
Cálculo de la carga poblacional para la Isla de San Andrés: un enfoque al turismo y la afectación en servicios básicos.

Calculation of the carrying capacity for the Island of San Andrés: an approach to
tourism and the impact of it on basic services.

Luz Mary Londoño^a
luzlondono@usantotomas.edu.co

Director: Edgar Sardi^b
e_sardip@telmex.net.co

Resumen

La carga poblacional está comprendida por la capacidad máxima que tiene un lugar o ecosistema para sustentar a sus habitantes de forma sostenible con los recursos que se encuentran a disposición, manteniendo un equilibrio con los recursos generados sin causar un choque ambiental con repercusiones a futuro. Este estudio aborda como tema global la estimación de la capacidad de carga, enfocada hacia el turismo y el impacto que este tiene sobre el consumo de recursos como el agua, y la descarga de residuos como basuras y vertimiento de aguas al alcantarillado.

Para la metodología utilizada a la hora de estimar la carga turística, se tuvo en cuenta la población isleña, esto con el fin de verificar el número de personas ajenas a las residentes que se espera arriben al lugar. A la hora de calcular la proyección poblacional, se utilizó la función logística y para el análisis de carga se tuvo en cuenta cuatro tipos de modelos, de los cuales el que mejor ajuste obtuvo de acuerdo a los datos y variables tenidas en cuenta, fue el modelo STIRPAT.

Palabras clave: Carga poblacional, Proyección, Función logística, modelo stirpat .

Abstract

The population load is comprised by the maximum capacity that a place or ecosystem has to maintain its habitants in a sustainable manner with the resources that are available, maintaining a balance with the generated resources without causing an environmental shock with repercussions for the future. This study addresses as a global issue the estimation of cargo capacity, focused on tourism and the impact it has on the consumption of resources such as water, and the discharge of waste such as garbage and dumping of sewage water.

For the methodology used when estimating the tourist load, the island population was taken into account, this in order to verify the density of people outside the residents who may arrive at the place. At the time of calculate the population projection, the logistic function was used and for load analysis four types of models were taken into account, of which the STIRPAT model was the one that best stimulated the data taken into account.

Key words: Population load, Projection, Logistic funtion, stirpat model.

^aEstudiante de estadística

^bMatemático-Demógrafo, Docente de la universidad Santo Tomás

1. Introducción

San Andrés Isla hace parte del Archipiélago Colombiano conformado por la isla de San Andrés, Providencia y Santa Catalina, con aproximadamente $27Km^2$; fundamenta su economía en el sector de Industria, Comercio y Turismo, siendo este último uno de los principales pilares de inversión dentro del presupuesto departamental con \$1.990 millones de pesos para el 2016 (DNP, 2016). La población proyectada para el año 2016 es de 71.946 habitantes según el Departamento Administrativo Nacional de Estadística (DANE) y una isla con más de 2000 habitantes por km^2 , se considera sobrepoblada (Marquéz, James, Márquez Pérez, Castellanos, & Taylor). A pesar de los controles que han sido ejercidos por la oficina de Control, Circulación y Residencia (OCCRE) mediante el Decreto 2762 DE 1991 (diciembre 13), la disminución de turistas que se han instalado en la isla aumentando la población residente, ha sido mínima.

La isla no cuenta con una metodología indicada para el tratamiento y el manejo de la población en comparación con los recursos existentes en el medio natural isleño, dando como resultados desabastecimientos en ciertas épocas del año, de servicios vitales como el agua (UNGRD, 2016), sobrepasar los metros cúbicos de aguas residuales causando rebosamientos en la parte céntrica de la isla; además de esto, según la secretaria de Servicios Públicos y Ambiente de San Andrés, el relleno sanitario le quedan aproximadamente tres años de vida útil.

Es por esto que utiliza el estudio de carga, como herramienta principal para identificar problemáticas en cuanto a recursos y crecimiento poblacional. De acuerdo a lo anterior, para este análisis se hará una estimación y proyección de la población residente y turística mediante la función logística. Para la estimación de impacto, se utilizará el mejor modelo que ajuste de acuerdo a los datos, tomando como variables explicativas la población estimada, metros cúbicos de consumo de agua residencial y comercial, kilogramos de basuras generadas y metros cúbicos de agua residual arrojados al alcantarillado.

Finalmente, el objetivo que se quiere alcanzar mediante este estudio, es dar cifras del total de arribo de turistas que se espera en la isla de San Andrés, en el período comprendido entre 2017-2020, en función del comportamiento de los recursos hídricos, la descarga de basuras y la fluctuación de los niveles de vertimiento de aguas residuales a los alcantarillados; siendo este un apoyo para próximas investigaciones.

2. Objetivos

2.1. Objetivo general

Predecir mediante herramientas estadísticas la capacidad de carga turística de la isla de San Andrés, teniendo en cuenta las variables población residente, consumo de agua, basuras y alcantarillado.

2.2. Objetivos específicos

1. Calcular el número de turistas y población residente, por medio de herramientas estadísticas.
2. Estimar por medio de modelos la capacidad turística de la isla de San Andrés, en función de la población residente, agua, basuras y alcantarillado.
3. Proyectar el comportamiento de la población, agua, basuras y alcantarillado entre los años 2017 y 2020.
4. Proyectar el total de turistas esperados entre los años 2017 y 2020, por medio de modelos estadísticos.
5. Comparar el modelo de capacidad de carga, con las ecuaciones y metodologías de estimación turística ya establecidos.

3. Marco de Referencia

La isla de San Andrés, luego del nombramiento de puerto libre en la década de 1950, por el expresidente Gustavo Rojas Pinilla, presentó un crecimiento del número de residentes no nacidos en la isla (inmigrantes); a raíz de la declaración de centro de comercio y turismo, arribaron a la isla numerosas familias y se asentaron en el archipiélago, a partir de esto, comenzaron a surgir problemas de sobrepoblación que aún se presenta en la isla cada año con mayor porcentaje de crecimiento, según el censo de 2005, la población raizal (nativa) solo ocupaba un 39,4% del total de los habitantes.

Para el control de población y migración, se constituyó la OCCRE, mediante el Decreto 2762 DE 1991 (diciembre 13), la disminución de población residente y turística, ha sido mínima. Teniendo una gran recepción de residentes de los departamentos de Bolívar y Atlántico (Triana & Londoño, 2016); además según la Secretaría de Turismo Departamental para el año 2015 el arribo de turistas nacionales e internacionales, aumentó en 14.04%.

En 2012, debido al Fallo de la Haya, se delimitó una nueva frontera marítima, donde gran parte del sustento económico se vio afectado, pues para muchos isleños la pesca era una de las principales fuentes de ingreso, por tal motivo el turismo asciende como pilar importante en la economía de San Andrés, aunque la mayoría de sus hoteles no tributen en la isla de acuerdo a la ley 1607 del 2012, sino en el lugar en el que las sedes principales se encuentren.

San Andrés, además de verse afectado en materia de recursos renovables y no renovables, también presenta incomodidades desde años atrás, reflejando para el censo de 2005 un índice de 41% a las necesidades básicas insatisfechas según el informe del Departamento Nacional de Planeación (DNP), y reflejando el deterioro y prestaciones de servicios básicos no óptimos para el año 2016.

Investigaciones más recientes, como el efecto de la densidad poblacional sobre las fuentes fluviales de reserva en el Archipiélago de San Andrés y Providencia, mediante un modelo STIRPAT, concluyó que:

” El gasto de agua del archipiélago siguió los mismos patrones que la población flotante, turistas que ingresan año a año, por lo que la población de San Andrés y Providencia no influyó de manera significativa en el gasto de este recurso.

se evidenció que la relación entre la densidad poblacional, más aún el crecimiento poblacional y las fuentes fluviales de reserva en el archipiélago de San Andrés y Providencia se vio fuertemente

afectado por los turistas. Tal relación se da en forma decreciente a lo largo de los 10 años de estudio.” (Acosta, 2016).

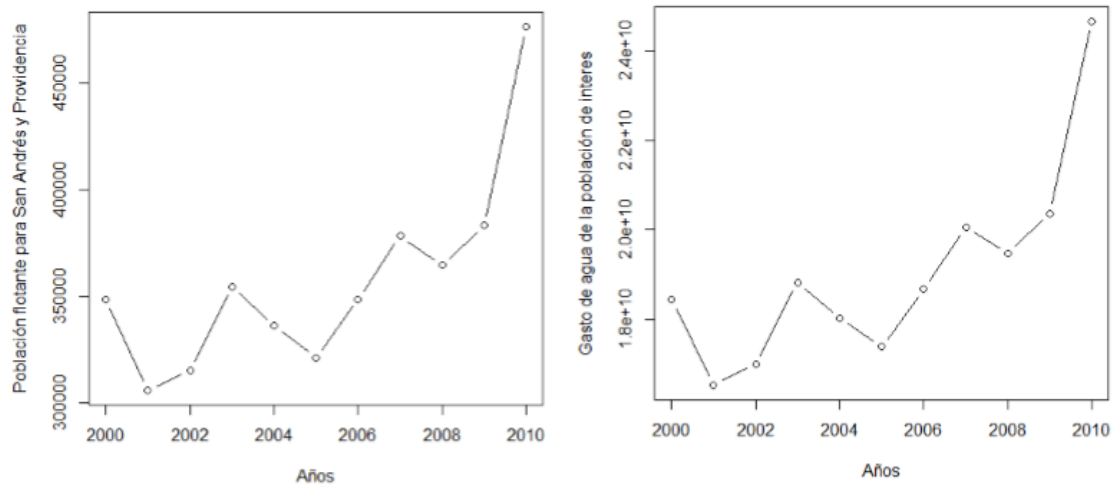


Figura 1: Población flotante de San Andrés y Providencia, años 2000 al 2010 y gasto de agua anual. Fuente: *Efecto de la Densidad Poblacional sobre las fuentes fluviales de reserva en el archipiélago de San Andrés y Providencia*

4. Marco Teórico

De acuerdo a los objetivos planteados para el avance de este estudio, se tendrán en cuenta una serie de teorías y ecuaciones matemáticas y estadísticas, para el respectivo análisis y desarrollo del proyecto.

4.1. Función logística

La función logística, normalmente se utiliza para modelar el crecimiento de cualquier población sujeta a factores limitantes.

En 1840 Pierre Verhulst plantea el crecimiento logístico como una forma mejorada de un crecimiento exponencial de Malthus, argumentando que la población (N) tenía un crecimiento exponencial hasta cierto punto, pero al acercarse a un límite comenzaba tener un crecimiento exponencial pero más lento ($L-N$). La asíntota superior de la curva, puede estar representado por algún factor. Entre más pequeño sea N , la tasa de crecimiento es más pequeña y va a tender a 0, por el contrario si N es muy grande este tenderá a L .

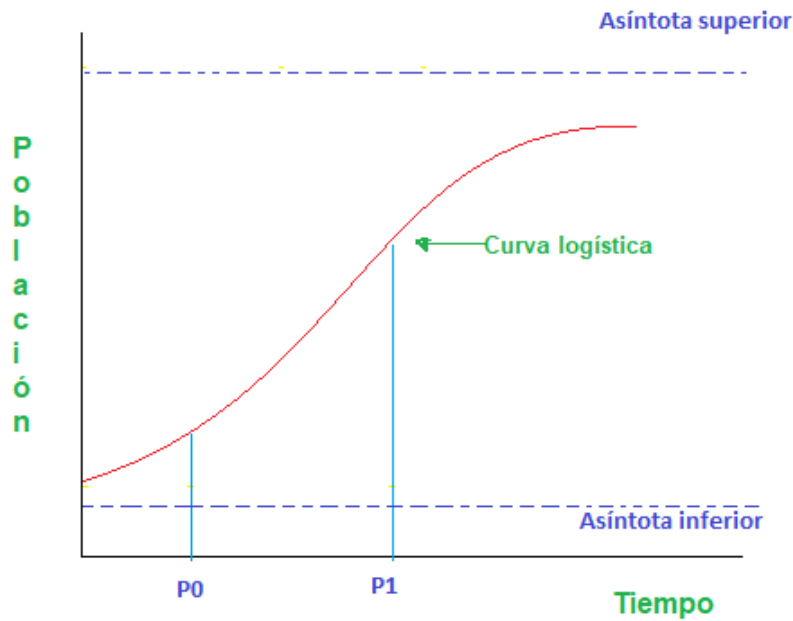


Figura 2: Función logística. Fuente: *Elaboración propia*

Los pivotes para la función logística son datos observados, para este caso son períodos de tiempo, en el cuál el valor de P_0 es el número de habitantes en el año base y P_1 el número de habitantes en el año que se quiera estimar.

La ecuación diferencial de Verhulst se define de la siguiente manera:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{L} \right) \quad (1)$$

donde:

- N: Número de habitantes.
- r: Constante relacionada al crecimiento poblacional (proporción) en cierta época.
- L: límite máximo de población o capacidad de soporte.
- t: tiempo.

Realizando el cálculo de la ecuación diferencial (1) se tiene que:

$$\frac{\frac{dN}{dt}}{N * \left(1 - \frac{N}{L} \right)} = r$$

Aplicando fracciones parciales:

$$\left(\frac{1}{N} + \frac{\frac{1}{L}}{1 - \frac{N}{L}} \right) \frac{dN}{dt} = r$$

Cálculando sus respectivas antiderivadas:

$$\ln(N) - \ln\left(1 - \frac{N}{L}\right) + C_1 = rt + C_2$$

$$\frac{N}{1 - \frac{N}{L}} = e^{rt+C_3}$$

$$\frac{1}{N} - \frac{1}{L} = e^{rt} C_3$$

$$N = \frac{L}{LC_3 e^{rt} + 1}$$

Como la multiplicación $L * C_3$ da una constante, entonces se nombrará a y este lo puedo tomar como e^a ya que es una constante cualquiera, entonces se puede escribir de las siguientes maneras:

$$N = \frac{L}{e^{rt+a} + 1} \quad (2)$$

$$N = \frac{L}{ae^{rt} + 1} \quad (3)$$

Ahora como se quiere observar la variación de la función dentro de las asíntotas A_i y A_s , siendo:

- A_i : asíntota inferior, el valor mínimo de población registrado.
- A_s : asíntota superior, el valor máximo que pueda tomar la población.

y la variación entre estas asíntotas cuando $t(0)$:

$$a = \frac{A_s - A_i}{A_s}$$

Ahora, reemplazando en (3)

$$N = \frac{L}{\frac{A_s - A_i}{A_s} e^{rt} + 1} \quad (4)$$

Esta ecuación también se puede reescribir como:

$$N_j = A_s + \frac{A_i - A_s}{e^{a+rt} + 1} \quad (5)$$

4.1.1. Validación de la función logística

Es necesario tener un argumento válido que confirme el buen ajuste y precisión de las herramientas estadísticas o matemáticas que se utilizan en la práctica. Una de las pruebas de bondad de ajuste es el coeficiente de determinación R^2 -ajustado, este coeficiente mide la proporción de la variable respuesta (N) explicada por las predictivas (r,L,t,a).

La ecuación determinada del R^2 -ajustado para evaluar la función logística es la siguiente:

$$R^2 = 1 - \frac{ECM}{S_p^2}$$

Siendo:

ECM: Error cuadrático medio y S_p^2 : Varianza poblacional

La interpretación de R^2 -ajustado esta dada por el porcentaje de variabilidad de N con respecto a las variables utilizadas para su estimación. Si R^2 -ajustado $\rightarrow 1$: significa que la función se ajusta muy bien a los datos, en cambio, si R^2 -ajustado $\rightarrow 0$: la función tiene problemas de falta de ajuste.

4.2. Método de bisección

El método de bisección es un algoritmo basado en el Teorema de Valores Intermedios, que a través de múltiples interacciones divide intervalos a la mitad en busca de raíces, la interacción acaba una vez el método consiga un cambio en el signo entre los intervalos. Este método es uno de los más sencillos a la hora de resolver ecuaciones, se tendrá en cuenta para el cálculo de la asíntota superior de la función logística.

El Teorema de Valores Intermedios, consiste en:

"Sea f una función continua en un intervalo $[a, b]$ y $f(a)f(b) < 0$, entonces existe al menos un $\alpha \in (a, b)$, tal que $f(\alpha) = 0$ "

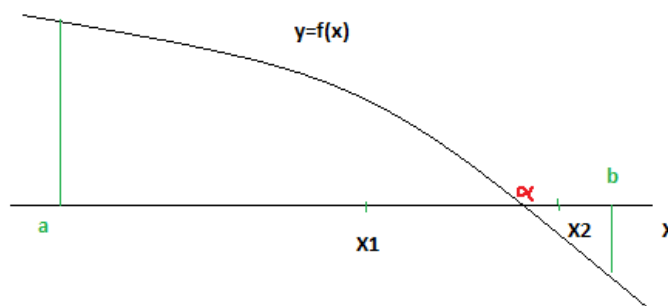


Figura 3: Representación del método de bisección. Fuente: *Elaboración propia*

4.3. Modelos a considerar para las estimaciones

De acuerdo a los objetivos planteados para el desarrollo del estudio, se tendrán en cuenta cuatro tipos de modelos que de acuerdo a su estructura, literatura y aplicabilidad se pueden utilizar para el tipo de datos que se requiere en este proyecto. Una vez establecido el mejor modelo de acuerdo a su ajuste, se utilizará para la estimación de la carga turística.

4.3.1. Modelo de regresión lineal múltiple

En general los modelos de regresión son una herramienta estadística empleada para analizar causa y efecto, es decir, identificar que variables o conjunto de variables independientes explican una variable dependiente.

Los modelos de regresión lineal múltiple son una extensión de los modelos de regresión lineal simple. se utilizan para relacionar un conjunto de variables cuantitativas X_1, X_2, \dots, X_k con una variable principal Y , obteniendo como ventaja una estimación más precisa. Este modelo se establece de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + e_i \quad i = 1, 2, 3, \dots, n. \quad (6)$$

de acuerdo a la ecuación (6) las n observaciones generan el sistema de ecuaciones:

$$\begin{array}{rcccccccc} Y_1 = & \beta_0 + & \beta_1 X_{11} + & \beta_2 X_{12} + & \beta_3 X_{13} + & \dots & \beta_k X_{1k} + & e_1 \\ Y_2 = & \beta_0 + & \beta_1 X_{21} + & \beta_2 X_{22} + & \beta_3 X_{23} + & \dots & \beta_k X_{2k} + & e_2 \\ Y_3 = & \beta_0 + & \beta_1 X_{31} + & \beta_2 X_{32} + & \beta_3 X_{33} + & \dots & \beta_k X_{3k} + & e_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_n = & \beta_0 + & \beta_1 X_{n1} + & \beta_2 X_{n2} + & \beta_3 X_{n3} + & \dots & \beta_k X_{nk} + & e_n \end{array}$$

Cuya forma matricial se puede escribir de la siguiente manera:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ 1 & X_{31} & X_{32} & \dots & X_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

Una equivalencia de la anterior matriz es denotarla de la siguiente forma:

$$Y = X\beta + e \quad (7)$$

Para estimar los parámetros por mínimos cuadrados, tenemos que:

$$\sum_{i=1}^n e_i^2 = (Y - X\beta)'(Y - X\beta) \quad (8)$$

Derivando e igualando a cero, se llega a lo siguiente:

$$(X'X)\beta = X'Y \quad (9)$$

y si $(X'X)$ no tiene inversa y sea una matriz no singular, se tiene que:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (10)$$

Para que el vector de los estimadores sea insesgados, se debe cumplir con los supuestos de:

- Normalidad.
- Homocedasticidad.
- Independencia.

4.3.2. Modelo STIRPAT

"*The Stochastic Impacts by Regression on Population, Affluence and Technology*". Este modelo permite estimar impactos medio-ambientales asociados a variables consideradas como causantes de estos efectos.

$$I_i = aP_i^b A_i^c T_i^d e_i$$

Esta ecuación también se puede escribir como un modelo aditivo de la forma:

$$\ln(I_i) = a + b * \ln(P_i) + c * \ln(A_i) + d * \ln(T_i) + e_i \quad (11)$$

donde,

- I : impacto ambiental.
- P : población.
- A : afluencia.
- T : tecnología.
- a : Coeficiente del modelo.
- b,c,d : parámetros del modelo a estimar.
- e : error.

Para que el modelo tenga un buen ajuste y los vectores de estimadores sean insesgados, los residuales deben cumplir con los supuestos de:

- Normalidad.
- Homocedasticidad.
- Independencia

4.3.3. Definición de los Supuestos

- Normalidad.

$$e_i \approx N(0, \sigma^2)$$

Al cumplir este supuesto, da pie para utilizar pruebas estadísticas como t , F y χ^2 , utilizados para el cálculo de bondad de ajuste. Según la literatura el test de Shapiro-Wilk, que se basa en la comparación entre los percentiles observados en los datos y los percentiles de la distribución normal, tiene un mejor desempeño y potencia en casi todos los escenarios.

$$W = \frac{1}{nS^2} \left(\sum_{j=1}^h a_{jn} (x_{n-i+1} - x_i) \right)^2$$

esta prueba tiene como hipótesis:

H_0 : Los datos provienen de una distribución normal Vs H_1 : Los datos no proviene de una distribución normal.

- Homocedasticidad.

Este supuesto, considera que la varianza de los errores es constante a lo largo de las observaciones. Los test más utilizados para este supuesto son: Bartlett, Levene, F-máx y Breusch-Pagan Test.

El sistema de hipótesis para este supuesto es:

H_0 : Varianza constante Vs H_1 : Varianza no constante.

- Independencia.

El supuesto de independencia, asume que los residuos son aleatorios y no están correlacionados. El test considerado para este supuesto es el de Durbin-Watson.

$$DW = \frac{\sum_n^{i=2} (\hat{r}_i - \hat{r}_{i-1})^2}{\sum_n^{i=1} \hat{r}_i^2}$$

Quien tiene como sistema de hipótesis:

H_0 : Independencia Vs H_1 : No independencia.

4.3.4. Modelos lineales generalizados (MLG)

En la práctica, no siempre la información o los datos obtenidos se distribuyen de manera normal, es decir, se pueden presentar variables de tipo binomial, poisson, binomial negativa, gamma, etc. Por lo cual, es necesario realizar pruebas con el fin de tener los parámetros adecuados para utilizar un MLG.

Los MLG se utilizan para el análisis y ajuste de las variables dependientes, cuya distribución pertenece a la familia exponencial, es por esto, que modelos de este tipo requieren de un marco

teórico distinto al de los anteriormente nombrados. Su estimación se realiza por medio del método de máxima verosimilitud.

Estos modelos se componen de la siguiente manera:

- Componente aleatoria: se define a partir de la familia exponencial.

$$f(y_i; \theta_i, \phi) = \exp\{a_i \phi^{-1} [y_i \theta_i - b(\theta_i)] + c(y_i; \phi)\}.$$

$a(\cdot)$, $b(\cdot)$ y $c(\cdot)$, funciones conocidas.

- Componente sistemática: ó predictor lineal

Involucra las covariables del sistema con las siguientes restricciones:

1. No correlación entre variables.
2. No aleatoriedad entre variables.

Descrita de la siguiente forma:

$$\eta = X\beta$$

Donde X es la matriz diseño y β el vector de parámetros.

- Función enlace: Esta función une la componente aleatoria con la componente sistemática; en otras palabras relaciona la media con el predictor lineal. Esta función es monótona y es la que caracteriza el modelo.

$$g(\mu_i) = \eta_i$$

4.3.5. Modelo Poisson

El modelo Poisson pertenece a la familia exponencial y hace parte del ramillete de los MLG, su variable respuesta tiene una distribución Poisson y depende solo del su parametro λ . Es apropiado utilizar este modelo para trabajar con datos de conteo en un período de tiempo, como los especificados para este proyecto.

Partiendo de la distribución Poisson, el modelo tiene la siguiente forma:

$$\begin{aligned} f(y) &= \frac{\lambda^y e^{-\lambda}}{y!} \\ &= \exp\{\ln(\lambda^y e^{-\lambda}) - \ln(y!)\} \\ &= \exp\{\ln(\lambda^y) + \ln(e^{-\lambda}) - \ln(y!)\} \\ &= \exp\{y \ln(\lambda) - \lambda - \ln(y!)\} \end{aligned}$$

Siendo $\theta = \ln(\lambda)$, $b(\theta) = e^\theta = \lambda$, $b'(\theta) = e^\theta = \lambda$, $a(\phi) = 1$, $c(y; \phi) = \ln(y!)$

4.3.6. Diagnóstico de Dispersión

En ciertos casos los modelos Poisson no son los apropiados para los datos que se quieren estimar, debido a que violan algunos supuestos, uno de ellos y el más fuerte es no presentar sobre ni subdispersión; en contexto, el supuesto de dispersión puede ser comparado con el supuesto de homocedasticidad en un modelo de regresión lineal. En la práctica se ha identificado que el problema más común a la hora de trabajar con este modelo es la sobredispersión, según la literatura algunas causas que pueden producir este inconveniente son:

- Variabilidad muy alta en los datos.
- No independencia de las variables a través del tiempo.
- Los datos no provienen de una distribución poisson.
- La función de enlace puede no ser la apropiada para el modelo.

La prueba de dispersión se realiza para verificar si en un modelo poisson se cumple que $E(y) = \mu$ y $Var(y) = \mu$. Cameron & Trivedi en 1985 propusieron: "En un modelo poisson donde la varianza y la media son iguales a μ , entonces la varianza se define $Var(y) = \mu + c * f(\mu)$ donde la constante $c < 0$ significa subdispersión y $c > 0$ sobredispersión, y la función $f(\cdot)$ es una función monótona".

En este caso se debe probar la siguiente hipótesis:

$$H_0 : c = 0 \quad Vs \quad H_1 : c > 0$$

4.3.7. Modelo Binomial Negativo

El modelo binomial negativo hace parte de los MLG y pertenece a la familia exponencial, se puede decir que es una generalización del modelo Poisson. Este modelo se utiliza cuando en un modelo Poisson se ha diagnosticado problemas de sobredispersión.

La estructura de este modelo es de la siguiente forma:

$$P(Y = y) = \binom{k + y - 1}{k - 1} \left(\frac{p}{p + 1} \right)^y \frac{1}{(p + 1)^k}$$

$$P(y \leq y) \approx e^{-\mu} \sum_{i=0}^y \frac{\mu^i}{i!} - \frac{(y - \mu)k e^{-\mu} \mu^y}{2(\mu + k)y!}$$

5. Metodología

Este estudio es de tipo predictivo y aplicativo, el propósito es estimar la población residente y turística de la isla de San Andrés desde el año 2017 hasta el año 2020, basado en los datos observados del período comprendido entre el año 2012 hasta el año 2016. Además de esto, se estima el número total de turistas esperados por año, que van a ingresar a la isla de acuerdo al consumo de agua, población residente, generación de basuras y vertimiento de aguas residuales.

5.1. Tratamiento de los datos

Los datos utilizados para este estudio corresponden a diversas fuentes institucionales encargadas de su administración y seguridad.

- Las cifras de población para la isla de San Andrés, se tomaron de la publicación de proyecciones del DANE hasta el año 2020, basadas en el censo de población realizado en el 2005 y ajustado por todos los errores.
- El registro administrativo de entradas y salidas en general fue aportado por la OCCRE, de acuerdo a los objetivos del estudio, se realizó una depuración a la base de datos manteniendo solo el registro de entrada y salida de turistas.
- La base de datos para consumo de agua residencial y comercial, generación de basuras y vertimientos de aguas residuales al alcantarillado, fueron suministrados por la Secretaría de Servicios Públicos y Ambiente del Archipiélago de San Andrés, Providencia y Santa Catalina, en respuesta a la solicitud número 171109 del 03 de Octubre de 2017.

A todas las fuentes antes mencionadas se les realizó la debida evaluación, análisis y tratamiento de la calidad de los datos, con el fin de identificar inconsistencia, datos faltantes y outliers; estos resultados se explican más adelante en cada uno de los desarrollos del estudio.

5.2. Variables de estudio.

Total de turistas: Variable estimada y proyectada por medio de una función logística, cuyos resultados se tienen en cuenta para el ajuste y predicción del modelo de capacidad de carga turística.

Población: Variable estimada y proyectada por medio de una función logística, los resultados de esta estimación, se toma como una variable explicativa para el modelo de capacidad de carga.

Consumo de agua por metros cúbicos: variable explicativa, que contiene el consumo en metros cúbicos de agua por habitante, esta información se encuentra desde el año 2012 hasta 2016. Los usuarios de este servicio se encuentran desagregados por tipo de consumo residencial y comercial. Solo se tuvo en cuenta estos dos aspectos a la hora de medir la capacidad de carga, basando este argumento en estudios realizados, cuya conclusión relaciona el incremento de turistas con el desabastecimiento de fuentes fluviales para el consumo de agua en la Isla.

Basuras recolectadas: variable explicativa, contiene total de toneladas de basuras que recolecta la empresa de aseo por año, estas cifras se encuentran a partir del año 2012 hasta 2016.

Vertimiento de aguas al alcantarillado: variable explicativa, que contiene el agua residual total recolectada en metros cúbicos por las tuberías de alcantarillado, conectada a viviendas residenciales, hoteles, posadas y comercio. Para este caso solo se tuvo en cuenta la red de alcantarillado comercial, debido a que la Isla no cuenta con cubrimiento total para este servicio, pero el área comercial donde se encuentra la mayoría de la zona hotelera si.

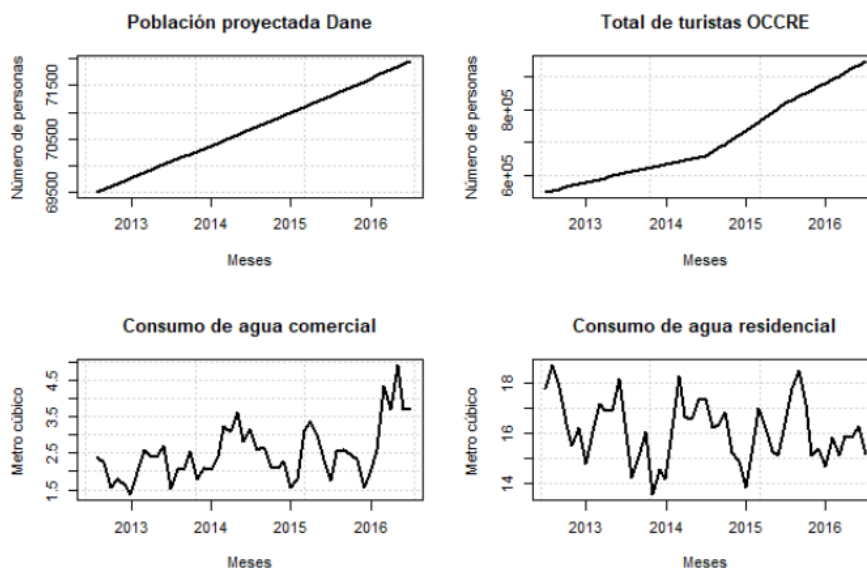
5.3. Método.

La metodología a considerar para la organización y el desarrollo de este documento son los siguientes:

- Realizar un análisis descriptivo de cada una de las series, para observar su comportamiento y poder tomar decisiones sobre ellas.
- Estimar y proyectar por medio de una función logística la población residente y turistas, estas variables se implementan en el modelo de carga turística.
- Validar la precisión y el ajuste de las estimaciones resultado de la función logística.
- Ajustar cada uno de los modelos propuestos para el estudio.
- Verificar la significancia de las variables empleadas en los modelos, por medio de la tabla ANOVA.
- Analizar los supuestos de los modelos, con el fin de obtener consistencia y evitar sesgos.
- De acuerdo a los resultados obtenidos en los ítems anteriores, se realiza las prueba de bondad de ajuste de los distintos modelos para la estimación de la carga de turistas, de acuerdo a las variables ya nombradas.
- Comparación, prueba y ajuste del mejor modelo.
- Proyección de total de turistas soportado por el mejor modelo.
- Resultados y conclusiones.

6. Resultados

6.1. Análisis descriptivo



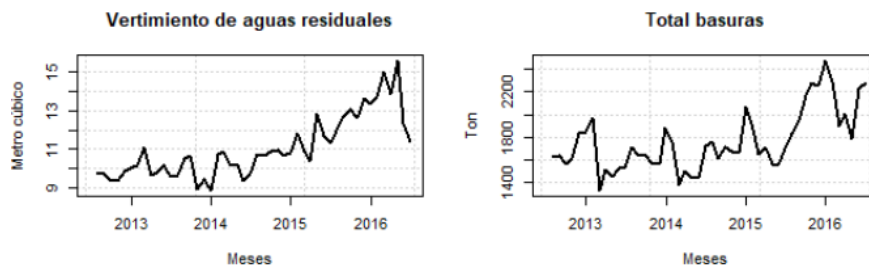


Figura 4: Gráfico del comportamiento de las series.

De acuerdo a los resultados de las gráficas anteriores, se observa que las variables como población y total de turistas tiene un comportamiento creciente muy marcado y es de esperarse, ya que el tiempo de análisis es muy corto y el crecimiento mensual de estas variables son de tipo exponencial. Las variables de consumo de agua comercial, vertimiento de aguas residuales y total de basura; reflejan una tendencia creciente de acuerdo a lo evidenciado en el comportamiento de las series, por el contrario el consumo de agua residencial, se puede observar que tiende a ser decreciente.

En general, para temporada altas (junio-julio y diciembre-enero) se evidencia puntos de niveles bajos en el consumo de agua, esto se puede deber al clima y a los movimientos migratorios de la época, tanto para residentes como turistas. Hay que tener en cuenta que con la entrada al mercado de vuelos a bajo costo para temporada baja, se ha incrementado la visita de turistas en otros períodos de tiempo durante el año.

En el caso de generación de basuras, se puede ver que cada comienzo de año, se registran picos muy altos en las toneladas de desechos que ingresan al relleno sanitario.

6.2. Prueba de normalidad

Con el fin de proporcionar una idea del tipo de distribución que puede tener los datos tomados para este estudio, se realiza una evaluación de normalidad, en este caso se aplica la prueba de Shapiro-Wilk, ya que esta prueba es una de las más sencibles y potentes para contraste de normalidad. La hipótesis planteada para esta prueba es: H_0 : "los datos provienen de una distribución normal"

Prueba Shapiro-Wilk	
Variable	P-value
Turistas	0.0007128
Población	0.06479
Cons agua residencial	0.8364
Cons agua comercial	0.006678
Basura	0.005296
Alcantarillado	0.001855

Tabla 1: Prueba S-W

Para analizar la anterior tabla se debe verificar que el p-valor de cada variable sea mayor a 0.05, si es así entonces se acepta la hipótesis de que los datos siguen una distribución normal, por el contrario, si el p-valor es menor, se rechaza a hipótesis planteada. Teniendo claro lo anterior, se puede evidenciar que las variables población residente y consumo de agua residencial siguen una distribución normal; por el contrario las variables turistas, consumo de agua comercial, vertimientos de aguas residuales y toneladas de basuras, no se distribuyen normal.

Este resultado nos puede dar un bosquejo del comportamiento de las series, para así tener un mejor idea de los modelos a tener en cuenta.

6.3. Estimación de la población residente

En el contexto del estudio se propone realizar un contraste entre el ingreso de turistas a la isla, con el impacto de servicios al ambiente de la misma; en este sentido, en la investigación se tiene en cuenta a la población residente, por pertenecer al entorno a analizar y ser susceptibles a toda situación que en él ocurra.

La Isla de San Andrés dada su formación y características, mantiene sus recursos acotados, es por esto que a la hora de trabajar con los datos de población no se utilizó el registro del Dane, sino los datos resultante por medio de una función logística. Esta función es muy usual en el área demográfica para estimar y proyectar el crecimiento de una población con algún límite de tolerancia.

Para esta evaluación, se tomaron los datos referentes a los censos de 1993 y 2005, ajustando primero la población para esa sección de tiempo y así pronosticar el total a partir del año 2005 hasta el año 2020, luego de esto poder contrastar con la proyección calculada por el Dane.

6.3.1. Método de Bisección

Es necesario estimar el número de tolerancia o factor limitante para este rango de tiempo, por lo cuál se aplica el método de bisección para hallar el límite máximo. Este método se utiliza teniendo en cuenta que el comportamiento de la curva, cuando su población es pequeña esta va a converger a la asíntota inferior y por el contrario si la población es cada vez más grande, va a tender a su límite, por lo que esta curva en alguna cifra va a tener un punto de inflexión.

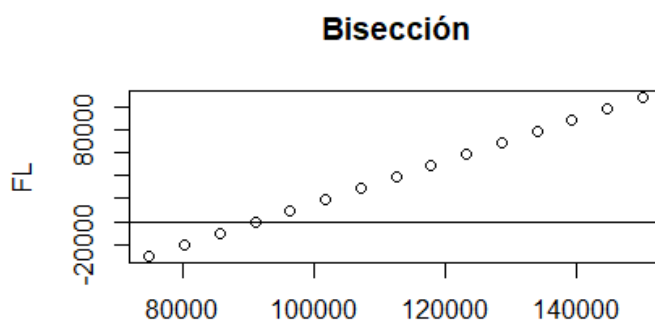


Figura 5: Gráfico de Bisección para la población residente

Se puede evidenciar que este método se ajusta bien a la función logística para hacer sus debidas interacciones, permite encontrar el valor de la asíntota superior de una manera muy precisa. Es de aclarar que para este ejercicio la curva logística no es tan pronunciada debido a el rango de tiempo escogido para el análisis.

Los valores encontrados fueron los siguientes:

Población Máxima	Raíz
75000.01	-29782.4208
80357.15	-19934.6007
85714.3	-10086.7805
91071.44	-238.9604
96666.68	9608.8597
101785.73	19456.6799
107142.87	29304.5
112500.01	39152.3202
117857.16	49000.1403
139285.73	88391.4208
144642.88	98239.241
150000.02	108087.061

Tabla 2: Método de Bisección

De acuerdo a los resultados arrojados por el método de bisección, el valor positivo más pequeño entre el punto de inflexión es 96666.68. teniendo en cuenta este resultado, se realiza una prueba gráfica para analizar el comportamiento de la población con cambios en las asíntotas alrededor del valor ya calculado.

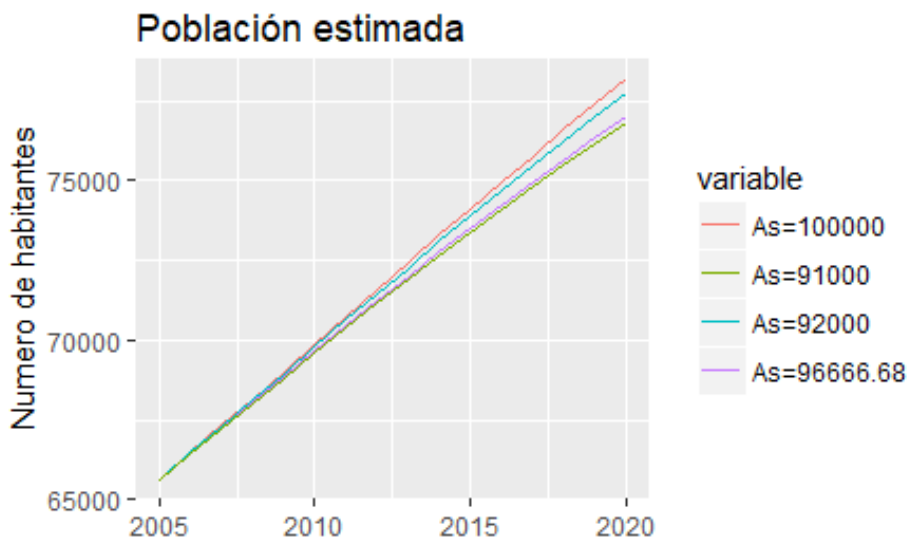


Figura 6: Prueba gráfica de estimaciones de proyección

Observando el comportamiento de las cifras de población para los diferentes valores de la asíntota superior, no se tiene una brecha muy amplia en las comparaciones de las series, es decir, cualquier punto que se tome entre esos valores, no va a generar un cambio influyente en la respuesta. Por lo tanto para el cálculo de población se va a utilizar como asíntota superior el valor generado por el método de bisección equivalente a 96666,68.

6.3.2. Aplicación de la función logística

- A_s : **96666.68** , este valor se tomó de acuerdo al resultado obtenido por el método de bisección.
- A_i : **39700**, el valor de la asíntota inferior se tomó de acuerdo al registro de población más pequeño conocido para la isla de San Andrés, este dato corresponde al censo de 1986 publicado por el DANE.
- a : **-0.8834217**, esta cifra corresponde a la variación entre las asíntotas de acuerdo a el período del estudio.
- r : **0.05862002** , esta variable explica la proporción del crecimiento de la población estimada en el período de tiempo estudiado.

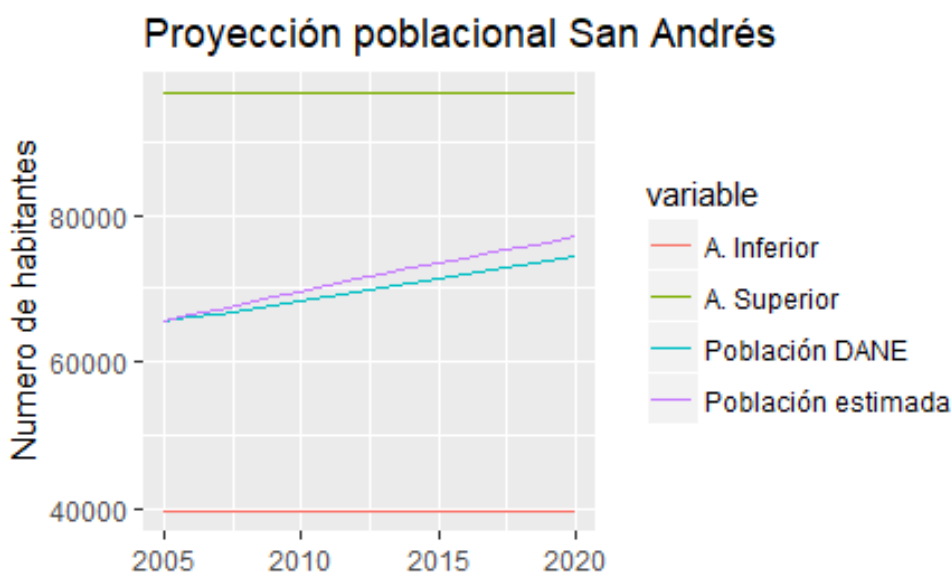


Figura 7: Comparación de la proyección estimada por el Dane Vs la estimada por función logística.

Al analizar la gráfica comparativa, queda en evidencia que ambas series tienen un comportamiento creciente, es posible observar que la población estimada por medio de la función logística, es un poco más alta que los datos proyectados por el DANE; esto puede estar explicado por el efecto que ejerce el valor de la asíntota superior, en la tasa de crecimiento implícita.

Los valores estimados en cada período de tiempo son los siguientes:

Año	Población DANE	Δ	Población Estimada	Δ_{est}	Error relativo	Error relativo ²
2005	65627	0.8 %	65627	1.3 %	0	0
2006	66125	0.8 %	66457.07	1.2 %	0.00502185	2.5219E-05
2007	66641	0.8 %	67290.09	1.2 %	0.0097401	9.487E-05
2008	67173	0.8 %	68124.64	1.2 %	0.014167	0.0002007
2009	67720	0.8 %	68959.3	1.2 %	0.01830035	0.0003349
2010	68283	0.9 %	69792.62	1.2 %	0.02210828	0.00048878
2011	68868	0.9 %	70623.19	1.2 %	0.02548629	0.00064955
2012	69463	0.9 %	71449.6	1.1 %	0.0285994	0.00081793
2013	70069	0.9 %	72270.48	1.1 %	0.03141874	0.00098714
2014	70684	0.9 %	73084.48	1.1 %	0.03396073	0.00115333
2015	71305	0.9 %	73890.31	1.1 %	0.03625706	0.00131457
2016	71946	0.9 %	74686.74	1.0 %	0.0380944	0.00145118
2017	72585	0.9 %	75472.58	1.0 %	0.03978205	0.00158261
2018	73221	0.9 %	76246.72	1.0 %	0.04132312	0.0017076
2019	73850	0.8 %	77008.13	1.0 %	0.04276412	0.00182877
2020	74466		77755.83		0.04417895	0.00195178

Tabla 3: Proyección poblacional

Segun los resultados de la tabla(3), se confirma la diferencia entre la variación de las estimaciones, pero si se analiza el error cuadrado de esas estimaciones, se puede observar que las desviaciones son muy pequeñas.

Los errores al cuadrado presentados en la tabla(3), son insumos para la obtención del coeficiente de determinación ajustado, aplicado en la validación de la función logística.

6.3.3. Validación de la función logística para estimación de población residente

Con el fin de obtener la precisión con que la función logística está estimando los datos, se aplica una prueba de bondad de ajuste, para este caso el método escogido es el coeficiente de determinación R^2 -ajustado.

Los resultados de la validación se especifican en la siguiente tabla:

Promedio Poblacional	69743.92
Varianza Poblacional	24481005
Error Cuadrático medio	0.0007285714
R^2-ajustado	0.99999999998

Tabla 4: Coeficientes

De acuerdo a los resultados obtenidos para la validación de la función logística, se puede apreciar que el R^2 -ajustado tiene una precisión muy cercana al 100 %, esto evidencia que para efectos de comportamiento de evolución poblacional, la función logística es importante y valida.

Con respecto a los análisis aplicados a la estimación de población residente por medio de la

función logística, se puede concluir que esta función con la evaluación de los debidos parámetros para su cálculo, tiene buen ajuste de predicción.

El uso de esta función siguiendo los resultados anteriores, no presenta ningún problema de precisión, por lo cual, se puede implementar con total confianza para hacer la posterior estimación y proyección de la población turística.

6.4. Estimación de Turistas

Siendo la variable número total de turistas la más relevante en la generalidad del proyecto y de acuerdo a la finalidad del estudio, se toma como referencia la conclusión aportada en la estimación de población residente, para proponer el uso de la función logística en la estimación y proyección del número de turistas esperado en el período de 2005 al 2020.

Es de aclarar que además de lo argumentado, se quiere evaluar el comportamiento y la evolución de esta variable en el área comprendida por la isla de San Andrés, teniendo en cuenta que se encuentra en un medio con recursos limitados.

Con la intención de estimar el número de turistas, se tomaron los datos registrados por la OCCRE del año 1993 y el año 2005. Para la primera parte se ajustó la función teniendo en cuenta ese período de tiempo, para luego predecir hasta el año 2020; en este caso, la proyección estimada desde el año 2016-2020 no es comparable con los datos suministrados por la OCCRE, puesto que estos no tienen algún cálculo de proyección, solo cuentan con los datos administrativos registrados hasta el año 2016.

Continuando con los pasos para aplicar la función logística se utiliza el método de bisección, esto con el objeto de encontrar el tope máximo o valor limitante para el rango de tiempo del estudio, que será utilizada en el posterior cálculo de la función.

6.4.1. Método de Bisección

Retomando la descripción de este método empleada para población residente, es necesario hallar la asíntota o límite superior para el ajuste de la función logística; el método de bisección es uno de los más fáciles por su aplicabilidad, el algoritmo interno se puede ajustar a la función que se requiera, dando así una mejor predicción del valor calculado.

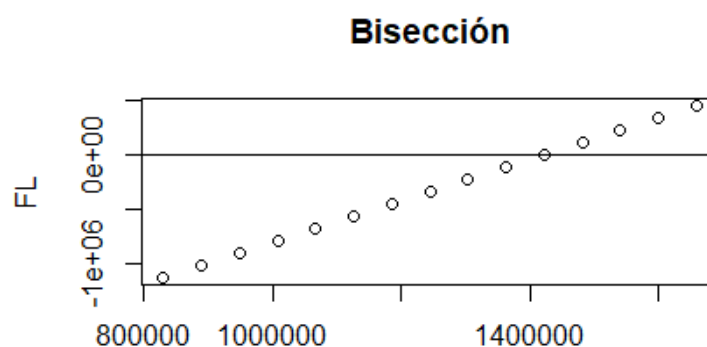


Figura 8: Gráfico de Bisección para Turistas

En este caso, como era de esperarse, se evidencia que el método de bisección se ajusta bien a la función logística, los valores resultantes de las iteraciones encontradas por este método fueron las siguientes:

Población Máxima	Raíz
830000	-1123617.56
948571.4	-898298.102
1007857.2	-785638.371
1185714.3	-447659.177
1245000	-334999.446
1304285.7	-222339.714
1363571.4	-109679.983
1400857.2	2979.748
1482142.9	115639.479
1541428.6	228299.211
1600714.3	340958.942
1660000	453618.673

Tabla 5: Método de Bisección

Según los resultados del método de bisección, el valor más pequeño y positivo entre el punto de inflexión es 1400857.2. De acuerdo a este resultado se realiza una prueba gráfica para observar el comportamiento del número de turistas alrededor de los valores cercanos del tope máximo.

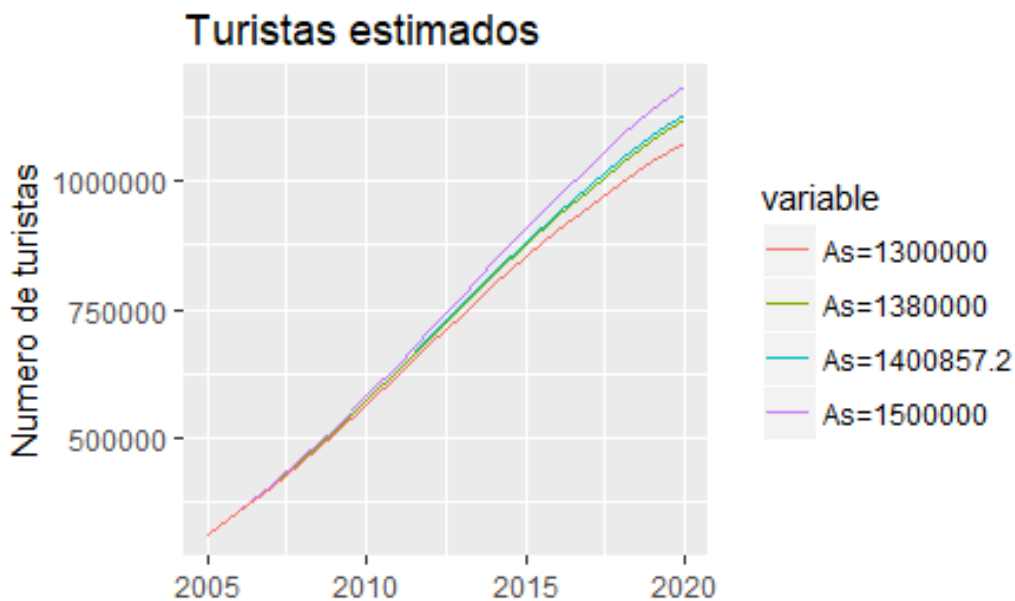


Figura 9: Prueba gráfica de estimaciones de proyección

Las series comparadas en la figura (9) son las cifras generadas por los valores de la estimación de turistas variando su asíntota superior. El rango que se tuvo en cuenta fue desde 1'300.000 hasta 1'500.000 turistas como máximo.

De acuerdo a la gráfica cualquier cifra escogida entre ese rango de valores máximo, no genera un cambio de mayor influencia en los resultados de la estimación del número de turistas, por lo cual se utiliza el valor arrojado por el método de bisección (1'400.857.2).

6.4.2. Aplicación de la función logística.

Al momento de aplicar la función se debe considerar los siguientes parámetros:

- A_s : **1400857.2**, valor resultante del método de bisección.
- A_i : **14413**, este dato es el menor valor observado de entradas de turistas, tomado de acuerdo al registro administrativo proporcionados por la OCCRE, correspondiente al año 1965.
- a : **-3.466706**, este valor corresponde a la variación entre las asíntotas, de acuerdo al periodo estimado de 1993 hasta 2005.
- r : **0.1809032**, valor explicado de la proporción del crecimiento de la población estimada entre 1993 a 2005.

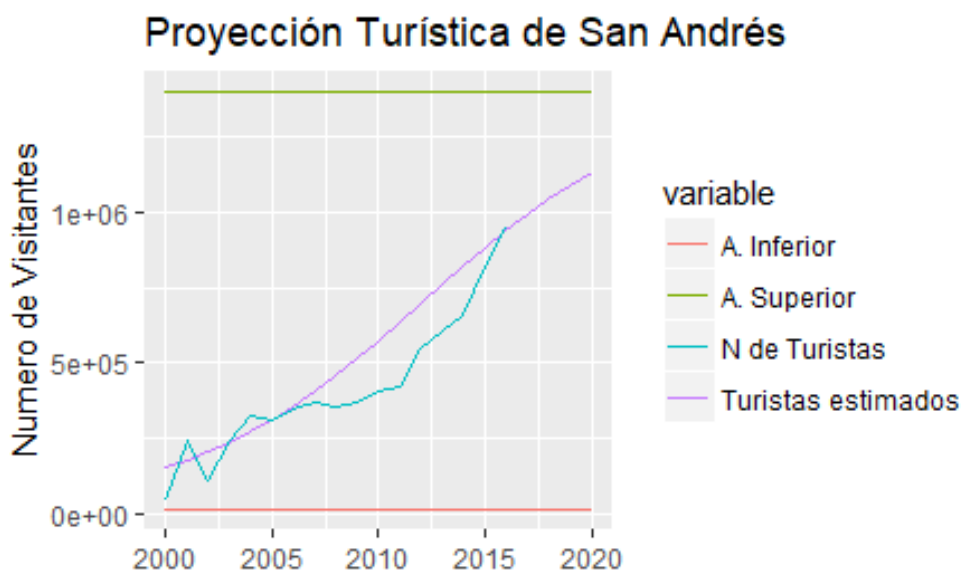


Figura 10: Comparación de la función logística Vs registros de la OCCRE

Al analizar la gráfica, se denota claramente el comportamiento de la curva logística pronunciada para los datos estimados, lo que nos da indicios del buen ajuste de la función hacia las observaciones.

Teniendo en cuenta la asíntota superior establecida, se evidencia que en el año 2016 los valores son muy parecidos, confirmando que el comportamiento de la entrada de turistas en la isla de San Andrés registrado por la OCCRE, se asemeja a lo expuesto por Pierre Verhulst.

El contenido de las salidas de cada una de las estimaciones por año es el siguiente:

Año	Turistas	Δ	Turistas estimados	Δ_{est}	Error relativo	Error relativo ²
2005	312122	11.6 %	312122	13.3 %	0	0
2006	350565	4.8 %	356578	12.8 %	0.01715231	0.0002942
2007	367699	-3.8 %	405289	12.2 %	0.10223036	0.01045105
2008	353833	4.5 %	457985	11.6 %	0.29435355	0.08664401
2009	369995	9.3 %	514217	10.9 %	0.38979446	0.15193972
2010	405960	3.7 %	573347	10.1 %	0.41232387	0.17001098
2011	421138	26.9 %	634575	9.4 %	0.50681012	0.2568565
2012	550961	10.0 %	696972	8.6 %	0.2650115	0.07023109
2013	608821	7.8 %	759536	7.8 %	0.24755224	0.06128211
2014	658471	21.9 %	821253	7.0 %	0.2472121	0.06111382
2015	819485	14.3 %	881164	6.3 %	0.07526556	0.0056649
2016	945545		938420	5.6 %	-0.00753534	5.6781E-05
2017			992331	4.9 %		
2018			1042381	4.3 %		
2019			1088246	3.7 %		
2020			1129775			

Tabla 6: Estimación y proyección turística.

Es evidente que entre los registros de turistas y la estimación por función logística, tienen una diferencia notable en las variaciones. Se puede observar que las variaciones de turistas estimados a lo largo de los años estudiados es decreciente, esto se explica por su cercanía al límite máximo.

De acuerdo a lo anterior, las diferencias en la variabilidad influye en los errores, el error relativo al cuadrado en ciertos valores son muy grandes, lo cual se evidencia en el comportamiento de la serie entre los años 2009 hasta 2014.

6.4.3. Validación de la función logística para estimación de turistas

Promedio Población flotante	493739.3
Varianza población flotante	62972270245
Error Cuadrático Medio	0.4507882
R^2-ajustado	0.99999999993

Tabla 7: Coeficientes

Siguiendo la interpretación descrita para la prueba de bondad de ajuste y con los cálculos obtenidos, se aprecia que la función logística para la estimación y proyección del total de turista es casi del 100 %, evidenciando que esta función al igual que en el caso de la población residente proporciona una precisión muy exacta.

Es valido concluir que la función logística es una buena herramienta para la estimación y proyección de población, sobre todo cuando se quiere medir en población con crecimiento amplio, lo cual se evidenció en la estimación del número de turistas, al adaptar su comportamiento y evolución a la curva logística. Siguiendo los argumentos propuestos por Verhulst, si el crecimiento de turistas registrado por la OCCRE de se mantiene de acuerdo a lo estimado por la función logística, el ingreso de turistas está llegando al valor del factor limitante.

El cálculo de las estimaciones de la población residente y turistas, son insumos bases para el ajuste de los modelos de capacidad de carga turística, quien es el objetivo general de este estudio.

6.5. Proceso de prueba de modelos propuestos.

En esta sección se aborda el tema fundamental de la investigación, de acuerdo al marco teórico, se tendrán en cuenta cuatro tipos de modelos que por su estructura se consideraron para el estudio, pues tienen mejor aplicabilidad a los datos obtenidos y el objetivo del análisis.

Cabe recordar que la finalidad del estudio, es predecir el número total de turistas esperados que arriban a la isla de San Andrés en función de la población residente, consumo de agua residencial (m^3), consumo de agua comercial (m^3), vertimiento de aguas residuales al alcantarillado comercial (m^3) y generación de basuras (Ton).

Para cada modelo se realiza su respectivo análisis de acuerdo a los parámetros requeridos para la prueba y ajuste de los residuales. La prueba común entre los modelos para evaluar la significancia de las variables, es el ANOVA o tabla de análisis de varianza, cuyo sistema de hipótesis es H_0 : "La variable explicativa no ejerce un efecto significativo en la variable respuesta ", esta hipótesis se prueba de acuerdo a los valores del estadístico F, la idea es que el p-valor sea menor que $\alpha = 0.05$.

6.5.1. Modelo de Regresión Múltiple

Este modelo es el más general de aplicar, puesto que solo requiere medir la variable respuesta que en este caso, es número total de turistas explicado por las variables antes mencionadas sin someterlas a alguna transformación. La ecuación del modelo de regresión múltiple se describe de la siguiente forma:

$$Turistas = \beta_0 + \beta_1 Poblacion + \beta_2 Caguacom + \beta_3 Caguaresid + \beta_4 Basura + \beta_5 Alcantarillado + e_i$$

6.5.2. Resultados prueba ANOVA

Los resultados arrojados por la prueba ANOVA para el modelo de regresión lineal múltiple son los siguientes:

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.718e+06	1.285e+04	-367.199	< 2e-16 ***
pobest	7.587e+01	1.855e-01	409.060	< 2e-16 ***
aguacom	-7.035e+02	1.625e+02	-4.328	8.81e-05 ***
aguaresid	1.262e+02	7.872e+01	1.603	0.116221
basura	-1.930e+00	4.862e-01	-3.970	0.000268 ***
alcom	-3.627e+02	1.000e+02	-3.627	0.000755 ***

Los coeficientes resultantes de la tabla ANOVA, muestra que las variables estudiadas están teniendo un efecto sobre el número de turistas, excepto el consumo de agua residencial y puede ser lógico, ya que en su mayoría los turistas no se hospedan en casas residenciales. Para mejorar la estimación del modelo se elimina la variable ya que esta, no está aportando a la variable respuesta.

Dado que la regresión lineal es la base general para los modelos propuestos, los resultados de la tabla ANOVA van a seguir el mismo comportamiento en la conclusión de la significancia en las variables, por lo tanto no se tendrá en cuenta para ajustar los posteriores modelos.

De acuerdo a lo anterior, la ecuación del modelo de regresión múltiple tendrá la siguiente forma:

$$Turistas = \beta_0 + \beta_1 Poblacion + \beta_2 Caguacom + \beta_3 Basura + \beta_4 Alcantarillado + e_i$$

Al aplicar el respectivo análisis de varianza, se tiene lo siguiente:

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-4.711e+06	1.228e+04	-383.762	< 2e-16	***
pobest	7.580e+01	1.831e-01	413.875	< 2e-16	***
aguacom	-6.326e+02	1.592e+02	-3.974	0.000259	***
basura	-2.041e+00	4.898e-01	-4.166	0.000142	***
alcom	-3.432e+02	1.010e+02	-3.398	0.001453	**

Es evidente que la estimación del modelo mejora y que las variables consideradas producen un efecto en el número de turistas.

■ Multicolinealidad.

Habitualmente a la hora de trabajar con un modelo de regresión múltiple, se puede presentar problemas de multicolinealidad, es decir, que el aporte de algunas variables explicativas es redundante para el modelo, debido a que están correlacionadas con la variable respuesta; un método para verificar la presencia de multicolinealidad es la prueba VIF (factor de inflación de la varianza).

Según la literatura, algunos autores proponen que un modelo tiene problemas de multicolinealidad si el VIF > 10. Si esto se cumple, se recomienda eliminar la variable, ya que es reiterativo a los resultados y el R^2 no tendría mayor cambio, de acuerdo con esto:

Variable	VIF
Población estimada	3.935338
Consumo de agua comercial	1.858085
Basura	2.269562
Alcantarillado	3.318247

Tabla 8: Multicolinealidad

De acuerdo a los resultados de la tabla(8), las variables consideradas para el modelo de regresión múltiple son independientes de la variable respuesta, en este caso número de turistas.

6.5.3. Verificación de los supuestos sobre los residuales.

Además de estimar un modelo se debe comprobar la validez que este tiene para ajustar a los datos, por lo tanto a continuación se presentan las respectivas pruebas para la verificación de los supuestos.

- Normalidad

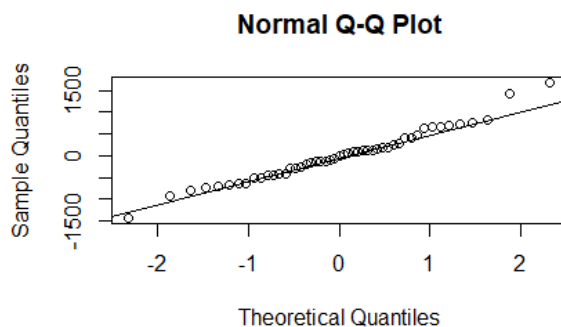


Figura 11: Gráfico de normalidad en los residuales.

Gráficamente se puede observar que los residuales se ajustan a la línea de normalidad, este argumento se justifica a continuación mediante la aplicación de la prueba de Shapiro-Wilk, esta prueba tiene como hipótesis H_0 : "Los residuales siguen una distribución normal".

Prueba	P-valor
Normalidad	0.5516

Tabla 9: Prueba de Shapiro-wilk

Según los resultados de la prueba S-W, se acepta la hipótesis nula confirmando que los residuales se distribuyen normal.

- Homocedasticidad

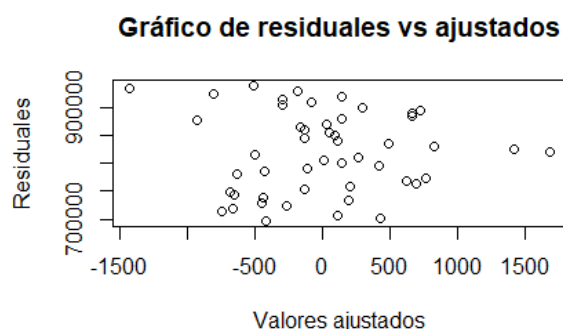


Figura 12: Gráfico de homocedasticidad.

Se puede observar que el gráfico (12) no refleja algún tipo de patrón definido, los valores ajustados y los residuales tienen un comportamiento aleatorio, dando indicio a que no se viola el supuesto de varianza constante. Para confirmar el resultado gráfico se aplica la prueba de Breusch-Pagan, la hipótesis que maneja esta prueba es H_0 : "Varianza constante".

Prueba	P-valor
Varianza Constante (ncvTest)	0.4749423

Tabla 10: Prueba de Breusch-Pagan

Los resultados de la prueba de Breusch-Pagan corroboran que los residuales no violan el supuesto de varianza constante.

- Independencia

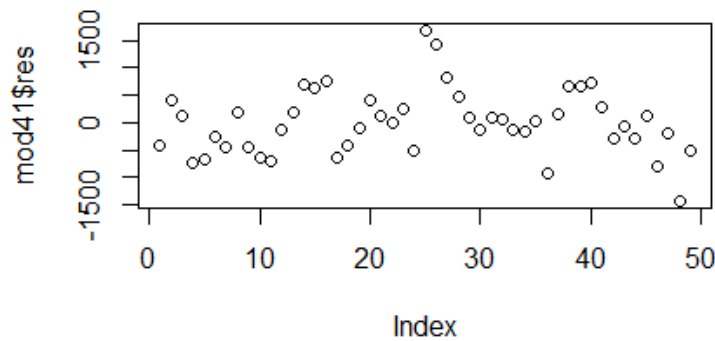


Figura 13: Gráfico de Independencia.

De acuerdo al gráfico de independencia, los residuales asemejan a una tendencia no muy marcada, para confirmar o rechazar este argumento, se aplica la prueba de Durbin-Watson, la hipótesis contenida en esta prueba es H_0 : "Los errores son independientes".

Prueba	P-valor
Independencia (D-W test)	0

Tabla 11: Prueba de Durbin-Watson

La prueba de D-W confirma la independencia de los errores, es decir que a pesar que la gráfica (13) parezca tener tendencia, los errores son aleatorios.

Los resultados de la evaluación de los supuestos del modelo de regresión lineal múltiple, indican que este tiene un buen ajuste cumpliendo con los parámetros de validez, por lo tanto, el modelo tiene una buena estimación y predicción de los datos.

6.5.4. Ecuación del modelo

Una vez verificado los supuestos y cada uno de los análisis para la validación del modelo de regresión múltiple propuesto, la estimación del número de turistas queda representada por medio de la siguiente ecuación:

$$Turistas = -4711204.241 + 75.798(Poblacion) - 632.649(Caguacom) - 2.041(Basura) - 343.192(Alcantarillado) \quad (12)$$

6.5.5. Modelo STIRPAT

El modelo STIRPAT es el segundo modelo propuesto para la estimación del número de turistas que pueden ingresar a la isla de San Andrés, teniendo en cuenta los efectos en las variables población residente, consumo de agua comercial, vertimiento de aguas residuales y generación de basuras. Este modelo se tuvo en cuenta debido a su utilidad para estimar impactos medio-ambientales.

La estructura lineal del modelo STIRPAT se describe de la siguiente forma:

$$\ln(Turistas) = \beta_0 + \beta_1 \ln(Poblacion) + \beta_2 \ln(Caguacom) + \beta_3 \ln(Basura) + \beta_4 \ln(Alcantarillado) + e_i$$

6.5.6. Resultados prueba ANOVA

Luego de ajustar el modelo con las variables mencionadas y aplicar la prueba ANOVA, se obtuvo como resultado:

Coefficients:					
	Estimate	Std. Error	t value		Pr(> t)
(Intercept)	-65.006714	0.727812	-89.318	< 0.0000000000000002	***
log(pobest)	7.035753	0.067031	104.963	< 0.0000000000000002	***
log(aguacom)	-0.004993	0.001982	-2.520	0.015456	*
log(alcom)	-0.018047	0.005656	-3.191	0.002618	**
log(basura)	-0.017241	0.004331	-3.981	0.000254	***

Se aprecia que los coeficientes de las variables, están teniendo buena significancia en el aporte a la variable respuesta, para afirmar el aporte de las variables en el modelo se realiza la prueba de multicolinealidad.

- **Multicolinealidad.**

Para verificar si el modelo presenta problemas de multicolinealidad se aplica la prueba de inflación de la varianza, se debe tener en cuenta a la hora de interpretar esta prueba que si el $VIF > 10$ la variable tiene problemas de multicolinealidad, es decir, su resultado es redundante para la variable respuesta.

Variable	VIF
Población estimada	4.231764
Consumo de agua comercial	1.781278
Basura	3.328645
Alcantarillado	2.266521

Tabla 12: Multicolinealidad

Se evidencia que ninguna de las variables consideradas para el modelo STIRPAT presenta problemas de multicolinealidad.

6.5.7. Verificación de los supuestos sobre los residuales.

Es necesario comprobar la validez de que tiene el modelo STIRPAT para estimar y predecir el número de turistas, para ello es preciso evaluar los supuestos de los residuales expuestos en el marco teórico.

- Normalidad

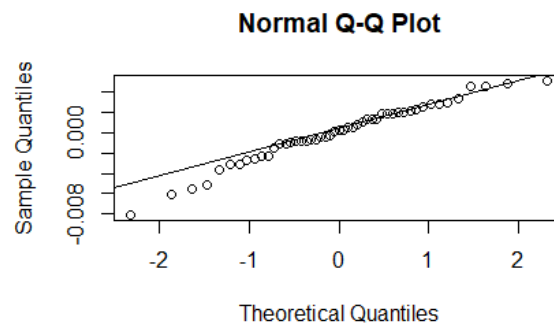


Figura 14: Gráfico de normalidad en los residuales

La prueba gráfica de normalidad evidencia que los residuales se adaptan a la línea de distribución normal, para argumentar esta afirmación se realiza el test de Shapiro-Wilks, esta prueba tiene como hipótesis H_0 : "Los residuales siguen una distribución normal".

Prueba	P-valor
Normalidad	0.2509

Tabla 13: Prueba de Shapiro-Wilk

El valor resultante de la prueba S-W indica que se acepta la hipótesis nula y se confirma que los residuales provienen de una distribución normal.

- Homocedasticidad

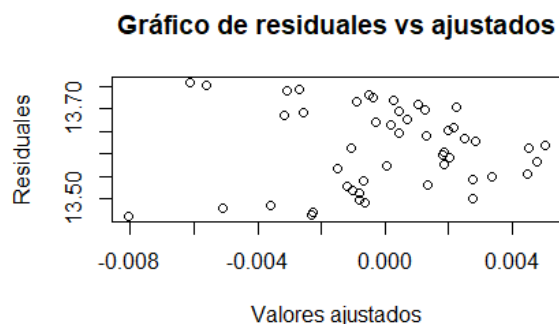


Figura 15: Gráfico de homocedasticidad

Gráficamente se observa que los residuales y los valores ajustados no tienen un patrón de comportamiento, se evidencia que estas observaciones son aleatorias, lo cual se puede afirmar que los residuales no presentan problemas de varianza constante. Para justificar este argumento se realiza la prueba de Breusch-Pagan, cuya hipótesis es H_0 : *Varianza constante*".

Prueba	P-valor
Varianza Constante (ncvTest)	0.3698544

Tabla 14: prueba de Breusch-Pagan

La prueba de B-P da fuerza a la afirmación de que los residuales no viola el supuesto de varianza constante.

- Independencia

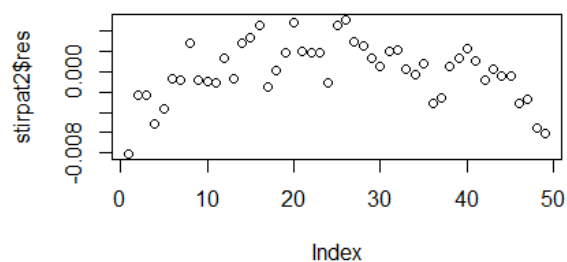


Figura 16: Gráfico de Independencia

Según el gráfico de independencia, se observa una especie de tendencia en los residuales, lo cual puede indicar correlación entre las variables, para corroborar los resultados de la gráfica (16) se evalúa la prueba de Durbin-Watson, la hipótesis contenida en esta prueba es H_0 : "Los errores son independientes".

Prueba	P-valor
Independencia (D-W test)	0

Tabla 15: Prueba de Durbin-Watson

De acuerdo a los resultados de la prueba D-W, los residuales son independientes.

Luego de realizar el análisis adecuado para comprobar el ajuste del modelo STIRPAT, se puede decir que este modelo es válido para estimar y predecir el número de turistas en función de las variables consideradas para el estudio.

6.5.8. Ecuación del modelo

El juicio anterior da pie para representar la ecuación para estimar el número de turistas:

$$\begin{aligned} \ln(\text{Turistas}) = & -65.006714 + 7.035753\ln(\text{Poblacion}) - 0.004993\ln(\text{Caguacom}) \\ & - 0.018047\ln(\text{Basura}) - 0.017241\ln(\text{Alcantarillado}) \end{aligned} \quad (13)$$

6.5.9. Modelo Poisson

El modelo Poisson fue considerado para este estudio por sus características, es un modelo cuya variable respuesta es de conteo y las variables en general pueden estar distribuidas diferente a la normal. El supuesto con más fuerza para evaluar el ajuste en este modelo es la sobredispersión, el cual se prueba en el desarrollo de esta sección.

La ecuación general para el modelo Poisson es la siguiente:

$$\text{Turistas} = e^{(\beta_0 + \beta_1(\text{Poblacion}) + \beta_2(\text{Caguacom}) + \beta_3(\text{Basura}) + \beta_4(\text{Alcantarillado}))}$$

6.5.10. Resultados prueba ANOVA

Los resultados suministrados por la prueba ANOVA para el MLG Poisson son:

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	6.5797992065	0.0223182434	294.817	<0.0000000000000002	***
pobest	0.0000968366	0.0000003322	291.465	<0.0000000000000002	***
aguacom	-0.0026704319	0.0002768295	-9.646	<0.0000000000000002	***
basura	-0.0000117652	0.0000008638	-13.621	<0.0000000000000002	***
alcom	-0.0015523529	0.0001760003	-8.820	<0.0000000000000002	***

Los coeficiente del MLG Poisson son muy significativos, lo que indica que las variables consideradas para el modelo están teniendo un efecto sobre el número de turistas.

6.5.11. Análisis de Dispersión

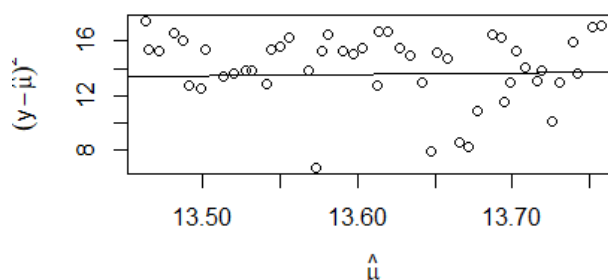


Figura 17: Gráfico de dispersión

Visiblemente la gráfica parece tener sobredispersión, dado que se evidencia más observaciones arriba de la línea media que por debajo, si este modelo no presentara problemas de dispersión, las observaciones estarían muy cercanas a la línea media. Para argumentar lo anterior, se aplica el test de dispersión cuya hipótesis se define de la siguiente forma: $H_0 : \alpha = 0$ vs $H_1 : \alpha > 0$

```
Overdispersion test
data: mod22
z = 4.1718, p-value = 1.511e-05
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
6.237925
```

Siguiendo los resultados de la prueba de dispersión, es evidente que el α generado es mayor a 0, lo que indica que este modelo tiene problemas de dispersión y el caso que se presenta es sobredispersión, es decir la varianza es mayor que la media, violando el supuesto de equidispersión.

6.5.12. Prueba de Bondad de Ajuste

La prueba para validar la precisión y ajuste del MLG Poisson, en este caso se realizará bajo la prueba Chi-cuadrado, según la literatura esta prueba es muy usual para la estructura de este modelo. La hipótesis a probar en este caso es H_0 : "El modelo tiene buen ajuste".

Prueba de Bondad de ajuste χ^2	
Estadístico de prueba	354.658
P-Valor	3.609108e-50

Tabla 16: Bondad de ajuste

Al realizar la prueba de bondad de ajuste, se evidencia a gran escala que este modelo para la variable turistas en función de los servicios básicos y población no es el adecuado, una muestra de ello es el problema de sobredispersión; es posible modificar los parámetros para obtener un mejor ajuste, lo anterior se puede tener en cuenta para futuros análisis.

6.5.13. Modelo Binomial Negativo

El modelo Binomial Negativo es una generalización del modelo Poisson, se utiliza para mejorar el ajuste de un MLG Poisson diagnosticado con problemas de sobredispersión. La ecuación de este MLG se define así:

$$\ln(\text{turistas}) = \hat{\beta}_0 + \text{poblacion}\hat{\beta}_1 + \text{caguacom}\hat{\beta}_2 + \text{basura}\hat{\beta}_3 + \text{alcantarillado}\hat{\beta}_4$$

6.5.14. Resultados prueba ANOVA

Los resultados basados en la prueba ANOVA del MLG Binomial Negativo son los siguientes:

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	6.567e+00	5.970e-02	109.993	< 2e-16	***
pobest	9.702e-05	8.904e-07	108.968	< 2e-16	***
aguacom	-2.559e-03	7.695e-04	-3.326	0.000881	***
basura	-1.156e-05	2.372e-06	-4.873	1.1e-06	***
alcom	-1.637e-03	4.884e-04	-3.352	0.000804	***

Para este modelo los resultados de los coeficientes son significativos, apuntando a que las variables aportan un efecto en la variable respuesta.

6.5.15. Prueba de Bondad de Ajuste

En este caso también se valida el ajuste del MLG Binomial Negativo con la prueba de bondad de ajuste Chi-cuadrado, vale la pena recordar que la hipótesis bajo esta prueba es H_0 : "El modelo tiene buen ajuste".

Prueba de Bondad de ajuste χ^2	
Estadístico de prueba	49.1
P-Valor	0.277

Tabla 17: Bondad de ajuste

La prueba de bondad de ajuste para el MLG Binomial Negativo, indica que el modelo si esta ajustando bien a los datos, por lo cual es un buen modelo a tener en cuenta para la estimación de número de turistas de acuerdo al objetivo del estudio.

7. Predicción.

De acuerdo a la metodología planteada para el desarrollo del estudio, se requiere escoger el mejor modelo entre los propuestos para estimar la carga turística, dos de las herramientas para comparar modelos es el MAPE (Media Absoluta del Error porcentual) y el AIC (criterio de información Akaike). Ambos miden la precisión en la calidad del modelo para predecir.

El criterio para escoger el mejor modelo se basa a partir del AIC y MAPE más pequeño:

Modelo	AIC	MAPE
Regresión Múltiple	775	0.0558
STIRPAT	-424	0.00222
Binomial Negativo	915.8	0.00232

Tabla 18: Método de comparación

Considerando los resultados de la tabla (18) el modelo justificado con el AIC y MAPE más pequeño es el modelo STIRPAT, este resultado indica que este modelo tiene mejor ajuste en comparación con los otros propuestos, para la predicción de la carga turística.

Teniendo en cuenta la finalidad del estudio a gran escala, el resultado obtenido de las predicciones mediante el modelo STIRPAT es el siguiente:

Fecha	Capacidad de Carga Turística	Consumo de agua comercial	Vertimiento alcantarillado	Basuras	Población residente estimado
2012	702.602,8 Tur	1,83 m^3	8,78 m^3	1.438,09 Ton	71.449,6 hab
2013	76.0059,2 Tur	1,53 m^3	9,61 m^3	1.527,16 Ton	72.270,48 hab
2014	817.555,8 Tur	3,16 m^3	9,71 m^3	1.720,08 Ton	73.084,48 hab
2015	883.427 Tur	1,77 m^3	11,34 m^3	1.697,81 Ton	73.890,31 hab
2016	944.180,7 Tur	3,74 m^3	11,39 m^3	2.282,79 Ton	74.686,74 hab

Tabla 19: Predicciones STIRPAT

En la tabla (19) se pueden encontrar la cantidad estimada de turistas por año que debieron ingresar a la isla de San Andrés, entre el año 2012 hasta 2016 de acuerdo a la evolución de las variables estudiadas. Una forma de corroborar el comportamiento de las estimaciones obtenidas por el modelo es compararlo con la estimación de la función logística para turistas.

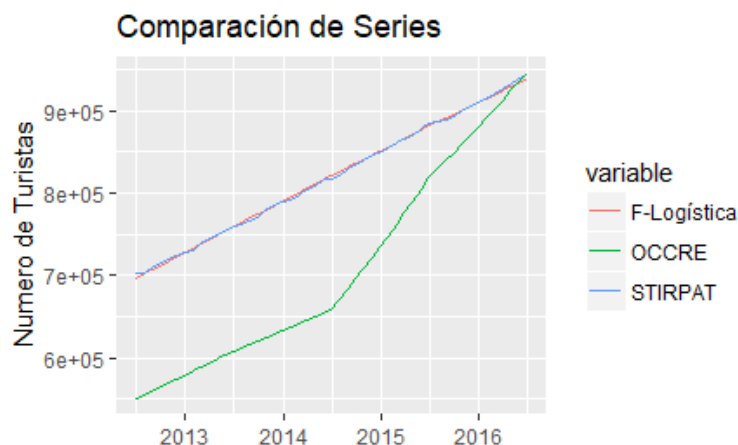


Figura 18: Gráfica de comparación entre la función logística, STIRPAT y OCCRE

De acuerdo al gráfico de comparación del modelo STIRPAT, la función logística y los registros de la OCCRE, las estimaciones de número de turistas se asemejan bastante en su predicción, lo que reafirma el buen ajuste de los turistas mediante la función logística; es notable que los datos registrados por la OCCRE a lo largo del tiempo tiende a acercarse al valor de los datos predichos.

7.1. Proyecciones

El objetivo final del estudio, es predecir a futuro la capacidad de carga turística desde el año 2017 hasta 2020.

Para hacer esta proyección se debe tener en cuenta cada uno de los valores a futuro de las variables explicativas, para esto se realiza la proyección de los servicios básicos por medio de un suavizamiento exponencial simple.

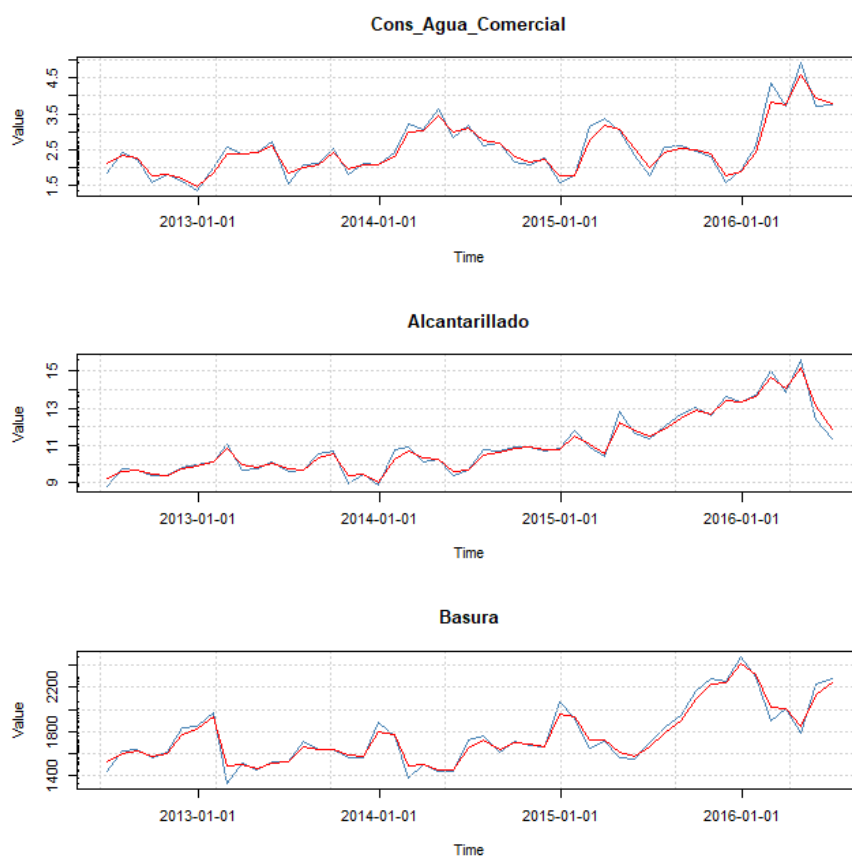


Figura 19: Suavizamiento Exponencial Simple

Teniendo en su totalidad los insumos para hacer la proyección de carga turística hasta el año 2020 por medio de el modelo STIRPAT, se procede realizar la estimación, obteniendo como resultado:

Año	Consumo de agua comercial	Vertimiento alcantarillado	Basuras	Población residente estimado	Capacidad de carga Turística
2017	3,6 m ³	11,58 m ³	2.226,28 Ton	75.472,58 hab	= 1'016.646 Tur
2018	3,8 m ³	11,32 m ³	2.402,55 Ton	76.246,72 hab	= 1'091.060 Tur
2019	3,9 m ³	11,96 m ³	2.382,32 Ton	77.008,13 hab	= 1'168.919 Tur
2020	4,2 m ³	12,24 m ³	2.593,59 Ton	77.755,83 hab	= 1'248.333 Tur

Tabla 20: Proyecciones STIRPAT

Las cifras reflejadas en la tabla (20) especifican la capacidad de turistas esperados que puede recibir la isla de San Andrés, de acuerdo al comportamiento de las variables población residente, consumo de agua comercial, vertimiento de aguas residuales al alcantarillado y generación de basuras; en el período comprendido entre 2017 y 2020.

Si bien, se evidencia una tendencia creciente en el consumo de los recursos y los desechos, pero el comportamiento de población residente y la capacidad de carga comienzan a disminuir en su variación de crecimiento, según la teoría utilizada para este estudio, puede ser indicio de acercarse al límite máximo de los recursos que ofrece la isla.

Es claro que la base de la economía del departamento proviene del sector turístico y que el crecimiento de éste, repercute en los ingresos de quienes se benefician de este sector productivo, pero así como genera utilidad para una proporción definida, impacta negativamente en los recursos limitados que puede ofrecer la isla a corto plazo.

El comportamiento creciente del ingreso de turistas se va a mantener con esta tendencia mientras se ofrezcan los vuelos a bajo costo, los paquetes turísticos todo incluido y la desmesurada aparición de posadas nativas o apartamentos turísticos. De acuerdo a los registros de la secretaria de turismo para el año 2016 el número de camas para disponibilidad por día eran 23.339 del total de hospedajes (hoteles, residencias, cabañas, etc) que se encuentran activos en la cámara de comercio, que al año representa una capacidad para albergar hasta 8'500.000 turistas.

Se hace necesario una intervención en el menor tiempo posible, para buscar alternativas en el corto y mediano plazo, que puedan sobrellevar el arribo de turistas y mantener un balance de los recursos y descarga de residuos en la isla de San Andrés.

8. Conclusiones.

1. De los resultados en general, se puede observar que las proyecciones de las series tienen una tendencia creciente a lo largo de los años, que si no se interviene en el menor tiempo posible, se pueden presentar impactos negativos en el ambiente y los recursos de la Isla. Es evidente que la proyección de turistas refleja un comportamiento verhulstiano y si el arribo de turistas mantiene la tendencia de acuerdo a los resultados de la investigación, se puede decir que está llegando al límite de los recursos más importantes ofrecidos por la Isla; ejemplo de ello, la isla de San Andrés en materia de desechos está agotando la capacidad del relleno sanitario, según la secretaría de servicios públicos y ambiente le restan aproximadamente 3 años de vida útil; esto sin nombrar los repetitivos rebosamientos de aguas residuales de los alcantarillados ubicados en la zona central-comercial.

Lo anterior es indicio de que el ajuste de la función logística para la estimación del total de turistas, está logrando su cometido y está representando de la mejor manera el comportamiento a futuro de la capacidad de carga turística.

2. A la hora de obtener una buena estimación de población para áreas con cierto tipo de limitaciones, la función logística puede ser una buena herramienta demográfica, debido a que tiene un buen ajuste para este tipo de datos. Se debe tener en cuenta el cálculo adecuado de los pivotes y del límite superior para obtener un ajuste óptimo de la población, el cual se aconseja por facilidad utilizar una calibración simultánea del método de bisección.
3. Para este análisis el modelo STIRPAT fue un buen aliado, además de su estructura para la estimación de impacto en función de variables ambientales, tuvo el mejor ajuste para predecir y proyectar la carga turística.
4. Algunos resultados como la influencia de los turistas en el consumo de agua residencial, no tuvo significancia en los modelos, corroborando la hipótesis planteada por Adrina Acosta en su ponencia *Efecto de la Densidad Poblacional sobre las fuentes fluviales de reserva en el archipiélago de San Andrés y Providencia*.
5. El modelo Poisson para este caso no tuvo un buen ajuste y se vio reflejado en la sobredispersión presentada; el modelo binomial negativo en cambio, si tuvo un mejor ajuste con los datos evaluados, este modelo se podría tener en cuenta para investigaciones futuras.
6. Las comparaciones entre las estimaciones del modelo STIRPAT y la función logística, fueron muy parecidas aportando al indicio de buen ajuste en predicción para cualquiera de los dos métodos.

9. Recomendaciones

La proyección del comportamiento de las variables consumo de agua, vertimiento de aguas residuales y basuras, se obtuvieron por medio de un suavizamiento exponencial simple; estos resultados fueron utilizados para estimar el número de turistas esperados para los años proyectados. Si se quiere tener la cifra óptima de capacidad de carga turística para un periodo de tiempo establecido, se puede ajustar el modelo con los valores predeterminados para cada una de las variables de acuerdo a parámetros de medición establecidos.

Este ejercicio se debe hacer en el menor tiempo posible, dado que las proyecciones de las series arrojan crecimientos en los consumos de recursos y en la población; lo cual cada vez va a reducir la capacidad que tiene la Isla para albergar turistas cada año, repercutiendo los efectos que se pueden producir en los residentes habituales.

10. Agradecimientos

Ante todo a Dios, a mis padres y a mi familia por la confianza y la oportunidad que me brindaron durante toda la carrera universitaria, a mi director de tesis por el apoyo, consejos y enseñanzas, a los profesores Wilmer Pineda y Felipe Ortíz por la invaluable ayuda y a mis compañeros Edinson, Dayana, Vanessa y Juan D., por la paciencia, colaboración y el cariño durante el proceso universitario.

11. Referencias

1. Acosta, A. (2016). *Efectos de la Densidad Poblacional sobre las fuentes fluviales de reserva en el archipiélago de San Andrés y Providencia*. Ponencia presentada en el seminario de Estadística. Universidad Nacional. Bogotá D.C.
2. Cameron A., Trivedi K., (1990), *Regression-Based Tests for Overdispersion in the Poisson Models*. Journal of Econometrics, 1990, Vol. 46(3), pp.347-364.
3. DANE. Departamento Administrativo Nacional De Estadística. Obtenido de <http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion>.
4. DNP. (2016). *Informe Archipelago de San Andrés, Providencia y Santa Catalina*.
5. Marquéz, G., James, J., Márquez Pérez, A. I., Castellanos, O., & Taylor, S(2011). *Consideraciones sobre desarrollo y sostenibilidad del Archipiélago de San Andrés, Providencia y Santa Catalina*. Revista Aguaita No.22 (Mayo 2011) pag.62.
6. Triana, D., Londoño L. (2016), *Archipiélago de San Andrés, Providencia y Santa Catalina, un análisis demográfico*.
7. UNGRD. (15 de abril de 2016). Portal de Gestion del Riesgo. Obtenido de <http://portal.gestiondelriesgo.gov.co/Paginas/Noticias/2016/Desplegada-operacion-en-la-isla-de-San-Andres-para-mitigar-falta-de-agua.aspx>