, semanal, diaria u otra.

Las series de tiempo pueden aplicarse en la mayoría de los campos de las ciencias, por ejemplo, en economía se puede mencionar el producto interno bruto trimestral, las exportaciones anuales, los precios de la canasta básica, entre otras. En meteorología se pueden encontrar series de lluvia anual, velocidad mensual del viento, temperaturas diarias, por mencionar algunas.

González (2009) indica que en el análisis de series de tiempo se deben tener en cuenta que las mismas tienen un orden cronológico, el cual es de suma importancia, y la independencia entre las observaciones no existe, ya que muchas veces las observaciones son dependientes.

Derivado de las características de las series temporales, hay que desarrollar modelos en los cuales se pueda aprovechar la dependencia entre las observaciones y permitan proyectar el comportamiento de las mismas en un futuro próximo. Al conjunto de técnicas que permiten obtener estos modelos se le denomina análisis de series temporales.

Antes de entrar en este análisis resulta necesario tener claro lo que es un proceso estocástico. Novales (1993) lo define como “una sucesión de variables aleatorias” (p. 414); por lo que las mismas pueden tomar cualquier valor en el tiempo, oscilando sus valores dentro del rango $(-\infty; \infty)$. Las observaciones de una serie de tiempo se denotan por $Y_1, Y_2, Y_3, Y_4 \dots Y_t$, donde Y_t es el valor que tiene la observación Y en el momento t .

Para fines de este trabajo de investigación se busca que el proceso estocástico sea estacionario débil o de segundo orden, el cual según Gujarati y Porter (2010) se presenta cuando la media y la varianza de las variables son constantes en el tiempo, añadiéndole que su covarianza solo es función del rezago. Según la literatura, la estacionalidad definida de manera rigurosa implica el cumplimiento de varias condicionantes, razón por la cual, en este trabajo al analizar un proceso estocástico estacionario, se entenderá por uno con estacionalidad débil o de segundo orden (Novales, 1993).

2.2.2.4. Modelos autoregresivos

Para modelar una serie de tiempo estacionaria existen diversos métodos, de los cuales en este trabajo de investigación se aplicaron los procesos autoregresivos (AR), en donde según Hurtado (2008): “los procesos se presentan como la suma ponderada de observaciones pasadas de la variable” (p. 17).

Gujarati y Porter (2010) explican que un proceso autoregresivo de primer orden o AR (1) es aquel en donde el valor de Y en la observación t depende del valor que haya tomado en el periodo anterior, es decir t-1, y de un valor aleatorio o ruido blanco, el cual debe tener media igual a cero ($\mu = 0$), varianza igual a 1 ($\sigma^2 = 1$) y deben ser independientes en el tiempo, según lo explica Novales (1993). Y_t en un proceso AR (1) se modela de la siguiente manera:

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \varepsilon_t \quad (\text{Ec.9})$$

Donde δ es la media de Y, α es una constante o alguna proporción del valor en el periodo anterior (t-1) y ε_t es el ruido blanco.

Ahora bien, si el proceso se modela de acuerdo con la expresión

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + \varepsilon_t \quad (\text{Ec.10})$$

El proceso autoregresivo es considerado de segundo orden o AR (2), ya que el valor de Y en el tiempo depende de los valores que haya tomado la observación en los dos periodos predecesores.

De manera general, un proceso autoregresivo de orden p, sigue el siguiente modelo:

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + \dots + \alpha_p (Y_{t-p} - \delta) + \varepsilon_t \quad (\text{Ec.11})$$

Una característica a tomar en cuenta en los modelos presentados con anterioridad, es que en ellos solo se incluyen observaciones pasadas y actuales de Y, es decir que no hay otras variables regresoras. Por lo que Gujarati y Porter (2010) afirman que “los datos hablan por sí mismos” (p. 776).

Al contar con una serie de tiempo no estacional, es necesario diferenciar “d” veces la variable Y_t para convertirla en una serie estacional, y que de esta manera se comporte como un proceso autoregresivo (AR) con una media móvil a lo largo del tiempo, obteniendo un proceso flexible y que puedan representar un gran número de series temporales. Lo anterior es conocido como un proceso ARIMA, según Antúnez (2011), el cual está definido por el siguiente modelo:

$$\Phi(L)\Delta^d Y_t = \delta + \theta(L)\varepsilon_t \text{ (Ec.12)}$$

3. PRESENTACIÓN DE RESULTADOS

3.1. Selección de las variables

3.1.1. Variables a incluir en el modelo de regresión lineal

Para la construcción de los modelos de regresión lineal de importación y exportación de contenedores en puerto Santo Tomás de Castilla, se utilizaron distintas variables, cuyo comportamiento se muestra en las tablas II y III.

Tabla II. **Variables a considerar en el modelo de importación de contenedores**

Año	Contenedores importación	Importación de bienes y servicios	Gastos de consumo final	Población
2001	51395	60756.37	137380.88	11899297
2002	57812	64568.69	150936.38	12190061
2003	55205	69941.34	163795.91	12484289
2004	57493	80180.25	179547.97	12781131
2005	53976	85156.34	199842.30	13080615
2006	60883	96270.20	220942.81	13381675
2007	73253	110811.52	251124.19	13684312
2008	91062	116593.99	290802.07	13988991
2009	89130	102031.16	295962.80	14296121
2010	105493	120943.49	321654.72	14605144
2011	119790	138605.36	354331.41	14916981
2012	119620	142549.42	380080.86	15231328
2013	118659	147359.20	410829.30	15548115
2014	123397	152377.82	439046.51	15867562
2015	126530	147150.23	465482.38	16190084
2016	138005	142642.65	495785.69	16515320

Fuente: elaboración propia.

Tabla III. **Variables a considerar en el modelo de exportación de contenedores**

Año	Contenedores exportación	Exportación de bienes y servicios	Ingreso nacional bruto per cápita
2001	54035	41458.71	1600
2002	54265	42737.57	1620
2003	48304	44857.56	1690
2004	50637	51372.16	1860
2005	51990	52045.83	2010
2006	57719	57302.11	2160
2007	68859	66919.66	2370
2008	87330	73134.26	2570
2009	89606	73836.25	2610
2010	103228	85957.25	2700
2011	117556	98783.40	2850
2012	115020	98162.47	3090
2013	117235	100357.47	3310
2014	123005	105151.44	3450
2015	119796	103290.11	3610
2016	131844	102045.56	3790

Fuente: elaboración propia.

Se elaboró el modelo de regresión correspondiente a la importación de contenedores, con las variables descritas anteriormente, obteniendo los resultados que se muestran en las tablas IV y V.

Tabla IV. **Análisis de regresión lineal para contenedores importación**

Variable	N	r²	r
Contenedores importación	16	0.96	0.98

Fuente: elaboración propia.

Tabla V. **Coefficientes de regresión y estadísticos asociados modelo contenedores importación**

Coefficiente	Estadístico	Error estándar	T	T teórico	p-valor
Constante	209440.49	166330.88	1.26		0.2319
Importaciones de bienes y servicios	0.32	0.22	1.47	2.09	0.1676
Gastos de consumo final	0.42	0.18	2.30	2.09	0.0404
Población	-0.02	0.02	-1.23	2.09	0.2439

Fuente: elaboración propia.

En la tabla V se verificó la significancia que cada una de las variables tiene dentro del modelo para las importaciones de contenedores, mediante la prueba de T de Student. Derivado de los resultados obtenidos se extrajeron las variables importaciones de bienes y servicios y población del mismo, por lo que se utilizó la variable gastos de consumo final para elaborar el modelo parsimonioso de importación de contenedores, del cual se presentan los resultados en las tablas VI, VII y VIII.

Tabla VI. **Análisis de regresión lineal para contenedores importación (modelo parsimonioso)**

Variable	N	r²	r
Contenedores importación	16	0.95	0.97

Fuente: elaboración propia.

Tabla VII. **Coefficientes de regresión y estadísticos asociados modelo contenedores importación (modelo parsimonioso)**

Coefficiente	Estadístico	Error estándar	T	T teórico	p-valor
Constante	12676.01	5039.25	2.52		0.0247
Gastos de consumo final	0.26	0.02	16.45	1.96	<0.0001

Fuente: elaboración propia.

Tabla VIII. **Análisis de varianza**

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F	F Teórico	p-valor
Modelo	14074160830.90	1	14074160830.90	270.74	4.60	<0.0001
Gastos de consumo final	14074160830.90	1	14074160830.90	270.74		<0.0001
Error	727783731.04	14	51984552.22			
Total	14801944561.94	15				

Fuente: elaboración propia.

Al obtener un modelo válido para los contenedores de importación, en lo referente a sus variables, se procedió a realizar el respectivo procedimiento para el caso de los contenedores de exportación, obteniendo los resultados que se muestran en las tablas IX y X.

Tabla IX. **Análisis de regresión lineal para contenedores exportación**

Variable	N	r²	r
Contenedores exportación	16	0.97	0.98

Fuente: elaboración propia.

Tabla X. **Coefficientes de regresión y estadísticos asociados modelo contenedores exportación**

Coefficiente	Estadístico	Error estándar	T	T teórico	p-valor
Constante	-9824.01	6292.56	-1.56		0.1425
Exportaciones de bienes y servicios	1.13	0.26	4.29	2.04	0.0009
Ingreso nacional bruto per cápita	4.85	8.77	0.55	2.04	0.5894

Fuente: elaboración propia.

En la tabla X se verifica la significancia que cada una de las variables tiene dentro del modelo para las exportaciones de contenedores, mediante la prueba T de Student. Derivado de los resultados obtenidos se extrajo la variable ingreso nacional bruto per cápita del mismo, por lo que se utilizó la variable exportaciones de bienes y servicios para elaborar el modelo parsimonioso de exportación de contenedores, del cual se presentan los resultados obtenidos en las tablas XI, XII y XIII.

Tabla XI. **Análisis de regresión lineal para contenedores exportación (modelo parsimonioso)**

Variable	N	r²	r
Contenedores exportación	16	0.97	0.98

Fuente: elaboración propia.

Tabla XII. **Coefficientes de regresión y estadísticos asociados modelo contenedores exportación (modelo parsimonioso)**

Coefficiente	Estadístico	Error estándar	T	T teórico	p-valor
Constante	-7809.25	5003.80	-1.56		0.1409
Exportaciones de bienes y servicios	1.27	0.06	19.84	1.96	<0.0001

Fuente: elaboración propia.

Tabla XIII. **Análisis de varianza**

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F	F Teórico	p-valor
Modelo	14172017247.41	1	14172017247.41	393.64	4.60	<0.0001
Exportaciones de bienes y servicios	14172017247.41	1	14172017247.41	393.64		<0.0001
Error	504035935.03	14	36002566.79			
Total	14676053182.44	15				

Fuente: elaboración propia.

Al obtener un modelo válido para los contenedores de exportación, en lo referente a sus variables, se procedió a realizar la comprobación de los supuestos que deben cumplir los modelos de regresión para ser válidos, obteniendo los resultados que se muestran en la siguiente sección.

3.1.2. Variables a incluir en el modelo generado por series de tiempo

Para la realización de un modelo de proyección mediante el análisis de series temporales se utilizaron la cantidad de contenedores de importación y exportación en puerto Santo Tomás de Castilla a partir del año 2000 hasta el 2016 con periodicidad mensual, por lo cual se tienen datos de 17 años esto hace un total de 204 registros, y se pueden encontrar en el apéndice 1.

3.2. Comprobación de los supuestos y propiedades del modelo

3.2.1. Supuestos modelo de regresión lineal

3.2.1.1. Linealidad

Para validar el supuesto de linealidad de los modelos de importación y exportación de contenedores se utilizó la prueba de significancia del coeficiente r . Los resultados obtenidos se muestran en la tabla XIV.

Tabla XIV. Prueba de significancia para el coeficiente r

Variable	Modelo importación	Modelo exportación
$n =$	16	16
Significancia =	0.05	0.05
Grados de libertad =	14	14
$r^2 =$	0.95	0.97
$r =$	0.97	0.98
T teórico =	2.14	2.14
T calculado =	16.31	21.28

Fuente: elaboración propia.

3.2.1.2. Normalidad

La normalidad de los residuos de los modelos, tanto de importación como de exportación de contenedores, se validó mediante la prueba de Shapiro-Wilks, los resultados de la misma se aprecian en la tabla XV.

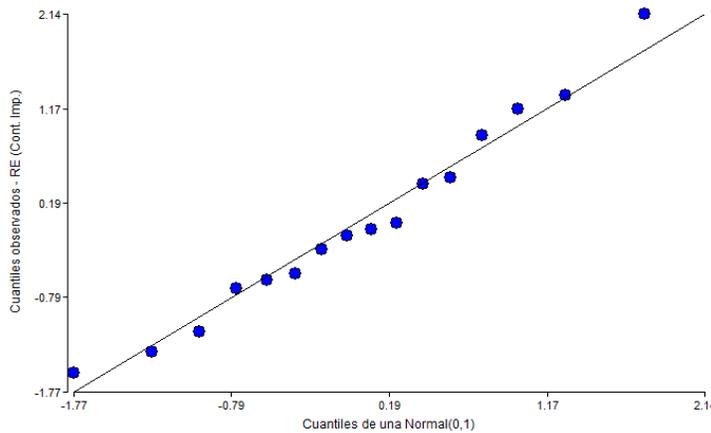
Tabla XV. **Prueba de Shapiro-Wilks para normalidad**

Variable	Modelo importación	Modelo exportación
n =	16	16
Media =	0.00	0.00
Significancia =	0.05	0.05
p valor =	0.8190	0.2662

Fuente: elaboración propia.

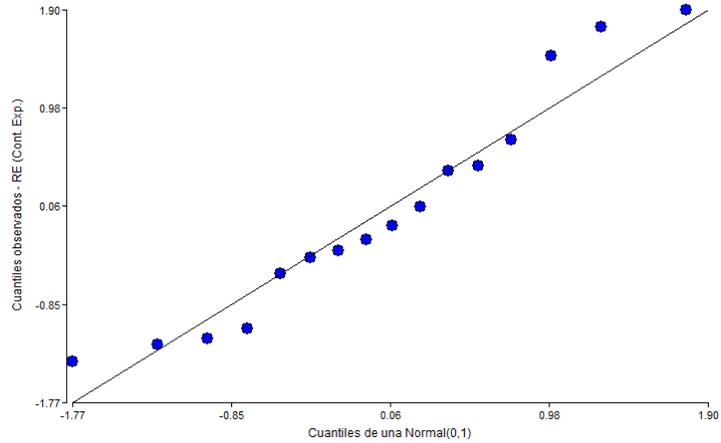
Otra prueba utilizada para comprobar la normalidad de los residuos en ambos modelos, fue mediante las gráficas Q-Q plot, tal y como se muestra en las figuras 4 y 5.

Figura 4. **Q-Q plot residuos modelo importación contenedores**



Fuente: elaboración propia.

Figura 5. **Q-Q plot residuos modelo exportación contenedores**



Fuente: elaboración propia.

3.2.1.3. Independencia

Para la validación del supuesto de independencia o aleatoriedad en los residuos de los modelos de importación y exportación de contenedores, se hizo uso de la prueba no paramétrica de rachas, ambos resultados se muestran en la tabla XVI.

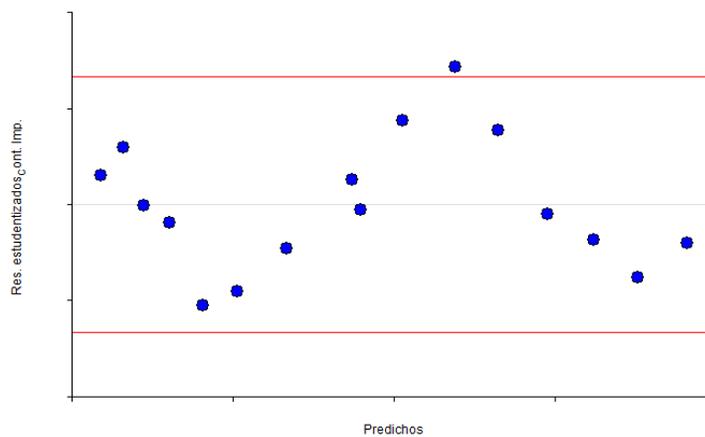
Tabla XVI. **Prueba de rachas para aleatoriedad**

Variable	Modelo importación	Modelo exportación
n =	16	16
Signos "+" =	6	7
Signos "-" =	10	9
Rachas =	6	5
M1 (teórico) =	4	4
M2 (teórico) =	13	14
Significancia =	0.05	0.05

Fuente: elaboración propia.

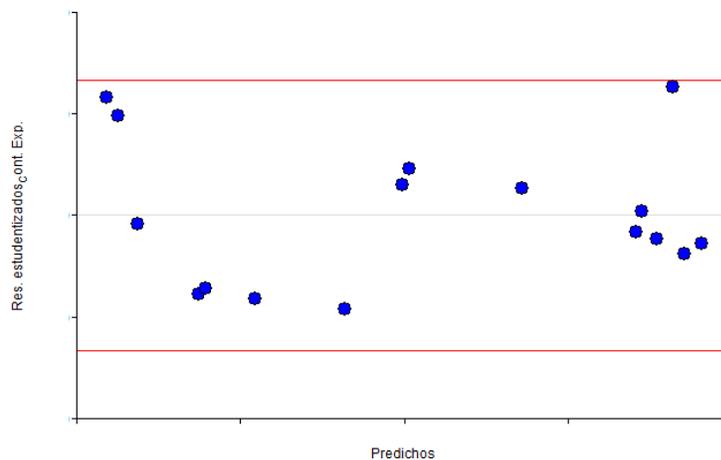
Otra prueba utilizada para comprobar la aleatoriedad de los residuos, en ambos modelos, fue mediante las gráficas denominadas correlogramas, tal y como se muestra en las figuras 6 y 7.

Figura 6. **Correlograma residuos y predichos modelo importación contenedores**



Fuente: elaboración propia.

Figura 7. **Correlograma residuos y predichos modelo exportación contenedores**



Fuente: elaboración propia.

3.2.1.4. Homogeneidad

El supuesto de homogeneidad de los residuos en los modelos de importación y exportación de contenedores fue validado mediante la prueba T de Student, los resultados de la misma se muestran en la tabla XVII.

Tabla XVII. Prueba T de Student para homogeneidad

Variable	Modelo importación	Modelo exportación
n =	16	16
n1 (n/2) =	8	8
n2 (n/2) =	8	8
Media 1 =	-1938.87	-1044.12
Varianza 1 =	35571121	49037262
Media 2 =	1938.87	1044.12
Varianza 2 =	59805507	20476028
Significancia =	0.05	0.05
Grados de libertad =	14	14
T teórico =	2.14	2.14
T calculado =	0.00014	0.00011

Fuente: elaboración propia.

3.2.1.5. Homocedasticidad

La validación del supuesto de homocedasticidad de los residuos se realizó mediante la prueba F de Snedecor, tanto para el modelo de importación como el de exportación de los contenedores. Los resultados de la prueba se encuentran en la tabla XVIII.

Tabla XVIII. Prueba F de Snedecor para homocedasticidad

Variable	Modelo importación	Modelo exportación
n =	16	16
n1 =	8	8
n2 =	8	8
Varianza 1 =	35571121	49037262
Varianza 2 =	59805507	20476028
Significancia =	0.05	0.05
Grados de libertad =	14	14
F teórico =	2.79	2.79
F calculado =	0.59	2.39

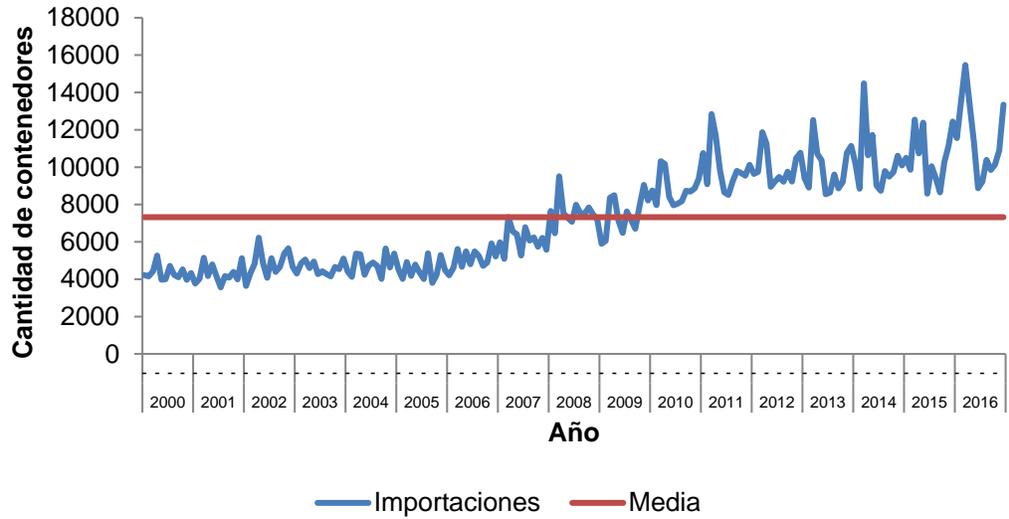
Fuente: elaboración propia.

3.2.2. Propiedades de las series de tiempo

3.2.2.1. Tendencia y aleatoriedad

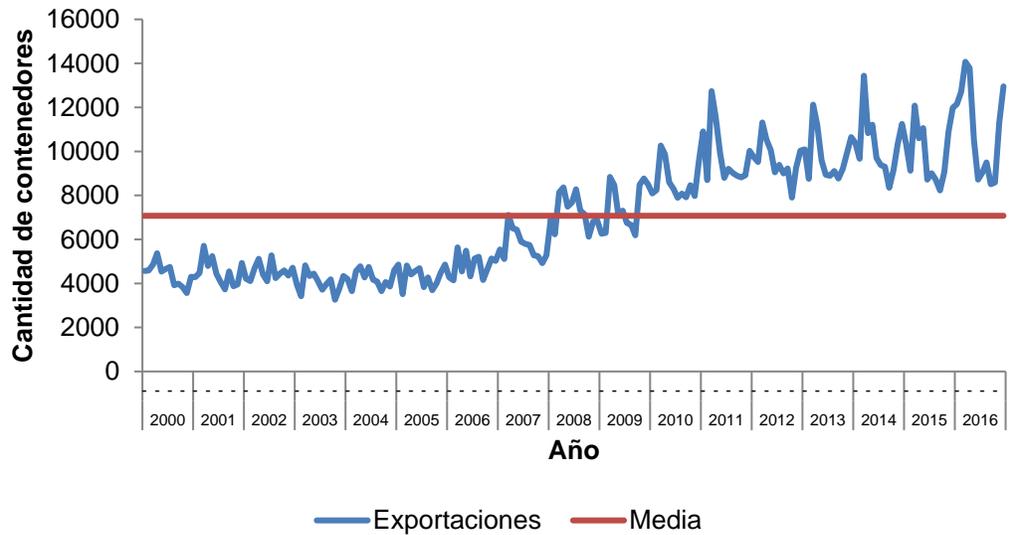
Con los datos mensuales de la cantidad de contenedores de importación y exportación se procedió a la generación de figuras para visualizar el comportamiento de las series y establecer si existe algún tipo de tendencia en las mismas, tal y como se muestra en las figuras 8 y 9.

Figura 8. **Serie de tiempo importación de contenedores**



Fuente: elaboración propia.

Figura 9. **Serie de tiempo exportación de contenedores**



Fuente: elaboración propia.

3.2.2.2. Estacionalidad

Para comprobar si las series de importación y exportación de contenedores, son estacionales, se hizo uso de la prueba de raíz unitaria de Dickey-Fuller, para la cual se obtuvieron los resultados presentados en la tabla XIX.

Tabla XIX. Prueba de raíz unitaria de Dickey-Fuller

Variable	Serie importación	Serie exportación
n =	204	204
Significancia =	0.05	0.05
Estadístico Dickey-Fuller =	0.1421	0.4530
Valor critico =	-2.8766	-2.8765
p valor =	0.9681	0.9847

Fuente: elaboración propia.

3.3. Construcción del modelo

3.3.1. Validación modelo de regresión lineal

Una vez se cumplen todos los supuestos de un modelo de regresión, queda realizar la validación de los mismos. Esto se realizó mediante el contraste global de los modelos, haciendo uso de la prueba F, la cual al ser aplicada, produjo los resultados que se describen en la tabla XX.

Tabla XX. **Contraste global del modelo, prueba de F**

Variable	Modelo importación	Modelo exportación
Significancia =	0.05	0.05
Grados de libertad =	14	14
F teórico =	4.60	4.60
F calculado =	270.74	393.64
p valor =	<0.0001	<0.0001

Fuente: elaboración propia.

Derivado de los resultados presentados con anterioridad, se obtiene el siguiente modelo para la importación de contenedores:

$$Y = 0.26 X + 12676.01 \quad (\text{Ec.13})$$

Donde Y corresponde a la cantidad de contenedores de importación y X es la variable gastos de consumo final. El intercepto de la misma viene dado por el valor 12676.01 y el 0.26 que acompaña a la variable X, indica que por cada millón de quetzales que aumenten los gastos de consumo final, la cantidad de contenedores de importación aumentará en 0.26 unidades.

Para el caso de las exportaciones, el modelo que se obtiene es el siguiente:

$$Z = 1.27 W - 7809.25 \quad (\text{Ec.14})$$

Donde Z es la cantidad de contenedores de exportación y W es la variable exportaciones de bienes y servicios. El intercepto de la misma viene dado por el valor 7809.25 y el 1.27 que acompaña a la variable W , indica que por cada millón de quetzales que aumenten las exportaciones de bienes y servicios, la cantidad de contenedores de exportación aumentará en 1.27 unidades.

Existen otro tipo de variables económicas que se pueden utilizar para verificar si pueden explicar o no, el comportamiento de las variables dependientes, dentro de ellas se pueden mencionar el tipo de cambio real, los precios internacionales de los principales productos comercializados y los ingresos de los principales socios comerciales de Guatemala; sin embargo, no se incluyeron dentro del presente análisis derivado de dos razones: la primera es que no se pudieron encontrar series históricas de datos, las cuales pudieran ser utilizadas para correlacionarlas con las series que se presentan en el presente trabajo; y la segunda, es que al ser variables económicas seguramente el comportamiento de las mismas se vería reflejado en el comportamiento de las series de datos de las unidades de contenedores de importación y exportación.

3.3.2. Validación del modelo generado por series de tiempo

Derivado que se utilizaron modelos ARIMA para las series de datos de importación y exportación de contenedores, los mismos generan residuos (los cuales se pueden encontrar en el apéndice), que deben cumplir con la normalidad e independencia para que los modelos obtenidos sean válidos.

Para la verificación de la normalidad se hizo uso de la prueba de bondad de ajuste de Kolmogórov-Smirnov, ya que en ambos casos se tienen más de cincuenta datos. Los resultados de la prueba se presentan en la tabla XXI.

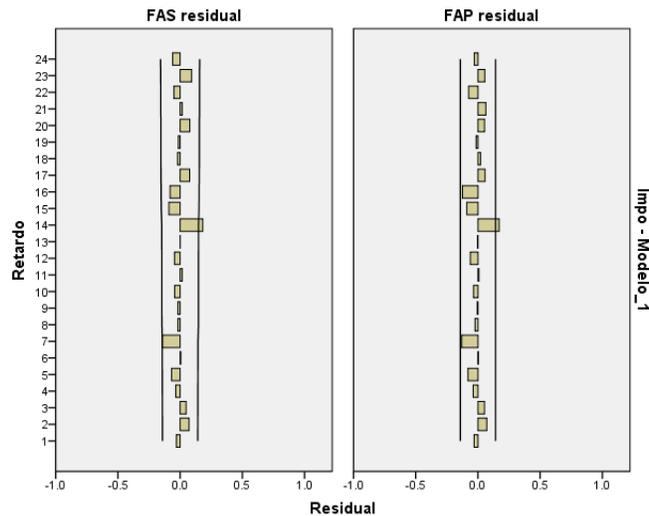
Tabla XXI. Prueba de bondad de ajuste Kolmogórov-Smirnov

Variable	Residuos importación	Residuos exportación
n =	191	191
Media =	28.54	45.64
Desviación estándar =	757.54	632.52
Significancia =	0.05	0.05
Estadístico D =	0.035	0.041
p valor =	0.20	0.20

Fuente: elaboración propia.

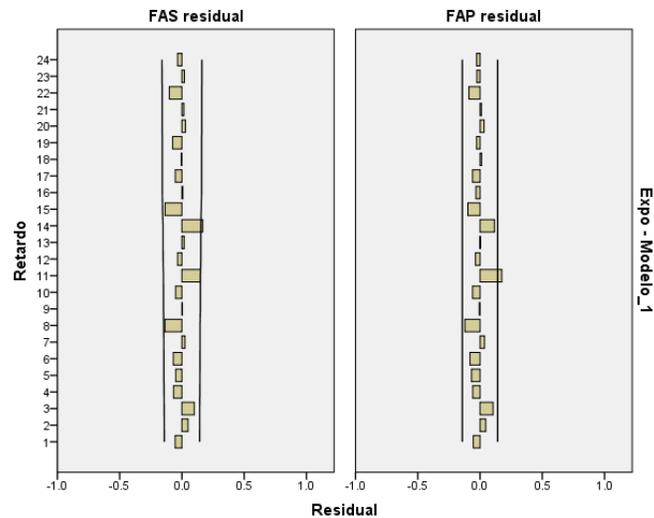
Para evaluar la independencia de los residuos para ambos modelos se realizaron los diagramas correspondientes a la función de auto correlación serial (FAS) y a la función de autocorrelación parcial (FAC), los cuales se muestran en las figuras 10 y 11.

Figura 10. FAS y FAC residuos modelo importación



Fuente: elaboración propia.

Figura 11. FAS y FAC residuos modelo exportación



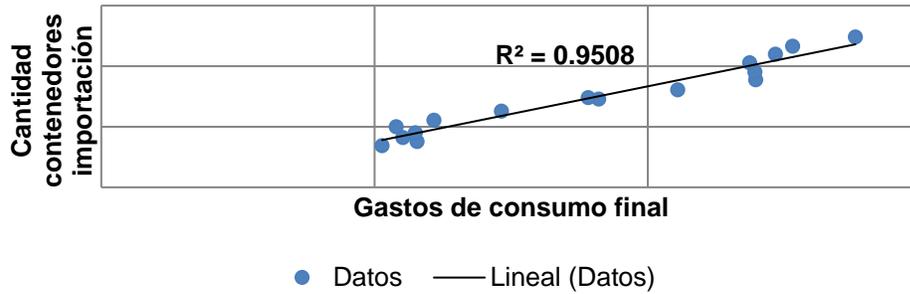
Fuente: elaboración propia.

3.4. Robustez del modelo

3.4.1. Ajuste del modelo de regresión lineal

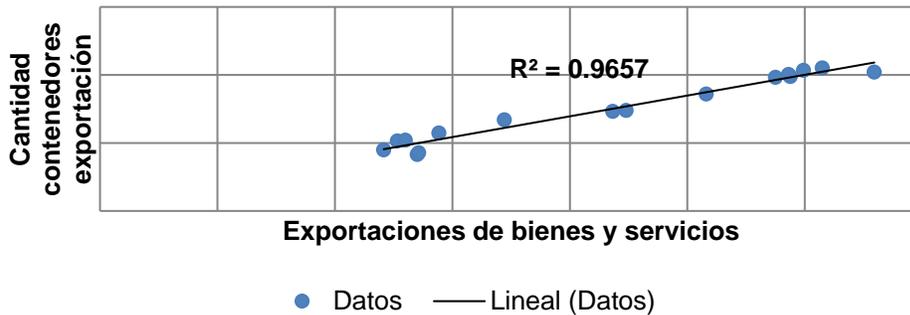
Para determinar el ajuste de los modelos de regresión para el caso de la importación y exportación de contenedores en puerto Santo Tomás de Castilla, se procedió a la realización de sus respectivos diagramas de dispersión, los cuales se presentan en las figuras 12 y 13.

Figura 12. **Diagrama de dispersión importación-gastos de consumo final**



Fuente: elaboración propia.

Figura 13. **Diagrama de dispersión exportación-exportaciones de bienes y servicios**

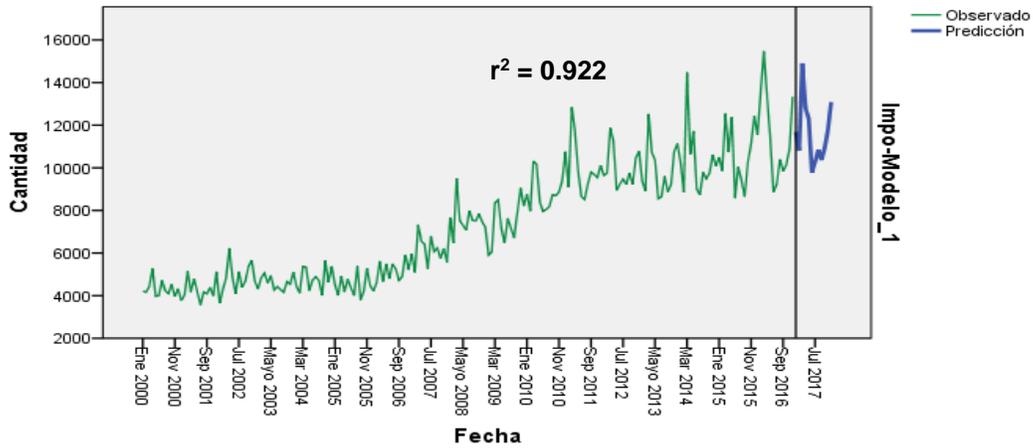


Fuente: elaboración propia.

3.4.2. Ajuste del modelo generado por series de tiempo

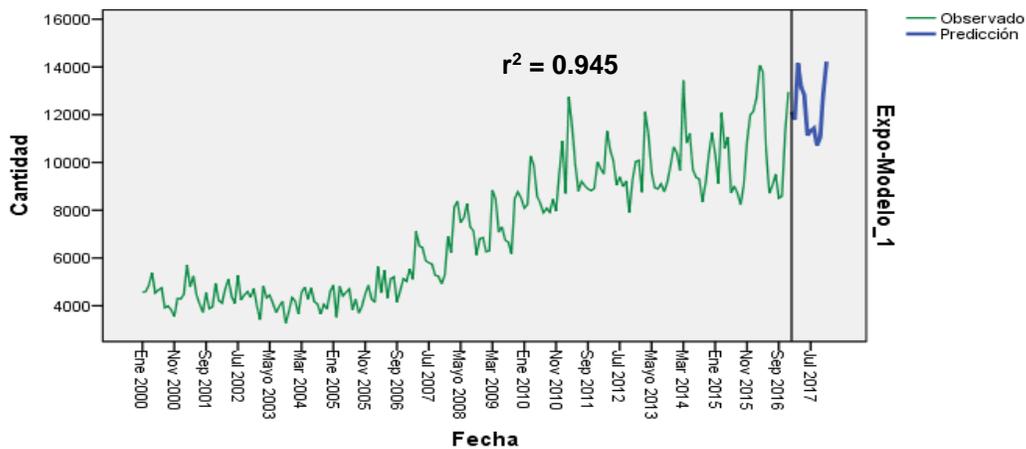
El ajuste de los modelos ARIMA aplicados a las series de tiempo de la cantidad de contenedores de importación y exportación de contenedores, se determinó por medio del gráfico de cada una de ellas, con su respectiva proyección para el año 2017, tal y como se muestra en las figuras 14 y 15.

Figura 14. Contenedores importación y proyección ARIMA 2017



Fuente: elaboración propia.

Figura 15. Contenedores exportación y proyección ARIMA 2017



Fuente: elaboración propia.

4. DISCUSIÓN DE RESULTADOS

4.1. Selección de las variables a incluir en el modelo

4.1.1. Variables a incluir en el modelo de regresión lineal

Como se muestra en las tablas I y II, se utilizaron distintas variables para elaborar los modelos de importación y exportación de contenedores.

Para el caso de la importación se seleccionaron dos variables macroeconómicas, provenientes del producto interno bruto del país y una variable demográfica, siendo ellas: la importación de bienes y servicios, los gastos de consumo final y la población. Con las variables definidas se realizó el respectivo análisis de regresión en donde se obtuvo un coeficiente de correlación r^2 de 0.96, lo cual se interpreta como que un 96 % de la variabilidad de los datos es debido al modelo de regresión lineal simple, considerándose un buen ajuste del modelo.

A su vez, para el modelo de exportación se seleccionaron dos variables macroeconómicas, siendo ellas el ingreso nacional bruto per cápita y la exportación de bienes y servicios. Con estas variables se realizó el análisis de regresión, obteniendo un coeficiente de correlación r^2 de 0.97, es decir, que el 97 % de la variabilidad de los datos se debe al modelo de regresión lineal simple, proporcionando un buen ajuste al mismo.

Para comprobar si las variables introducidas al modelo tienen significancia estadística, se realiza la prueba de contraste T de Student. Partiendo esta de la hipótesis nula en donde el valor de las variables es igual a cero, es decir que no tiene significancia estadística dentro del modelo; contra la hipótesis alternativa que señala que el valor de las mismas es distinto a cero y por ende la significancia estadística existe.

Lo anterior se verifica comparando el valor T de Student teórico contra el valor T de la variable en el modelo, aceptando la hipótesis nula si este último es menor que T teórico y rechazándola si es mayor.

Según los resultados que se muestran en la tabla IV, se determinó que las variables población e importaciones de bienes y servicios, no son significativas para el modelo de importación, ya que, según la regla de decisión descrita con anterioridad, se debe aceptar la hipótesis nula, ya que el T calculado es menor que el T teórico. Por su parte, para la variable gastos de consumo final se decide rechazar la hipótesis nula, por lo que la misma se incluyó en el modelo de importación de contenedores.

Para el caso del modelo de exportación, se obtiene que la variable ingreso nacional bruto per cápita no es significativa, ya que se acepta la hipótesis nula derivado que el T calculado es menor que el T teórico, según se puede observar en la tabla IX. Continuando con el análisis, se decide rechazar la hipótesis nula para la variable exportaciones de bienes y servicios, razón por la cual es incluida en el modelo de exportación de contenedores.

Es importante resaltar que las variables significativas, en cada uno de los modelos, son parte del producto interno bruto nacional, es decir que las mismas son utilizadas para su cálculo, lo cual es muy similar a lo descrito en los antecedentes de este informe; sin embargo, hay que tomar en cuenta que, en los casos mencionados en esa sección, los investigadores utilizan el valor total del PIB como variable a introducir en el modelo de regresión lineal.

Una vez definidas las variables a incluir en los modelos, se realizó nuevamente el análisis de regresión, obteniendo un coeficiente de correlación r^2 de 0.95 y 0.97, para el modelo de importación y exportación respectivamente, lo cual se interpreta como que un 95 % y un 97 % de la variabilidad de los datos es debido a los modelos de regresión lineal simple, considerándose como un buen ajuste de los mismos.

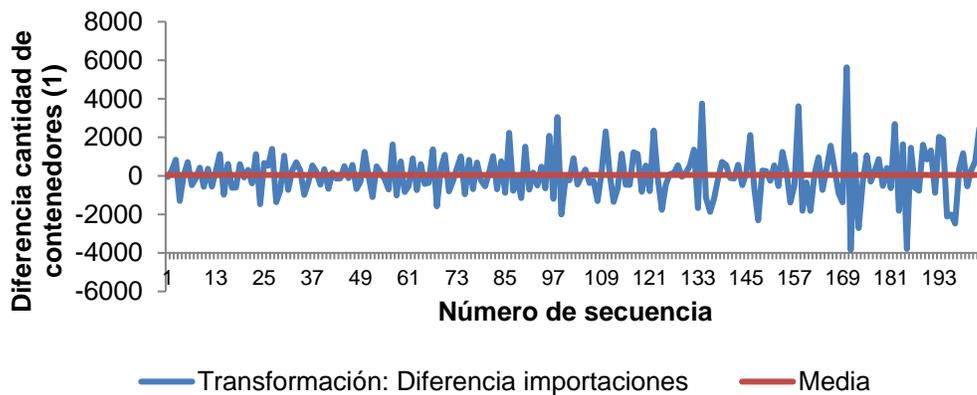
De igual forma que con los análisis anteriores, se verificó la significancia de las variables a introducir en ambos modelos mediante la prueba T de Student, en donde se aceptó la hipótesis alternativa, según se observa en las tablas VI y XI, razón por la cual ambas variables tienen una significancia estadística en los modelos.

4.1.2. Variables a incluir en el modelo generado por series de tiempo

Como se ha mencionado, para generar un modelo por medio de series temporales para la cantidad de contenedores de importación y exportación en puerto Santo Tomás de Castilla, se utilizaron únicamente los datos mensuales de contenedores movilizados desde el año 2000 hasta el 2016.

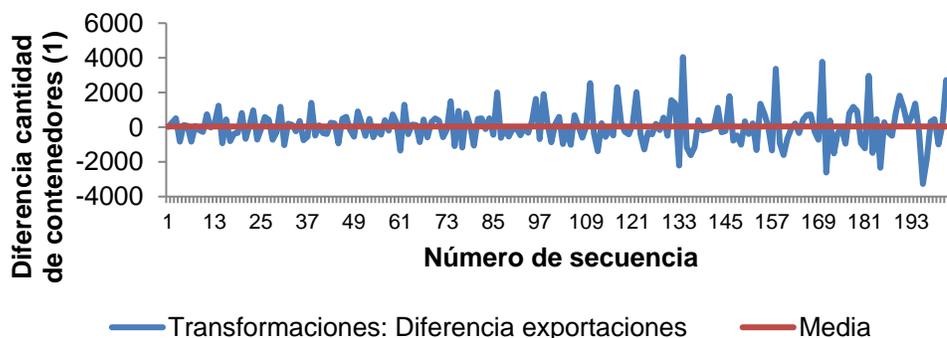
Con estos datos se realizaron modelaciones ARIMA (p,d,q), obteniendo un modelo ARIMA (0,1,1) para ambos casos, por lo que las variables a incluir en el mismo son producto de la diferenciación (d) que se le debe hacer a la serie para hacerla estacional, ya que según lo que se muestra en las figuras 8 y 9, las series tienen tendencia; por lo tanto, las mismas no son estacionales. Para ambos casos se diferenció la serie una vez (d=1) para hacer que la misma no presentara tendencia, lo cual se puede visualizar en las figuras 16 y 17.

Figura 16. **Diferencia contenedores importación (d=1)**



Fuente: elaboración propia.

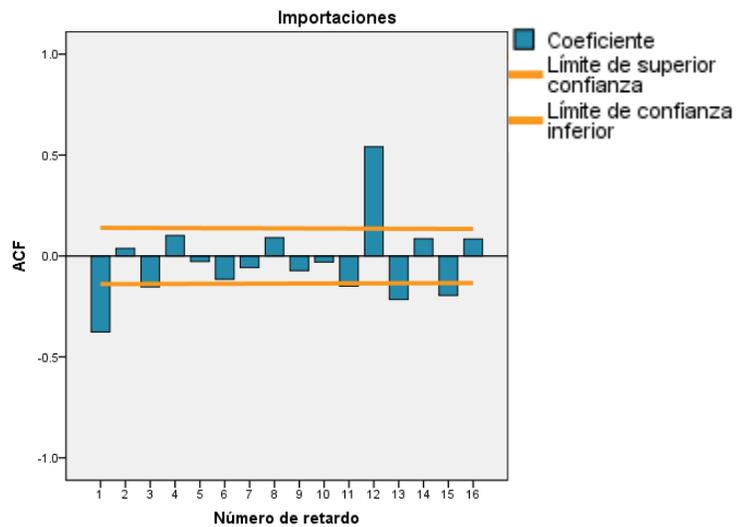
Figura 17. **Diferencia contenedores exportación (d=1)**



Fuente: elaboración propia.

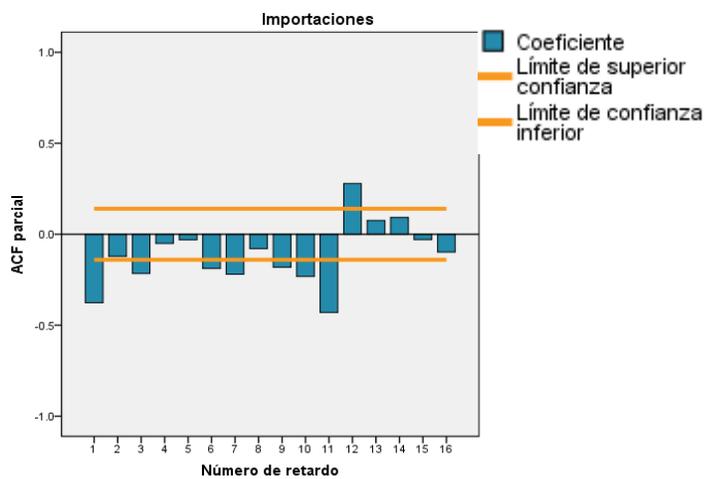
De igual manera se hizo con el componente de media móvil (MA), analizando la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) de la primera diferencia a partir de las figuras 18, 19, 20 y 21.

Figura 18. **FAC importaciones (d=1)**



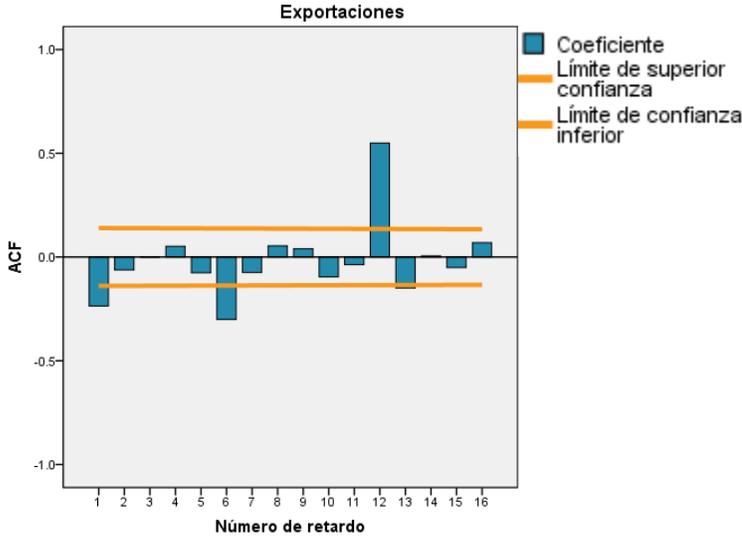
Fuente: elaboración propia.

Figura 19. **FACP importaciones (d=1)**



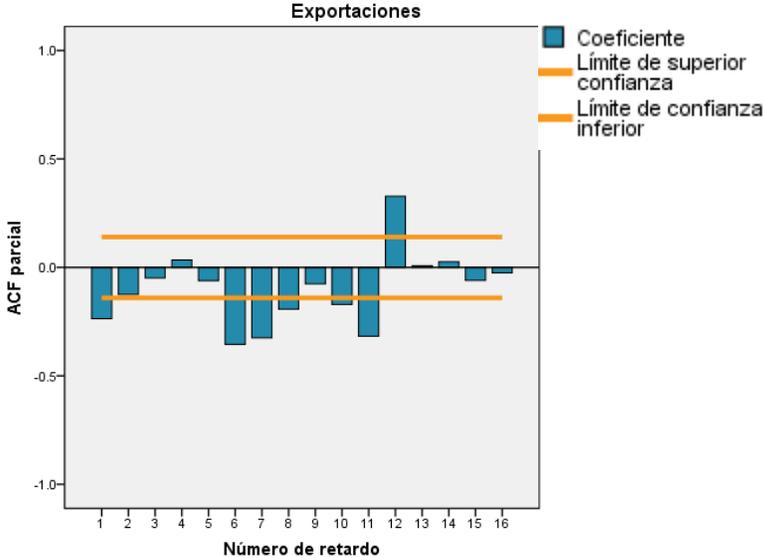
Fuente: elaboración propia.

Figura 20. **FAC exportaciones (d=1)**



Fuente: elaboración propia.

Figura 21. **FACP exportaciones (d=1)**



Fuente: elaboración propia.

Lo que se puede observar en las funciones de autocorrelación y autocorrelación parcial, para los casos de la importación y exportación de contenedores, de la primera diferencia de la serie es un comportamiento irregular en los retardos con comportamiento sinusoidales, por lo que existe evidencia de una componente de media móvil, que debe ser incluida en el modelo ($q=1$), para ambos casos.

Con los análisis realizados anteriormente, el modelo ARIMA con operador de rezagos del proceso autoregresivo (p) igual a 0, diferenciación (d) con valor 1 y operador de rezagos del proceso de media móvil (q) 1, queda expresado de la siguiente manera:

$$(1-L)(-L^{12})X_t = (1-\theta_1L)(1-\beta_{12}L^{12})\varepsilon_t \text{ (Ec.15)}$$

4.2. Comprobación de los supuestos y propiedades del modelo

4.2.1. Supuestos modelo de regresión lineal

Una vez se tienen definidas las variables a incluir en los modelos de importación y exportación de contenedores, se procedió a comprobar los supuestos que debe cumplir un modelo de regresión lineal para ser válido.

4.2.1.1. Linealidad

Para comprobar el supuesto de linealidad en los modelos de regresión lineal para la importación y exportación de contenedores en puerto Santo Tomás de Castilla, se utilizó la prueba de significancia aplicada al coeficiente de correlación de Pearson, también conocido como r , y cuyos resultados se presentan en la tabla XIII. En ellos se observa que los valores de T calculados con la Ec.2, tanto para el caso de la importación como la exportación, son menores que los valores teóricos de la misma, razón por la cual se toma la decisión de rechazar la hipótesis nula, concluyendo que el coeficiente de correlación es significativo, es decir que el valor de r es diferente de cero, razón por la cual se cumple con el supuesto de linealidad.

4.2.1.2. Normalidad

Para el cumplimiento del supuesto de normalidad se hizo uso de la prueba de Shapiro-Wilks, en donde se contrasta un valor crítico contra el valor de significación elegido, partiendo de la hipótesis nula de que los datos siguen una distribución normal contrastada contra la hipótesis alternativa de que los mismos no siguen la distribución de normalidad.

Según los datos observados en la tabla XIV se decide aceptar la hipótesis nula, derivado de que el valor crítico de p es mayor que el nivel de significancia, por lo que se concluye que los residuos siguen una distribución normal. Tanto el modelo de importación como en el de exportación cumplen con el supuesto de normalidad de los residuos.

Otra prueba utilizada para verificar el supuesto de normalidad, es la prueba gráfica Q-Q plot, la cual se muestra en las figuras 4 y 5, en ellas se observa que los datos tienen un buen ajuste a la recta probabilística normal, el cual se presenta cuando los datos están más cercanos a la misma, esto confirma los resultados obtenidos en la prueba de Shapiro-Wilks.

4.2.1.3. Independencia

Para verificar el supuesto de independencia de los errores en los modelos de regresión se utilizó la prueba de rachas, para la cual se tiene una hipótesis nula que expresa existencia de independencia en los residuos, y la hipótesis alternativa de que no existe independencia en ellos.

Para esta prueba se debe hacer uso de las tablas M1 y M2 (anexo), las cuales proporcionan los valores críticos para el número de rachas cuando la significancia es del 5 %.

Según los datos que se encuentran en la tabla XVI, las rachas para los residuos de los modelos de importación y exportación de contenedores se encuentran dentro de los valores M1 y M2, razón por la cual se acepta la hipótesis nula, concluyendo que existe independencia de los errores, cumpliéndose así, este supuesto.

Este supuesto, también fue comprobado mediante la prueba gráfica de la distribución de los residuos versus los valores predichos, en donde se esperaría que en el gráfico no se evidenciará ningún patrón, lo cual puede visualizarse en las figuras 6 y 7.

4.2.1.4. Homogeneidad

La comprobación de este supuesto se realizó mediante la prueba T de Student, para ambos modelos. Esta prueba tiene como hipótesis nula la no existencia de homogeneidad en los residuos, es decir que son heterogéneos, contrastada contra la hipótesis alternativa de homogeneidad.

En la tabla XVI se observa que los valores de T calculados mediante la Ec.3 son menores, en ambos casos, al valor de T teórico; razón por la cual se decide rechazar la hipótesis nula y aceptar la alternativa, cumpliéndose el supuesto de homogeneidad en los residuos para los modelos de regresión de la importación y exportación de contenedores.

4.2.1.5. Homocedasticidad

El último supuesto a comprobar es la homocedasticidad en los residuos de los modelos. Para ello se utilizó la prueba F de Snedecor, la cual parte de la hipótesis nula de que las varianzas son iguales, y por lo tanto existe homocedasticidad; mientras que la hipótesis alternativa dice que las varianzas son diferentes, y por lo que no se cumple con este supuesto.

Los datos de la tabla XVII reflejan valores calculados de F menores a los teóricos, en ambos modelos, razón por la que se decide aceptar la hipótesis nula, y por lo tanto se concluye que existe homocedasticidad en los residuos, es decir que sus varianzas no presentan una diferencia significativa.

4.2.2. Propiedades de las series de tiempo

4.2.2.1. Tendencia y variabilidad

Tal y como se muestra en las figuras 8 y 9, correspondientes a la importación y exportación de contenedores respectivamente, se puede visualizar que ambas series tienen tendencia y variabilidad. Lo mismo se comprueba derivado de que los valores de la media a lo largo del tiempo no son iguales y a su vez existe una variabilidad o componente estacional dentro de los periodos, en este caso, algunos meses que repiten el comportamiento cada año, y son debidos a las temporadas o componentes estacionales (será verificado en el siguiente apartado), pero que, al solo observar estas figuras, se puede deducir que se cuenta con una serie que no tiene estacionalidad en todos sus datos.

4.2.2.2. Estacionalidad

Para comprobar estadísticamente con qué tipo de datos se está trabajando, es fundamental conocer si los mismos constituyen una serie estacional o no estacional. Lo anterior se determinó mediante la prueba de raíz unitaria de Dickey-Fuller, la cual parte de una hipótesis nula correspondiente a la no estacionalidad y una hipótesis alternativa de estacionalidad.

Según los resultados observados en la tabla XIX se puede apreciar que se tiene un valor crítico de la prueba, en ambos modelos, menor que el estadístico de Dickey-Fuller, razón por la cual se acepta la hipótesis nula, y por lo tanto ambas series son no estacionales.

Es importante resaltar que, derivado a que no existe estacionalidad en los datos, es necesario diferenciar la serie en “d” retardos, hasta obtener estacionalidad y eliminar de esta manera la tendencia de las series analizadas.

4.3. Construcción del modelo

4.3.1. Validación modelo de regresión lineal

Con las variables del modelo definidas y con el cumplimiento de los supuestos, se procedió a validar el modelo mediante el contraste global del mismo. Para ello se hizo uso de la prueba F, y cuyos resultados se pueden apreciar en la tabla XVIII.

Cabe mencionar que la prueba F tiene como hipótesis nula la no validez del modelo, mientras que la hipótesis alternativa es la validez del mismo. Según los datos de la tabla mencionada con anterioridad, se decide la aceptación de la hipótesis alternativa, ya que los valores calculados de F son mayores a los teóricos, razón por la cual se concluye que el modelo es válido, tanto para el caso de la importación como para la exportación de contenedores.

4.3.2. Validación del modelo generado por series temporales

Una vez se tiene un modelo de series de tiempo y se sabe de las propiedades de las mismas, es necesario comprobar la normalidad e independencia de los residuos con la finalidad de darle validez al mismo.

Para el caso de la normalidad, y derivado de que se cuenta con más de 50 datos, 191 para ser precisos, se descarta la utilización de la prueba de Shapiro-Wilks, ya que esta se recomienda emplearse con menos de 50 registros. Es por ello que se utilizó la prueba de bondad de ajuste de Kolmogórov-Smirnov, y cuyos resultados se pueden apreciar en la tabla XXI. Esta prueba parte de la hipótesis nula de normalidad versus la hipótesis alternativa de que los datos no siguen una distribución normal. Para ello al realizar el contraste del p-valor y verificar que es mayor que el nivel de significancia, se decide la aceptación de la hipótesis nula, y por lo tanto se concluye que los residuos de los modelos, para la cantidad de contenedores importación y exportación generados por series de tiempo, siguen una distribución normal.

En lo correspondiente a la verificación de independencia en los residuos estimados del modelo, se hizo uso de las pruebas gráficas, basado en las funciones de autocorrelación serial (FAC) y parcial (FACP) presentados en las figuras 10 y 11, en donde se observa que el comportamiento de los mismos se mantiene dentro de las bandas de confianza, lo cual indica que no existen valores significativos; por lo que el modelo cumple con este supuesto.

Derivado del cumplimiento de ambos supuestos, en los residuos estimados de los modelos realizados, se concluye que ambos son válidos, y pueden ser utilizados para proyectar los datos de la cantidad de contenedores, tanto de importación como de exportación, a movilizar en puerto Santo Tomás de Castilla.

4.4. Robustez del modelo

4.4.1. Ajuste del modelo de regresión lineal

Tal y como se puede apreciar en las figuras 12 y 13, el ajuste de los datos a la recta de regresión es bueno, confirmándose lo mismo con el coeficiente de correlación, el cual tiene un valor de 0.95 y 0.97, para el modelo de importación y exportación respectivamente.

Derivado de lo anterior se considera que ambos modelos explican de manera confiable los eventos que han sucedido durante el periodo de estudio, pudiendo realizar proyecciones con los mismos; sin embargo, es importante mencionar que la función principal de un modelo de regresión es explicar el comportamiento de los datos a lo largo de determinado periodo de tiempo y no para ser utilizado como método de proyección a futuro.

4.4.2. Ajuste del modelo generado por análisis de series temporales

Después de comprobar las propiedades de las series de tiempo de importación y exportación de contenedores, y del cumplimiento de la normalidad e independencia de los residuos, se definió que para ambos casos la modelación ARIMA es la más adecuada, ya que cuenta con las propiedades y cumple con los supuestos para validar el mismo. Adicionalmente se conoce que el coeficiente de correlación serial para el caso de la importación es de 0.922, mientras que para la exportación es de 0.945, razón por la cual el ajuste de los datos al modelo es considerado bueno.

Derivado de lo anterior se puede obtener la cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla a corto, mediano y largo plazo. En las figuras 14 y 15 se puede observar la gráfica de las series de tiempo de importación y exportación de contenedores hasta el año 2016, mientras que lo correspondiente al 2017 son los resultados obtenidos mediante el modelo ARIMA, y cuyos resultados hasta el año 2030 se encuentran en el apéndice de este informe.

En ambas figuras se puede observar que la cantidad de contenedores a movilizar durante el año 2017 obtenidas con el análisis de series de tiempo, mantienen un comportamiento similar al de los años anteriores, con picos elevados durante un primer periodo comprendido por los meses de marzo, abril y mayo; correspondientes a la temporada alta de fruta. Por su parte, el segundo periodo es a finales de año, siendo temporada alta por época navideña y de Año Nuevo, presentándose durante los meses de noviembre y diciembre.

Esta modelación ARIMA brinda una herramienta importante en la búsqueda de la solución a la problemática inicial, ya que al desarrollar la misma, se cuenta con la metodología necesaria para realizar el análisis de las series de tiempo, la cual permita proyectar a futuro la cantidad de cantidad de contenedores de importación y exportación a movilizar en el puerto.

CONCLUSIONES

General

La construcción de modelos mediante el análisis de regresión lineal y series de tiempo, requiere que el investigador tenga pleno conocimiento de los datos a estudiar. Dentro del presente informe se ha logrado desarrollar una metodología que permite explicar el comportamiento de los datos, determinando su ajuste para posteriormente proyectarlos a futuro, haciendo énfasis en el análisis previo y cumplimiento de supuestos para que sean válidos; por lo que puede ser puesta en práctica para el análisis de cualquier tipo de series temporales.

Específicas

1. Las variables a incluir en los modelos de importación y exportación de contenedores, mediante el análisis de regresión, se definen con la prueba de contraste T de Student, buscando que no exista colinealidad entre ellas, eligiendo las que mejor expliquen el comportamiento de los datos. Para el caso de la generación de un modelo mediante la utilización de series de tiempo, se incluyeron las variables o componentes de diferenciación para hacer la serie estacional ($d=1$) y el componente de media móvil ($q=1$).
2. Los supuestos de un modelo de regresión lineal se cumplen para las unidades de importación y exportación de contenedores, por lo que los mismos son válidos y explican cada una de las variables independientes.

En lo correspondiente a las propiedades de las series temporales las mismas cuentan con tendencia y variabilidad; mas no con estacionalidad, razón por la cual se tuvo que realizar la transformación correspondiente para proyectar la cantidad de contenedores a movilizar.

3. Los modelos obtenidos por análisis de regresión lineal son $Y = 0.26 X + 12676.01$; y $Z = 1.27 W + 7809.25$, donde las variables dependientes son la cantidad de contenedores a movilizar en importación (Y) y exportación (Z); mientras que las variables independientes son gastos de consumo final (x) y exportaciones de bienes y servicios (W). El modelo a utilizar en el análisis de series de tiempo para la cantidad de contenedores a movilizar es el autoregresivo integrado de medias móviles o ARIMA (0,1,1).
4. La robustez de los modelos obtenidos está definida por el coeficiente de correlación r, el cual explica qué porcentaje de la variabilidad de los datos es debida a cada uno de ellos. En este sentido los resultados obtenidos fueron de un 95 y un 97 % para los modelos de regresión lineal de importación y exportación; mientras que de los modelos ARIMA se obtuvo un coeficiente de correlación serial de 92 y 95 % respectivamente, considerando los mismos como un buen ajuste.

RECOMENDACIONES

1. Tomar en cuenta que, cuando se realizan proyecciones se trabaja con datos que están expuestos a verse afectados por factores externos, por lo que en determinado momento estas proyecciones tienden a fallar. Es por ello que se debe monitorear el valor real de los datos, tratando de establecer las causas de algún error en las estimaciones, para tenerlas en cuenta a la hora de calibrar el modelo o elaborar uno nuevo.
2. Utilizar los modelos obtenidos por medio de análisis de regresión lineal para describir el comportamiento de los datos en el tiempo, y si se quieren realizar proyecciones, realizarlo por medio del análisis de series de tiempo buscando el modelo que mejor se ajuste a las características de los mismos.
3. Continuar monitoreando los datos proyectados, cotejándolos con valores reales; para determinar el grado de error que los modelos tienen con el objetivo de realizar las calibraciones y análisis que lo minimicen.
4. Implementar el análisis de series de tiempo para la cantidad de contenedores de importación y exportación para puerto Barrios y Puerto Quetzal; así como realizar el mismo análisis para los diferentes tipos de carga que se movilizan en los puertos (carga general, granel sólido y líquido) mencionados anteriormente, adicionando al análisis al puerto Santo Tomás de Castilla.

BIBLIOGRAFÍA Y REFERENCIAS

1. Agencia de Cooperación internacional de Japón. (1998). *Estudio sobre el proyecto de desarrollo del puerto Santo Tomás de Castilla en la República de Guatemala*. Guatemala: Empresa Portuaria Nacional Santo Tomás de Castilla.
2. Antunez, C. (2011). Análisis de series de tiempo. Perú. Recuperado de <http://www.eumed.net/ce/2011a/chai.htm>.
3. Autoridad Portuaria Nacional de Perú. (2012). Plan nacional de desarrollo portuario. Perú. Recuperado de <http://www.apn.gob.pe/site/plan-nacional-de-desarrollo-portuario.aspx>.
4. Castellanos, A. (2015). Logística comercial internacional. Colombia: Universidad del Norte.
5. Cendrero, B., y Truyols, S. (2008). El transporte: aspectos y tipología. España: Delta publicaciones universitarias.
6. Comisión Portuaria Nacional de Guatemala. (2016). El sistema portuario en apoyo al comercio exterior 2015. Guatemala. Recuperado de <http://cpn.gob.gt/institucion/asesoria-y-estudios/estadisticas-portuarias/>.

7. Conferencia de las Naciones Unidas para el Comercio y Desarrollo. (2009). Programa de capacitación portuaria: módulo 2 La organización de un sistema portuario. Ginebra.
8. Díaz, A. (2009). Diseño estadístico de experimentos. Colombia: Universidad de Antioquia.
9. Gavarrete, J., y Fernández, M. I. (2001). Estudio centroamericano de transporte (informe técnico). Guatemala: Secretaría de Integración Económica Centroamericana. Guatemala. Recuperado de <http://www.sieca.int/PortalData/Documentos/CA565DC2-4459-458F-A103-1E9208BFD6E6.pdf>
10. Gómez, J., Aparicio, J., y Patiño, C. (2010). Manual de análisis de frecuencias en hidrología. México: Instituto Mexicano de Tecnología del Agua.
11. González, P. (2009). Análisis de series temporales: Modelos ARIMA. España: Universidad del País Vasco.
12. Guàrdia, J., Freixa, M., Però, M., y Turbany, J. (2008). *Análisis de datos en psicología*. España: Delta publicaciones.
13. Guisande, C. (2006). Tratamiento de datos. España: Universidad de Vigo.
14. Gujarati, D., y Porter, D. (2010). Econometría. México: McGraw-Hill.

15. Hair Jr., Joseph F., Anderson, Rolph E., Tatham, Ronald L., y Black, William C. (1999). Análisis multivariante. España: Prentice Hall.
16. Hurtado, C. (2008). Series de tiempo. Chile: Universidad de Chile.
17. Iztapa Management, S.A. (2003). Estudio del plan maestro de desarrollo de puerto Quetzal. Guatemala: Empresa Portuaria Quetzal.
18. Lind, D., Marchal, W., y Wathen, S. (2012). Estadística aplicada a los negocios y la economía. México: McGraw-Hill.
19. Marques, M. (2001). Estadística básica: Un enfoque no paramétrico. México: Universidad Nacional Autónoma de México.
20. Marí, R., De Souza, A., Martín, J., y Rodrigo, J. (2003). El transporte de contenedores: terminales, operatividad y casuística. España: Universitat Politècnica de Catalunya.
21. Ministerio de Océanos y Pesca de la República de Corea. (2015). Plan maestro portuario de Guatemala. Guatemala.
22. Navidi, W. (2006). Estadística para ingenieros y científicos. México: McGraw-Hill.
23. Novales, A. (1993). Econometría. España: McGraw-Hill.
24. Puerto Valparaíso. (2015). Plan maestro puerto Valparaíso. Chile. http://www.puertovalparaiso.cl/img/media/1452629139_01.2016%20Plan%20Maestro%202015%20Final.pdf.

25. Rodríguez, T. (2015). Predicción de tráfico de contenedores a corto plazo mediante técnicas de minería de datos: redes neuronales artificiales y redes bayesianas (tesis de doctorado). Universidad Politécnica de Madrid. España. Recuperado de http://oa.upm.es/38183/1/Tomas_Rodriguez_Garcia.pdf.
26. Royal Haskoning. (2008). Plan maestro para el complejo portuario Limón-Moín. Costa Rica. Recuperado de <http://www.cnc.go.cr/content/documentos/contenidos%20TCM/Plan%20Maestro%20Limon-Moin.pdf>.
27. Sánchez, R. (2004). Puertos y transporte marítimo en América Latina y el Caribe: un análisis de su desempeño reciente. Chile: Comisión Económica para América Latina y el Caribe. Recuperado de <http://www.cepal.org/es/publicaciones/6452-puertos-transporte-maritimo-america-latina-caribe-un-analisis-su-desempeno>.
28. Subsecretaría de Transportes de Chile. (2013). Plan nacional de desarrollo portuario. Chile. Recuperado de https://www.mtt.gob.cl/wp-content/uploads/2014/02/desarrollo_portuario_06_02_14_1.pdf
29. Walpole, R., Myers, R., Myers, S., y Ye, K. (2012). Probabilidad y estadística para ingeniería y ciencias. México: Pearson Educación.
30. Webster, A. (2000). Estadística aplicada a los negocios y la economía. Colombia: McGraw-Hill.

APÉNDICES

Apéndice 1. **Series de tiempo importación y exportación**

AÑO	Mes	Importaciones	Exportaciones
2000	E	4221	4568
	F	4157	4593
	M	4427	4857
	A	5268	5372
	M	3971	4537
	J	3999	4658
	J	4714	4747
	A	4231	3915
	S	4102	3990
	O	4531	3831
	N	3962	3561
	D	4328	4304
2001	E	3767	4284
	F	4020	4473
	M	5149	5711
	A	4172	4790
	M	4790	5241
	J	4173	4436
	J	3563	4063
	A	4172	3731
	S	4082	4546
	O	4390	3876
	N	3995	3957
	D	5122	4927

2002	E	3644	4216
	F	4306	4105
	M	4828	4698
	A	6225	5118
	M	4855	4400
	J	4079	4101
	J	5128	5277
	A	4394	4241
	S	4663	4447
	O	5371	4590
	N	5652	4357
	D	4667	4715
	2003	E	4310
F		4852	3416
M		5060	4820
A		4600	4338
M		4943	4446
J		4271	4107
J		4429	3713
A		4285	3969
S		4157	4189
O		4656	3259
N		4536	3744
D		5106	4338
2004		E	4423
	F	4124	3650
	M	5371	4563
	A	5340	4779
	M	4239	4274
	J	4732	4751
	J	4890	4172
	A	4717	4081
	S	4013	3650
	O	5645	4060
	N	4626	3861
	D	5373	4599

2005	E	4538	4858
	F	4014	3513
	M	4922	4812
	A	4179	4413
	M	4785	4560
	J	4366	4695
	J	4007	3832
	A	5384	4274
	S	3804	3692
	O	4200	3990
	N	5292	4492
	D	4485	4859
	2006	E	4212
F		4611	4142
M		5615	5641
A		4662	4550
M		5491	5490
J		4804	4322
J		5491	5132
A		5255	5213
S		4716	4154
O		4895	4632
N		5913	5135
D		5218	5026
2007		E	5973
	F	5093	5108
	M	7331	7113
	A	6557	6503
	M	6414	6440
	J	5265	5897
	J	6784	5800
	A	6068	5744
	S	6237	5275
	O	5741	5239
	N	6218	4920
	D	5572	5276

2008	E	7652	6915
	F	6461	6231
	M	9514	8137
	A	7515	8362
	M	7305	7490
	J	7074	7678
	J	7987	8278
	A	7534	7308
	S	7510	7139
	O	7840	6123
	N	7463	6809
	D	7207	6860
	2009	E	5900
F		6053	6295
M		8362	8838
A		8497	8465
M		7143	7080
J		6480	7308
J		7630	6754
A		7170	6659
S		6702	6185
O		7925	8490
N		9050	8767
D		8218	8504
2010		E	8755
	F	7970	8244
	M	10317	10263
	A	10164	9873
	M	8404	8594
	J	7957	8302
	J	8048	7892
	A	8186	8084
	S	8739	7917
	O	8696	8465
	N	8863	7971
	D	9394	9534

2011	E	10756	10906
	F	9086	8698
	M	12843	12740
	A	11710	11567
	M	9836	9953
	J	8652	8800
	J	8516	9210
	A	9243	9023
	S	9799	8898
	O	9684	8819
	N	9545	8912
	D	10120	10030
	2012	E	9636
F		9753	9524
M		11869	11314
A		11258	10536
M		8949	10057
J		9227	9053
J		9474	9393
A		9222	8998
S		9761	9221
O		9227	7898
N		10473	9252
D		10771	10036
2013		E	9385
	F	8909	8752
	M	12524	12122
	A	10709	11201
	M	10368	9595
	J	8548	8944
	J	8643	8891
	A	9600	9105
	S	8866	8770
	O	9202	9206
	N	10771	9914
	D	11134	10646

2014	E	10228	10387
	F	8849	9664
	M	14485	13437
	A	10633	10827
	M	11726	11210
	J	9016	9702
	J	8732	9378
	A	9792	9306
	S	9492	8350
	O	9743	9160
	N	10613	10333
	D	10088	11251
	2015	E	10496
F		9857	9120
M		12545	12083
A		10735	10598
M		12378	11059
J		8583	8721
J		10045	9006
A		9423	8718
S		8647	8227
O		10263	9062
N		11121	10882
D		12437	11984
2016		E	11556
	F	13582	12713
	M	15467	14072
	A	13358	13786
	M	11344	10516
	J	8872	8715
	J	9218	9038
	A	10391	9500
	S	9849	8506
	O	10134	8588
	N	10893	11309
	D	13341	12956

Fuente: elaboración propia.

Apéndice 2. **Cantidad de contenedores a movilizar en puerto Santo Tomás de Castilla durante los años 2017-2030**

Año	Mes	Importaciones	Exportaciones
2017	E	11683	12123
	F	10801	11801
	M	14903	14153
	A	12813	13156
	M	12297	12842
	J	9778	11142
	J	10276	11322
	A	10842	11436
	S	10370	10719
	O	10928	11053
	N	11777	12969
	D	13087	14228
2018	E	12118	13769
	F	11236	13447
	M	15337	15798
	A	13248	14801
	M	12732	14487
	J	10213	12788
	J	10711	12968
	A	11276	13082
	S	10805	12364
	O	11363	12698
	N	12212	14614
	D	13521	15873

2019	E	12553	15414
	F	11671	15092
	M	15772	17444
	A	13683	16447
	M	13167	16133
	J	10648	14433
	J	11146	14613
	A	11711	14727
	S	11240	14010
	O	11798	14344
	N	12647	16260
	D	13956	17519
	2020	E	12988
F		12106	16738
M		16207	19089
A		14118	18092
M		13601	17778
J		11083	16079
J		11580	16259
A		12146	16373
S		11675	15655
O		12232	15989
N		13082	17905
D		14391	19164
2021		E	13423
	F	12540	18383
	M	16642	20735
	A	14553	19738
	M	14036	19424
	J	11518	17724
	J	12015	17904
	A	12581	18018
	S	12110	17301
	O	12667	17635
	N	13517	19551
	D	14826	20810

2022	E	13858	20351
	F	12975	20029
	M	17077	22381
	A	14988	21383
	M	14471	21069
	J	11953	19370
	J	12450	19550
	A	13016	19664
	S	12545	18946
	O	13102	19280
	N	13952	21196
	D	15261	22455
2023	E	14293	21996
	F	13410	21674
	M	17512	24026
	A	15423	23029
	M	14906	22715
	J	12388	21015
	J	12885	21196
	A	13451	21309
	S	12980	20592
	O	13537	20926
	N	14387	22842
	D	15696	24101
2024	E	14728	23642
	F	13845	23320
	M	17947	25672
	A	15857	24674
	M	15341	24360
	J	12823	22661
	J	13320	22841
	A	13886	22955
	S	13415	22238
	O	13972	22571
	N	14821	24487
	D	16131	25747

2025	E	15163	25288
	F	14280	24965
	M	18382	27317
	A	16292	26320
	M	15776	26006
	J	13257	24307
	J	13755	24487
	A	14321	24600
	S	13850	23883
	O	14407	24217
	N	15256	26133
	D	16566	27392
	2026	E	15598
F		14715	26611
M		18817	28963
A		16727	27965
M		16211	27651
J		13692	25952
J		14190	26132
A		14756	26246
S		14285	25529
O		14842	25863
N		15691	27778
D		17001	29038
2027		E	16032
	F	15150	28257
	M	19252	30608
	A	17162	29611
	M	16646	29297
	J	14127	27598
	J	14625	27778
	A	15191	27891
	S	14719	27174
	O	15277	27508
	N	16126	29424
	D	17436	30683

2028	E	16467	30224
	F	15585	29902
	M	19686	32254
	A	17597	31257
	M	17081	30943
	J	14562	29243
	J	15060	29423
	A	15625	29537
	S	15154	28820
	O	15712	29154
	N	16561	31070
	D	17870	32329
2029	E	16902	31870
	F	16020	31548
	M	20121	33899
	A	18032	32902
	M	17516	32588
	J	14997	30889
	J	15495	31069
	A	16060	31183
	S	15589	30465
	O	16147	30799
	N	16996	32715
	D	18305	33974
2030	E	17337	33515
	F	16455	33193
	M	20556	35545
	A	18467	34548
	M	17950	34234
	J	15432	32534
	J	15930	32714
	A	16495	32828
	S	16024	32111
	O	16582	32445
	N	17431	34361
	D	18740	35620

Fuente: elaboración propia.

Apéndice 2. **Residuos modelos ARIMA**

Año	Mes	Residuos importación	Residuos exportación
2001	F	317	164
	M	1011	1041
	A	-1187	-928
	M	1102	814
	J	143	-508
	J	-1221	-723
	A	190	127
	S	180	806
	O	13	-96
	N	184	302
	D	899	382
	2002	E	-389
F		287	-487
M		-29	-471
A		1559	487
M		-2	-369
J		-443	-80
J		750	1311
A		-305	185
S		148	-198
O		458	493
N		1095	90
D		-959	-469

2003	E	-126	-603
	F	84	-857
	M	-377	254
	A	-1284	-391
	M	133	301
	J	-35	173
	J	-334	-775
	A	-127	637
	S	-289	189
	O	-238	-665
	N	-174	290
	D	464	116
2004	E	351	484
	F	-480	-59
	M	430	-141
	A	176	338
	M	-688	-162
	J	573	742
	J	253	-278
	A	231	80
	S	-514	-686
	O	736	567
	N	-345	-44
	D	258	103
2005	E	46	688
	F	-622	-615
	M	-346	7
	A	-1055	-382
	M	438	267
	J	51	257
	J	-582	-520
	A	1138	389
	S	-423	-357
	O	-900	219
	N	914	626
	D	-568	22

2006	E	52	-469
	F	577	420
	M	599	598
	A	-228	-600
	M	773	744
	J	148	-853
	J	794	867
	A	-95	455
	S	222	-556
	O	-393	120
	N	563	351
	D	-200	-464
	2007	E	1157
F		-96	378
M		1260	915
A		726	442
M		127	-183
J		-620	-164
J		768	-257
A		-310	-265
S		648	9
O		-518	-281
N		-425	-804
D		-636	-312
2008		E	1617
	F	335	528
	M	1849	579
	A	33	1116
	M	-288	-465
	J	282	452
	J	330	833
	A	-10	-559
	S	317	94
	O	474	-1092
	N	-521	71
	D	-187	-211

2009	E	-2296	-1455
	F	-893	-147
	M	-474	733
	A	983	240
	M	-591	-974
	J	-583	-73
	J	-130	-854
	A	-256	-121
	S	-445	-152
	O	698	615
	N	1510	264
	D	676	-317
	2010	E	1083
F		361	80
M		424	-27
A		817	-109
M		-575	-568
J		-297	-571
J		-1101	-610
A		-321	152
S		632	335
O		-186	792
N		-517	-393
D		707	1369
2011		E	1694
	F	120	-986
	M	1590	1476
	A	505	-76
	M	-433	-662
	J	-983	-1359
	J	-1481	-57
	A	-224	-146
	S	348	111
	O	-199	-222
	N	-677	23
	D	180	450

2012	E	-1027	-608
	F	341	497
	M	-503	-880
	A	-259	-529
	M	-1107	512
	J	245	-160
	J	129	209
	A	-361	-166
	S	16	367
	O	-678	-1212
	N	566	714
	D	552	281
	2013	E	-1171
F		-818	-659
M		439	602
A		-805	125
M		828	-621
J		-845	-205
J		-818	-335
A		327	283
S		-860	-158
O		-180	890
N		820	565
D		755	190
2014		E	114
	F	-782	91
	M	2053	979
	A	-1165	-1289
	M	1403	939
	J	-700	-308
	J	-1012	-559
	A	-103	-318
	S	-293	-955
	O	-32	428
	N	-161	745
	D	-926	513

2015	E	353	-607
	F	495	-672
	M	-1024	-632
	A	-316	-209
	M	1651	872
	J	-877	-826
	J	802	-42
	A	-695	-254
	S	-1124	-111
	O	633	485
	N	382	1192
	D	1653	871
	2016	E	567
F		162	2115
M		1561	-675
A		1137	917
M		-1694	-39
J		-1191	-201
J		-1151	151
A		170	697
S		7	-133
O		-458	-516
N		-497	1165
D		1563	1290

Fuente: elaboración propia.

ANEXOS

Anotar número de página consultada

Anexo 1. Tabla M1 y M2 para la prueba de rachas

Tabla M Valores críticos de r en la prueba de rachas

La tabla M1 y M2 contienen varios valores críticos de r para varios valores de n_1 y n_2 . Para la prueba de rachas de una sola muestra, todo valor de r que sea igual o menor que el mostrado en la tabla M1 o igual o mayor que el mostrado en la tabla M2 es significativo al nivel del 0.05.

Tabla M1

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2											2	2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
4				2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5			2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5
6		2	2	3	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	8	8	8	8	9	9	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14	14

Tabla M2

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2																				
3																				
4			9	9																
5		9	10	10	11	11														
6		9	10	11	12	12	13	13	13	13										
7		11	12	13	13	14	14	14	14	14	15	15	15							
8		11	12	13	14	14	15	15	16	16	16	16	16	17	17	17	17	17	17	17
9			13	14	14	15	16	16	16	17	17	18	18	18	18	18	18	18	18	18
10			13	14	15	16	16	17	17	18	18	18	18	19	19	19	19	19	19	19
11			13	14	15	16	17	17	18	18	19	19	19	20	20	20	20	20	21	21
12			13	14	16	16	17	18	19	19	20	20	20	21	21	21	21	22	22	22
13				15	16	17	18	19	19	20	20	21	21	22	22	22	23	23	23	23
14				15	16	17	18	19	20	20	21	22	22	23	23	23	23	23	24	24
15				15	16	18	18	19	20	21	22	22	23	23	24	24	24	24	24	25
16					17	18	19	20	21	21	22	23	23	24	25	25	25	25	25	25
17					17	18	19	20	21	22	23	23	24	25	25	26	26	26	26	26
18					17	18	19	20	21	22	23	24	25	25	26	26	27	27	27	27
19					17	18	20	21	22	23	23	24	25	26	26	27	27	27	27	27
20					17	18	20	21	22	23	24	25	25	26	27	27	27	27	27	28

Fuente: Webster, A. (2000). *Estadística aplicada a los negocios y la economía*. p. 624.