

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

DIVISIÓN DE AGRONOMÍA

DEPARTAMENTO FORESTAL



Aplicación de Transformación Box Cox en Modelos Lineales Para La Estimación de  
Biomasa en Siete Especies Forestales

Por:

**EMMANUEL DE JESÚS DE LA CRUZ GARCÍA**

TESIS

Presentada como requisito parcial para obtener el título de:

**INGENIERO FORESTAL**

Saltillo, Coahuila, México

Noviembre 2019

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

DIVISIÓN DE AGRONOMÍA

DEPARTAMENTO FORESTAL

Aplicación de Transformación Box Cox en Modelos Lineales Para La Estimación de  
Biomasa en Siete Especies Forestales

Por:

**EMMANUEL DE JESÚS DE LA CRUZ GARCÍA**

TESIS

Presentada como requisito parcial para obtener el título de:

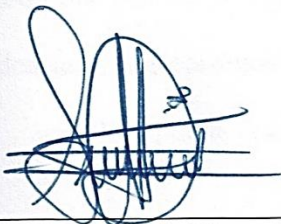
**INGENIERO FORESTAL**

Aprobada por el Comité de Asesoría:



Dr. Jorge Méndez González

Asesor Principal



Ing. Cristóbal de Jesús Flores Hernández

Coasesor



Dr. Genaro Esteban García Mosqueda

Coasesor



Dr. José Antonio González Fuentes

Coordinador de la División de Agronomía



Saltillo, Coahuila, México

Noviembre 2019

## **DEDICATORIA**

A Dios por darme la vida, salud, valor, fuerza y sobre todo por acompañarme en todo momento para conseguir este logro parte de mi carrera profesional.

### **A mi abuelita:**

Luisa López Sánchez por todo su amor y sus grandes consejos que en todo momento me brindó y por el apoyo incondicional que en todo momento nunca me negó.

### **A mis padres:**

A mi Madre Cerilda García Gutiérrez por todo su amor y cariño incontable que siempre me ha regalado y por todos sus sabios consejos que fueron el sustrato en donde mis sueños sembré y hoy rinden frutos a ti mamita te dedico este logro.

A mi padre Francisco de la Cruz López por todo su amor incondicional que me ha brindado y por todos sus sabios consejos y enseñanzas, por nunca rendirse y trabajar de sol a sol para que hoy pueda concluir mis estudios, a ti te dedico este logro Papá.

### **A mis hermanos:**

Gaudencio Cruz García, María del Rosario Cruz García, José Luis De La Cruz García, Ramona De La Cruz García, Juan G. De La Cruz García, por todo su apoyo incondicional que siempre me ofrecieron y por sus buenos consejos que nunca me negaron y sobre todo por estar siempre ahí cuando me sentía solo.

**A mis tíos:**

A mis tíos Enrique De La Cruz, Crisóforo De La Cruz, Diego De La Cruz, Aristarco López y a mis tías, Gudelia De La Cruz, Elena De La Cruz, Cecilia De La Cruz, por todo el apoyo incondicional y sus consejos que nunca me negaron

**A mis amigos:**

Liliana Sánchez salinas, Alexis Maldonado, Claudia Arispe Vázquez, Frankser Abdiel Lucas Rodríguez, Ivan Hernández Maldonado Edmar Oseyner Ruiz López, Ernesto Gayosso Espíndola, Gamadiel Raymundo, Urbano Gómez García, Edgar Guzmán Arreola, José Luis González Castañeda, Alejandro Ramírez Pérez, Jesús Pérez Ortiz, y a todos mis compañeros de la generación.

## **AGRADECIMIENTOS**

A mi alma mater, la Universidad Autónoma Agraria Antonio Narro por brindarme la oportunidad de estudiar una carrera y por acogerme durante cinco años, en los cuales me brindó conocimientos, me hice de amigos y las mejores experiencias.

A mi asesor, el Dr. Jorge Méndez González, por guiarme durante todo el desarrollo del presente proyecto de tesis, brindándome sus conocimientos, con paciencia y dedicación, gracias por su compromiso y por ser pieza fundamental en mi desarrollo profesional.

A mis coasesores el Ing. Cristóbal de Jesús Flores Hernández y el Dr. Genaro Esteban García Mosqueda, por todo su conocimiento y su apoyo incondicional e interés que siempre demostraron.

A los profesores del Departamento Forestal que con sus enseñanzas y dedicación influyeron en mi desarrollo profesional.

## ÍNDICE DE CONTENIDO

ÍNDICE DE CUADROS .....	III
RESUMEN .....	IV
ABSTRACT.....	V
INTRODUCCIÓN .....	1
Objetivos .....	4
Hipótesis.....	4
REVISIÓN DE LITERATURA .....	5
Generalidades de las especies .....	5
<i>Pinus cembroides</i> Zucc (1832).....	5
<i>Pinus halepensis</i> Miller.....	5
<i>Pinus greggii</i> Engelm. ....	6
<i>Prosopis glandulosa</i> Torr.....	6
<i>Atriplex canescens</i> (Pursh) Nutt.....	6
<i>Yucca filifera</i> Chaubad.....	7
<i>Acacia constricta</i> Benth.....	7
Normalidad.....	8
Homogeneidad .....	8
Box Cox .....	9
R studio .....	10

Paquete Trafo: otras transformaciones.....	10
Estudios afines .....	11
MATERIALES Y MÉTODOS .....	13
Descripción del área de estudio.....	13
Método directo para obtener biomasa de las especies .....	14
ANÁLISIS ESTADÍSTICOS .....	16
Modelos de regresión para cuantificar biomasa en las siete especies .....	16
Prueba de normalidad.....	17
Prueba de homogeneidad de varianza .....	18
Procesamiento de datos .....	19
Método de transformación de variables Box Cox (1964) .....	19
Recalculando $R^2$ ajustada.....	20
La librería Trafo: otras transformaciones.....	21
RESULTADOS .....	22
Resultados con la transformación Box Cox (1964) .....	28
Resultados de la librería de transformaciones (Trafo) .....	33
DISCUSIÓN .....	36
CONCLUSIONES.....	41
BIBLIOGRAFÍA .....	42

## ÍNDICE DE CUADROS

<b>Cuadro 1.</b> Características generales del área de estudio.-----	13
<b>Cuadro 2.</b> Especies utilizadas en el estudio.-----	14
<b>Cuadro 3.</b> Modelos matemáticos a evaluar biomasa aérea de siete especies en su forma logarítmica. -----	16
<b>Cuadro 4.</b> Resumen estadístico para la prueba de normalidad. -----	23
<b>Cuadro 5.</b> Resumen estadístico de la prueba de homogeneidad. -----	24
<b>Cuadro 6.</b> Coeficiente de determinación $R^2$ ajustada de los modelos por especie. -----	26
<b>Cuadro 7.</b> Error estándar de los modelos (RSCE) de regresión lineal por especie. -----	27
<b>Cuadro 8.</b> Resultados de la prueba de normalidad residual después de la transformación. ----	28
<b>Cuadro 9.</b> Resultados de homogeneidad de varianza de los modelos p/esp después de la transformación. -----	29
<b>Cuadro 10.</b> Valores del coeficiente de determinación ( $R^2$ ajustada) de los modelos por especie al aplicar la transformación. -----	30
<b>Cuadro 11.</b> Resultados del error estándar de los modelos por especie aplicando la transformación. -----	31
<b>Cuadro 12.</b> Transformaciones de la librería (Trafo) para normalidad y homogeneidad de R Studio para los modelos por especie. -----	34



## RESUMEN

El estudio se llevó a cabo con datos de siete especies: *Acacia constricta* Benth; *Pinus cembroides* Zucc; *Pinus greggii* Engelm; *Pinus halepensis* Miller; *Prosopis glandulosa* Torr; *Yucca filiera* Chaubad; *Atriplex canescens* (Pursh) Nutt, teniendo como objetivos: a) Demostrar que aplicando el método de Box Cox se obtiene un mejor ajuste en los modelos de regresión lineal eliminando problemas de normalidad y de heterocedasticidad de los residuales de los modelos, b) sugerir alternativas de cómo proceder ante la distribución de datos no normales y ante la presencia de heterocedasticidad en modelos de regresión lineal. Para la estimación de biomasa se utilizaron seis modelos de regresión lineal, utilizando el diámetro normal, diámetro de copa, diámetro basal y la altura total como variables independientes. Los modelos que mejores resultados demostraron después de la transformación fueron los modelos 2, 3 y 6 para el supuesto de normalidad ya que cumplen el supuesto para todas las especies. Caso contrario de homogeneidad en donde solo los modelos 3 y 6 cumplen esto para todas las especies y el modelo 2 solo se ajusta para cinco especies, de igual manera el modelo 6 presenta un valor promedio de  $R^2$  ajustada = 0.73 y el modelo 3 presenta una  $R^2$  ajustada = 0.79, así como un error estándar promedio de 0.283 kg. Por esto se entiende que el modelo 3 es el que mejor ajuste demostró en todas las especies. El método de transformación de Box Cox resulta de gran ayuda para cumplir los supuestos de normalidad y homogeneidad, ya que se obtuvo un 86.27 % de ajuste para normalidad y un 90.2 % para homogeneidad.

**Palabras clave:** Biomasa, Normalidad, Homogeneidad, Heterocedasticidad, Box Cox.

## ABSTRACT

The study was carried out with data from seven species: *Acacia constricta* Benth; *Pinus cembroides* Zucc; *Pinus greggii* Engelm; *Pinus halepensis* Miller; *Prosopis glandulosa* Torr; *Yucca filiera* Chaubad; *Atriplex canescens* (Pursh) Nutt, with the following objectives: a) to demonstrate that applying the Box Cox method provides a better fit in the linear regression models eliminating problems of heterocedasticity in the distribution of the data. b) suggest alternatives of how to proceed with the distribution of non-normal data and the presence of heterocedasticity in linear regression models. To estimate aboveground biomass of these species, six linear regression models were tested, using the normal diameter, cup diameter, basal diameter and the total height as independent variables. The models that showed the best results after the transformation were models 2, 3 and 6 for the assumption of normality and success for all species. Otherwise the homogeneity where only models 3 and 6 expected the assumption for all species and model 2 is only adjusted for 5 species, in the same way model 6 has an adjusted average value of  $R^2 = 0.73$  and model 3 presents an adjusted  $R^2 = 0.79$ , as well as an average standard error of 0.283 kg. For this it is understood that model 3 is the best demonstration fit in all species. The Box Cox transformation method was a great help to meet the assumptions of normality and homogeneity, since an 86.27 % adjustment for normality and 90.2 % for homogeneity is obtained.

**Keywords:** Biomass, Normality, Homogeneity, Heterocedasticity, Box Cox.

## INTRODUCCIÓN

En México los bosques y selvas ocupan una superficie de 56' 851,500 ha, además de ofrecer una gran diversidad de productos maderables y no maderables son considerados como reservorio de carbono contribuyendo en la mitigación del calentamiento global (Ordóñez y Masera, 2001).

Se estima que cada año se pierden alrededor de 155 mil ha de bosques, además se registra que entre 250 y 300 mil ha por año, están siendo degradadas (CONAFOR, 2012).

Las principales causas del incremento del CO<sub>2</sub> atmosférico, se les atribuye a procesos naturales, además de ello también se presenta por las actividades humanas, dado que la tala desmedida de los bosques y la quema de combustibles fósiles como el carbón, petróleo y gas, han causado un incremento de CO<sub>2</sub> contribuyendo al calentamiento global (Caballero *et al.*, 2007).

De igual manera el cambio climático se debe a causas como, cambio de uso del suelo por deforestación, que en México se estiman, 155 mil ha por año (CONAFOR, 2012).

La biomasa forestal desarrolla un papel importante en el manejo forestal sostenible, es por ello que se necesitan estimaciones correctas y precisas basados en los datos específicos de la especie (Cutini *et al.*, 2013), su determinación adecuada es de gran importancia con base a esta podemos determinar la acumulación de carbono y algunos otros de sus componentes químicos (Schlegel, 2001). Existen diferentes métodos para estimar la biomasa arbórea, entre ellos destacan las ecuaciones matemáticas (Pimienta *et al.*, 2007).

Las ecuaciones matemáticas cuantifican la biomasa por medio de modelos de regresión con variables de fácil medición en árboles en pie como son el diámetro normal, diámetro de base, diámetro de copa y altura total (Brown, 1997).

En la mayoría de los estudios para la selección de un modelo se toma como referencia la calidad de ajuste de la variable  $R^2$ , esta se interpreta como la razón entre la varianza explicada por el modelo y la varianza total (Picard *et al.*, 2012). Cancino (2012) menciona que para seleccionar un modelo existen supuestos o pruebas que se deben cumplir en los que destacan las pruebas de normalidad y homogeneidad de varianza.

El supuesto de normalidad es el más importante, se define basándose en que los datos disponibles sigan una distribución de forma normal, que dicho supuesto sea cumplido es de vital importancia ya que es una forma de garantizar la validez de los resultados que se obtengan, ya que de no ser así los estimadores planteados no serían adecuados (González y Abad, 2006).

Para evaluar la normalidad existen muchas formas entre ellas los histogramas y los test de normalidad. El test de Jarque-Bera, utiliza un estadístico en la prueba el cual involucra la kurtosis y la asimetría, además es viable cuando se tratan de muestras grandes, esto con la finalidad de observar la distribución (Jarque y Bera, 1997). El test o prueba de Breusch-Pagan se utiliza para probar la heteroscedasticidad en un modelo de regresión lineal, la prueba consiste en demostrar si las varianzas de los errores residuales dependen de las variables independientes, de ser así la heteroscedasticidad existe en el modelo (García, 2017).

Cuando estos supuestos de normalidad y homogeneidad fallan se tienen diferentes alternativas, las cuales pueden ser, ignorar la violación de dichos supuestos y continuar con

el modelo o encontrar un nuevo modelo que cumpla con los supuestos, con base a una transformación a los datos o a los que sean atípicos de la muestra (Castaño, 2012). El método de transformación de variables Box Cox demuestra su potencial, cuando se presentan variables de respuesta cuya distribución tiene un comportamiento sesgado (heterosedasticidad), que pretenden ser explicadas mediante un modelo lineal (Castaño, 2012).

Gutiérrez (2012), utiliza el método de transformación Box Cox para solucionar problemas de normalidad obteniendo como resultado una considerable mejora en el ajuste del modelo lineal y una mejor homogeneidad en la distribución de los residuales. Por su parte Peña y Peña (1986), realizaron un estudio de los casos en los que se pueden utilizar los diferentes test de normalidad y el uso de la transformación Box Cox para contrarrestar los casos de distribución anormal, obteniendo como resultado que el método es altamente potente y supera en todos los aspectos a los diferentes test de normalidad ya que proporciona alternativas para hacer que las variables se apeguen a una distribución normal y por ende el modelo se ajuste con mayor uniformidad.

En el presente estudio se realizó con la finalidad aplicar el método de transformación de variables Box Cox en seis modelos de regresión lineal, para obtener una ecuación que estime de forma precisa la biomasa aérea, además comprobar que estos cumplan con los supuestos iniciales que hacen válido un modelo y sus parámetros para otros estudios afines. Además, se pretende comprobar la eficiencia de una serie de transformaciones del paquete 'Trafo' de R para solucionar problemas de normalidad y homogeneidad de varianza en modelos lineales y compararlas con la de Box Cox.

## **Objetivos**

- a) Demostrar que aplicando el método de Box Cox se obtiene un mejor ajuste en los modelos de regresión lineal eliminando problemas de heteroscedasticidad en la distribución de los datos.
  
- b) Comprobar alternativas de cómo proceder ante la distribución de los datos no normales y la presencia de heteroscedasticidad en modelos de regresión lineal.

## **Hipótesis**

Ho: La transformación de Box Cox no es eficiente y no corrige problemas de normalidad y homogeneidad de varianza en modelos lineales para estimar biomasa en siete especies forestales.

Ha: La transformación de Box Cox es eficiente y corrige problemas de normalidad y homogeneidad de varianza en modelos lineales para estimar biomasa en siete especies forestales.

## REVISIÓN DE LITERATURA

### Generalidades de las especies

*Pinus cembroides* Zucc (1832)

Nombres comunes: pino piñonero, piñón, pino.

Es un Árbol de 5 a 10 m y hasta 15m de altura, con un diámetro normal de 30 cm y hasta 70 cm. Es de tronco corto y ramas ascendentes, delgadas y distribuidas irregularmente en el tallo, es originaria de México, y se extiende al sur de los Estados Unidos, se caracteriza por ser uno de los pinos con un amplio rango de distribución en México (19 estados), desarrolla masas puras en la sierra madre oriental al norte del trópico de cáncer. Las mayores poblaciones se encuentran en Chihuahua, Durango, Coahuila, Nuevo León, Hidalgo, y Zacatecas. Altitud: 1,350 a 2,800 m (Montoya, 1990).

*Pinus halepensis* Miller

Nombre común: pino carrasco, pino halepo.

Son árboles de 15 a 20 m de altura y hasta 70 cm de diámetro normal. Es una especie introducida en varios estados de la República, tiene su origen en la zona mediterránea de Europa, Asia y África se desarrolla a una altitud sobre el nivel del mar desde 2,200 msnm con una precipitación que va desde los 4000 mm a 8000 mm, es utilizado para proteger suelos deficientes de humedad, en la fijación de dunas, terrenos erosionados y con un alto potencial para cortinas rompe vientos de corta altura, ornamental para parques y jardines (Bravo *et al.*, 2004).

*Pinus greggii* Engelm.

Nombres comunes; pino greggii, palo prieto, pino prieto, pino ocote.

Especie nativa de la sierra madre oriental en los estados de Coahuila, Nuevo León, San Luis Potosí e Hidalgo, árbol pequeño de 10 a 25m de altura, de rápido crecimiento, se desarrolla sobre una altitud máxima de 3,000 a 2,550 msnm es asociada con bosque de *Quercus* y bosque de coníferas, su madera es destinada a la industria de celulosa y aserrío, para la fabricación de muebles, durmientes pilones, vigas, Etc., también se utiliza como especie ornamental y en algunas localidades se utiliza como árbol navideño (Rodríguez *et al.*, 2013)

*Prosopis glandulosa* Torr

Nombre común: Mezquite dulce.

Es una especie con porte mediano a pequeño, normalmente alcanza de 5 a 9 m de altura y en ocasiones puede alcanzar los 14 m, su floración es de marzo a noviembre, con espigas pálidas, amarillas elongadas. Es de origen nativo de la república, presentándose principalmente en los estados; Coahuila, Durango, Chihuahua, Sonora, Nuevo León y norte de Tamaulipas, se usa en el noreste de México como fin alimenticio, y como condimento para comidas entre otros (Carmona *et al.*, 2007).

*Atriplex canescens* (Pursh) Nutt

Nombres comunes; Cenizo, Costilla de vaca, Chamizo, saladillo. Es una especie originaria de las zonas áridas de América del Norte, es un arbusto de 0.15 a 2 m de alto, con un diámetro del tallo hasta 8 cm, especie dioica perenne y leñosa, generalmente erecto, densamente ramificado desde la base, especie de activo crecimiento con frecuencia presenta



raíces gruesas, y presenta asociación con micorrizas del tipo vesículo arbuscular. Se presenta en todos los hábitats áridos y semiáridos, en áreas perturbadas, generalmente su crecimiento se da formando pequeños manchones, en ocasiones forma masas puras de consideración, es utilizada como un forraje natural en regiones áridas del mundo principalmente en la alimentación y engordamiento de cabras (Urrutia *et al.*, 2014).

*Yucca filifera* Chaubad.

Nombres comunes; Izote, Palma corriente, Palma grande y Palma china.

Es una especie nativa de México se distribuye desde el norte hasta el centro del país, alcanza una altura de 10 m y una categoría diamétrica de hasta 90 cm, el tallo es monopodico y robusto en su parte inicial posteriormente se ramifica y se cubre y protege en gran parte por las hojas muertas, se ve afectada por competencia interespecífica con nanofanerófitas, caméfitas, y hemicriptófitas incluyendo *Larrea divaricata*, *Flourenzia cernua*, *Prosopis glandulosa*, entre otras, debido a que esta especie crece en las partes altas y medias de las laderas. Principalmente la flor y el fruto se utiliza como alimento, la inflorescencia se utiliza como forraje, al igual se extrae aceite comestible de la semilla y como celulosa en la fabricación de papel kraft (Nava *et al.*, 1980).

*Acacia constricta* Benth

Nombre común: Acacia de espina blanca, Huizache. Es una planta o arbusto de 3 a 4 m de altura con espinas pareadas en nudos que envejecen, se distribuye desde el norte hasta el sur de México se desarrolla principalmente en laderas secas, lavados, áreas desérticas planas y mesas, a menudo en suelos de caliche poco profundos, elevaciones medias y altas (Rico, 2001).

## **Normalidad**

El supuesto de normalidad es considerado quizá uno de los más importantes, ya que para su cumplimiento los datos disponibles deben seguir en forma conjunta una distribución normal. Dicho cumplimiento garantiza al investigador la validez de su estudio y por ende la validez de los resultados que obtenga. Para ello se puede estudiar de diferentes maneras, una de ellas es realizando una inspección visual con ayuda de histogramas y de esta forma observar si la distribución de los datos se asemeja a la campana de Gauss (campaniforme y simétrica). Más, sin embargo, los histogramas presentan deficiencias al estimar normalidad ya que estos dependen del número y amplitud de los intervalos que se consideren, en este sentido para observar normalidad desde una perspectiva más directa, es necesario realizar test de normalidad entre ellos el test de Jarque-Bera, Kolmogorov-Smirnov, Lilliefors y Shapiro -Wilks (González y Abad, 2006).

La distribución normal es la piedra angular de las aplicaciones estadísticas. Los datos en muchos casos no se ajustan a dicho criterio, en los casos en que el supuesto no se ha cumplido, los investigadores en muchas ocasiones recurren a transformar los datos no normales a datos que sigan una distribución normal, el método de transformación Box Cox ofrece un método simple para obtener la transformación de potencia más adecuada para el estudio (Olivier y Norberg, 2010).

## **Homogeneidad**

Uno de los supuestos que se requieren como prioridad en aplicaciones estadísticas, tales como el análisis de varianza, análisis de regresión, etc., es el de homogeneidad de varianza. Dicho supuesto es de vital importancia para garantizar la calidad de los resultados que se

obtengan en el estudio estadístico utilizado en pruebas de hipótesis y construcción de modelos matemáticos de composición lineal (Correa *et al.*, 2006). El test de Breush-Pagan es utilizado en estadística para probar la heteroscedasticidad en un modelo de regresión lineal. El test prueba si la varianza de los errores de una regresión depende de los valores de las variables independientes, de ser así la heteroscedasticidad está presente en el modelo y se recurre a la utilización de transformaciones que logren cumplir la homogeneidad de varianza en el modelo. Dicho test es una prueba de Chi Cuadrado, prueba la hipótesis nula planteada si el valor de la Chi Cuadrado es significativo con valores de “p” por debajo de un umbral apropiado entonces se asume la heteroscedasticidad (García, 2017).

### **Box Cox**

En ocasiones el análisis de datos supone que las observaciones tienden a seguir una distribución normal de manera independiente con varianza constante y con expectativas especificadas por un modelo lineal para un conjunto de parámetros. Para el análisis de modelos como son la varianza y el análisis de regresión múltiple generalmente se justifican con los supuestos de que existe una constancia de la varianza del error, la normalidad de las distribuciones y la independencia de los datos. Cuando dichos supuestos son violados o incumplidos, con el uso de una transformación no lineal de la variable “y” puede mejorar las cosas en gran medida, en estos casos lo preocupante no solo se basa en encontrar una transformación que justifique los supuestos, sino más bien encontrar cuando es posible una métrica en términos del cual los hallazgos puedan ser expresados de manera correcta y precisa, que hagan válido los resultados del estudio y justifique el uso del modelo lineal en cuestión (Box y Cox, 1964).

## **R studio**

Es un lenguaje de programación interpretado, su distribución es libre bajo licencia GNU. Fue creado por Ross Ihaka y Robert Gentleman en 1992 (Team, 2013) para realizar análisis estadísticos y gráficos, posee muchas funciones en las que destaca por permitir la visualización de sus gráficos de manera inmediata en su propia ventana, además de ser exportados en diferentes formatos, tiene una interfaz agradable y fácil de usar, su principal característica es su gran flexibilidad. Algunos programas clásicos arrojan los resultados de un análisis estadístico de forma directa, R guarda los resultados como un objeto, permitiendo realizar muchos análisis y ejecutar comandos continuamente antes de ver los resultados (Santana y Mateos, 2014).

### **Paquete Trafo: otras transformaciones**

Los modelos de regresión lineal se han utilizado para describir, predecir e inferir fines esenciales, los cuales se basan en un grupo de supuestos que no siempre se cumplen cuando se tratan de datos empíricos. Para mejorar la validez de dichos supuestos el uso de transformaciones es el método más simple. La paquetería 'Trafo' de R ofrece una manera sencilla y práctica de cómo elegir la mejor transformación que realmente es adecuada acoplándose a las necesidades del usuario, además combina y extiende las características proporcionadas por cualquier otro paquete de transformación. Así mismo algunas otras ventajas es que realiza una verificación inicial lo cual ayuda a la selección de la transformación, realizando la comparación del modelo en su forma normal y su forma transformada y finalmente proporciona un diagnóstico para verificar si la transformación es útil y cumple con los supuestos del modelo (Medina *et al.*, 2017).

## Estudios afines

Olivier y Norberg (2010) en el estudio Positively Skewed Data: Revisiting the Box-Cox Power Transformation. Comparan la transformación de potencia de Box Cox y la distribución ex Gaussiana con la finalidad de discutir y poner en manifiesto cuál de los métodos es más simple de aplicar y de interpretar. Concluyen que el método de transformación de potencia Box Cox consigue normalizar los datos y es de fácil aplicación e interpretación en comparación con la distribución ex gaussiana. Además, mencionan que no es la mejor transformación más sin embargo es vasta y sencilla de aplicar.

Avanza et al. (2003) en un estudio de aplicación de transformaciones para el cumplimiento de los supuestos de normalidad y homosedasticidad, a concentraciones foliares en, N, P y K en mandarino (*Citrus reticulada*, Blanco), cuyas variables en el estudio son proveniente de 12 muestras fructíferas, evaluaron la normalidad utilizando el test de Shapiro-Wilk y el test de Shapiro-Wilk modificado por Mahibbur y Govindarajulu (1997) y para el caso de homogeneidad utilizaron los test de Bartlett y Levene, probando las transformaciones, logaritmo natural, raíz cuadrada y potencia de Box y Cox. Concluyendo que las transformaciones de potencia de Box y Cox son las que logran establecer la normalidad para N y K, en cambio para P no se consiguió normalidad con ninguna de las transformaciones probadas. Al igual para el caso de homogeneidad Box Cox, sigue demostrando mejores resultados ya que es el único que cumple con ambos supuestos con valores de p. value = 0.2 a 0.3 para el supuesto de normalidad y de 0.19 a 0.5 para homogeneidad para N y K respectivamente. Concluyendo que la transformación de Box Cox con valores del parámetro  $\lambda$  negativos o cercanos a cero, se obtienen mejores resultados.

Peña y Peña (1986) llevan a cabo el estudio de contraste de normalidad basado en la transformación Box-Cox con la finalidad de ver las posibilidades de utilizar las transformaciones de Box Cox para conseguir la normalidad de una variable aleatoria. Concluyendo que Box Cox supera a todos sus oponentes en el caso de distribuciones claramente asimétricas para todos los tamaños de muestra, además presenta la ventaja que no tiene ninguno de sus oponentes y es que facilita una transformación de los datos, lo cual permite trabajar con estos como si fuesen aproximadamente normales en la mayoría de los casos.

## MATERIALES Y MÉTODOS

### Descripción del área de estudio

El presente trabajo se desarrolló en tres localidades, ubicadas en el norte del país (Arteaga, Saltillo. Coahuila y Mazapil en Zacatecas) los datos (temperatura, precipitación, clima y tipo de suelo), fueron tomados de INEGI (1990), INIFAP y CONABIO (1995).

El muestreo se desarrolló en plantaciones forestales: El Centro Agrícola Experimental Sierra de Arteaga (CAESA) perteneciente a la Universidad Autónoma Agraria Antonio Narro (UAAAN) y la Reforestación también perteneciente a la UAAAN; a diferencia de la localidad de Mazapil, Zacatecas donde las especies se distribuyen de forma natural.

**Cuadro 1.** Características generales del área de estudio.

Sitio	Especies	Altitud	Tm	TM	TP	Pp	S	C
Mazapil.	<i>Aci; Acm; Atc;</i>							
Zac.	<i>Pgs; Pcn; Yfl</i>	2200	7.5	24.6	22	500	Litosol Feozem	BS1K(X')
UAAAN <sup>1</sup>	<i>Pgi; Pcp</i>	2230	7.1	23.4	14	262	calcárico Xerosol	BS1K(X')
UAAAN <sup>2</sup>	<i>Phs</i>	1820	11.2	25.6	20	262	cálcico	Bsohw

Donde: *Aci* = *Acacia constricta* Benth (ind); *Acm* = *Acacia constricta* Benth (manchón); *Atc* = *Atriplex canescens* (Pursh) Nutt; *Pgs* = *Prosopis glandulosa* Torr; *Pcn* = *Pinus cembroides* Zucc (natural); *Yfl* = *Yucca filifera* Chaubad; *Pgi* = *Pinus greggii* Engelm; *Pcp*=*Pinus cembroides* Zucc (plantación); *Phs* = *Pinus halepensis* Miller; Tm =

temperatura mínima (°C); TM = temperatura máxima (°C); Tp = temperatura promedio (°C); Pp = precipitación (mm); S = tipo de suelo; C = tipo de clima;

UAAAN<sup>1</sup> (Universidad Autónoma Agraria Antonio Narro) = Instalaciones del CAESA (Centro Agrícola Experimental Sierra de Arteaga); UAAAN<sup>2</sup> = Reforestación (Saltillo, Coahuila).

### Método directo para obtener biomasa de las especies

Para realizar este estudio se siguió la metodología propuesta por Díaz et al. (2007) y Nívar (2011), para ello se seleccionaron 531 ejemplares cuyas especies y cantidad se resumen en el Cuadro 2.

**Cuadro 2.** Especies utilizadas en el estudio.

Espece	No. de árboles
<i>Acacia constricta</i> Benth (Ind.)	49
<i>Acacia constricta</i> Benth (Manchón)	14
<i>Pinus cembroides</i> Zucc (Natural)	29
<i>Pinus cembroides</i> Zucc (Plantación)	48
<i>Pinus greggii</i> Engelm	250
<i>Prosopis glandulosa</i> Torr.	30
<i>Yucca filifera</i> Chaubad	31
<i>Atriplex canescens</i> (Pursh)Nutt	51
<i>Pinus halepensis</i> Miller	29
Total	531



De estos ejemplares se registraron las variables de peso verde del fuste (PVF), peso verde en ramas (PVR) y el peso verde total (PVT). De las especies arbóreas se hizo la selección cuidando que el árbol no mostrara deformidades ni enfermedades, de las arbustivas que estuviesen bien conformadas, procurando evaluar todas las categorías diamétricas de diámetros y alturas presentes en los diferentes predios.

Antes del derribo se tomaron los datos de las variables dendrométricas como el diámetro a la altura del pecho (DAP) a 1.30 m sobre el suelo y el diámetro basal (DB) a 0.20 m, con ayuda de una cinta diamétrica; posteriormente al derribo de los árboles se procedió a medir la altura total (HT) de los individuos con la ayuda de una cinta métrica de 30 m. Se utilizó un flexómetro de 8 m de longitud para las especies de porte menor o arbustivas; el diámetro de copa se midió utilizando una cinta métrica de 30 m de longitud.

Obtenidas las mediciones dendrométricas de cada individuo de las especies arbóreas, se separaron en componentes de biomasa: fuste y ramas, para ello los fustes de los árboles fueron seccionados a una longitud de 2.55 m, siendo las últimas de 1.25 m considerando las medidas aceptadas para el comercio. Para el pesado se utilizó una báscula romana con capacidad de 120 kg, cada troza se pesó en fresco para determinar el PVF (sumando los pesos de cada troza), PVR (todas las ramas juntas) y para el PVT (suma de PVF+ PVT) Díaz et al. (2007) y Návar (2011).

Para el caso de las especies arbustivas la biomasa se pesó sin seccionarse, los componentes de biomasa verde sin excepción se obtuvieron en el sitio del derribo.

## ANÁLISIS ESTADÍSTICOS

### Modelos de regresión para cuantificar biomasa en las siete especies

Para realizar las pruebas y aplicación del método de transformación de variables Box Cox (1964) se seleccionaron seis modelos alométricos en su forma logarítmica propuestos por Picard et al. (2012), por ser modelos lineales de fácil estimación de parámetros. Además, estos modelos ya han sido probados en otro estudio, por ejemplo, Soriano et al., (2015) utilizan algunos de estos modelos para la estimación de biomasa aérea por componente estructural.

**Cuadro 3.** Modelos matemáticos a evaluar biomasa aérea de siete especies en su forma logarítmica.

M	Tipo de ajuste	Ecuación	Varianza
M1	Lineal simple	$\text{Ln}(\text{BT}) = b_0 + b_1 \text{Ln}(\text{DN}) + \varepsilon$	$\text{Var}(\varepsilon) = \sigma^2$
M2	Lineal simple	$\text{Ln}(\text{BT}) = b_0 + b_1 \text{Ln}(\text{DN}^2 \text{HT}) + \varepsilon$	$\text{Var}(\varepsilon) = \sigma^2$
M3	Polinomial	$\text{Ln}(\text{BT}) = b_0 + b_1 \text{Ln}(\text{DN}) + b_2 \text{Ln}(\text{DN}^2) + \varepsilon$	$\text{Var}(\varepsilon) = \sigma^2$
M4	Múltiple	$\text{Ln}(\text{BT}) = b_0 + b_1 \text{Ln}(\text{DN}) + b_2 \text{Ln}(\text{HT}) + \varepsilon$	$\text{Var}(\varepsilon) = \sigma^2$
M5	Lineal ponderada	$\text{BT} = b_0 + b_1 \text{DN}^2 \text{HT} + \varepsilon$	$\text{Var}(\varepsilon) = \alpha D^{2c}$
M6	Polinomial ponderada	$\text{BT} = b_0 + b_1 \text{DN} + b_2 \text{DN}^2 + \varepsilon$	$\text{Var}(\varepsilon) = \alpha D^{2c}$

Donde: M = modelo de regresión lineal, Ln = logaritmo natural, BT = Biomasa total (kg), DN = Diámetro normal (cm), HT = altura (m),  $b_0$ ,  $b_1$  y  $b_2$  = Coeficientes de regresión,  $\varepsilon$  = error,  $\sigma$  = varianza,  $\alpha$  y  $c$  = coeficientes de ponderación.

Para realizar el ajuste de los modelos se utilizaron las variables de diámetro normal en (cm) y diámetro de copa en (cm) y altura total en (m) respectivamente en cada especie, los parámetros  $b_0$  representan la ordenada al origen y  $b_1$  y  $b_2$ , representan la pendiente del modelo de regresión respectivamente para cada una de las especies (Acosta *et al.*, 2002). La pendiente establece la relación de la tasa de crecimiento de la variable respuesta con respecto a la variable explicativa para un tiempo determinado (Niklas, 2004).

Utilizando el programa R Studio versión 8.12 (Ihaka y Gentleman, 1996) para Windows, se realizaron los ajustes para cada uno de los modelos de regresión, con el comando `linear model (lm)` por ser modelos linealizados en su forma logarítmica, comparando todas las variables dependientes con la independiente según fuese el caso de cada especie.

### **Prueba de normalidad**

Posterior a ello se procedió a comprobar el primer supuesto (normalidad residual) con el paquete 'tseries' (Trapletti *et al.*, 2009) en R Studio. Dicha paquetería contiene el test de normalidad de Jarque Bera, cabe mencionar que se optó por el test de Jarque Bera, ya que es uno de los mejores para testar la normalidad de los residuos de un modelo lineal, esto debido a que utiliza un estadístico en la prueba el cual involucra la Kurtosis y la Asimetría (Jarque y Bera, 1987). Además, es viable cuando se tratan de muestras grandes. El planteamiento de la hipótesis de normalidad se define en:

Ho: La muestra no proviene de una distribución normal.

Ha: La muestra proviene de una distribución normal.

Para pruebas de normalidad siempre se plantean de ésta manera las hipótesis.

El criterio que define si la hipótesis se rechaza o se acepta se resume en el valor de p, si p. value < 0.05 se rechaza la hipótesis nula, y se entiende que los residuales no siguen una distribución normal; si p. value > 0.05 los residuales siguen una distribución normal, se acepta la hipótesis.

### **Prueba de homogeneidad de varianza**

De igual manera se realiza la prueba de homogeneidad de varianzas la cual asume que la varianza de los valores de “y” es una constante igual a  $\sigma^2$ , independiente del valor de “x”. Cancino (2012) define dos formas de ajustar modelos heterogéneos.

1. Modelar la heterogeneidad del modelo, utilizando regresión ponderada o cuadrados mínimos ponderados.
2. Ignorar la heterogeneidad usando métodos de ajuste que resistan la heterogeneidad.

Para realizar la prueba de homogeneidad y comprobar si existe heterogeneidad se optó por el método de Breusch-Pagan con el paquete ‘lmtest’ (Zeileis y Hothorn, 2002), este comando realiza la prueba de homogeneidad de varianza del modelo seleccionado.

De igual forma la homogeneidad de varianza está dada por la variable de p. value y se define en:

Si Valor p > 0.05 existe homogeneidad de varianza, los datos siguen una varianza constante.

Si Valor p < 0.05 no existe homogeneidad de varianza, los datos no tienen una varianza constante.

## **Procesamiento de datos**

Los resultados fueron capturados en una base de datos de Excel 2013, almacenando las variables de interés del modelo las cuales fueron: los coeficientes de regresión ( $b_0$ ,  $b_1$  y  $b_2$ ) según corresponda a cada modelo, de igual manera se capturó el grado de significancia de cada coeficiente (valor  $p$ ), si los coeficientes de regresión son significativamente diferentes de cero, planteándose en la siguiente hipótesis.

Si valor  $p > 0.05$  el coeficiente no es significativamente diferente de cero.

Si valor  $p < 0.05$  el coeficiente es significativamente diferente de cero.

La desviación estándar residual (RSCE), el coeficiente de determinación ( $R^2$ ajustada), la prueba de Fisher (F) y el valor P y las pruebas de normalidad y homogeneidad para cada modelo.

Para el caso de los modelos logarítmicos, fue necesario recalcular el coeficiente de determinación ( $R^2$  ajustada) ya que el estimado por el programa está en logaritmo natural y no es un valor real, para regresar los estimados a su escala normal es necesario aplicar el antilogaritmo a los valores y realizar el análisis de varianza nuevamente para así obtener el valor de  $R^2$  ajustada real.

### **Método de transformación de variables Box Cox (1964)**

Para la aplicación del método de transformación Box Cox se pueden transformar tanto la variable dependiente como las variables explicativas (DN, DC, HT). En el presente trabajo solo se transformó la variable de respuesta en este caso biomasa total (BT), siguiendo la metodología de transformación de variables propuestas por Box y Cox (1964). Que Se

estima en primera instancia el valor de lambda ( $\lambda$ ) con una serie de comandos, por lo general el valor estimado de lambda se encuentra de (-1 a 1). Una vez obtenido el valor de lambda se procedió a elevar la BT al valor obtenido y correr el modelo para obtener los residuales y realizar las pruebas de normalidad y homogeneidad de varianza, esto con la finalidad de observar si existe una mejora en el modelo ajustado con la transformación. De igual forma se almacenaron los resultados en una base de datos de Excel 2013.

### **Recalculando $R^2$ ajustada**

Para el caso del coeficiente de determinación ( $R^2$ ajustada) de igual forma se necesita hacer la conversión de los valores predichos por el modelo a su escala original y volver a estimar la  $R^2$ ajustada con el análisis de varianza, para dicha conversión se utilizó la siguiente fórmula:

$$(\text{Predichos})^{\frac{1}{\lambda}}$$

Donde: Predichos = predichos del modelo,  $\lambda$  = valor estimado de lambda para el modelo,

Nota: el valor de lambda es diferente para cada modelo y por supuesto para cada especie.

De esta manera se convierten los estimados con Box Cox a estimados reales a su escala normal y posterior recalculan el coeficiente de determinación ( $R^2$  ajustada). Con esto es posible evaluar la calidad de ajuste de cada modelo considerando la  $R^2$  ajustada en conjunto con las pruebas de normalidad y homogeneidad de varianza.

## **La librería Trafo: otras transformaciones**

Los modelos lineales que no presentaron mejoría después de aplicar la transformación Box Cox, se presenta una alternativa de cómo proceder y de esta forma encontrar una solución antes de desechar el modelo. Dicha alternativa consta de la paquetería de transformaciones 'Trafo' de R studio (Medina *et al.*, 2017) la cual realiza 13 transformaciones diferentes incluidas Box Cox. Para ello solo se les aplicó a los modelos que no pasaron los dos supuestos que se buscan cumplir aplicando la transformación de variables Box Cox.

## RESULTADOS

En el presente estudio se realizó el ajuste de seis modelos de regresión lineal para la estimación de biomasa aérea, con la finalidad de evaluar el método de transformación de variables propuesto por Box y Cox (1964). Dicho método de transformación es utilizado para hacer cumplir los supuestos de normalidad y homogeneidad de varianza que hacen válido un modelo lineal, ya que la normalidad es considerada como la piedra angular de cualquier modelo que pretenda explicar la relación que existe entre una o más variables.

En primera instancia se procesaron los modelos en su forma linealizada con Ln (logaritmo natural) (Cuadro 3), esto con la finalidad de observar y determinar los modelos y las especies que no cumplieran normalidad y homogeneidad. Con base a estos resultados poder compararlo con los resultados obtenidos después de aplicarles la transformación Box Cox y con ello comprobar la hipótesis planteada, la cual define que con la transformación Box Cox se obtienen resultados favorables que hacen cumplir los supuestos de normalidad y homogeneidad residual. Para los modelos logarítmicos y no logarítmicos se obtuvieron los siguientes resultados.



**Cuadro 4.** Resumen estadístico para la prueba de normalidad.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Total
M1	1	1	1	1	1	1	1	0	1	8
M2	1	1	1	1	NAP	0	1	0	1	6
M3	1	1	1	1	1	1	1	0	1	8
M4	1	1	1	1	NAP	0	1	0	1	6
M5	0	1	1	1	NAP	0	0	1	0	4
M6	0	1	1	0	0	0	1	0	0	3
Total	4	6	6	5	2	2	5	1	4	35

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica). Nota: el valor 1 indica que el modelo cumple el supuesto de normalidad y el 0 indica que no se cumple el supuesto.

Como se puede observar en el Cuadro 4, los modelos 1 y 3 cumplen el supuesto de normalidad para 8 de las 9 especies seguidos de M2 y M4 con 6 de 9 especies el modelo y los modelos 5 y 6 solo se cumple el supuesto en 4 y 3 especies respectivamente, de las 9 especies para *Acm* y *Atc* cumplen el supuesto de normalidad para todos los modelos caso contrario de la especie *Pgs* la cual solo cumple este supuesto para el modelo 5. Cabe señalar que las especies con más problemas de normalidad fueron, *Pgi* y *Pgs* los cuales solo se ajustan en 2 y 1 modelos respectivamente. Por su parte que para la especie *Pcp* solo se aplicaron los modelos 1, 3 y 6, debido a que solo se cuenta con datos de diámetro y

biomasa y no para HT, el cual es necesario para los modelos 2, 4 y 5 tal y como se observa en la ecuación matemática respectivamente de cada modelo (Cuadro 3). De los aplicados solo 2 cumplen el supuesto de normalidad.

Resumiendo, las posibilidades de cumplir el supuesto de normalidad residual son 51, de las cuales solo se cumple en 35, representando el 68.63 % del total.

De igual manera se obtuvieron los siguientes resultados para homogeneidad de varianza los cuales se resumen en el cuadro 5.

**Cuadro 5.** Resumen estadístico de la prueba de homogeneidad.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Total
M1	1	1	1	1	1	1	1	1	1	9
M2	1	1	1	1	NAP	1	1	1	1	8
M3	1	1	1	1	1	1	1	1	1	9
M4	1	1	1	1	NAP	1	1	1	1	8
M5	0	1	1	0	NAP	0	0	0	1	3
M6	0	1	0	0	0	0	0	1	0	2
Total	4	6	5	4	2	4	4	5	5	39

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica). Nota: el valor 1 indica que el modelo cumple el supuesto de homogeneidad y el 0 indica que no se cumple el supuesto.

Como se observa en el cuadro anterior, los modelos 1 y 3 cumplen el supuesto para las 9 especies, seguidos por los modelos 2 y 4 con 8 especies, caso contrario de los modelos 5 y 6, los cuales se ajustan para 3 y 2 especies únicamente. Resumidos por especie tenemos que para *Acm* pasa la prueba para todos los modelos, para *Atc*, *Pgs* y *Yfl* se ajustan en 5 modelos, más sin embargo para *Aci*, *Pcn*, *Pgi* y *Phs* solo se cumple el supuesto para 4 de 6 modelos.

De la misma forma se calcula las posibilidades de cumplir el supuesto de homogeneidad en un total de 51 posibilidades, pero solo se cumple en 39 de ellas, lo cual representa el 76.47 % del total.

Para evaluar la calidad de ajuste de un modelo autores como Nívar et al. (2001); Da Cunha y Guimares (2009) y Curiel et al. (2007) utilizan el valor observado del coeficiente de determinación ( $R^2$  ajustada).

Martínez (2005) demuestra con ejemplos el grave error que se comete cuando se utiliza el coeficiente de determinación  $R^2$  como única medida de ajuste o grado de fiabilidad de un modelo para un conjunto de datos. Dicho coeficiente se interpreta como la razón entre la varianza explicada por el modelo y la varianza total (Picard *et al.*, 2012), el cual presenta valores que se encuentran en los rangos de 0 a 1. Un  $R^2$  igual a 1 indica un ajuste lineal perfecto, ya que la suma de cuadrados total es igual a la suma de cuadrados del error ( $STC = SEC$ ), el cual se interpreta como la variación total de la variable “Y” es explicada por el modelo de regresión.

Un valor de cero indica que el modelo no puede explicar nada de la variación, al igual se observa el valor del error estándar residual del modelo, el cual se busca que este sea el mínimo en los estudios con modelos de regresión. En este sentido, se hicieron los cálculos

correspondientes para estimar dichos parámetros y de esta manera observar el ajuste de cada modelo, para posteriormente observar el comportamiento después de aplicar la transformación. Los resultados se observan en el Cuadro 6.

**Cuadro 6.** Coeficiente de determinación  $R^2$ ajustada de los modelos por especie.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Promedio
M1	0.89	0.56	0.53	0.37	0.84	0.74	0.95	0.89	0.84	0.74
M2	0.79	0.66	0.39	0.39	NAP	0.77	0.94	0.79	0.80	0.69
M3	0.91	0.57	0.60	0.62	0.86	0.76	0.94	0.91	0.92	0.79
M4	0.85	0.77	0.57	0.39	NAP	0.77	0.95	0.85	0.88	0.75
M5	0.89	0.69	0.53	0.48	NAP	0.76	0.74	0.89	0.94	0.74
M6	0.91	0.61	0.64	0.64	0.86	0.76	0.95	0.91	0.94	0.80
Promedio	0.88	0.64	0.54	0.48	0.86	0.76	0.91	0.88	0.89	0.75

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica).

En resumen, el valor promedio de  $R^2$  ajustada de modelos por especie son: M1 = 0.74, M2 = 0.69, M3 = 0.79, M4 = 0.75, M5 = 0.74, M6 = 0.80, como se puede observar en el cuadro 6 los valores promedio de  $R^2$ ajustada van de 0.69 para el modelo 2 y 0.80 para el modelo 6.

Continuando con los resultados se obtuvo el error estándar promedio de los modelos por todas las especies obteniéndose los resultados presentados en el Cuadro 7.

**Cuadro 7.** Error estándar de los modelos (RSCE) de regresión lineal por especie.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Promedio
M1	0.411	0.760	0.581	0.404	0.193	0.365	0.168	0.363	0.421	0.407
M2	0.337	0.642	0.629	0.397	NAP	0.351	0.174	0.346	0.303	0.397
M3	0.335	0.784	0.572	0.317	0.183	0.359	0.166	0.366	0.425	0.390
M4	0.329	0.570	0.572	0.405	NAP	0.351	0.168	0.346	0.276	0.377
M5	0.120	3.896	0.807	3.978	NAP	3.553	0.091	0.075	0.022	1.568
M6	0.132	1.063	0.635	0.098	0.077	0.112	0.051	4.495	2.836	1.055

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica).

Como se puede apreciar en el Cuadro 7, el valor promedio más bajo para el error estándar (RSCE) lo presenta el M4 con un valor de RSCE = 0.377 kg y el valor promedio más alto lo presenta el M5 con un valor de 1.568 kg dicho coeficiente demuestra la calidad de ajuste de los modelos a mayor valor presenta el error estándar menor es el ajuste del modelo, por lo consiguiente se espera que los modelos tengan un margen de error bajo para tener un mejor ajuste en los modelos lineales. la especie *Acm* presenta el error más alto y *Phs* demuestra el error más bajo de todas las especies, esto podría deberse al tamaño de la muestra en cada uno de los casos, por lo que a menor tamaño maestral mayor error y viceversa.

### Resultados con la transformación Box Cox (1964)

Los resultados indicaron que los modelos presentaron una mejoría en la distribución de los residuales después de aplicar la transformación Box Cox (1964), los resultados para normalidad se resumen en el Cuadro 8.

**Cuadro 8.** Resultados de la prueba de normalidad residual después de la transformación.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Total
M1	1	1	1	1	1	1	1	0	0	7
M2	1	1	1	1	NAP	1	1	1	1	8
M3	1	1	1	1	1	1	1	0	1	8
M4	1	1	1	1	NAP	1	1	0	1	7
M5	1	1	1	1	NAP	0	1	1	0	6
M6	1	1	1	1	1	1	1	0	1	8
Total	6	6	6	6	3	5	6	2	4	44

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica). Nota: el valor 1 indica que el modelo cumple el supuesto de normalidad y el 0 indica que no se cumple el supuesto.

Los modelos que mejor resultados obtuvieron fueron: M2, M3 y el M6, ya que se aprobaron los supuestos de normalidad en 8 de las 9 especies, presentando valores de  $p = 0.27$  a  $0.97$  para el modelo 2, de  $0.27$  a  $0.71$  para el modelo 3 y de  $0.23$  a  $0.63$  para el modelo 6. De igual manera los modelos: M1 y M4, se ajustaron a 7 de las 9 especies, el

modelo que menos mejoría presentó fue el M5, ya que solo se ajustó en 6 de las 9 especies. De las 9 especies únicamente para *Aci*: (*A. constricta* ind.); *Acm*: (*A. constricta* manchón); *Atc*: (*A. canescens*); *Pcn*: (*P. cembroides* natural) y *Phs*: (*P. halepensis*), se aceptaron los supuestos de normalidad para todos los modelos, a excepción de *Pgs*: (*P. glandulosa*), la cual solo mejoro en 2 de los 6 modelos estudiados.

A continuación, se presenta el cuadro de resumen de los resultados obtenidos para la prueba de homogeneidad de varianza de los modelos por especie, después de aplicar la transformación de potencia Box Cox.

**Cuadro 9.** Resultados de homogeneidad de varianza de los modelos por especie después de la transformación.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Total
M1	0	1	1	1	1	1	0	1	1	7
M2	0	1	1	1	NAP	0	0	1	1	5
M3	1	1	1	1	1	1	1	1	1	9
M4	1	1	1	1	NAP	1	1	1	1	8
M5	1	1	1	1	NAP	1	1	1	1	8
M6	1	1	1	1	1	1	1	1	1	9
Total	4	6	6	6	3	5	4	6	6	46

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta* manchón), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica). Nota: el valor 1

indica que el modelo cumple el supuesto de homogeneidad y el 0 indica que no se cumple el supuesto.

Los modelos de regresión lineal (6 y 3) presentaron mejoría para las 9 de las especies, obteniéndose valores de  $p = 0.11$  a  $0.99$  para el modelo 3 y  $0.21$  a  $0.67$  para el modelo 6. Los modelos (4 y 5) en 8 especies. En general la mayoría de los modelos cumple el supuesto de homogeneidad, a excepción del modelo 2 el cual solo se hace presente en 5 de 9 especies. Ahora bien, las especies *Acm*, *Atc*, *Pcn*, *Pgs* y *Yfl*, cumplen el supuesto de homogeneidad para todos los modelos aplicados, sin embargo, las especies *Aci* y *Phs*, solo se cumplen para 4 de 6 modelos. De igual manera se estima también el coeficiente de determinación  $R^2$  ajustada para los modelos transformados.

**Cuadro 10.** Valores del coeficiente de determinación ( $R^2$  ajustada) de los modelos por especie al aplicar la transformación.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Promedio
M1	0.91	0.59	0.62	0.47	0.86	0.76	0.94	0.91	0.93	0.78
M2	0.91	0.67	0.44	0.65	NAP	0.78	0.94	0.91	0.61	0.74
M3	0.91	0.58	0.62	0.64	0.86	0.76	0.95	0.91	0.91	0.79
M4	0.92	0.88	0.66	0.50	NAP	0.80	0.95	0.92	0.98	0.83
M5	0.77	0.59	0.39	0.45	NAP	0.57	0.85	0.77	0.73	0.64
M6	0.87	0.41	0.60	0.64	0.80	0.55	0.94	0.87	0.90	0.73

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica).



Como se puede observar en el Cuadro 10, los valores promedio de  $R^2$  ajustada van de 0.64 para el modelo 5 y con el valor más alto de 0.83 en este caso lo obtiene el modelo 4.

El modelo que mejor se ajustó en casi todas las especies fue el M4, y el modelo que menor ajuste demostró fue el M5, ya que es el que presenta el valor promedio más bajo entre las especies con valores de  $R^2$  ajustada de 0.40 en algunas especies.

**Cuadro 11.** Resultados del error estándar de los modelos por especie aplicando la transformación.

Modelo	<i>Aci</i>	<i>Acm</i>	<i>Atc</i>	<i>Pcn</i>	<i>Pcp</i>	<i>Pgi</i>	<i>Phs</i>	<i>Pgs</i>	<i>Yfl</i>	Promedio
M1	0.075	1.081	0.235	0.039	0.086	0.117	0.181	0.745	1.777	0.482
M2	0.176	2.306	0.259	0.033	NAP	0.953	1.598	3.009	0.362	1.087
M3	0.007	0.663	0.293	0.023	0.012	0.050	0.012	1.238	0.250	0.283
M4	0.084	1.275	0.230	0.040	NAP	0.111	0.074	0.312	1.350	0.434
M5	0.026	1.283	0.115	0.001	NAP	0.012	0.114	0.005	0.0002	0.195
M6	0.069	0.094	0.273	18.550	0.008	0.000	6E-05	3E-04	0.040	2.115

Donde: *Aci* = (*Acacia constricta* ind.), *Acm* = (*Acacia constricta manchón*), *Atc* = (*Atriplex canescens* (Pursh) Nutt.), *Pcn* = (*Pinus cembroides* Zucc natural), *Pcp* = (*Pinus cembroides* Zucc plantación), *Pgi* = (*Pinus greggii* Engelm), *Phs* = (*Pinus halepensis* Miller), *Pgs* = (*Prosopis glandulosa*), *Yfl* = (*Yucca filifera* Chaubad), NAP = (No aplica).

Resumiendo, el Cuadro 11, se puede apreciar que el error estándar promedio más bajo lo presenta el M5 con un valor promedio del error estándar = 0.195 kg el cual indica que la relación entre las variables observadas y las estimadas presentan margen de error bajos o casi nulos lo cual representa una calidad de ajuste del modelo bastante bueno, caso contrario para el M6 el cual el valor promedio del error estándar de 2.115 kg siendo este el

que presenta margen de error más elevado. De igual manera *Acm* presenta los valores de error estándar más elevados, y *Aci* obtiene los valores más bajos al igual que en el caso anterior esto puede deberse al tamaño de muestra de cada especie.

En resumen, las pruebas de normalidad residual de modelos por especie, propone 51 posibilidades de ajuste, de los cuales solo 44 de ellas cumplen con el supuesto de normalidad. Expresado en porcentaje el 86.27 % de los modelos mostraron mejoría después de aplicar la transformación de potencia Box Cox.

De igual manera para el supuesto de homogeneidad de varianzas se presentan 51 posibilidades de ajuste, de los cuales solo 46 de ellos cumplen con el supuesto de homogeneidad, expresado en porcentaje representan el 90.2 % de los modelos, demostrando mejoría después de la transformación.

El supuesto de normalidad residual antes de la transformación se hacía presente en un 68.63 %, posterior mente aplicando la transformación Box Cox, se obtiene un 86.27 % de modelos que si cumplen el supuesto de normalidad, lo que indica que después de la transformación mejoraron en un 17.64 % y el 13.73 % de los modelos aún sigue sin mostrar mejoría. De igual manera para homogeneidad de varianza antes de la transformación el 76.47 % cumple dicho supuesto, al aplicar la transformación los modelos cumplen el supuesto en el 90.2 % de los modelos lineales, demostrando mejoría en el 13.73 % de los modelos, y el 9.8 % de los modelos lineales sigue sin presentar mejoría. Con ello se puede intuir que en general la transformación de variables Box Cox, si cumple y satisface los supuestos de normalidad y homogeneidad de varianza en su forma transformada y aumenta en algunos casos el coeficiente de determinación de  $R^2$  ajustada.

### **Resultados de la librería de transformaciones (Trafo)**

Como se puede observar en los resultados la transformación Box Cox, solo demostró mejoría en el 86.27 % de los modelos de regresión lineal y un 90.2 % en homogeneidad de varianza. Los modelos que con la transformación Box Cox, siguen sin cumplir los supuestos (normalidad y homogeneidad) los cuales se resumen por especie de la siguiente manera: Para, *Aci* (M1 y M3), *Pgi* (M2 y M5), *Phs* (M1 y M2), *Pgs* (M1, M3, M4 y M6) y *Yfl* (M1 y M5), donde en algunos casos, las especies si cumplen el supuesto de normalidad, pero no el de homogeneidad y viceversa.

Es por ello que se hace necesario buscar alternativas que hagan cumplir los dos supuestos, y con ello avalar el modelo lineal. Para ello se seleccionaron las mejores transformaciones por cada modelo como se puede ver en el Cuadro 12.

**Cuadro 12.** Transformaciones de la librería Trafo para normalidad y homogeneidad de modelos de biomasa aérea por especie.

M	Sp	Test supuestos de normalidad			Test supuestos de Homogeneidad		
		Transf.	Shapiro_W	Shapiro_p	Transf.	Breusch Pagan_V	Breusch Pagan_p
M1	<i>Aci</i>	Dual	0.98	0.42	Dual	0.32	0.57
M2		Log	0.96	0.13	Log	3.79	0.05
M2	<i>Pgi</i>	Sqrtshift	0.99	0.19	Sqrtshift	3.78	0.05
M5		Sqrtshift	0.99	0.19	Sqrtshift	3.78	0.05
M1	<i>Phs</i>	Neglog	0.95	0.16	Neglog	0.97	0.58
M2		Neglog	0.97	0.58	Neglog	0.03	0.85
M1	<i>Yfl</i>	Sqrtshift	0.96	0.25	Sqrtshift	1.11	0.29
M5		Logshiftopt	0.97	0.43	Logshiftopt	0.00	0.99

Donde: M = (modelo), Sp = (especie), Shapiro\_W = (prueba de normalidad del estadístico), Shapiro\_P = (prueba de normalidad residual), Breusch Pagan\_V = (prueba de homogeneidad del estadístico), Breusch Pagan\_P = (prueba de homogeneidad del modelo).

Como se puede observar, las especies y modelos, después de aplicar la paquetería 'Trafo' se logró encontrar la mejor transformación para cada uno de los modelos con los cuales alcanzan a cumplir los supuestos de normalidad y homogeneidad de varianza. Las diferentes transformaciones presentan valores de  $p = 0.13$  a  $0.43$ , para las pruebas de normalidad y valores de  $0.05$  a  $0.99$  para los test de homogeneidad de varianza.

De las transformaciones aplicadas, resalta con más repeticiones la transformación (Sqrtshift), ya que demuestra mejor ajuste que las demás en dos modelos para *Pgi* y un modelo para *Yfl*. Para normalidad no se presentan problemas ya que los valores son altos y sobrepasan por mucho el criterio de aceptación (valor  $p > 0.05$ ) usado para definir la prueba de normalidad residual. Por el contrario, para homogeneidad dicha transformación no es muy representativa ya que demuestra valores de  $p = 0.05$ . De igual manera para *Neglog* se repite como el mejor para la especie *Phs* y tanto las pruebas de normalidad como homogeneidad presenta valores bastante elevados y por tanto más representativos. Para el caso de *Pgs* ninguna transformación logró cumplir ambos supuestos.

## DISCUSIÓN

La conservación y regeneración de los bosques y reservorios naturales juegan un papel importante como medida de mitigación del cambio climático, debido a que los árboles almacenan dióxido de carbono en la biomasa (Chou y Espeleta, 2013), es por esto que se hace necesario la realización de estudios sobre los factores que puedan ayudar a determinar la cantidad de producción de biomasa (Camacho *et al.*, 2014)., siempre y cuando los supuestos de los modelos se cumplan.

Se han establecido diferentes métodos para la estimación de la biomasa, los cuales se clasifican en métodos directos e indirectos. El primero también conocido como destructivo, haciendo alusión a que es necesario derribar el árbol y realizar el pesaje de la biomasa de sus diferentes componentes de manera directa y posteriormente determinar su peso en seco, por el contrario, el método indirecto se realiza por medio de ecuaciones y modelos matemáticos, obtenidos mediante el análisis de regresión entre variables recabadas en inventarios y muestreos (Brown, 1997).

Los análisis de regresión se definen como un método estadístico el cual estudia la interdependencia entre variables, centra su estudio con el objetivo de predecir valores de una o más variables dependientes, partiendo de variables predictoras y asume que esta puede ser manipulada por el experimentador (Cancino, 2012). La validación de los modelos se puede dar de diversas formas, cuando se tienen un conjunto de muestras grandes, se pueden dividir de manera aleatoria la muestra en dos grupos y con ellos obtener dos modelos y comparar los resultados entre sí para ver si se obtienen resultados similares, el modelo de regresión lineal se estima mediante los mínimos cuadrados y para la regresión logística se utiliza el método de máxima verosimilitud (Peláez, 2016).

Además, Olivier y Norberg (2010), Cancino (2012), Castaño (2012) mencionan que el método matemático o análisis de regresión lineal, se fundamenta en que el error es aditivo, sigue una tendencia de distribución normal y tiene una varianza constante. Cuando uno de los supuestos no se cumple se buscan alternativas para proceder con el modelo antes de desecharlo, algunos optan por el camino de rediseñar el modelo no omitiendo las características de importancia del modelo original, y por medio de una transformación adecuada de los datos haga cumplir los supuestos al igual, que hacer una exhaustiva revisión de los datos para encontrar valores aberrantes que impliquen un problema para el ajuste del modelo (Castaño, 2012).

Los modelos de regresión lineal en ocasiones son muy sensibles a las propiedades estadísticas en los términos de error aleatorio, en el análisis de varianza se incluyen errores independientes, con distribución normal y varianzas homogéneas, ante la falta de cumplimiento de los supuestos se pueden seguir por otros caminos, una es ignorar el incumplimiento de los supuestos y continuar con el estudio considerando que los resultados no son tan exactos y la otra es buscar técnicas alternativas que no exijan el cumplimiento de los supuestos (Avanza *et al.*, 2003).

Olivier y Norberg (2010) hacen referencia a las aplicaciones estadísticas sustentan sus resultados teniendo como punto de apoyo la distribución normal de los datos, más sin embargo no siempre se ajustan a estos criterios, por lo que se hace necesario realizar una transformación de los datos no normales a datos aproximadamente normales.

Algunos autores como Peña y Peña (1986), Teugels y Vanroelen (2004), Olivier y Norberg (2010), Osborne (2010), y Castaño (2012), mencionan que Box y Cox (1964) propusieron una familia de transformaciones para dar soluciones a los problemas de aditividad, errores

no normales y presencia de heterocedasticidad. El método de transformación de potencia Box Cox presenta un método sencillo para elegir la transformación de potencia más apropiada, dicho método destaca por su sencillez de aplicación e interpretación y por ser muy eficiente para dar solución al incumplimiento de los supuestos básicos de una regresión lineal.

Castaño (2012) realizó el experimento de comparación en la estimación del valor de lambda, en muestras contaminadas con observaciones atípicas y observa el comportamiento de la transformación de Box Cox, concluyendo que el procedimiento de dicho método presenta sensibilidad ante la presencia de datos atípicos en la variable respuesta, por el contrario, cuando se tienen muestras sin observaciones atípicas es muy eficiente. Es por ello que se recomienda realizar la búsqueda de datos atípicos en la muestra antes de realizar la transformación de Box Cox.

Peña y Peña (1986), presentan un estudio poniendo a prueba y pretendiendo sustentar si el método de transformación de Box Cox resulta ser factible para testar la normalidad, comparándolo con otros, en el cual encuentra que Box Cox claramente supera a sus competencias en casos de distribuciones asimétricas para todos los tamaños de muestras. Lo cual lo hace un método muy potente, superando incluso a los test direccionales. Concluyendo que Box Cox es excelente cuando se presenta asimetría, discreto en casos de apuntamientos, y malo para casos de distribuciones contaminadas. Por el contrario, casi siempre presenta una ventaja adicional que todos sus rivales no poseen y esto es que facilita una transformación de variables que ayudan a tratar a estas como aproximadamente normales.



Al realizar un análisis exploratorio antes del ajuste de los modelos lineales con ello se pretende observar la relación y tendencia que siguen nuestras variables independientes con relación a la variable de respuesta (Olivier y Norberg, 2010). Cuando no existe relación significativa entre las variables, ajustar un modelo lineal no es recomendable ya que no cumpliría los supuestos de una regresión lineal, por consiguiente, es necesario buscar métodos de transformación de variables que aproximen los datos a una distribución normal.

Camacho et al. (2014) utilizan el coeficiente de determinación  $R^2$ , la prueba de Akaike (AIC), la prueba de Bayes (BIC) y la estadística de Mallows (CP) para seleccionar el mejor modelo. Mas sin embargo estas pruebas son insuficientes, se hace necesario la verificación y cumplimiento de los supuestos del modelo de regresión y aunque estos se cumplen el modelo presenta un ajuste del 95 %, lo cual no se considera como modelo final, debido a que la estimación de la variable predictora representa costes elevados y con ayuda de equipos sofisticados.

El método de transformación de variables Box Cox no normaliza fácilmente los datos asimétricos, además de que al normalizar los datos puede llegar a observarse un impacto dramático en los análisis. Muchos programas como SAS y SPSS incorporan rutinas potentes, que hacen relativamente simple utilizar el script para encontrar y determinar rápidamente los valores de lambda y de esta manera obtener una transformación óptima. Por lo tanto, se debe tener cuidado al momento de interpretar los resultados (Osborne, 2010).

El uso de Box Cox para minimizar el sesgo en la distribución residual es factible y aproxima los datos a una distribución normal, para el caso de la varianza es bien conocido entre los analistas que es posible mantenerlo bajo control tanto el tamaño de la muestra sea razonable, caso contrario del sesgo que, en muchas ocasiones, perjudica a las propiedades y validez de los estimadores (Teugels y Vanroelen, 2004).

En la actualidad con el desarrollo de la informática permite usar métodos complejos para el procesamiento de datos, las transformaciones continúan siendo una alternativa viable de cumplir con los supuestos de un modelo de regresión lineal, el paquete 'Trafo' facilita una amplia recopilación de transformaciones y una amplia colección de métodos de estimación para parámetros de transformación, 'Trafo' es el único paquete R que permite decidir fácilmente que transformación (Medina *et al.*, 2017).

## CONCLUSIONES

Después de la transformación, los modelos demostraron mejor ajuste del coeficiente de determinación, en el cual el modelo 4 destaca de los demás presentando valores altos de  $R^2$  ajustada que van desde 0.89 a 0.99, lo cual indica que el modelo puede predecir en un 90 % la biomasa. Los modelos que cumplen con los supuestos de normalidad y homogeneidad en la mayoría de las especies son los modelos 2, 3, 4 y 6 por lo que se podría seleccionar alguno de estos modelos para predecir biomasa en todas las especies. Se concluye que la transformación Box Cox en la mayoría, soluciona problemas de normalidad y homogeneidad en variables que provienen de muestras que no lo son, logrando que los modelos transformados puedan predecir la biomasa con mayor precisión. Con la paquetería 'Trafo' se logra obtener una transformación adecuada la cual se ajusta a nuestras necesidades, ya que consta de una serie de transformaciones proporcionando detalles específicos de los supuestos, además de automáticamente realizar la comparación con el modelo no transformado, se puede decir que tiene muchas ventajas que lo hacen un método sencillo y práctico de aplicar, más sin embargo las transformaciones que lo integran son inferiores a la transformación de Box Cox ya que ante valores atípicos en la muestra simplemente no logran normalizar y Box Cox los hace aproximadamente normales.

## BIBLIOGRAFÍA

- Acosta, M. M., Vargas, H. J., Velázquez, M. A., y Etchevers, B. J. D. (2002).** Estimación de la biomasa aérea mediante el uso de relaciones alométricas en seis especies arbóreas en Oaxaca, México. *Agrociencia*. 36(6): 725-736.
- Avanza, M., Mazza, S., Martínez, G., y Giménez, L. (2003).** Aplicación de transformaciones para el cumplimiento de los supuestos de normalidad y homocedasticidad, a concentraciones foliares de N, P y K en mandarino. *Agrotecnia*. (11), 18-23.
- Box, G. E., y Cox, D. R. (1964).** An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*. 26(2): 211-243.
- Bravo, O. A., Del Río, M., y Montero, G. (2004).** Site index curves and growth model for Mediterranean maritime pine (*Pinus pinaster* Ait.) in Spain. *For. Ecol. Manage.* (201): 187-197.
- Brown, S. (1997).** Estimating biomass and biomass change of tropical forests. *FAO. Forestry Paper*. 134(1): 42-45.
- Ordóñez, B. J. A., y Maser, O. (2001).** Captura de carbono ante el cambio climático. *Madera y Bosques*. 7(1): 3-12.
- Caballero, M., Lozano, S., y Ortega, B. (2007).** Efecto invernadero, calentamiento global y cambio climático: una perspectiva desde las ciencias de la tierra. *Revista Digital Universitaria*. 8(10): 2-12.

- Camacho, R. N., Segura, E. E., Elizondo, E. B., y Aguilar, D. A. (2014).** Estimating aboveground biomass for *Eucalyptus saligna* Sm. and *Eucalyptus camaldulensis* Dehn in the center region of Costa Rica. *Revista Forestal Mesoamericana Kurú*. 11(27): 22-33.
- Cancino, J. O. (2012).** Dendrometría básica. Universidad de Concepción. Facultad de Ciencias Forestales. Departamento Manejo de Bosques y Medio Ambiente. (ed.) Santiago de Chile. 5-162
- Carmona, H. D., Trejo, C. R., Esquivel, A. O., Arreola, A. J. G., y Flores, H. A. (2007).** Evaluación de un método para medir fotosíntesis en mezquite (*Prosopis glandulosa* Torr. *Revista Chapingo Serie Zonas Áridas*, Universidad Autónoma de Chapingo Durango. México. 5(2): 185-190.
- Castaño, E. (2012).** Una estimación no paramétrica y robusta de la transformación Box-Cox para el modelo de regresión, lecturas de economía. Medellín. Universidad de Antioquia. (75): 89-106.
- Comisión Nacional Forestal (2012).** Inventario nacional forestal y de suelos. Informe 2004-2009. CONAFOR. 2-22.
- Correa, J. C., Iral, R., y Rojas, L. (2006).** Estudio de potencia de pruebas de homogeneidad de varianza, Universidad Nacional de Colombia, escuela de estadística, Medellín. *Revista Colombiana de Estadística*. 29(1): 57-76.
- Curiel, R. A., Espinoza, E. J. R., Cruz, H. J. P., y Almaguer, V. G. (2007).** Modelos de regresión para estimar área de la hoja y biomasa de órganos en árboles jóvenes de manzano. *Revista Chapingo Serie Horticultura*. 13(2): 171-177.

- Cutini, A., Chianucci, F., y Manetti, M. C. (2013).** Allometric relationships for volume and biomass for stone pine (*Pinus pinea* L.) in Italian coastal stands. *Forest.* 6(6): 331-337.
- Chou, S. W., y Espeleta, E. E. G. (2013).** Ecuación para estimar la biomasa arbórea en los bosques tropicales de Costa Rica. *Tecnología en Marcha.* 26(2): 41-54.
- Da Cunha, T. A., y Guimares, C. A. (2009).** Modelo de regresión para estimar el volumen total con corteza de los árboles de *Pinus taeda* L. En el sur de Brasil, Kuru: Revista Forestal (Costa Rica). 6(16): 1-15.
- Díaz, R., Acosta, M. M., Carrillo, A. F., Buendía, R. E., Flores, A. E., y Etchevers, B. J. D. (2007).** Determinación de ecuaciones alométricas para estimar biomasa y carbono en *Pinus patula* Echl. et Cham. *Madera y Bosques.* 13(1): 25-34.
- García, C. M. F (2017).** Una nueva prueba para el problema de igualdad de varianzas. Universidad Santo Tomas. Primer Claustro Universitario de Colombia. 1-35
- González, N., y Abad, J. (2006).** Normalidad y otros supuestos en análisis de covarianzas en Lévy. (Ed.). *Modelización con estructuras de covarianzas.* Netbiblo. La Coruña. 31-57.
- Gutiérrez, E. E. (2012).** Ecuación para estimar la biomasa arbórea en los bosques tropicales de Costa Rica. *Tecnología en Marcha.* 26(2): 41-54.
- Ihaka, R., y Gentleman, R. (1996).** R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics.* (5): 299-314.
- Instituto Nacional de Estadística, Geografía e Informática. (1990).** INEGI, Lugo, H. J., Vidal, Z. R., Fernández, E. A., Gallegos, G. A., Zavala, H. J. y otros,

'Hipsometría'. Extraído de Hipsometría y Batimetría, I.1.1. Atlas Nacional de México. Vol. I. Escala 1:4000000. Instituto de Geografía, UNAM. México.

**Instituto Nacional de Investigaciones Forestales y Agropecuarias.** (1995). INIFAP, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), 'Edafología'. Escalas 1:250000 y 1:1000000. México.

**Jarque, C. M., y Bera, A. K. (1987).** A Test for normality of observations and regression residuals, *International Statistical Review*. 55(2): 163-172.

**Martínez, R. E. M. (2005).** Errores frecuentes en la interpretación del coeficiente de determinación lineal. *Anuario Jurídico y Económico Escorialense*. (38): 315-331.

**Medina, L., Rojas, P. N., Kreutzmann, A., y Castro, P. (2017).** The R package *trafo* for transforming linear regression models. R package version 1.0.0. 2-31.

**Montoya, O. J. M. (1990).** El pino piñonero. *Agroguías mundi-prensa*, Ediciones Mundi-Prensa, Castello, 37, 28001 Madrid. 01-85.

**Nava, R., De Luna, R. V., Reynaga, R. V., y García, R. L. (1980).** *Ecocultivo de Yucca filifera* en las zonas áridas de México. In *Yucca*, serie el desierto. Centro Invest. Química Aplicada Saltillo, Coahuila. (3): 145-171.

**Návar, J., González, N., y Graciano, J. (2001).** Ecuaciones para estimar componentes de biomasa en plantaciones forestales de Durango, México, Simposio internacional medición y monitoreo de la captura de carbono en ecosistemas forestales Valdivia, Chile. (18): 2-12.

- Návar, J. (2011).** Plasticity of biomass component allocation patterns in semiarid Tamaulipan thornscrub and dry temperate pine species of northeastern México. *Polibotánica*. 31: 121 - 141.
- Niklas, K. J. (2004).** Plant allometry: is there a grand unifying theory, *Biological Reviews*. 79(4): 871-889.
- Olivier, J., y Norberg, M. M. (2010).** Positively skewed data: revisiting the Box Cox power transformation. *International Journal of Psychological Research*. 3(1): 69-78.
- Osborne, J. W. (2010).** Improving your data transformations: Applying the Box - Cox transformation. *Practical Assessment, Research & Evaluation*. 15 (12): 1-9.
- Peña, D., y Peña, S. De R. J. I. (1986).** Un contraste de normalidad basado en la transformación Box-Cox, departamento de econometría, Universidad Autónoma de Madrid, *estadística española*. (110): 33-46.
- Peláez, I. M. (2016).** Modelos de regresión: lineal simple y regresión logística. *Revista SEDEN*. 36(4): 195-214
- Picard, N., Saint, A. L., y Henry, M. (2012).** FAO, Rome (Italy). Manual de construcción de ecuaciones alométrica para estimar el volumen y la biomasa de los árboles: del trabajo de campo a la predicción. Depto. Forestal Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement. Montpellier (France). 1-223.



- Pimienta, de la T. D. De J., Domínguez, C. G., Aguirre, C. O., Hernández, F.J., y Jiménez, P.J. (2007).** Estimación de biomasa y contenido de carbono de *Pinus cooperi* Blanco, en Pueblo Nuevo, Durango, Madera y Bosque. 13(1): 35-46.
- Rico, A, M. de L. (2001).** El género *Acacia* (leguminosae, mimosoideae) en el Estado de Oaxaca, México (parte A). Anales del Jardín Botánico de Madrid. 58(2): 251-275.
- Rodríguez, L. R., Razo, Z. R., Valencia, M. S., y Meza, R. J. (2013).** Características dasométricas de *Pinus greggii* Engelm. ex Parl. var. *greggii* de nueve procedencias en galeana, nuevo león. Revista Mexicana de Ciencias Forestales. 4(18): 116-124.
- Santana, S. S., y Mateos, F. E. (2014).** El arte de programar en R: un lenguaje para la estadística. Instituto Mexicano de Tecnología del Agua. UNESCO. Comité Nacional Mexicano del Programa Hidrológico Internacional. México.1-182.
- Soriano, L. M. D. L. Á., Ángeles, P. G., Martínez, T. T., Plascencia, E. F. O., y Razo, Z. R. (2015).** Estimación de biomasa aérea por componente estructural en Zacualtipán, Hidalgo, México. Agrociencia. 49(4): 423-438.
- Schlegel, B. (2001).** Estimación de la biomasa y carbono en bosques del tipo forestal siempre verde. Simposio Internacional Medición y Monitoreo de la Captura de Carbono en los Ecosistemas Forestales. 1-13
- Team, R. C. (2013).** R: A language and environment for statistical computing. 01-16.
- Teugels, J. L., y Vanroelen, G. (2004).** Box–Cox transformations and heavy-tailed distributions. Journal of Applied Probability. 41(1): 213-227.

**Trapletti, A., Hornik, K., y Lebaron, B. (2009).** Time series analysis and computational finance. R Package Versión 1. 0. 10-22.

**Urrutia, M. J., Gámez, V. H. G., Beltrán, L. S., y Díaz, G. M. O. (2014).** Utilización de *Atriplex canescens* y *Opuntia ficus indica* en la alimentación de cabras lactantes durante la sequía. *Agronomía Mesoamericana*. 25(2): 01-11.

**Zeileis, A., y Hothorn, T. (2002).** Diagnostic checking in regression relationships. *R News*. 2(3): 7-10.