

**ESTIMACIÓN DEL CONSUMO POTENCIAL DE ENERGÍA ELÉCTRICA
MEDIANTE EL USO DE UN MODELO MATEMÁTICO PARA EL
DEPARTAMENTO DEL VICHADA DURANTE EL PERIODO 2019-2039**

KAREN AVILAN TRUJILLO

FERNANDO ALEISY GONZÁLEZ

**UNIVERSIDAD DE LOS LLANOS
FACULTAD DE CIENCIAS HUMANAS Y DE LA EDUCACIÓN
ESCUELA DE PEDAGOGÍA Y BELLAS ARTES
PROGRAMA DE LICENCIATURA EN MATEMÁTICAS Y FÍSICA
VILLAVICENCIO
2019**

**ESTIMACIÓN DEL CONSUMO POTENCIAL DE ENERGÍA ELÉCTRICA
MEDIANTE EL USO DE UN MODELO MATEMÁTICO PARA EL
DEPARTAMENTO DEL VICHADA DURANTE EL PERIODO 2019-2039**

**KAREN AVILAN TRUJILLO
Código 141002225**

**FERNANDO ALEISY GONZÁLEZ
Código 141002403**


**Trabajo de grado modalidad EPI como requisito para optar por el título de
Licenciado(a) en Matemáticas y Física**

Directores:

**Ivonne Amparo Londoño Agudelo
Magister en Docencia de las Matemáticas**

**Cristóbal Lugo López
Magister de Planificación y manejo ambiental de cuencas hidrográficas**

**UNIVERSIDAD DE LOS LLANOS
FACULTAD DE CIENCIAS HUMANAS Y DE LA EDUCACIÓN
ESCUELA DE PEDAGOGÍA Y BELLAS ARTES
PROGRAMA DE LICENCIATURA EN MATEMÁTICAS Y FÍSICA
VILLAVICENCIO
2019**

	UNIVERSIDAD DE LOS LLANOS	CÓDIGO: FO-DOC-97
		VERSIÓN: 02 PÁGINA: 3
	PROCESO DOCENCIA	FECHA: 02/09/2016
	FORMATO AUTORIZACION DE DERECHOS	VIGENCIA: 2016

**FACULTAD DE CIENCIAS HUMANAS Y DE LA EDUCACIÓN
UNIVERSIDAD DE LOS LLANOS**

AUTORIZACIÓN

Yo Fernando Aleisy Gonzalez mayor de edad, identificado con la Cédula de Ciudadanía No. 1.119.887.249 de Cumaral-Meta, actuando en nombre propio en mi calidad de autor del trabajo de tesis, monografía o trabajo de grado denominado Estimación del consumo de energía eléctrica mediante el uso de un modelo matemático para el departamento de Vichada durante en el periodo 2019 - 2039, hago entrega del ejemplar y de sus anexos de ser el caso, en formato digital o electrónico (CD-ROM) y autorizo a la **UNIVERSIDAD DE LOS LLANOS**, para que en los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia, con la finalidad de que se utilice y use en todas sus formas, realice la reproducción, comunicación pública, edición y distribución, en formato impreso y digital, o formato conocido o por conocer de manera total y parcial de mi trabajo de grado o tesis.

Yo Fernando Aleisy Gonzalez, Como autor, manifiesto que el trabajo de grado objeto de la presente autorización, es original y se realizó sin violar o usurpar derechos de autor de terceros; por tanto, la obra es de mi exclusiva autoría y poseo la titularidad sobre la misma; en caso de presentarse cualquier reclamación o acción por parte de un tercero en cuanto a los derechos de autor sobre la obra en cuestión, como autor, asumiré toda la responsabilidad, y saldré en defensa de los derechos aquí autorizados, para todos los efectos la Universidad actúa como un tercero de buena fe.


Para constancia, se firma el presente documento en dos (2) ejemplares del mismo valor y tenor en Villavicencio - Meta, a los dieciséis (16) días del mes de septiembre de dos mil diecinueve (2019).

EL AUTOR – ESTUDIANTE

Firma

Nombre: Fernando Aleisy Gonzalez

C.C. No. 1.119.887.249 de Cumaral-Meta

	UNIVERSIDAD DE LOS LLANOS	CÓDIGO: FO-DOC-97
		VERSIÓN: 02 PÁGINA: 4
	PROCESO DOCENCIA	FECHA: 02/09/2016
	FORMATO AUTORIZACION DE DERECHOS	VIGENCIA: 2016

**FACULTAD DE CIENCIAS HUMANAS Y DE LA EDUCACIÓN
UNIVERSIDAD DE LOS LLANOS**

AUTORIZACIÓN

Yo Karen Liced Avilan Trujillo mayor de edad, identificado con la Cédula de Ciudadanía No. 1.121.862.639 de Santa Fe de Bogotá, actuando en nombre propio en mi calidad de autor del trabajo de tesis, monografía o trabajo de grado denominado Estimación del consumo de energía eléctrica mediante el uso de un modelo matemático para el departamento de Vichada durante en el periodo 2019 - 2039, hago entrega del ejemplar y de sus anexos de ser el caso, en formato digital o electrónico (CD-ROM) y autorizo a la **UNIVERSIDAD DE LOS LLANOS**, para que en los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia, con la finalidad de que se utilice y use en todas sus formas, realice la reproducción, comunicación pública, edición y distribución, en formato impreso y digital, o formato conocido o por conocer de manera total y parcial de mi trabajo de grado o tesis.

Yo Karen Liced Avilan Trujillo, Como autor, manifiesto que el trabajo de grado o tesis objeto de la presente autorización, es original y se realizó sin violar o usurpar derechos de autor de terceros; por tanto, la obra es de mi exclusiva autoría y poseo la titularidad sobre la misma; en caso de presentarse cualquier reclamación o acción por parte de un tercero en cuanto a los derechos de autor sobre la obra en cuestión, como autor, asumiré toda la responsabilidad, y saldré en defensa de los derechos aquí autorizados, para todos los efectos la Universidad actúa como un tercero de buena fe.

Para constancia, se firma el presente documento en dos (2) ejemplares del mismo valor y tenor en Villavicencio - Meta, a los dieciséis (16) días del mes de septiembre de dos mil diecinueve (2019).

EL AUTOR – ESTUDIANTE

Firma

Nombre: Karen Liced Avilan Trujillo

C.C. No. 1.121.862.639

de Villavicencio

AUTORIDADES ACADÉMICAS

PABLO EMILIO CRUZ CASALLAS
Rector

MARIA LUISA PINZÓN ROCHA
Vicerrectora académica

GIOVANNY QUINTERO REYES
Secretario general

LUZ HAYDEÉ GONZÁLEZ OCAMPO
Decana de la Facultad Ciencias Humanas y de la Educación

BEATRIZ AVELINA VILLARRAGA BAQUERO
Directora de la Escuela de Pedagogía y Bellas Artes

NASLY YANIRA MARTINEZ VELASQUEZ
Director del Programa de Licenciatura en Matemáticas y Física

NOTA DE ACEPTACIÓN

Aprobado en cumplimiento de los requisitos exigidos por la Universidad de los Llanos para optar al título de Licenciado(a) en Matemáticas y Física En constancia de lo anterior, firman:

FREDY LEONARDO DUBEIBE MARIN

Director centro de investigaciones de la Facultad de Ciencias Humanas y Educación

NASLY YANIRA MARTINEZ VELASQUEZ

Director de Programa

FRANCISCO JAVIER GUTIÉRREZ

Evaluador

CESAR HUERTAS CHAPARRO

Evaluador

IVONNE AMPARO LONDOÑO AGUDELO

Director de la opción de grado

CRISTÓBAL LUGO LÓPEZ

Director de la opción de grado

Villavicencio, 16 de septiembre de 2019

AGRADECIMIENTOS

Yo Fernando Aleisy Gonzalez agradezco a la Profesora Ivonne Amparo Londoño Agudelo y al ingeniero Cristóbal Lugo López, directores de mi trabajo de grado, por los momentos dedicados para aclarar cualquier tipo de duda que me surgiera, por acompañarme en el proceso de aprendizaje que se dio durante la elaboración de este trabajo y por apoyarme de manera personal e institucional.

También agradezco a Salomé González Acuña, mi mamá, por cada consejo que me ha dado a lo largo de mi vida, consejos que han forjado mi carácter, que me han guiado en mi aprendizaje académico y en la vida. Por enseñarme que el conocimiento es el bien más grande que una persona puede tener.

Yo Karen Liced Avilan Trujillo agradezco a los directores del trabajo de grado, al ingeniero Cristóbal Lugo López y la profesora Ivonne Amparo Londoño Agudelo por ser una guía durante en el desarrollo de este trabajo, por sus valiosas asesorías y comentarios.

Gracias a mis padres, Catalina Trujillo y Pablo Avilan, por la confianza puesta en mí, por el apoyo que siempre me han brindado, por desear y anhelar lo mejor para mi vida, también les agradezco por cada consejo y por cada una de sus palabras que me guiaron durante toda mi existencia.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN	13
1. MARCO REFRENCIAL	15
1.1. ESTADO DEL ARTE	15
1.2. MARCO TEORICO CONCEPTUAL.....	16
1.3. ANÁLISIS DE HIPÓTESIS Y TABLA F.....	24
1.4. COEFICIENTE DE DETERMINACIÓN.....	25
1.5. COEFICIENTE DE DETERMINACIÓN AJUSTADO.....	25
1.6. ANÁLISIS DE RESIDUALES.....	26
2. METODOLOGÍA	33
2.1. ENFOQUE METODOLÓGICO	33
2.2. ESTRATEGIA DE INVESTIGACIÓN.....	33
2.3. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN.....	34
2.4. ANÁLISIS Y SISTEMATIZACIÓN.....	34
3. RESULTADOS Y ANÁLISIS DE RESULTADOS	36
3.1. MODELOS PARA ESTIMAR LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA.....	38
3.2. MODELOS PARA ESTIMAR CONSUMO DE ENERGÍA ELÉCTRICA ...	49
4. CONCLUSIONES	62
5. RECOMENDACIONES	64
BIBLIOGRAFÍA.....	65
ANEXOS.....	67
RESUMEN ANALITICO ESPECIALIZADO.....	82

LISTA DE GRÁFICAS

	Pág.
GRÁFICA 1. DE PROBABILIDAD NORMAL: A) GRÁFICA IDEAL, B) COLAS GRUESAS, C) DISTRIBUCIÓN CON COLAS DELGADAS D) ASIMETRÍA POSITIVA Y E) ASIMETRÍA NEGATIVA.....	31
GRÁFICA 2. MODELOS DE GRÁFICAS DE RESIDUALES EN FUNCIÓN DEL ÍNDICE ESTIMADO (\hat{Y}_i): A) SITUACIÓN IDEAL, B) EN EMBUDO, C) DOBLE ARCO Y D) NO LINEAL	32
GRÁFICA 3 DISPERSIÓN DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA, SEGÚN CENSO DEL DANE, EN FUNCIÓN DEL TIEMPO EN AÑOS.....	39
GRÁFICA 4. CRECIMIENTO DE LA POBLACIÓN EN LOS PERIODOS 1985-1993, 1993-2005 Y 2005-2018	39
GRÁFICA 5. RESIDUALES DE LA REGRESIÓN LINEAL SIMPLE PARA LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN DANE.....	40
GRÁFICA 6. REGRESIÓN LINEAL SIMPLE DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA A PARTIR DE CENSOS DEL DANE	41
GRÁFICA 7. RESIDUALES DE LA REGRESIÓN EXPONENCIAL PARA LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN DANE.....	43
GRÁFICA 8. REGRESIÓN EXPONENCIAL DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA A PARTIR DE CENSOS DEL DANE	43
GRÁFICA 9. RESIDUALES DE LA REGRESIÓN PARABÓLICA PARA LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN DANE.....	45
GRÁFICA 10. REGRESIÓN PARABÓLICA DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA A PARTIR DE CENSOS DEL DANE	45
GRÁFICA 11. RESIDUALES DE LA REGRESIÓN POBLACIONAL LOGÍSTICA PARA LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN DANE.....	47
GRÁFICA 12. REGRESIÓN POBLACIONAL LOGÍSTICA DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA A PARTIR DE CENSOS DEL DANE	47
GRÁFICA 13. MATRIZ DE DISPERSIÓN DE LAS VARIABLES AÑO, USUARIOS, POBLACIÓN Y CONSUMO DEL DEPARTAMENTO DEL VICHADA	50
GRÁFICA 14. RESIDUALES DE LA REGRESIÓN LINEAL SIMPLE PARA EL CONSUMO DEL DEPARTAMENTO DEL VICHADA SEGÚN SUI.....	51
GRÁFICA 15. REGRESIÓN LINEAL SIMPLE DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA A PARTIR DE DATOS DEL SUI.....	52
GRÁFICA 16. RESIDUALES DE LA REGRESIÓN EXPONENCIAL PARA EL CONSUMO DEL DEPARTAMENTO DEL VICHADA SEGÚN SUI.....	53

GRÁFICA 17. REGRESIÓN EXPONENCIAL DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA A PARTIR DE DATOS DEL SUI.....	54
GRÁFICA 18. RESIDUALES DE LA REGRESIÓN PARABÓLICA PARA EL CONSUMO DEL DEPARTAMENTO DEL VICHADA SEGÚN SUI.....	55
GRÁFICA 19. REGRESIÓN PARABÓLICA DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA A PARTIR DE DATOS DEL SUI.....	56
GRÁFICA 20. RESIDUALES DE LA REGRESIÓN LOGÍSTICA PARA EL CONSUMO DEL DEPARTAMENTO DEL VICHADA SEGÚN SUI.....	57
GRÁFICA 21. REGRESIÓN LOGÍSTICA DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA A PARTIR DE DATOS DEL SUI.....	58

LISTA DE TABLAS

	Pág.
TABLA 1. TABLA F, TABLA QUE PERMITE ESTABLECER EL VALOR F DE FISHER PARA EL ANÁLISIS DE VARIANZA (ANOVA).....	24
TABLA 2. CENSOS DEL DEPARTAMENTO DEL VICHADA, FUENTE DANE	36
TABLA 3. CONSUMO DE ENERGÍA POR VIVIENDA, FUENTE SUI	36
TABLA 4. CONSUMO DE ENERGÍA ELÉCTRICA DEL VICHADA AÑO 2018, FUENTE PERS ORINOQUIA	37
TABLA 5. RESUMEN DEL CONSUMO DE ENERGÍA EN EL VICHADA, FUENTE DANE, PERS ORINOQUIA Y SUI	37
TABLA 6. ANOVA DE LA REGRESIÓN LINEAL SIMPLE DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN CENSOS DEL DANE	40
TABLA 7. ANOVA DE LA REGRESIÓN EXPONENCIAL DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN CENSOS DEL DANE	42
TABLA 8. ANOVA DE LA REGRESIÓN PARABÓLICA DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN CENSOS DEL DANE	44
TABLA 9. ANOVA DE LA REGRESIÓN LOGÍSTICA POBLACIONAL DE LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA SEGÚN CENSOS DEL DANE.....	46
TABLA 10. RESUMEN DEL COEFICIENTE DE DETERMINACIÓN Y RESIDUAL PRESS DE CADA MODELO APLICADO A LA POBLACIÓN DE DEPARTAMENTO DEL VICHADA SEGÚN DATOS DEL DANE	48
TABLA 11. POBLACIÓN ESTIMADA PARA EL PERIODO 2013-2016 DEL DEPARTAMENTO DEL VICHADA POR MEDIO DEL MODELO LINEAL SIMPLE.....	49
TABLA 12. MATRIZ DE CORRELACIÓN DE LAS VARIABLES AÑO, USUARIOS, POBLACIÓN Y CONSUMO DEL DEPARTAMENTO DEL VICHADA	50
TABLA 13. ANOVA DE LA REGRESIÓN LINEAL SIMPLE DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA SEGÚN EL SUI.....	51
TABLA 14. ANOVA DE LA REGRESIÓN EXPONENCIAL DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA SEGÚN EL SUI.....	52
TABLA 15. ANOVA DE LA REGRESIÓN PARABÓLICA DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA SEGÚN EL SUI.....	54
TABLA 16. ANOVA DE LA REGRESIÓN LOGÍSTICA DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA SEGÚN EL SUI.....	56
TABLA 17. RESUMEN DEL COEFICIENTE DE DETERMINACIÓN Y RESIDUAL PRESS DE CADA MODELO APLICADO AL CONSUMO DE ENERGÍA	

ELÉCTRICA DE DEPARTAMENTO DEL VICHADA SEGÚN DATOS DEL SUI.....	58
TABLA 18. CONSUMOS DE ENERGÍA ELÉCTRICA ESTIMADOS CON LOS MODELOS LINEAL SIMPLE, EXPONENCIAL, PARABÓLICO Y LOGÍSTICO, DEL DEPARTAMENTO DEL VICHADA A PARTIR DE DATOS RECOLECTADOS DEL SUI.	59
TABLA 19. ESCENARIOS FUTUROS DEL CONSUMO DE ENERGÍA ELÉCTRICA DEL DEPARTAMENTO DEL VICHADA EN EL PERIODO 2019-2039.....	60

INTRODUCCIÓN

El Plan de Energización Rural Sostenibles de la Orinoquia (PERS Orinoquía) es un plan que nace de reconocimiento de la problemática asociada a la falta de expansión de la cobertura de la energía eléctrica en las zonas rurales apartadas y que, a partir de los elementos regionales más relevantes en términos de condiciones socioeconómicas, de recurso, de oferta y demanda energética, busca identificar las fuentes locales aprovechables para el suministro de energía mediante la estructuración y elaboración de proyectos integrales y sostenibles en el corto plazo y en un horizonte de mínimo 15 años, que además de generar energía, apoyen el crecimiento y el desarrollo de las comunidades rurales de las regiones objetivas.

PERS Orinoquía será la formulación de una política pública energética que vaya en sintonía con el entorno y la visión de desarrollo regional con emprendimientos y productividad local, a partir de la generación de energía eléctrica que apoye el crecimiento y el mejoramiento de las condiciones de las comunidades locales en los departamentos de Arauca, Casanare, Meta y Vichada.

Con este trabajo se busca aportar a PERS Orinoquía puesto que la estimación del consumo de energía eléctrica entre los años 2019 y 2039 permitirá que se establezca, entre la política pública, la formulación de proyectos de energización en donde se haga uso de energías alternativas a partir de la realidad del departamento del Vichada. Estos proyectos permitirán que la población del departamento del Vichada tenga acceso a los fondos de apoyo financiero, que brinda el gobierno.

El departamento del Vichada presenta gran interés para PERS Orinoquía puesto que no se encuentra en el sistema interconectado nacional (SIN), se estableció que, a partir de los datos del DANE y del SUI, que el departamento tiene un déficit en prestación del servicio de 54,3 %. Los departamentos de la periferia son los más afectados principalmente por lo costoso que sería llevar la infraestructura eléctrica. Según la IPSE, al 31 de diciembre de 2018, el 94% de la energía consumida en Puerto Carreño correspondía a la energía suministrada por la interconexión con Venezuela, el 6 % restante se genera desde la central DIESEL de Puerto Carreño.

El departamento del Vichada está situado en el extremo oriental de Colombia y se encuentra conformado por los municipios de Cumaribo, La primavera, Puerto Carreño y Santa Rosalía

Para obtener información energética en el Vichada, se realizó una investigación con enfoque cuantitativo, en el que se estableció el tipo de correlación entre distintas variables, tales como, la población total, el número de viviendas con energía eléctrica, con la variable consumo en kWh/mes en el periodo 2005-2018. Lo anterior

permitió identificar el modelo adecuado para estimar el consumo de energía eléctrica para el departamento en el periodo 2019-2039 en diversos escenarios.

Como estrategia de investigación se usó el estudio de caso. Se analizó la información existente sobre la población y el consumo de energía eléctrica del departamento de Vichada en las bases del Sistema único de información (SUI), y DANE, junto con el consumo de energía eléctrica del 2018 del Departamento de Vichada según los resultados obtenidos en el proyecto PERS Orinoquía.

Luego de tener la información recopilada, se diseñó y se aplicó un script de R Project por cada modelo de regresión, a los datos y se obtuvo modelos ajustados con los que se estimó los consumos de energía eléctrica en el periodo del 2005 al 2018 en el departamento. Los valores estimados se usaron para los análisis de varianza (ANOVA), cálculo del coeficiente de determinación o determinación ajustado y de residuales. Luego de determinar el modelo más adecuado, se procedió a estimar los consumos de energía eléctrica para el periodo del 2019-2039 del departamento, con el modelo que mejor se ajustó a los datos recolectados.

En este proyecto se encontró que el consumo de energía eléctrica para el periodo 2019-2039 está entre 2201982 kWh/mes y 3117186 kWh/mes; pero para que toda la población tenga acceso a la energía eléctrica se requiere una producción de energía de 4818341.36 kWh/mes para el año 2019 y 6820975.93 kWh/mes para el año 2039.

El presente trabajo se encuentra estructurado de la siguiente manera: el marco referencial o revisión bibliográfica, la metodología, los resultados de aplicar los diferentes modelos a los datos, el análisis de estos resultados, las conclusiones y los anexos.

1. MARCO REFERENCIAL

1.1. ESTADO DEL ARTE

El término regresión¹ fue utilizado por primera vez en la estadística por Sir Francis Galton, el cual designó la palabra regresión como el proceso general de predecir una variable a partir de otra. Francis Galton realizó experimentos con semillas y observaciones a estaturas de padres e hijos.

Actualmente el análisis de regresión es la herramienta útil para los investigadores de las diversas ciencias porque permite entender los fenómenos estudiados, bien sea para analizar la asociación o relación entre variables o para usarla con fines predictivos.

En múltiples ocasiones se ha usado modelos de regresión para pronosticar la demanda de energía eléctrica, como Orellana Romero² en su tesis para optar al grado de maestro en estadística, identificó que el modelo ARIMA estacional es más preciso que el modelo econométrico utilizado por la Unidad de Transacciones, S.A. (UT), para realizar el pronóstico de la demanda de energía eléctrica de Mediano Plazo de El Salvador. Otros autores que han usado modelos para pronosticar la demanda de energía eléctrica son Barrientos, Olaya y González³ quienes realizaron pronósticos para el departamento del Valle del Cauca, al usar el modelo spline obtuvieron como resultado que este es apropiado cuando se cuenta con información reciente, por lo tanto es preferible su uso frente al modelo Auto Regresivo de Media Móvil (ARIMA).

Zetina y Ríos⁴ utilizaron modelos poblacionales para estimar la población de langostas *Panulirus argus* de las costas de Yucatán México, al realizar la proyección de la población concluyeron que el nivel de explotación se puede mantener durante algunos años. Por otro lado, Ricardo Ocaña⁵ analizó los modelos markovianos

¹ CARDONA, Diego, *et al.* Inferencia estadística. Módulo de regresión lineal simple. Documento de investigación No. 147. Bogotá D.C. 2013.

² ORELLANA, José Luis. Modelación y Pronóstico de la Demanda de Energía Eléctrica de Mediano Plazo de El Salvador. Ciudad Universitaria, El Salvador. 2012

³ BARRIENTOS, Andres Felipe. OLAYA, Javier y GONZÁLEZ, Victor Manuel. Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, 30(2), 2007 pp 187-202.

⁴ ZETINA, Carlos y RÍOS, Gloria. Estimación del tamaño de la población de langostas *Panulirus Argus* en las costas de Yucatan, usando diferentes modelos de evaluación. 1998.

⁵ OCAÑA, Ricardo. Modelos de Markov aplicados a la investigación en Ciencias de la Salud. *Revista Interciencia*. 2009. pp. 157-162

utilizados en las ciencias de la salud, centrándose en los métodos de estimación, la interpretación de resultados y diversos softwares disponibles; abordando cuatro aspectos fundamentales para llevar a cabo análisis de datos basados en modelos de Markov, los cuales son: definición de los estados del proceso y los mecanismos de transición entre ellos, selección del modelo más apropiado, determinación de las probabilidades de transición entre estados y descripción de la evolución temporal del proceso.

Costa, Boj del Val y Fortiana⁶ estudiaron los criterios para la elección de un punto de corte adecuado en el modelo de regresión logística basado en distancias. Todo ello con aplicación al problema de credit scoring. Los criterios de calidad de ajuste se analizaron con el coeficiente Kolmogorov-Smirnov y el índice de Gini, junto con la representación gráfica ROC. También calcularon las probabilidades de mala clasificación y unas funciones de coste del error.

No obstante, hay situaciones donde no se ha utilizado un modelo matemático para pronosticar el consumo de energía eléctrica, como es el caso de Medina y García⁷, quienes, por medio dos técnicas de inteligencia artificial pronosticaron la demanda de energía eléctrica mensual en Colombia con el objetivo de encontrar un método que permita una disminución importante en los errores de predicción porque esto da beneficios importantes para todos los agentes que operan en el sector eléctrico. Jorge Barrientos y Mónica Toro⁸. también utilizaron la técnica de inteligencia artificial de redes neuronales junto con técnicas de vectores de corrección de errores para hacer pronósticos de precios de energía eléctrica en Colombia a largo plazo y obtuvieron mejores resultados con la segunda técnica.

1.2. MARCO TEORICO CONCEPTUAL

A diario los investigadores científicos se encuentran con variables que pueden estar relacionadas, que les permite hacer predicciones de una de ellas a partir de la otra o las otras variables, por ello es necesario establecer si en realidad existe una relación entre ellas, para lo cual se recurre al procedimiento de análisis de regresión y correlación.

Según Ciro Martínez⁹ la relación que puede existir entre dos variables puede ser de: dependencia causa unilateral, cuando una variable influye a la otra, pero no al contrario; interdependencia, cuando la influencia entre las variables es recíproca;

⁶ COSTA, Teresa., BOJ DEL VAL, Eva y FORTIANA, José. Bondad de ajuste y elección del punto de corte en regresión logística basada en distancias. Aplicación al problema de "credit scoring". *Dialet*(18), 2012. Pp 19-40.

⁷ MEDINA, Santiago, y GARCÍA Josefina. Predicción de demanda de energía en Colombia mediante un sistema de inferencia difuso neuronal. *Energética. Revista Energética*(33). 2005. pp 15-24

⁸ BARRIENTO, Jorge, y TORO, Mónica. Análisis de los fundamentales del precio de la energía eléctrica: evidencia empírica para Colombia. *revista de economía del caribe*(19). 2017. pp 33-59.

⁹ MARTÍNEZ, Ciro. *Estadística y muestreo*. 13 ed. Bogotá D.C.: Ecoe ediciones. 2012.

de dependencia indirecta, en la que dos a más variable pueden mostrar una correlación a través de una tercera variable que influye en ellas; de concordancia, si dos variables independientes se les puede determinar correlación; covarianza casual si la correlación que se presenta entre las dos variables es totalmente casual o accidentada.

1.2.1. Modelos de regresión

Un modelo de regresión es un modelo matemático en el que permite relacionar una variable respuesta con una o múltiples variables regresoras en donde la variable respuesta es la que se quiere predecir a partir de la o las variables regresoras. Se suele representar a la variable respuesta con una y y a las variables regresoras con $x_1, x_2, x_3, \dots, x_p$ donde p es el número de variable regresoras.

En la regresión simple se analiza la asociación y relación entre un conjunto en los que están involucrados dos variables de la cuales una es la regresora y la otra la variable respuesta.

Para trabajar los modelos con mínimos cuadrados se suele utilizar la linealización en la que se busca expresar la ecuación del modelo en forma de una ecuación lineal, en otras palabras, la variable respuesta depende de una combinación lineal de funciones de la o las variables regresoras.

1.2.2. Modelo de regresión lineal simple

La regresión lineal simple es un modelo con una sola variable regresora x que tiene una relación con una respuesta y , donde la relación es una línea recta. El modelo de regresión lineal es:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

Donde la ordenada al origen β_0 y la pendiente β_1 son constantes desconocidas, y ε es un componente aleatorio de error. Se supone que los errores tienen promedio cero y varianza σ^2 desconocida. Además, se supone que los errores no están correlacionados. Esto quiere decir que el valor de un error no depende del valor de cualquier otro error.

El modelo de regresión ajustado para la variable regresora corresponde a la ecuación (2).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \varepsilon \quad (2)$$

Bajo el supuesto de que los errores tienen promedio cero el valor esperado o valor promedio de la variable respuesta \hat{y} (estimado de la variable respuesta) está dado por la ecuación (3)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (3)$$

La estimación de los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$ se realiza bajo el supuesto de mínimos cuadrados del cual se deduce las ecuaciones (4) y (5)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (5)$$

Dónde:

x_i es el valor de la variable regresora en la observación i .

y_i es el valor de la variable respuesta en la observación i .

\bar{x} es la media de la variable regresora.

\bar{y} es la media de la variable respuesta.

n es el número total de observaciones.

1.2.3. Modelo de regresión Lineal Múltiple

En la RM (Regresión Múltiple) se analiza la asociación y relación entre las variables de un conjunto de datos en los que están involucrados más de dos variables de las cuales una es la dependiente y las otras las independientes. A la ecuación que describe cómo está relacionada la variable dependiente y con las variables independientes $x_1, x_2, x_3, \dots, x_p$. le conoce como modelo de RM. El modelo de RM toma la forma siguiente:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (6)$$

En el modelo de RM, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son parámetros, donde p es el número de variables regresoras y el término del error ϵ es una variable aleatoria. Los parámetros β_p hacen de este modelo que sea lineal.

El modelo (1) puede ser expresado de forma matricial:

$$Y = XB + \epsilon \quad (7)$$

$$\text{Con } Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ 1x_{n1} & \dots & x_{nk} \end{bmatrix}, \quad B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad y \quad \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

donde Y es el vector de observaciones de orden $n \times 1$, X es la matriz de las variables independiente ($n \times p$) donde n es el número de observaciones o el tamaño de la muestra, B es el vector de los coeficientes de la regresión ($p \times 1$) y ε es el vector de los errores aleatorios ($n \times 1$).

La estimación de los parámetros asociados al modelo (2) se realiza a través de mínimos cuadrados con el fin de minimizar el error:

$$L = \sum_{i=1}^N \varepsilon_i^2 = \varepsilon' \varepsilon = (Y - XB)'(Y - XB) \quad (8)$$

Mediante procedimientos matemáticos y derivando parcialmente L en (4) con respecto a cada uno de los parámetros de interés se llega a la expresión (5), el vector de parámetros estimados $\hat{\beta}$. El proceso de estimación se encuentra planteado en Montgomery *et al*¹⁰.

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (9)$$

Finalmente, el modelo de regresión ajustado para las variables regresoras corresponde a la expresión (6) y de manera matricial es equivalente al modelo (7).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p \quad (10)$$

$$\hat{Y} = X\hat{B} \quad (11)$$

con $\hat{B} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$ el vector de coeficientes de regresión estimados.

1.2.4. Hipótesis y supuestos de la regresión

El análisis de la significancia de la ecuación estimada (6) se realiza a partir del conjunto de hipótesis dado en (8), con el fin de determinar si existe una relación lineal entre la variable respuesta y y cualquiera de las p variables regresoras.

¹⁰ Montgomery, D. C., Peck, E. A., & Geoffrey, V. G. *Introducción al análisis de regresión lineal* (3 ed.). Compañía editorial continental. México D.F. 2006.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_a: \beta_j \neq 0 \text{ al menos para un } j \quad (12)$$

El rechazo de H_0 implica que al menos una de las variables $x_1, x_2, x_3, \dots, x_p$ contribuye al modelo de manera significativa, Montgomery *et al*¹¹.

En la indagación de la significancia del modelo se lleva a cabo los siguientes supuestos a cerca del mismo.

- a. “el término del error $\varepsilon = y - \hat{y}$ es una variable aleatoria cuya media o valor esperado es cero, es decir $E(\varepsilon) = 0$. Para valores dados de x_1, x_2, \dots, x_p , el valor esperado o valor promedio de y está dado por:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p \quad (13)$$

- b. La varianza de ε se denota σ^2 y es la misma para todos los valores de las variables independientes (homocedasticidad) $x_1, x_2, x_3, \dots, x_p$. Por lo que la varianza de y respecto a la línea de regresión es σ^2 y es la misma para todos los valores de $x_1, x_2, x_3, \dots, x_p$.
- c. Los valores de ε son independientes. El valor de ε para un determinado conjunto de valores de las variables independientes no está relacionado con el valor de ε de ningún otro conjunto de valores.
- d. El término del error ε es una variable aleatoria distribuida normalmente y que refleja la desviación entre el valor de y y el valor esperado de y dado por $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$. Como $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son constantes para los valores dados de $x_1, x_2, x_3, \dots, x_p$, la variable dependiente “ y ” es también una variable aleatoria distribuida normalmente.

1.2.5. Modelo Exponencial

En la regresión exponencial (RE) se analiza el crecimiento de las poblaciones bajo el supuesto de que la tasa de crecimiento de la población es proporcional a la población, la tasa de natalidad y mortalidad en el instante, lo que se expresa de la siguiente forma:

¹¹ Montgomery, D. C., Peck, E. A., & Geoffrey, V. G. *Introducción al análisis de regresión lineal* (3 ed.). Compañía editorial continental. México D.F. 2006.

$$\frac{dP}{dt} = rP \quad (14)$$

Donde P es la población y r el potencial biótico o tasa de crecimiento en el instante t . La tasa de crecimiento se calcula a partir de la tasa de natalidad y de mortalidad:

$$r = N - M \quad (15)$$

Donde N es la tasa de natalidad y M es la tasa de mortalidad.

Pero las poblaciones de los países son más complejas, requiriendo que se tenga en cuenta la tasa de migración Mi y la tasa de inmigración I modificándose así la ecuación (5) como se encuentra a continuación.

$$r = N - M + I - Mi \quad (16)$$

En el modelo Exponencial (ME) la tasa de crecimiento es constante, partiendo de este principio se desarrolla la ecuación (14) se llegando al ME.

$$\frac{dP}{P} = r dt$$

$$\int_{P_0}^{P_f} \frac{dP}{P} = r \int_{t_0}^t dt$$

Donde:

P_f es la población final, la que se desea estimar

P_0 es la población inicial del espacio muestral

t el instante del cual se quiere estimar la población. Se cuenta a partir del momento en que se da la población inicial

t_0 es el instante en que se da la población inicial. Para fines prácticos se toma como 0 (cero)

$$\int_{P_0}^{P_f} \frac{dP}{P} = r \int_0^t dt$$

$$\ln|P_f| - \ln|P_0| = r(t - 0)$$

$$\ln \left| \frac{P_f}{P_0} \right| = rt$$

$$\frac{P_f}{P_0} = e^{rt}$$

$$P_f = P_0 e^{rt} \quad (16)$$

Este modelo presenta la desventaja que se debe usar cuando el tamaño de la población es pequeño en comparación con las dimensiones del ecosistema donde se aplique, debido a que no se tiene en cuenta las limitaciones del lugar donde se encuentra la población estudiada.

1.2.6. Modelo Poblacional Logístico o modelo de Verhulst

El modelo poblacional logístico (MPL) es un ajuste del ME en el que se tiene en cuenta las limitaciones del ecosistema tales como la escasez de los recursos y el espacio disponible dando una población máxima que puede llegar a darse en el ecosistema, teniendo como consecuencia que la tasa de crecimiento de la población es cero.

$$\lim_{t \rightarrow \infty} \frac{dP}{dt} = 0; \quad 0 \leq P(t) \leq K \quad (17)$$

Queda establecido que K es la población máxima que puede darse en el ecosistema.

$$\lim_{t \rightarrow \infty} P = K \quad (18)$$

Una ecuación que satisface las condiciones de las ecuaciones (17) y (18) es una modificación del modelo de la tasa de crecimiento poblacional exponencial.

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right) \quad (19)$$

A continuación, una breve demostración de que el modelo dado en la ecuación anterior cumple con el supuesto de la regresión poblacional logística:

$$\lim_{t \rightarrow \infty} \frac{dP}{dt} = \lim_{t \rightarrow \infty} rP \left(1 - \frac{P}{K} \right)$$

$$\lim_{t \rightarrow \infty} \frac{dP}{dt} = \lim_{t \rightarrow \infty} \left(rP - \frac{rP^2}{K} \right)$$

$$\lim_{t \rightarrow \infty} \frac{dP}{dt} = r \lim_{t \rightarrow \infty} P - \frac{r}{K} \lim_{t \rightarrow \infty} P^2$$

$$\lim_{t \rightarrow \infty} \frac{dP}{dt} = rK - \frac{r}{K} K^2$$

$$\lim_{t \rightarrow \infty} \frac{dP}{dt} = rK - rK = 0$$

Para obtener el modelo Poblacional Logístico se desarrolla la ecuación (19) (ver demostración en anexo A).

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right)$$

El modelo queda planteado como se muestra en la ecuación (20)

$$P(t) = \frac{KP_0 e^{rt}}{K + P_0(e^{rt} - 1)} \quad (20)$$

Este modelo es también conocido como el modelo de Verhulst. La ecuación (20) se puede ajustar con la ecuación (21):

$$A = \frac{k - P_0}{P_0} \quad (21)$$

$$AP_0 = k - P_0$$

$$P(t) = \frac{KP_0 e^{rt}}{K + P_0 e^{rt} - P_0}$$

$$P(t) = \frac{KP_0 e^{rt}}{(K - P_0) + P_0 e^{rt}}$$

Se reemplaza $K - P_0$ por AP_0 :

$$P(t) = \frac{KP_0 e^{rt}}{AP_0 + P_0 e^{rt}}$$

$$P(t) = \frac{K e^{rt}}{A + e^{rt}}$$

Y por último se divide al numerador y denominador por e^{rt} para obtener la ecuación (22)

$$P(t) = \frac{K}{1 + Ae^{-rt}} \quad (22)$$

Para estimar el potencial de población o k se debe utilizar la ecuación (23)

$$k = \frac{2P_0P_1P_2 - P_1^2(P_0 + P_2)}{P_0P_2 - P_1^2} \quad (23)$$

La tasa de crecimiento se calcula por medio de la ecuación (24)

$$r = \frac{1}{n} \ln \frac{P_0(k - P_1)}{P_1(k - P_0)}, \text{ con } n = t_2 - t_1 = t_1 - t_0 \quad (24)$$

1.3. ANÁLISIS DE HIPÓTESIS Y TABLA F

El análisis de hipótesis se puede usar para verificar los supuestos de las regresiones y en este caso se usará la tabla f. Para poder realizar el análisis de hipótesis por medio de la tabla f se usa la ANOVA, para ello se calcula la suma de los cuadrados debido a la regresión (SCR) y al error (SCE), la suma total de los cuadrados (SCT), el cuadrado medio de la regresión (CMR), el cuadrado medio del error (CME) y por último el valor de f. La ANOVA tiene la siguiente estructura:

Tabla 1. Tabla f, tabla que permite establecer el valor f de Fisher para el Análisis de varianza (ANOVA)

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Valor de f
Regresión	SCR	p	CMR	$\frac{CMR}{CME}$
Error	SCE	$n - 1 - p$	CME	
Total	STC	$n - 1$		

Recordar que p , el grado de libertad de la SCR, es el número de categorías observadas (variables regresoras) a cada sujeto de estudio y n la cantidad de sujetos observados o el número de observaciones realizadas.

A continuación, se muestra las ecuaciones que permiten el cálculo de las sumas de cuadrados y los cuadrados medios

$$SCR = \sum (\hat{y}_1 - \bar{y})^2 \quad (25)$$

$$SCE = \sum (y_1 - \hat{y}_1)^2 \quad (26)$$

$$STC = \sum (y_1 - \bar{y})^2 \quad (27)$$

También tenemos que la STC se puede expresar en función de SCR y SCE:

$$STC = SCR + SCE \quad (28)$$

Para los cuadrados medios:

$$CMR = \frac{SCR}{p} \quad (29)$$

$$CME = \frac{SCE}{n - 1 - p} \quad (30)$$

1.4. COEFICIENTE DE DETERMINACIÓN

El coeficiente de determinación (R^2) permite medir que tan bien se ajusta la ecuación estimada a los datos observados. Este coeficiente se calcula a partir de los datos obtenidos del ANOVA. El R^2 se calcula a partir de la siguiente ecuación

$$R^2 = \frac{SCR}{STC} \quad (31)$$

1.5. COEFICIENTE DE DETERMINACIÓN AJUSTADO

En el análisis multivariado (AM) se usa el coeficiente de determinación ajustado (R_{aj}^2) ya que al aumentar las variables regresoras o independientes el valor de R^2 también aumenta independientemente de la contribución de cada variable. Para

evidenciar la necesidad del uso del R^2_{aj} en vez del R^2 se procede a analizar la ecuación de este último, ecuación (31):

$$R^2 = \frac{SCR}{STC}$$

Si se desglosa esta ecuación obtenemos que el R^2 también se puede expresar de la siguiente forma:

$$R^2 = 1 - \frac{SCE}{SCR + SCE}$$

al fijarse en el denominador del segundo término se puede observar que al aumentar las variables regresoras tanto la SCR como SCE aumentan, por lo que este denominador aumentará más rápido que el numerador del mismo término debido a que este solo se incrementa por el aumento de la SCE , provocando que la fracción disminuya su valor al incrementar las variables, acercando así a cero independientemente del aporte que haga cada una de ellas.

El R^2_{aj} se calcula de la siguiente forma:

$$R^2_{aj} = 1 - (1 - R^2) \frac{(n - 1)}{(n - 1 - k)} \quad (32)$$

Reemplazado R^2 y simplificando la ecuación (26) queda:

$$R^2_{aj} = 1 - \frac{SCE(n - 1)}{STC(n - 1 - k)} \quad (33)$$

Esta ecuación tiene una estructura similar a la de la ecuación (29) solo que al aumentar las variables regresoras (k), disminuye el valor del denominador, lo que puede provocar que el término aumente y R^2_{aj} disminuya en algunos casos en donde el aporte de las nuevas variables no sea significativo, dicho de otra forma, el error generado por cada nueva variable sea muy grande, haciendo que el numerador crezca más que el denominador de este término.

1.6. ANÁLISIS DE RESIDUALES

El estadístico F y R^2_{aj} no pueden asegurar que se cumplan los supuestos de la regresión por lo que se recurre a los residuales del modelo y a su análisis para detectar incumplimiento de las premisas de la RM. Los residuales se definen como sigue:

$$e_i = y_i - \hat{y}_i \quad \text{con } i = 1, 2, \dots, n \quad (34)$$

Donde

y_i es la i -ésima observación

\hat{y}_i es el valor calculado para la misma observación o y_i estimado y

n el número de observaciones realizadas, como se vio anteriormente.

Se puede ver a los residuales como el error de los valores estimados con respecto a los observados, entonces por medio de los residuales se puede detectar el incumplimiento de los supuestos de las regresiones. Entre las propiedades de los residuales se tiene que la media de estos es cero y varianza promedio aproximada se calcula como:

$$\frac{\sum_{i=1}^n (e_i - \bar{e}_i)^2}{n-p} = \frac{\sum_{i=1}^n e_i^2}{n-p} = \frac{SC_{Res}}{n-p} = S^2 M_{Res} \quad (35)$$

Como se puede observar SC_{Res} es la suma de los cuadrados debido a los residuales y $S^2 M_{Res}$ es la varianza media de los residuales. Este valor tendrá poca significancia si el número de observaciones (n) es muy pequeño con respecto al número de variables (p).

1.6.1. Residuales escalados

El escalamiento de residuales es una forma de cambiar la escala de cada residual para que coincida con una distribución normal de media cero y desviación o varianza aproximada a uno.

Ya se ha dicho que los residuales permiten detectar el incumplimiento los supuestos de la regresión, ahora se verá cómo se logra esto. Lo primero que se busca son los residuales o errores que sean muy grandes con respecto a los demás (observaciones atípicas o valores extremos), para ello se recurre a los residuales estandarizados, residuales estudentizados, residuales PRESS y el R de Student.

1.6.2. Residuales estandarizados

La importancia que tiene los residuales estandarizados es que estos tienen media cero y varianza aproximadamente igual a 1, por lo que valores mayores que 3 y menores que -3, por ejemplo, se podrían considerar una observación atípica. La forma de calcular los residuales estandarizados es hallando la relación entre cada

uno de los residuales y su desviación estándar. Esta última también se suele calcular hallando la raíz cuadrada de la varianza.

Como es de esperar, en este caso se utiliza la S^2M_{Res} para calcular la desviación de los residuales.

$$d_i = \frac{e_i}{\sqrt{S^2M_{Res}}}, \quad i = 1, 2, \dots, n \quad (36)$$

1.6.3. Residuales estudentizados

Con los residuales estandarizados se logra solo una aproximación al escalamiento ya que se usa el S^2M_{Res} . La idea de los residuales estudentizados es utilizar varianza de cada residual para así lograr un mejor escalamiento. Montgomery *et al*¹² establecen que el vector de residuales se puede escribir de la siguiente forma:

$$e = (I - H)y \quad (37)$$

Donde

$$H = X(X'X)^{-1}X' \quad (38)$$

También cabe aclarar que la matriz de la variable dependiente y se puede escribir de la siguiente forma:

$$\begin{aligned} y &= \hat{y} + \varepsilon \\ y &= X\hat{B} + \varepsilon \end{aligned} \quad (26)$$

Uniendo las ecuaciones (24), (25) y (26) se llega a la siguiente ecuación:

$$e = (I - H)\varepsilon \quad (27)$$

Y la matriz de covarianza es:

$$Var(e) = \sigma^2(I - H)$$

Las particularidades de la matriz H es que es simétrica, es decir que $H = H'$ y que es idempotente, lo que quiere decir que $HH = H$. La matriz $I - H$ también hereda

¹² MONTGOMERY, Douglas, PECK, Elizabeth, y VINING, Geoffrey. *Introducción al análisis de regresión lineal* (3 ed.). Compañía editorial continental. México D.F. 2006.

estas características, además esta matriz no es diagonal por lo que las varianzas de los residuales están correlacionadas. Entonces la varianza entre el residual e_i es:

$$Var(e_i) = \sigma^2(I - h_{ii}) \quad (39)$$

La covarianza entre los residuales e_i y e_j es:

$$Cov(e_i, e_j) = -\sigma^2 h_{ij} \quad (40)$$

Ahora bien, volviendo al residual estandarizado se tiene que este es la relación entre el residual i -ésimo y la i -ésima desviación estándar, la cual se calcula por medio de la raíz cuadrada de la varianza del i -ésimo residual:

$$r_i = \frac{e_i}{\sqrt{\sigma^2(I - h_{ii})}}$$

$$r_i = \frac{e_i}{\sqrt{S^2 M_{Res}(I - h_{ii})}} \quad (41)$$

1.6.4. Residuales PRESS

Si la i -ésima observación es atípica, entonces el modelo de predicción será muy influenciado por esta observación. La idea del residual PRESS es analizar los residuales obtenidos a partir de la regresión de todos los datos a excepción de la observación i -ésima. Si dicha observación es atípica, el residual normal será pequeño en comparación con el nuevo residual. A estos nuevos residuales se les llama residuales PRESS porque se suelen utilizar para calcular la suma de cuadrados de error de predicción (por sus siglas en ingles)

El valor estimado para la i -ésima observación o y_i se denotará como $y_{(i)}$ y el valor para el nuevo residual será:

$$e_{(i)} = y_i - y_{(i)} \quad (31)$$

Una vez más utilizando la matriz H , ecuación (38) cada residual PRESS se puede calcular a partir de la ecuación (42)

$$e_{(i)} = \frac{e_i}{1 - y_{ii}}, \quad i = 1, 2, \dots, n \quad (42)$$

Un punto donde hay una gran diferencia entre el residual normal y el PRESS indicara que la observación correspondiente es necesaria para que el modelo y sin esta n se puede hacer buenas predicciones.

La varianza de los residuales PRESS y los residuales PRESS estandarizados son respectivamente:

$$Var(e_{(i)}) = \frac{\sigma^2}{1 - h_{ii}} \quad (43)$$

$$r_{PRESS} = \frac{e}{\sqrt{\sigma^2(1 - h_i)}} \quad (44)$$

1.6.5. Gráfica de residuales

Se recomienda utilizar los residuales estudentizados para investigar la adecuación del ajuste del modelo de la regresión porque estos tienen varianzas constantes, pero también se puede o se recomienda graficar tanto los residuales originales como otros residuales escalados.

Como el modelo depende de las suposiciones de normalidad, la falta de una distribución normal genera que los intervalos de confianza y la capacidad predictora del modelo disminuya, lo que indica que se debe optar por otro modelo que no sea el de mínimos cuadrados. Una manera de comprobar los supuestos de normalidad es realizando la gráfica de probabilidad normal de los residuales.

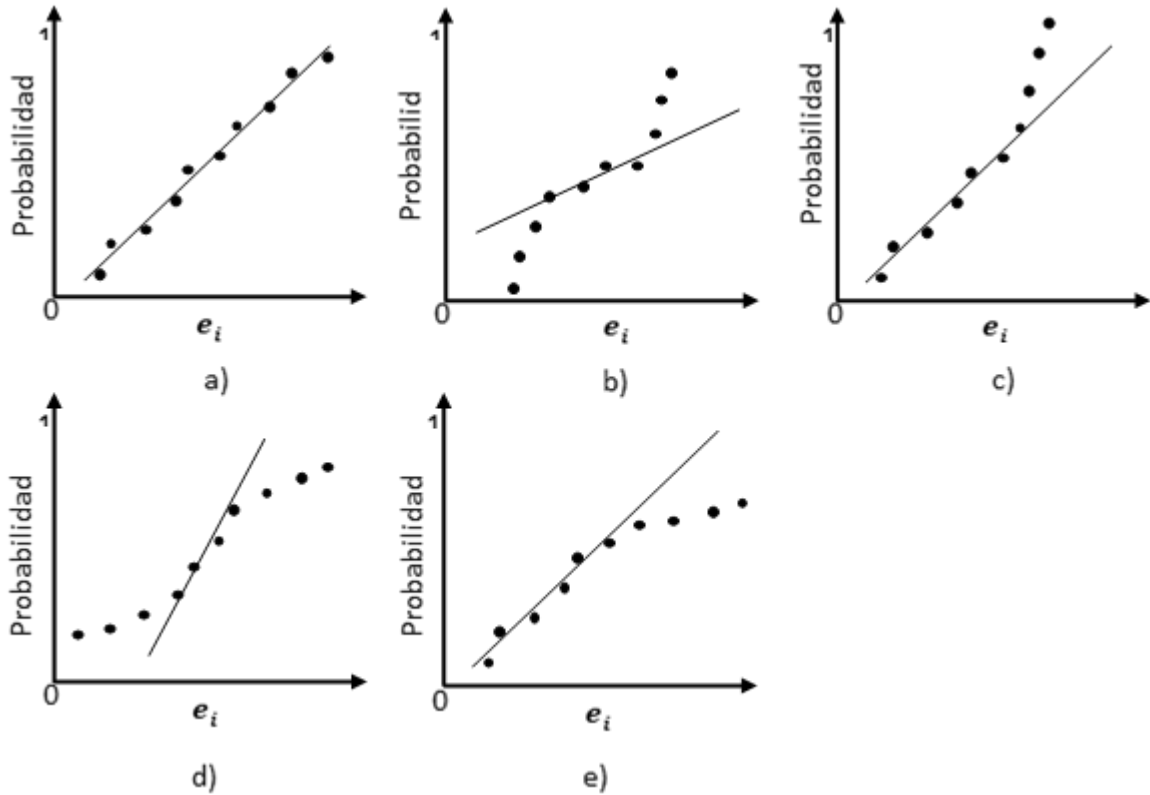
Al realizar la gráfica de probabilidad se puede ubicar los residuales ordenados en orden creciente en el eje horizontal o vertical. En este caso se analiza las posibles gráficas que se obtiene al ubicar los residuales en el eje horizontal.

En la gráfica de probabilidad normal de los residuales se puede utilizar el valor normal esperado $\Phi^{-1} \left[\left(i - \frac{1}{2} \right) / n \right]$, donde Φ representa la distribución acumulada normal estándar.

1.6.6. Gráfica de probabilidad normal

A continuación, se encuentran las estructuras que se puede obtener al realizar la gráfica de probabilidad normal de los residuales, teniendo en cuenta que se les da prioridad a los valores centrales de los resídales.

Gráfica 1. de probabilidad normal: a) gráfica ideal, b) colas gruesas, c) distribución con colas delgadas d) asimetría positiva y e) asimetría negativa.



Cabe aclarar que la recta que pasa por los valores centrales se realiza de forma visual y aproximada, por lo que este método solo sirve cuando las diferencias entre los unos extremos y la recta son muy notables, por lo que se puede dar el caso de que al ser estas diferencias muy pequeñas se opta por determinar que la regresión cumple con los supuestos de normalidad cuando en realidad no es así.

En la figura 1a se puede ver la forma aproximada que debe tener la gráfica de probabilidad normal de los residuales, en ella todos los puntos se encuentran cerca de la recta que pasa aproximadamente por los valores centrales.

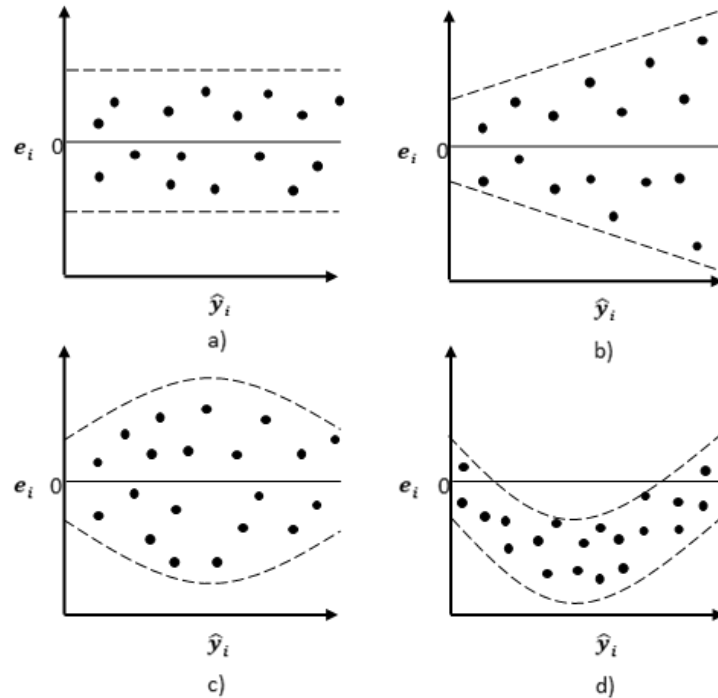
En una situación que corresponda a la figura 1b se puede deducir que la regresión no cumple con los supuestos de normalidad. La figura 1c representa una distribución con los extremos más delgados que los de una distribución normal

La figura 1d muestra patrones asociados con asimetría positiva y la 1e con asimetría negativa.

1.6.7. Grafica de residuales en función de valores ajustados (\hat{y}_i)

Ahora nos centramos en la adecuación del modelo, para ello se grafica los residuales e_i o de los residuales escalados en función de los valores estimados para y o y estimados (\hat{y}_i) con el fin de evidenciar alguna inadecuación del modelo.

Gráfica 2. Modelos de graficas de residuales en función del i -ésimo "y" estimado (\hat{y}_i): a) situación ideal, b) en embudo, c) doble arco y d) no lineal



En la figura 2a, en donde la gráfica se puede encerrar en una banda horizontal, indica que no hay defectos visibles del modelo. Esta sería la situación a esperar al realizar la gráfica de los e_i , o de los residuales escalados d_i , r_i y r_{PRESS} en función de \hat{y}_i . La figura 2b indica que la varianza es creciente con respecto al \hat{y}_i , también puede darse el caso contrario, el cual sería una varianza decreciente. La figura 2c muestra una gráfica que se obtiene con frecuencia cuando ρ es una proporción entre 0 y 1. La gráfica 2b y 2c indica que se debe hacer un ajuste al modelo por medio de transformación a la variable regresora o a la de respuesta o usar el método de mínimos cuadrados ponderados.

La figura 2d indica que no hay linealidad, esto puede darse por falta de involucrar alguna o múltiples variables en la elaboración del modelo. También se puede dar solución utilizando el cuadrado de una de las variables regresoras o de las variables respuesta.

2. METODOLOGÍA

2.1. ENFOQUE METODOLÓGICO

El trabajo se sustentó bajo un enfoque cuantitativo, pues buscó establecer el tipo de correlación entre distintas variables, tales como, la población total, el número de hogares que tienen acceso a energía eléctrica y el tiempo en el que tienen acceso a esa energía con la variable consumo de energía eléctrica en el periodo 2005-2018.

2.2. ESTRATEGIA DE INVESTIGACIÓN

Se usó el estudio de caso como estrategia de investigación. Se analizó la información existente sobre la población y el consumo de energía eléctrica del departamento de Vichada en las bases de datos del DANE y del Sistema único de información de servicios públicos (SUI), junto con el consumo de energía eléctrica del 2018 del Departamento de Vichada establecido por PERS Orinoquía.

El departamento de Vichada está situado en el extremo oriental de Colombia, localizado entre los 06°19'34" y 02°53'58" de latitud norte y 67°25'1" y 71°07'10" de longitud oeste. Cuenta con una superficie de 98.970 km² lo que representa el 8.6 % del territorio nacional. Limita por el Norte con el río Meta que lo separa de los departamentos de Casanare, Arauca y la República de Venezuela; por el Este con el río Orinoco que lo separa de la República de Venezuela, por el Sur con el río Guaviare que lo separa de los departamentos de Guainía y Guaviare y por el Oeste con los departamentos de Meta y Casanare. Este departamento se encuentra conformado por los municipios de Cumaribo al sur del departamento, La primavera al norte, Puerto Carreño al noreste y Santa Rosalía al norte.

El alcance de este estudio fue correlacional ya que se tuvo como hipótesis las existencias de correlación entre cada una de las siguientes variables a saber, la población total, el número de hogares que tienen acceso a energía eléctrica y el tiempo con el que cuentan con energía eléctrica estos hogares con la variable consumo de energía eléctrica, pero se buscó establecer qué tipo de correlación existe entre estas variable ya establecida con el consumo de energía eléctrica, si esta responde a un modelo lineal, exponencial, logístico, entre otros o una combinación de algunos de ellos.

Como herramientas estadísticas se utilizó los modelos de regresión, como el lineal simple, el múltiple, el exponencial y el poblacional logístico, el ANOVA, el cálculo del coeficiente de determinación y determinación ajustado y el análisis de residuales.

Figura 1. Captura de pantalla del mapa, departamento del vichada / Mapa de Ciudades, Municipios y Parques / 2017 © www.colombiamania.com, en el que se muestra la ubicación del departamento de Vichada dentro de la Republica de Colombia.



Fuente: Mapa departamento del vichada / Mapa de Ciudades, Municipios y Parques / 2017 © www.colombiamania.com

Como herramienta tecnológica se usó el software de uso libre R Project. Se creó un script (ver anexos) en un archivo “.R” para cada modelo de regresión y sus respectivos ANOVA, cálculo del coeficiente de determinación o determinación ajustado, según fue el caso y el análisis de residuales.

2.3. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN.

Para establecer qué tipo de correlación existente entre la población total, se recopiló información de documentación de mediciones del DANE, SUI y PERS Orinoquia de la siguientes forma, la información correspondiente al número de habitantes, la cantidad de hogares o personas, en promedio, por vivienda en la base de datos del DANE, el número de viviendas que tienen acceso a energía eléctrica, el tiempo en el que tienen acceso a esa energía y el consumo de energía eléctrica en las bases de datos del SUI, junto con el consumo de energía eléctrica del 2018 del Departamento de Vichada establecido por PERS Orinoquia.

2.4. ANÁLISIS Y SISTEMATIZACIÓN

Luego de tener la información recopilada, se creó y aplicó un script de R Project, por cada modelo a los datos de la población de los censos 1985, 1993, 2005 y 2018 y con estos modelos ajustados se estimó la población correspondiente a cada uno de los cuatro censos ya nombrados. Estos valores estimados se usaron realizar el análisis de varianza (ANOVA), el cálculo del coeficiente de determinación o

determinación ajustado y el análisis de residuales. A partir de los resultados de estos análisis se determinó el modelo que mejor se ajusta a los datos de los censos del DANE. Con el modelo seleccionado se estimó la población del departamento en el periodo 2003-2016.

Teniendo la información completa de cada una de las variables, se creó un script de R Project por cada modelo. Con los modelos de regresión se estimó los consumos de energía eléctrica en el periodo del 2003 al 2018 en el departamento. Estos valores estimados se usaron para los análisis de varianza (ANOVA), del coeficiente de determinación o determinación ajustado y de residuales. Estos análisis se utilizaron para determinar el modelo que mejor se ajusta a los datos del SUI.

Luego se determinó el modelo más adecuado y se procedió a estimar los consumos de energía eléctrica para tres escenarios futuros (pesimista, optimista y transicional), para el periodo 2019-2039 del departamento del Vichada. El escenario pesimista corresponde al consumo que tendría el departamento si se mantiene estable el porcentaje de la población que tiene acceso al servicio de energía eléctrica. El escenario optimista corresponde al consumo que tendría el departamento si el 100 % de la población llega a tener acceso al servicio de energía en los primeros años del periodo 2019-2039. El escenario transicional corresponde al consumo si el porcentaje de la población que tiene acceso al servicio de energía eléctrica va incrementando gradualmente hasta llegar al 100 % para el año 2039.

3. RESULTADOS Y ANÁLISIS DE RESULTADOS

Se analizó la información existente sobre la población y el consumo de energía eléctrica del departamento de Vichada en las bases de datos del SUI, el cual reúne la información brindada por las electrificadoras y el DANE, junto con el consumo de energía eléctrica del 2018 del Departamento de Vichada establecido por PERS Orinoquía.

Se tomaron los censos de 1985, 1993, 2005 y 2018 con los cuales se establecieron las viviendas, la población y las viviendas que contaban con el servicio de energía eléctrica en cada año. Estos censos no presentan información de cantidad de locales comerciales ni del sector oficial por lo que solo se pudo obtener información del sector residencial del departamento de Vichada. A continuación, se encuentra la información recolectada de los cuatro censos.

Tabla 2. Censos del departamento del Vichada, fuente DANE

Año del censo	Población censo	Viviendas	Viviendas con energía eléctrica
1985	13770	2284	934
1993	36336	7087	1196
2005	44592	8959	4730
2018	77276	ND	ND

El DANE no ha publicado un informe completo y detallado del censo de 2018, por lo que solo se contará con 3 datos para la estimación de la población en el departamento.

Del SUI se obtuvo informes mes a mes desde el año 2003 hasta el año 2017, en los cuales se especifica el consumo de energía eléctrica en kWh al mes, las viviendas, los establecimientos oficiales y comerciales que tienen acceso al servicio de energía eléctrica. Esta información se compiló en la siguiente tabla.

Tabla 3. Consumo de energía por vivienda, fuente SUI

Año	Usuarios / Consumidores			Consumo kWh/mes
	Viviendas	Oficiales	Comerciales	
2003	4053	90	828	824597.85
2004	4421	42	677	632950.88
2005	4470	96	721	874600.01
2006	4291	114	674	936406.65
2007	4385	112	612	985738.07
2008	4856	139	681	1018047.45

Año	Usuarios / Consumidores			Consumo kWh/mes
	Viviendas	Oficiales	Comerciales	
2009	4865	152	648	1071410.08
2010	4773	144	606	1083677.25
2011	5404	149	698	1246892.33
2012	5316	138	673	1323990.54
2013	6068	145	689	1447545.61
2014	6442	165	781	1724770.73
2015	7034	192	825	1905781.42
2016	7399	204	830	1972123.79
2017	7653	197	826	1310878.77

Finalmente se recopiló la información de los resultados obtenidos por estadística de PERS Orinoquia

Tabla 4. Consumo de energía eléctrica del Vichada año 2018, fuente PERS Orinoquia

Año	Viviendas	Usuarios/Consumidores			Consumo kWh/mes	
		total	viviendas	Oficiales		Comerciales
2018	11356	8985	7893	85	1007	2648791.58

En el siguiente cuadro se tiene el resumen de los datos recolectados del SUI, DANE y de PERS Orinoquia.

Tabla 5. Resumen del consumo de energía en el Vichada, fuente DANE, PERS Orinoquia y SUI

Año	Población censo	Viviendas	Usuarios/Consumidores			Consumo de energía eléctrica	
			total	viviendas	Oficiales		Comerciales
1985	13770	2284	934	934	0	0	ND
1993	36336	7087	1196	1196	0	0	ND
2003	ND	ND	4971	4053	90	828	824597.85
2004	ND	ND	5140	4421	42	677	632950.88
2005	44592	8959	5547	4730	96	721	874600.01
2006	ND	ND	5079	4291	114	674	936406.65
2007	ND	ND	5109	4385	112	612	985738.07
2008	ND	ND	5676	4856	139	681	1018047.45
2009	ND	ND	5665	4865	152	648	1071410.08
2010	ND	ND	5523	4773	144	606	1083677.25
2011	ND	ND	6251	5404	149	698	1246892.33
2012	ND	ND	6127	5316	138	673	1323990.54
2013	ND	ND	6902	6068	145	689	1447545.61

Año	Población censo	Viviendas	Usuarios/Consumidores				Consumo de energía eléctrica
			total	viviendas	Oficiales	Comerciales	
2014	ND	ND	7388	6442	165	781	1724770.79
2015	ND	ND	8051	7034	192	825	1905781.42
2016	ND	ND	8433	7399	204	830	1972123.79
2017	ND	ND	8676	7653	197	826	1310878.77
2018	77276	11356	8985	7893	85	1007	2648791.58

Fuente: elaboración propia con información DANE, SUI y PERS Orinoquia

Se buscó establecer qué tipo de correlación existe entre estas variables ya establecidas con el consumo de energía eléctrica, si esta responde a un modelo lineal, exponencial, logístico, entre otros o una combinación de algunos de ellos.

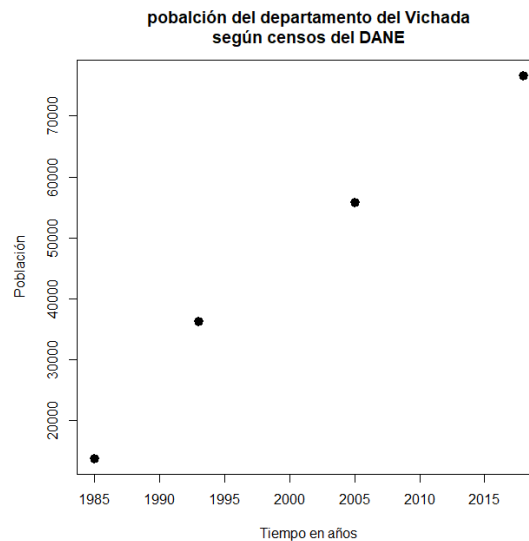
Como herramienta tecnológica se usó el software de uso libre R Project. Se creó un script en un archivo “.R” para cada modelo de regresión y sus respectivos ANOVA, cálculo del coeficiente de determinación o determinación ajustado, según sea el caso y el análisis de residuales. Inicialmente se aplicó el modelo lineal, exponencial, parabólico y poblacional logístico a las variables año y población para tener una herramienta estadística adecuada para estimar la población del 2003 al 2016 para complementar los datos recolectados y seguir con el análisis correlacional.

3.1. MODELOS PARA ESTIMAR LA POBLACIÓN DEL DEPARTAMENTO DEL VICHADA

Se analizaron cuatro modelos que se suelen ser usados para la estimación de poblaciones a través del tiempo, el modelo lineal, el exponencial, el parabólico y el poblacional logístico con el fin de contar con una herramienta que permita estimar la población en los periodos 2003-2016 y 2019-2039.

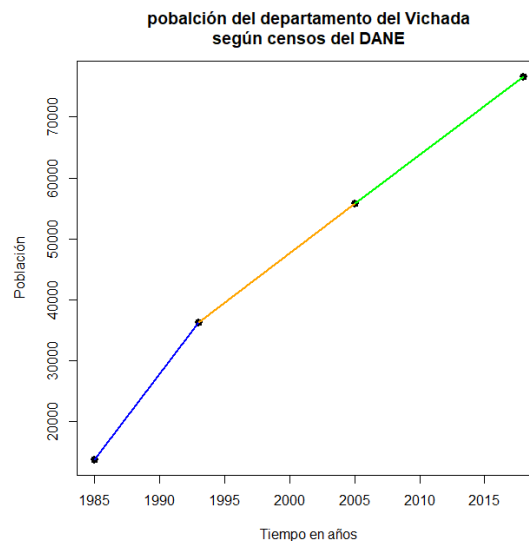
En un primer acercamiento a los datos se procedió a analizar la gráfica de dispersión, gráfica 3, para establecer el tipo de correlación que hay entre el año o el tiempo y la población del departamento del Vichada.

Gráfica 3 Dispersión de la población del departamento del Vichada, según censo del DANE, en función del tiempo en años



En la gráfica 4 de dispersión se puede observar que la tasa de crecimiento poblacional es variable, gráfica 4, el cual puede ser modelado por medio de una regresión logística poblacional, exponencial y parabólica, aunque una regresión lineal también puede ser útil.

Gráfica 4. Crecimiento de la población en los periodos 1985-1993, 1993-2005 y 2005-2018



A continuación, se muestra la aplicación de los scripts de cada modelo poblacional en R Project a los datos de la población del departamento del Vichada.

3.1.1. Modelo lineal simple

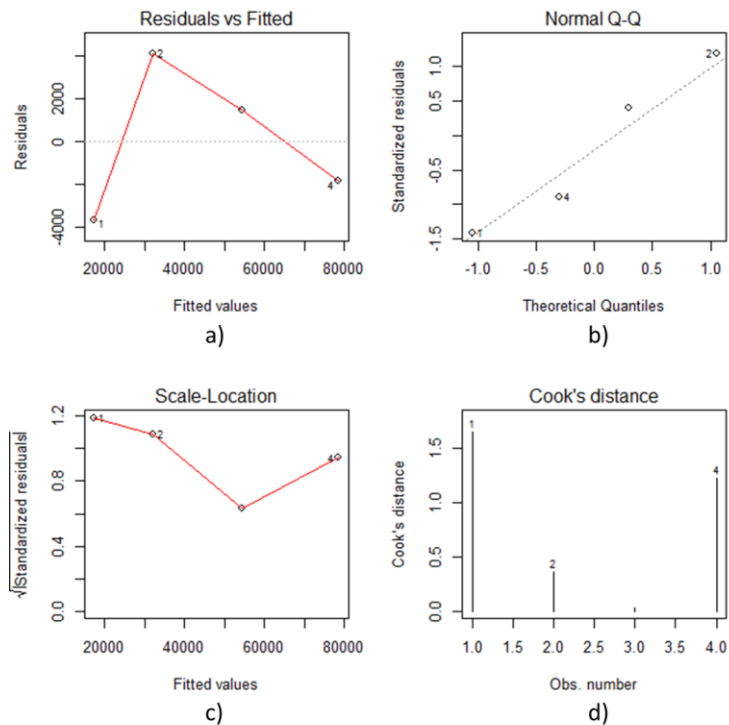
Tabla 6. ANOVA de la regresión lineal simple de la población del departamento del Vichada según censos del DANE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Pobl2\$ t	1	2132366543	2132366543	119.42	0.00827
Residuals	2	35711701	17855850		

En el análisis de Varianza o ANOVA se llegó a un P valor de 0.00827 o 0.8 % lo que indica que se rechaza la hipótesis nula en el que $\hat{\beta}_0 = \hat{\beta}_1 = 0$ con un nivel de significancia del 95 %, por lo que se acepta la hipótesis alternativa, en el cual algún $\hat{\beta}_i \neq 0$. La hipótesis nula y alternativa fueron establecidas en el marco referencial para la regresión lineal. Se concluye que este modelo se ajusta a los datos de forma significativa.

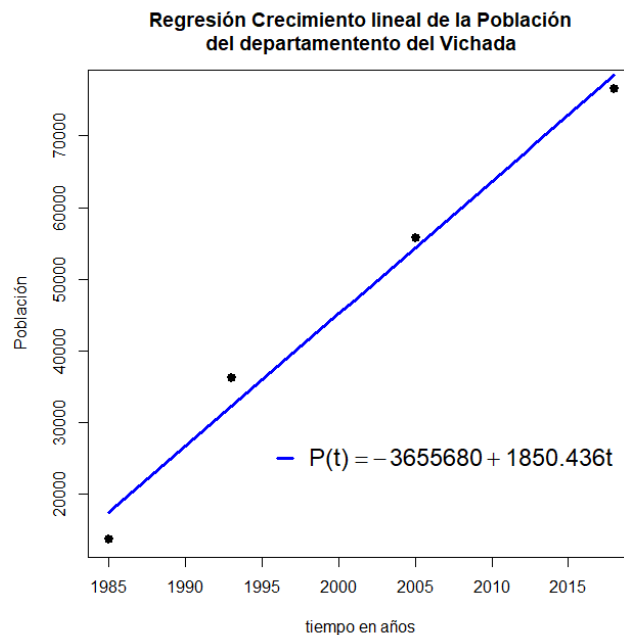
A continuación, se analiza las gráficas de los residuales para determinar si se cumplen los supuestos de la regresión.

Gráfica 5. Residuales de la regresión lineal simple para la población del Departamento del Vichada según DANE.



En la gráfica 5a se observa que los datos están distribuidos de forma aleatoria cumpliendo con el primer supuesto de la regresión. En la gráfica 5b se observa que los residuales cumple con el supuesto de normalidad debido a que los puntos se encuentran alrededor de la recta. En la gráfica 5c se observa que los residuales tienen la misma distribución, aunque el residual 3 tiene un valor de varianza menor al de los otros residuales por lo que se podría tener heterocedasticidad. En la gráfica 5d el residual 1 puede ser considerado atípico, aunque por ser pocos datos no se puede eliminar para la estimación del modelo. Se concluye que el modelo puede ser estimado por medio de todos los datos ya que se cumple con los supuestos de la regresión, aunque la gráfica 5c indica que se debería usar un ajuste u otro modelo de regresión, por ello se debe utilizar otros análisis como el residual PRESS antes de ajustar el modelo.

Gráfica 6. Regresión lineal simple de la población del departamento del Vichada a partir de Censos del DANE



En la gráfica 6 se observa que la recta de la regresión se ajusta a los datos de los censos del DANE, esto se comprueba con el coeficiente de determinación (R^2). Este coeficiente fue de 0.9835 lo que indica que 98,3 % de los datos estimados se ajustan a los datos reales, por lo que este modelo es un buen candidato para ser usado en la estimación de la población del periodo 2019-2039 del departamento del Vichada. Sin embargo, para obtener mayor certeza de si este modelo es el adecuado para las estimaciones de la población en el periodo requerido es necesario el residual PRESS, teniendo en cuenta que este debe ser analizado en conjunto con los residuales de los otros modelos.

3.1.2. Modelo exponencial

Tabla 7. ANOVA de la regresión exponencial de la población del departamento del Vichada según censos del DANE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Pobl2\$ <i>t</i>	1	1.45374	1.45374	13.26	0.06783
Residuals	2	0.21927	0.10964		

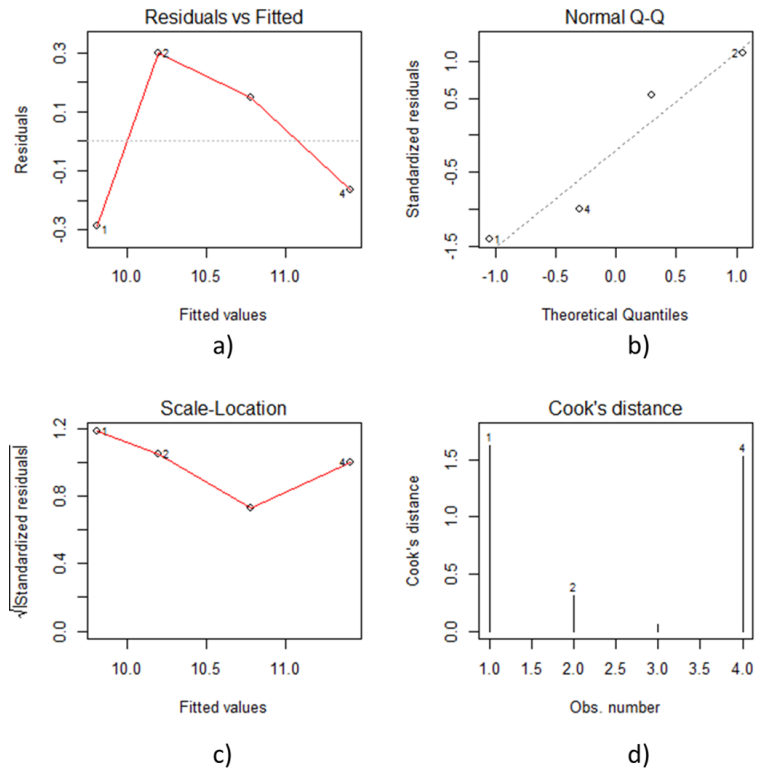
En el análisis de Variación o ANOVA se llegó a un P valor de 0.067 o 6.7 % lo que indica que se acepta la hipótesis nula en el que $\hat{\beta}_0 = \hat{\beta}_1 = 0$ con un nivel de significancia del 95 %, por lo que se rechaza la hipótesis alternativa, en cual algún $\hat{\beta}_i \neq 0$. Se concluye que este modelo no se ajusta a los datos de forma significativa a los datos de la población del Vichada.

El análisis de hipótesis indicó que el modelo exponencial no se ajusta de forma significativa a los datos de la población pero el coeficiente de determinación (R^2) fue de 0.8895 lo que indica que 88,9 % de los datos estimados se ajustan a los datos reales, siendo esta una correlación muy alta por lo que el modelo no puede ser descartado del todo aun. Se recurre, entonces al análisis de residuales.

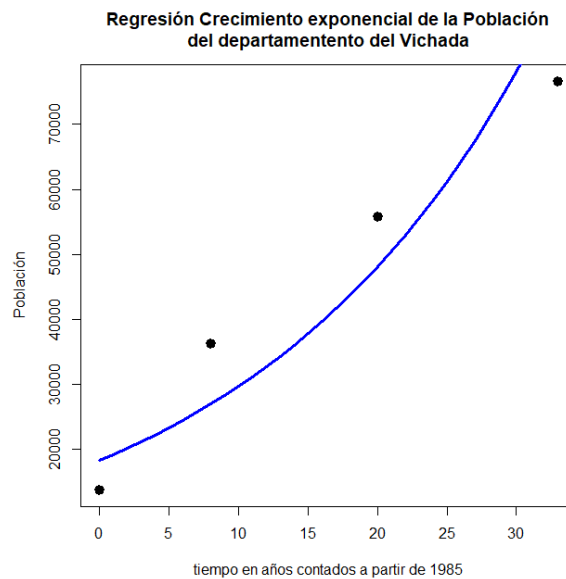
A continuación, se analiza las gráficas de los residuales para determinar si se cumplen los supuestos de la regresión.

En la gráfica 7a los datos están distribuidos de forma aleatoria. En la gráfica 7b se observa que los residuales cumple con el supuesto de normalidad porque los puntos se encuentran alrededor de la recta, aunque el residual 3 se encuentra alejado de ella. En la gráfica 7c se observa que los residuales tienen la misma distribución, aunque el residual 3 tiene un valor de varianza menor al de los otros residuales porque puede darse heterocedasticidad, aunque la diferencia de este residual escalado es menor que en el caso de la regresión lineal. En la gráfica 7d el residual 1 puede ser considerado atípico, pero al igual que en el caso de la regresión lineal, por ser pocos datos, no se puede eliminar para la estimación del modelo. Se concluye que el modelo puede ser estimado por medio de todos los datos debido a que se cumple con los supuestos de la regresión.

Gráfica 7. Residuales de la regresión exponencial para la población del Departamento del Vichada según DANE.



Gráfica 8. Regresión exponencial de la población del departamento del Vichada a partir de Censos del DANE



En la gráfica 6 se observa que la recta de la regresión se ajusta a los datos de los censos del DANE, lo cual es corroborado por el coeficiente de determinación (R^2) que indica que 88,9 % de los datos estimados se ajustan a los datos reales por lo que la regresión lineal es un buen candidato para la estimación del consumo de energía eléctrica del departamento del Vichada para el periodo 2019-2039.

3.1.3. modelo parabólico

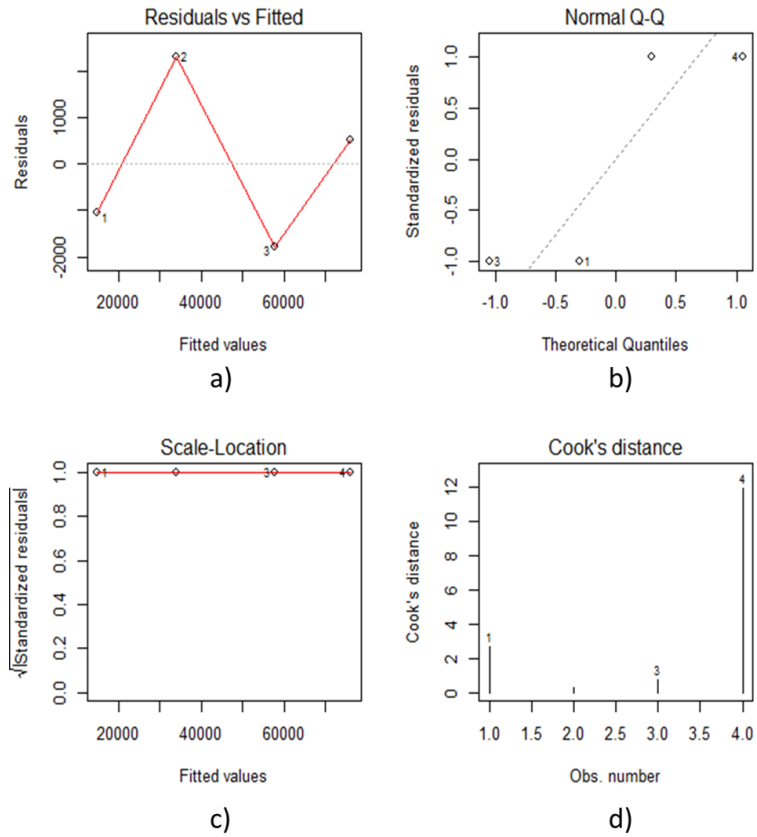
Tabla 8. ANOVA de la regresión parabólica de la población del departamento del Vichada según censos del DANE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Pobl2\$t	1	2132366543	2132366543	218.305	0.04302
Pobl2\$t2	1	25943869	25943869	2.6561	0.35037
Residuals	1	9767831	9767831		

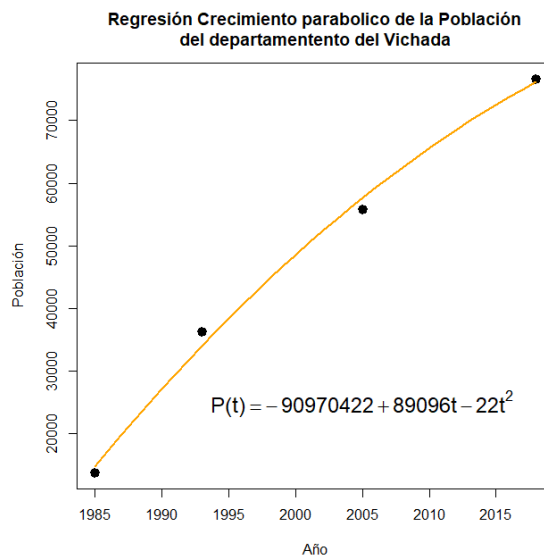
En la tabla 8 se observa que se rechaza la hipótesis nula para el β_1 o coeficiente de la variable tiempo, con un nivel de significancia de 95 %, visto que el valor de aceptación de la hipótesis nula para el estimado de este coeficiente fue de 0.0430 o 4,3 %. Por otro lado, no se puede descartar la hipótesis nula para el β_2 porque su p-valor es de 0.350 o 35,0 % por lo que se acepta la hipótesis de que $\beta_2 = 0$. Este resultado indica que el modelo parabólico no es el indicado para describir el comportamiento de la población en función del tiempo.

En la gráfica de 9a se observa que el modelo parabólico, al ser linealizado, cumple con el supuesto de que la media de los residuales es 0 o de que los residuales siguen una distribución aleatoria. La gráfica 9b se observa que los residuales no cumple con el supuesto de normalidad porque los residuales estandarizados se encuentran lejos de la gráfica de normalidad. La gráfica 9c indica que el supuesto de homocedasticidad se cumple para el modelo parabólico superando al modelo lineal y exponencial en este aspecto. Finalmente, en la gráfica 9d se establece que los residual 1 y 4 son atípicos y deben ser eliminados lo cual no es posible que se reduciría la cantidad de datos a la mitad.

Gráfica 9. Residuales de la regresión parabólica para la población del Departamento del Vichada según DANE.



Gráfica 10. Regresión parabólica de la población del departamento del Vichada a partir de Censos del DANE



En la gráfica 10 se observa que la curva de regresión se ajusta a los datos, lo que se puede comprobar con el coeficiente de determinación ajustado R_{aj}^2 (el modelo consta de 2 variables regresoras), este valor fue de 0.987 lo que indica que el 98,7 % de los datos de la población del departamento del Vichada, según el DANE, son explicados por el modelo parabólico.

3.1.4. Modelo Poblacional logístico.

El modelo poblacional logístico, en su estructura cuenta con un solo término y depende exclusivamente de los valores de K (la población máxima a la que puede albergar el ecosistema) y r (la tasa de crecimiento inicial). Inicialmente se calculó los valores del potencial de la población K y la tasa de crecimiento r , con estos valores procedió a realizar el análisis de varianza, tabla 9, en el cual se indica que la hipótesis nula se rechaza con un valor de 0,013 o 1,3 % por lo que se acepta la hipótesis alternativa, y se establece que el modelo ajustado a una regresión lineal simple es significativo.

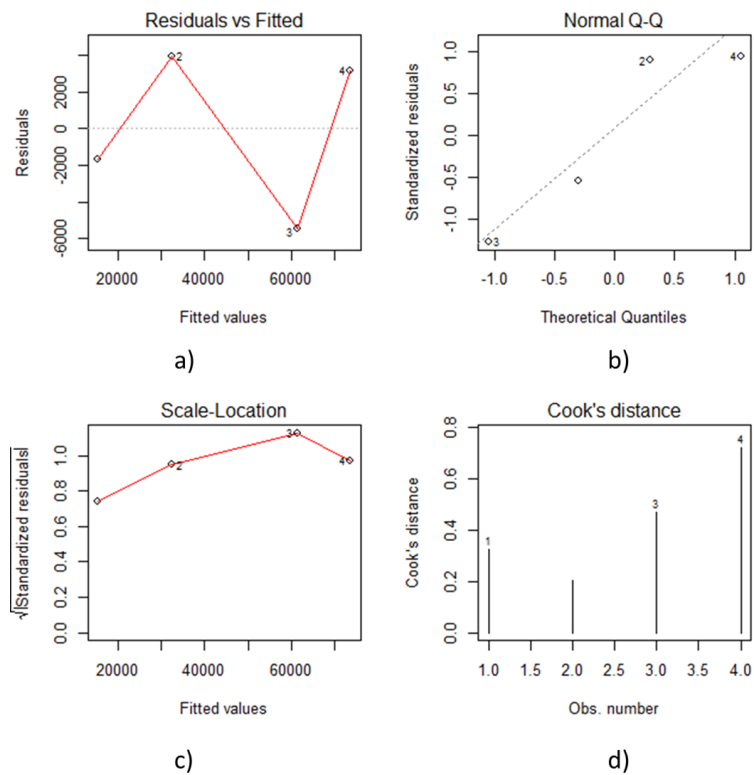
Tabla 9. ANOVA de la regresión logística poblacional de la población del departamento del Vichada según censos del DANE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Pobl2\$P.est1	1	2110190657	2110190657	72.906	0.01344
Residuals	2	57887587	28943793		

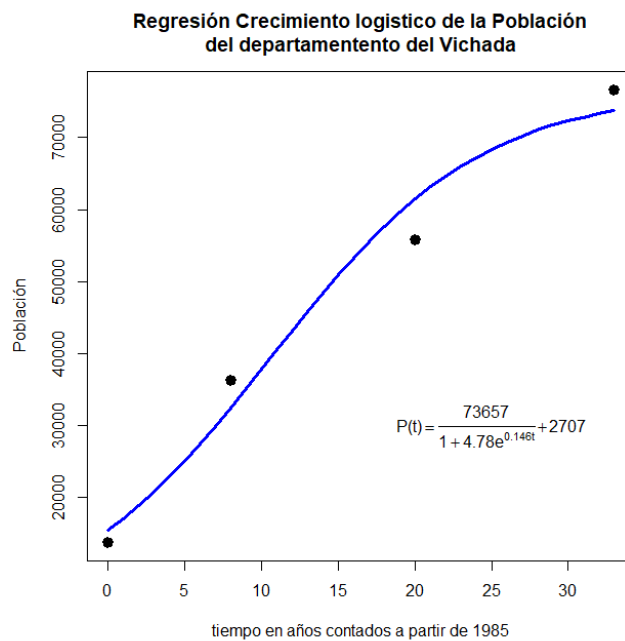
A continuación, se analiza las gráficas de los residuales para determinar si se cumplen los supuestos de la regresión.

En la gráfica 11a los datos están distribuidos de forma aleatoria, aunque el residual 3 tiene un valor muy alejado del 0 por lo que no se cumpliría el primer supuesto de la regresión en el que la media de los residuales debe ser 0. En la gráfica 11b se observa que los residuales cumple con el supuesto de normalidad puesto que los puntos se encuentran alrededor de la recta, aunque el residual 2 se encuentra un poco alejado de ella. En la gráfica 11c se observa que los residuales tienen la misma distribución, aunque el residual escalado tres tiene un valor de varianza mayor al de los otros residuales o el residual escalado dos tiene un valor menor al de los demás residuales, por lo que puede tener heterocedasticidad. En la gráfica 11d no se observa residuales o datos atípicos. Se concluye que el modelo ajustado a la regresión lineal simple cumple con los supuestos de la regresión.

Gráfica 11. Residuales de la regresión poblacional logística para la población del Departamento del Vichada según DANE.



Gráfica 12. Regresión poblacional logística de la población del departamento del Vichada a partir de Censos del DANE



Finalmente, en la gráfica 12 se observa que la curva de regresión se ajusta a los datos, lo que se puede comprobar con el coeficiente de determinación (R^2) cuyo valor fue de 0.973 lo que indica que el 97,3 % de los datos de la población del departamento del Vichada, según el DANE, son explicados por el modelo poblacional logístico.

3.1.5. Elección del modelo para estimar la población en el departamento del Vichada

Tabla 10. resumen del coeficiente de determinación y residual PRESS de cada modelo aplicado a la población de departamento del Vichada según datos del DANE

Modelo	Lineal	Parabólico	Exponencial	Logístico
Pr(>F)	0.00827	0.35037	0.06783	0.01344
R^2	0.9835	0.8895	0.9955	0.9733
PRESS	194618342	2214925224	496504604	203820152

En la tabla 10 tiene un resumen de los datos que permitió determinar el modelo más adecuado para la estimación de la población en los periodos 2004-2016 y 2019-2039 con el fin de contar con la información suficiente para establecer el modelo con el que se estimara el consumo de energía eléctrica del departamento del Vichada en el periodo 2019-2039.

Con un nivel de significancia de 95 % o un $\alpha = 0,05$ se estableció que algún β_i del modelo parabólico no es significativo, con p-valor de 0,35037 menor a α el modelo se descartó como posible candidato para la estimación de la población del departamento del Vichada. Según la prueba F de Fisher, los modelos que se ajustan a los datos del DANE son el lineal, con un p-valor de 0.00827, el exponencial con un p-valor de 0,06783 y el logístico con p-valor de 0,01344. El p-valor de los modelos lineal, exponencial y poblacional logístico son menores al valor de α , en estos casos se acepta los β_i de estos modelos como significativos por lo que estos modelos no pueden ser descartados por este medio.

Con el valor del coeficiente de determinación (R^2) se estableció que los estimados que mejor se ajustan a los datos del DANE son los calculados por medio del modelo exponencial, con un R^2 de 0.9955 frente a 0.9835 del modelo lineal y 0.9733 del modelo poblacional logístico. En este caso no se pudo descartar ningún modelo puesto que el comportamiento de la población esta explicado por los modelos lineal, exponencial y poblacional logístico.

Finalmente se buscó determinar el modelo más adecuado para la estimación de la población por lo que se acudió al residual PRESS. En este caso el modelo más adecuado es el lineal debido a que su residual PRESS es de 194618342 siendo el

residual de menor valor frente al PRESS de 496504604 del modelo exponencial y 203820152 del modelo logístico.

Se concluye que el modelo más adecuado para estimar la población del departamento del Vichada en los periodos 2003-2016 y 2019-2039 es el lineal simple.

En la tabla 11 se encuentra la estimación de la población para el periodo 2003-2016.

Tabla 11. Población estimada para el periodo 2013-2016 del departamento del Vichada por medio del modelo lineal simple

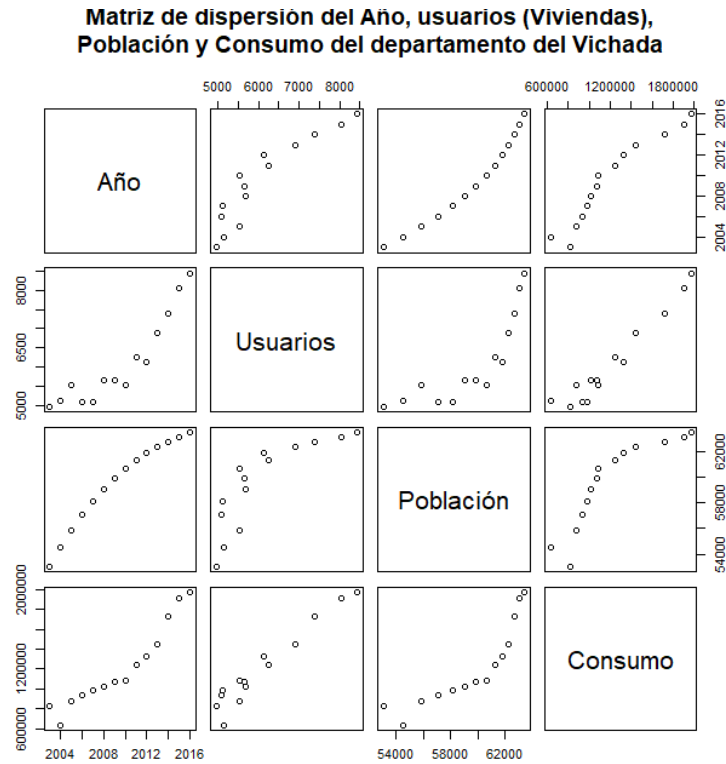
Año	Población
2003	50744
2004	52594
2005	54445
2006	56295
2007	58145
2008	59996
2009	61846
2010	63697
2011	65547
2012	67398
2013	69248
2014	71098
2015	72949
2016	74799

Con estos datos se analizó la correlación entre el consumo de energía según el SUI y la población del departamento del Vichada en el periodo 2003-2016.

3.2. MODELOS PARA ESTIMAR CONSUMO DE ENERGÍA ELÉCTRICA

Se buscó explicar el consumo de energía eléctrica por medio de la población, las viviendas consumidoras y el tiempo, por lo que se realizó la matriz de dispersión y de correlación de la población estimada a partir de los censos del DANE, las viviendas consumidoras y el consumo de energía eléctrica según el SUI. En este diagrama, gráfica 13, se encontró que hay correlación entre todas las variables.

Gráfica 13. Matriz de dispersión de las variables Año, Usuarios, Población y Consumo del departamento del Vichada



Para determinar la variable que se tomará se construyó la matriz de correlación, tabla 12. Las variables que más se encuentran correlacionadas con la variable consumo son el año y los usuarios, con una correlación de 0.9526563 y 0.9702971. Se ha escogido la variable Año ya que esta no requiere ser modelada.

Tabla 12. Matriz de correlación de las variables Año, Usuarios, Población y Consumo del departamento del Vichada

cor(Consumo2)				
	Año	Usuarios	Población	Consumo
Año	1	0.9144725	0.9761965	0.9526563
Usuarios	0.9144725	1	0.8178419	0.9702971
Población	0.9761965	0.8178419	1	0.8777748
Consumo	0.9526563	0.9702971	0.8777748	1

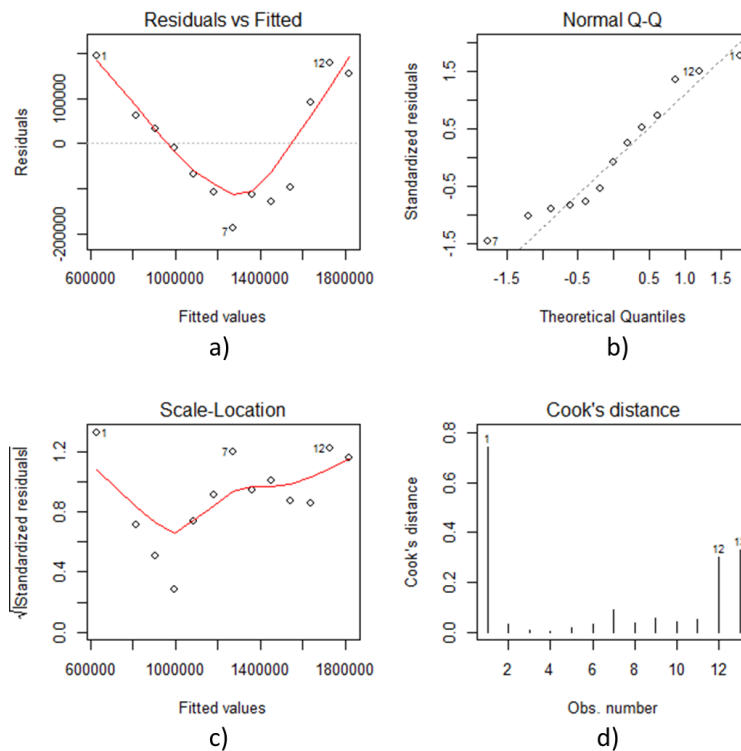
3.2.1. Modelo lineal simple

Tabla 13. ANOVA de la regresión lineal simple del consumo de energía eléctrica del departamento del Vichada según el SUI

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Consumo2\$ <i>t</i>	1	1.6265E+12	1.6265E+12	91.147	1.172E-06
Residuals	11	1.9629E+11	1.7844E+10		

En el análisis de Varianza o ANOVA, tabla 13, se llegó a un P valor de $1,17 * 10^{-6}$ por lo que se rechaza la hipótesis nula para un nivel de significancia del 95 %, aceptándose la hipótesis alternativa. Se concluye que los β_i . Son significativos y se acepta el modelo como una opción para determinar el consumo de energía eléctrica en el departamento del Vichada para el periodo 2019-2039.

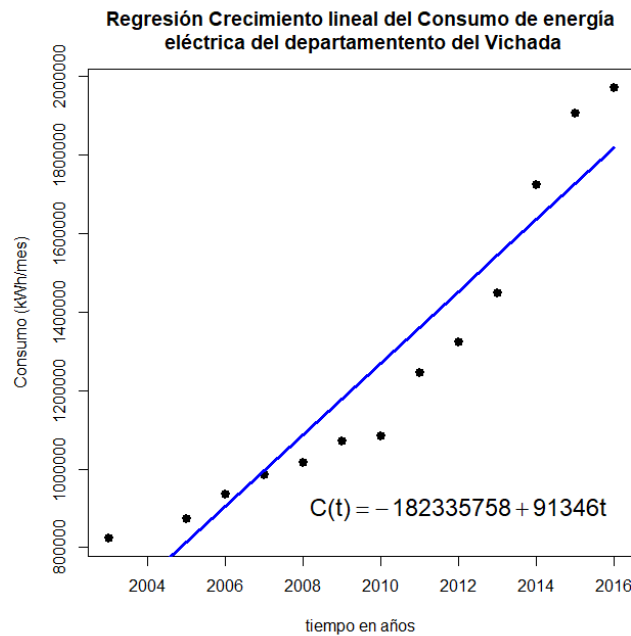
Gráfica 14. Residuales de la regresión lineal simple para el consumo del Departamento del Vichada según SUI.



En la gráfica 14a se observa que los datos no están distribuidos de forma aleatoria dado que los residuales tienen una tendencia, lo que sugiere el uso de otro modelo, aunque la media o promedio de los residuales es 0 siendo esta una razón para no descartar el modelo. En la gráfica 14b se observa que los residuales cumplen con el

supuesto de normalidad porque los puntos se encuentran alrededor de la recta de normalidad, aunque el residual 7 puede ser eliminado para mejorar la normalidad de los residuales. En la gráfica 14c se observa que los residuales tienen la misma distribución, aunque los residuales escalados 1, 4 y 5 se alejan del valor medio de los demás residuales por lo que se podría tener heterocedasticidad. En la gráfica 14d no se encuentra datos atípicos ya que todas las distancias son menores a 1.

Gráfica 15. Regresión lineal simple del consumo de energía eléctrica del departamento del Vichada a partir de datos del SUI.



En la gráfica 15 se observa que la recta de la regresión no se ajusta a los datos del SUI, pero el coeficiente de determinación (R^2) fue de 0.8923 o 89,2 % de los datos estimados se ajustan a los datos reales, siendo este un valor a la hora de estimar el ajuste del modelo a los datos. Se establece que este modelo es un buen candidato para ser usado en la estimación del consumo de energía eléctrica del periodo 2019-2039 del departamento del Vichada. Sin embargo, la elección del modelo se realizó con el análisis del residual PRESS.

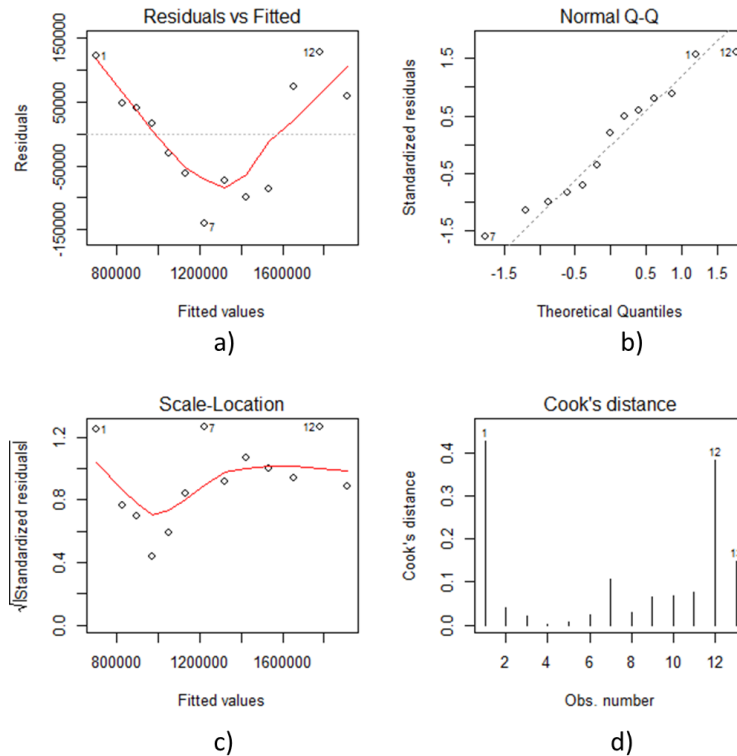
3.2.2. modelo exponencial

Tabla 14. ANOVA de la regresión exponencial del consumo de energía eléctrica del departamento del Vichada según el SUI

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Consumo2\$C.est	1	1.7315E+12	1.7315E+12	208.7	1.692E-08
Residuals	11	9.1262E+10	8296546441		

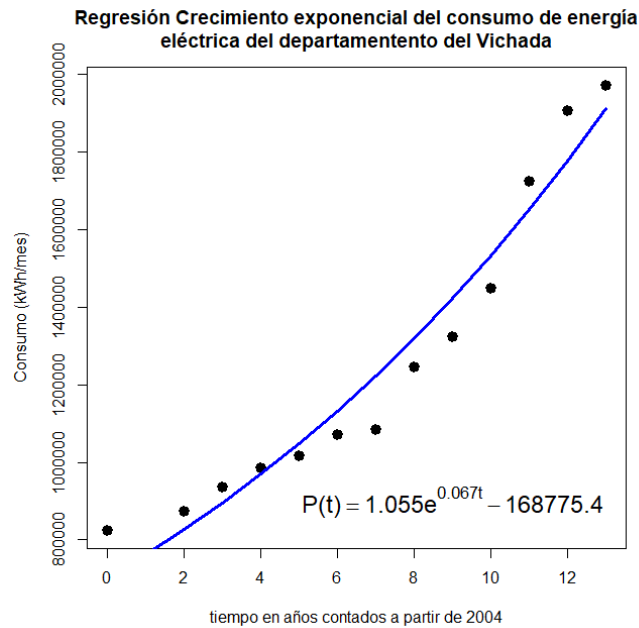
En el análisis de Varianza o ANOVA, tabla 14, se llegó a un P valor de $1,69 * 10^{-8}$ por lo que se rechaza la hipótesis nula para un nivel de significancia del 95 %, aceptándose la hipótesis alternativa. Se concluye que los β_i . Son significativos y se acepta el modelo como una opción para determinar el consumo de energía eléctrica en el departamento del Vichada para el periodo 2019-2039.

Gráfica 16. Residuales de la regresión exponencial para el consumo del Departamento del Vichada según SUI.



En la gráfica 16a se observa la misma situación que en la regresión lineal, los datos no están distribuidos de forma aleatoria porque los residuales tienen una tendencia, lo que sugiere el uso de otro modelo, aunque la media o promedio de los residuales es 0 siendo esta una razón para no descartar el modelo. En la gráfica 16b se observa que los residuales cumplen con el supuesto de normalidad debido a que los puntos se encuentran alrededor de la recta de normalidad. En la gráfica 16c se observa que los residuales tienen la misma distribución, aunque los residuales escalados 1, 4, 7 y 12 se alejan del valor medio de los demás residuales por lo que puede tener heterocedasticidad. En la gráfica 16d no se encuentran datos atípicos ya que todas las distancias son menores a 1.

Gráfica 17. Regresión exponencial del consumo de energía eléctrica del departamento del Vichada a partir de datos del SUI.



En la gráfica 17 se observa que la línea de la regresión se ajusta a los datos del SUI, y con un coeficiente de determinación (R^2) de 0.9499 o 95,0 % de los datos estimados se ajustan a los datos reales. Se establece que este modelo es un buen candidato para ser usado en la estimación del consumo de energía eléctrica del periodo 2019-2039 del departamento del Vichada.

3.2.3. Modelo parabólico

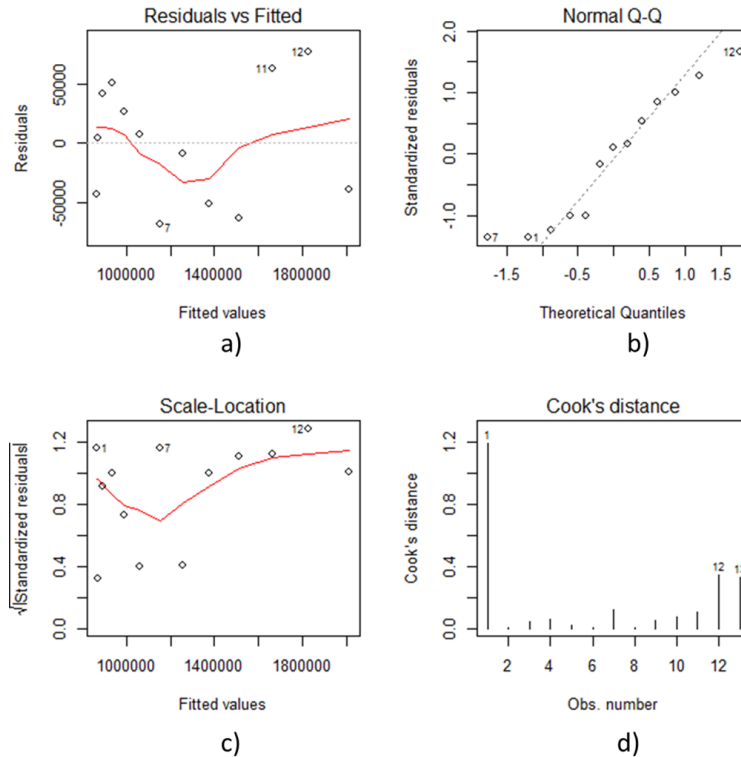
Tabla 15. ANOVA de la regresión parabólica del consumo de energía eléctrica del departamento del Vichada según el SUI

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Consumo2\$t	1	1.6265E+12	1.6265E+12	547.71	4.60E-10
Consumo2\$t2	1	1.6659E+11	1.6659E+11	56.1	2.09E-05
Residuals	10	2.9696E+10	2969554592		

En el análisis de Varianza o ANOVA, tabla 15, se llegó a un P valor de $4,60 * 10^{-10}$ por lo que se rechaza la hipótesis nula para un nivel de significancia del 95 %, aceptándose la hipótesis alternativa. Se concluye que los β_i . Son significativos y se

acepta el modelo como una opción para determinar el consumo de energía eléctrica en el departamento del Vichada para el periodo 2019-2039.

Gráfica 18. Residuales de la regresión parabólica para el consumo del Departamento del Vichada según SUI.

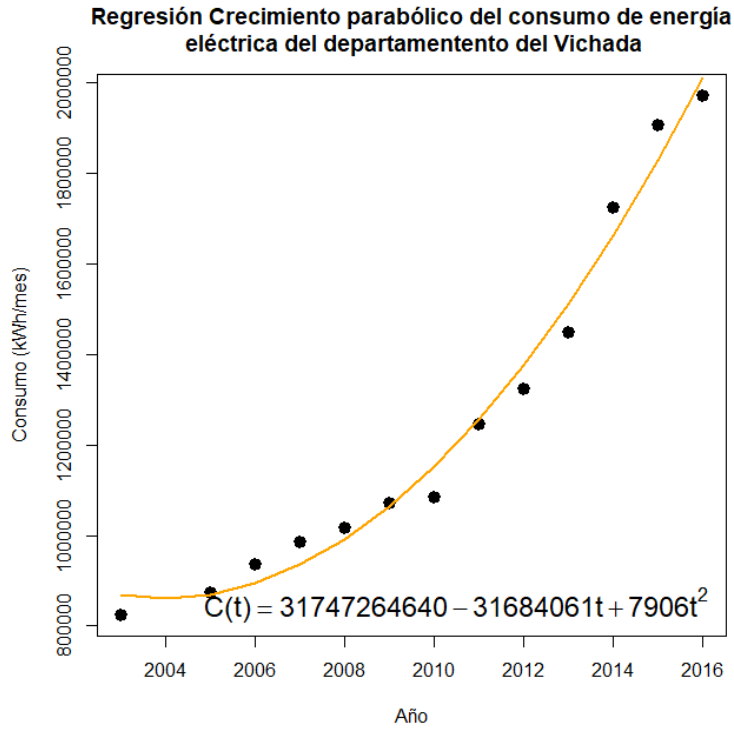


Este es el primer modelo aplicado al consumo de energía eléctrica que cumple a plenitud con sus residuales se distribuyan de forma aleatoria, se observa en la gráfica 18a que, además de que los residuales están distribuidos de forma aleatoria, también se puede establecer que la media o promedio de los residuales es 0. En la gráfica 18b se observa que los residuales cumple con el supuesto de normalidad porque los puntos se encuentran alrededor de la recta de normalidad, aunque el residual 8 deberá ser excluido. En la gráfica 18c se observa que los residuales tienen la misma distribución, aunque hay 3 residuales no especificados en la gráfica que se alejan del valor medio de los demás residuales por lo que puede tener heterocedasticidad. En la gráfica 18d se encuentra que el dato 1 puede ser atípico.

En la gráfica 19 se observa que hay más aleatoriedad en los residuales que para los dos primeros modelos corroborando el resultado de la gráfica 14a. también se observa que la recta de la regresión se ajusta a los datos del SUI, lo cual es confirmado con el coeficiente de determinación (R^2) de 0,9805 o 98,0 % de los datos estimados se ajustan a los datos reales, siendo este el coeficiente de determinación más alto hasta el momento. Hasta el momento este es el mejor modelo para ser

usado en la estimación del consumo de energía eléctrica del periodo 2019-2039 del departamento del Vichada.

Gráfica 19. Regresión parabólica del consumo de energía eléctrica del departamento del Vichada a partir de datos del SUJ.



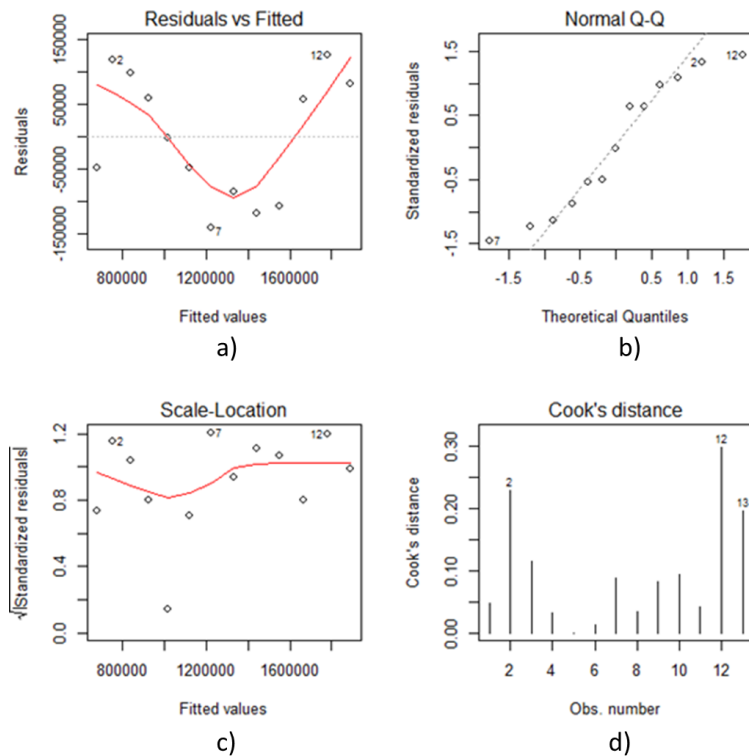
3.2.4. modelo logístico

Tabla 16. ANOVA de la regresión logística del consumo de energía eléctrica del departamento del Vichada según el SUJ

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Consumo2\$C.est1	1	1.9152E+12	1.9152E+12	192.66	2.572E-08
Residuals	11	1.0935E+11	9940969872		

En el análisis de Varianza o ANOVA, tabla 16, se llegó a un P valor de $2,572 * 10^{-8}$ por lo que, al igual que para los 3 modelos anteriores, se rechaza la hipótesis nula para un nivel de significancia del 95 %, aceptándose la hipótesis alternativa. Se concluye que los β_i son significativos y se acepta el modelo como una opción para determinar el consumo de energía eléctrica en el departamento del Vichada para el periodo 2019-2039.

Gráfica 20. Residuales de la regresión logística para el consumo del Departamento del Vichada según SUI.

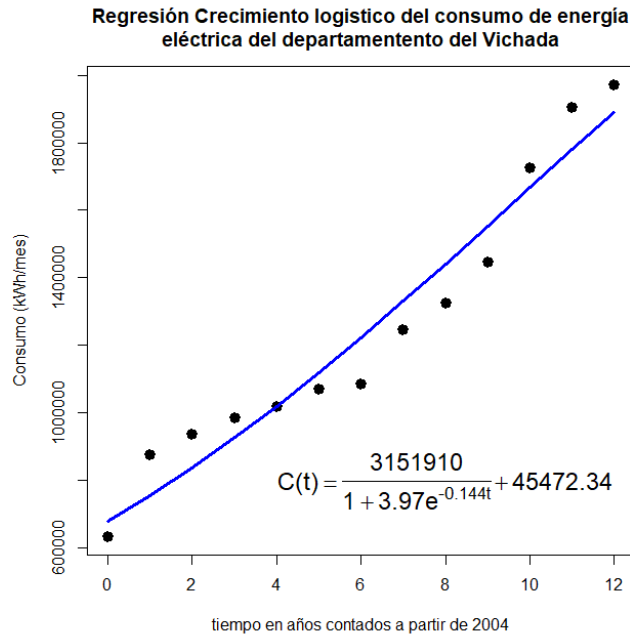


Otra vez se ha obtenido que los residuales no tienen una distribución de forma aleatoria, se observa en la gráfica 20a que se puede observar que la media o promedio de los residuales es 0. En la gráfica 20b se observa que los residuales cumple con el supuesto de normalidad ya que los puntos se encuentran alrededor de la recta de normalidad, aunque los residuales 7 y 12 deberán ser excluidos. En la gráfica 20c se observa que los residuales tienen la misma distribución, aunque hay 1 residuales no especificados en la gráfica que se alejan del valor medio de los demás residuales, pero en general no se ve problema de heterocedasticidad. En la gráfica 20d no se encuentra datos atípicos.

Finalmente en la gráfica 19 se observa que la recta de la regresión se ajusta a los datos del SUI, lo cual es confirmado con el coeficiente de determinación (R^2) de 0,946 o 94,6 % de los datos estimados se ajustan a los datos reales. Este es un mejor modelo para ser usado en la estimación del consumo de energía eléctrica del periodo 2019-2039 del departamento del Vichada.

A continuación, se analizó la información de los modelos completa para escoger el modelo más apropiado para estimar el consumo de energía eléctrica del departamento del Vichada para el periodo 2019-2039.

Gráfica 21. Regresión logística del consumo de energía eléctrica del departamento del Vichada a partir de datos del SUI.



3.2.5. Elección del modelo para estimar el consumo de eenergía eléctrica en el departamento del Vichada

Tabla 17. Resumen del coeficiente de determinación y residual PRESS de cada modelo aplicado al consumo de energía eléctrica de departamento del Vichada según datos del SUI.

Modelo	Lineal	Parabólico	Exponencial	Logístico
Pr(>F)	1.17E-06	1.69E-08	2.09E-05	2.57E-08
R ²	0.8923	0.9499	0.9805	0.946
PRESS	3.04857E+11	6.1534E+10	1.32849E+11	1.54921E+11

En la tabla 17 se tiene que el p-valor de los 4 modelos es menor al α de 0,05. A un nivel de significancia de 95 %, por lo que no se puede descartar modelos por con la ANOVA. El p-valor más pequeño es el del modelo parabólico, 1,69E-08, seguido por el del modelo logístico, con un valor de 2,57E-08.

Por otro lado, el coeficiente de determinación es significativo para los 4 casos, siendo los consumos estimados por medio del modelo exponencial los que más se ajustan a los datos del SUI, con un coeficiente de 0,9805 o 98,1 %, seguido del modelo parabólico con coeficiente de 0,9499 o 95,0 %. En este caso se puede pensar en descartar el modelo lineal simple debido a que tiene un coeficiente de 0,8923 o 89,2 %, pero esta no es una razón suficiente para prescindir del modelo.

Finalmente, al igual que para el modelo poblacional, se buscó determinar el modelo más adecuado para la estimación utilizando el residual PRESS. En este caso el modelo más adecuado es el parabólico puesto que su residual PRESS es de $6,1534E+10$, el modelo con el segundo residual PRESS más pequeño es el exponencial con un valor de $1,32849E+11$ y en tercer lugar se encuentra el modelo logístico con un residual de $1,54921E+11$.

De los cuatro modelos el más óptimo es el parabólico y el menos indicado es el lineal simple. Si se analizan los consumos estimados con el modelo parabólico, según la tabla 18, se obtiene un consumo de 10580786 kWh/mes o 10,6 millones de kWh/mes. Considerando que, según la tabla 5, el consumo para el 2003 fue de 824597.85 kWh/mes o 0,8 millones de kWh/mes y en el 2016 fue de 1972123.79 kWh/mes o 1,9 millones de kWh/mes en promedio, lo que representa un aumento de 1147525,94 kWh/mes o 1,1 millones en kWh/mes en 13 años. Se puede concluir que, a pesar de que el parabólico es el que más se ajustó a los datos y tiene mejores residuales, no se puede usar para estimar el consumo de energía eléctrica en el departamento del Vichada debido que por medio de este se estima un incremento del consumo mayor que el registrado históricamente. Este problema se puede observar con el modelo exponencial, con un estimado de 10,5 millones de kWh/mes para el año 2039. La estimación con el modelo logístico para el año 2039 es de 3117186 o 3,1 millón de kWh/mes, este valor representa un incremento de 1145062,21 o 1,1 millones de kWh/mes en un periodo de 23 años (2016-2039).

Se determinó que el modelo a usar para estimar el consumo de energía eléctrica es el logístico ya que con este se estima un crecimiento del consumo de energía eléctrica consistente con los datos histórico del SUI, además de que este tipo de modelos es especial para realizar estimaciones en largos periodos de tiempo.

Tabla 18. Consumos de energía eléctrica estimados con los modelos Lineal simple, Exponencial, Parabólico y Logístico, del departamento del Vichada a partir de datos recolectados del SUI.

Año	Consumo			
	Lineal	Exponencial	Parabólico	Logístico
2019	2091878	2375576	2654869	2201982
2020	2183224	2552091	2900961	2297114
2021	2274570	2740852	3162864	2386573
2022	2365916	2942709	3440578	2470014
2023	2457262	3158569	3734103	2547252
2024	2548608	3389404	4043439	2618248
2025	2639954	3636254	4368586	2683084
2026	2731300	3900229	4709543	2741947
2027	2822646	4182517	5066312	2795100
2028	2913992	4484390	5438891	2842866
2029	3005338	4807205	5827282	2885603

2030	3096684	5152415	6231483	2923692
2031	3188030	5521574	6651495	2957521
2032	3279376	5916343	7087318	2987473
2033	3370722	6338501	7538953	3013921
2034	3462068	6789945	8006398	3037218
2035	3553414	7272708	8489654	3057696
2036	3644760	7788963	8988720	3075662
2037	3736107	8341034	9503598	3091399
2038	3827453	8931405	10034287	3105163
2039	3918799	9562732	10580786	3117186

Con los consumos estimados por medio del modelo logístico se establecieron 3 escenarios principales, tabla 19, 1) el pesimista, que corresponde a los datos estimados directamente del modelo lo que representa un promedio del 45,7 % de la población con acceso al servicio de energía eléctrica. 2) el optimista, que corresponde al consumo que tendría el departamento del Vichada si toda la población (100 %) tuviese acceso al servicio de energía eléctrica. 3) el transicional, que corresponde a un aumento gradual del porcentaje de la población con acceso al servicio de energía eléctrica, partiendo de 45,7 % hasta llegar al 100 % para el 2019, como se muestra en la tabla 19.

Tabla 19. Escenarios futuros del consumo de energía eléctrica del departamento del Vichada en el periodo 2019-2039

Año	Consumo (kWh/mes)		
	Pesimista	Optimista	Transicional
2019	2201982	4818341.36	2201982.00
2020	2297114	5026507.66	2361221.15
2021	2386573	5222260.39	2523604.74
2022	2470014	5404844.64	2689035.98
2023	2547252	5573855.58	2857585.82
2024	2618248	5729207.88	3029503.04
2025	2683084	5871080.96	3205213.24
2026	2741947	5999884.03	3385328.72
2027	2795100	6116192.56	3570643.84
2028	2842866	6220713.35	3762146.50
2029	2885603	6314229.76	3961019.90
2030	2923692	6397575.49	4168663.29
2031	2957521	6471599.56	4386711.66
2032	2987473	6537140.04	4617066.69
2033	3013921	6595013.13	4861947.09

2034	3037218	6645991.25	5123944.33
2035	3057696	6690800.88	5406110.33
2036	3075662	6730113.79	5712066.12
2037	3091399	6764549.23	6046154.90
2038	3105163	6794667.40	6413638.34
2039	3117186	6820975.93	6820975.93

4. CONCLUSIONES

Se determinó la población del departamento de Vichada en los años 1985, 1993, 2005 y 2018 a partir de los datos del DANE 2005 y 2019. A estos datos se le aplicó los modelos lineal simple, parabólico, exponencial y poblacional logístico. Por medio del análisis de varianza o ANOVA, del coeficiente de determinación y el análisis de residuales se estableció que el modelo lineal es el más indicado para hacer estimaciones de la población del departamento del Vichada. Con el modelo indicado se estimó la población para el periodo 2003-2016.

Se obtuvo el consumo de energía eléctrica por mes desde el año 2003 hasta el año 2017 de la base de datos del SUI, pero no se tuvo en cuenta el consumo mensual promedio del año 2017 porque la información del SUI estaba incompleta y generó un dato atípico de 1310878.77 kWh/mes, cuando en los dos años anteriores el consumo fue de 1905781.42 kWh/mes para el 2015 y 1972123.79 kWh/mes para el 2016. También, del SUI, se estableció la cantidad de viviendas que tienen acceso a la energía eléctrica en promedio en el periodo 2003-2017.

Con la información obtenida del DANE y del SUI se determinó que, en la actualidad, el departamento del Vichada presenta un déficit en la prestación del servicio de energía eléctrica del 54,3 %.

A partir de la matriz de correlación (ver tabla 12), se estableció que los parámetros que influyen en el consumo de energía eléctrica son la población, la cantidad de viviendas que tienen acceso al servicio de energía eléctrica y el año; pero haciendo un análisis a la matriz de dispersión (ver gráfica 13) se encontró que estas variables también están correlacionadas entre sí, en consecuencia, al usar las tres variables en los modelos se genera problema de multicolinealidad, por lo tanto, se escogió solo el año para la aplicación de los modelos, al consumo de energía eléctrica.

El modelo lineal aplicado a los datos de consumo de energía eléctrica, fue descartado por tener el residual PRESS más grande, $3.04857 * 10^{11}$, frente a $6.1534 * 10^{10}$ del modelo parabólico, $1.32849 * 10^{11}$ del modelo exponencial y $1.54921 * 10^{11}$ del modelo Poblacional logístico; en consecuencia, se asume que el modelo lineal genera más errores que los otros modelos, al predecir el consumo de energía eléctrica en el departamento del Vichada.

Además, el modelo lineal presenta el coeficiente de determinación más bajo, de 0.8923 frente a los modelos parabólico, exponencial y logístico los cuales fueron de 0.9499, 0.9805 y 0.946 respectivamente (ver tabla 17); este hecho indica que el modelo lineal es el que menos sirve para hacer estimaciones del consumo de energía eléctrica en el departamento del Vichada.

A pesar de que el análisis de varianza o ANOVA, el coeficiente de determinación y el análisis de residuales indica que el modelo parabólico es el más indicado para las estimaciones, al realizar el análisis de los consumos estimados de los modelos se encontró que, el incremento del modelo parabólico y exponencial no son consistente con los datos histórico del SUI.

Según información obtenida del SUI (ver tabla 5), el consumo para el 2003 fue de 824597.85 kWh/mes o 0,8 millones de kWh/mes y en el 2016 fue de 1972123.79 kWh/mes o 1,9 millones de kWh/mes en promedio, lo que representa un aumento 1145062,21 o 1,1 millones de kWh/mes en 13 años. Según la tabla 18, para el 2039, el estimado con el modelo parabólico es de 10580786 o 10,6 millones de kWh/mes, con el modelo exponencial es de 9562732 o 9,5 millones de kWh/mes y con el modelo logístico es de 3117186 o 3,1 millón de kWh/mes.

Teniendo en cuenta los datos anteriores, se evidencia que en un periodo de 23 años (2016-2039), el incremento del consumo de energía eléctrica según los modelos parabólico, exponencial y poblacional logístico, respectivamente serán de 9756188,15 kWh/mes, 8516436,15 y 1145062,21 kWh/mes; en consecuencia, el modelo poblacional logístico es más consistente con el crecimiento según los datos del SUI.

Lo anterior quiere decir que para el periodo 2019-2039 el consumo mínimo de energía eléctrica estimado para el departamento del Vichada estará entre 2201982 y 3117186 kWh al mes; y el consumo de energía eléctrica máximo estimado estará entre 4818341.36 y 6820975.93 kWh al mes.

Teniendo en cuenta los análisis realizados, se determinó que el modelo que debe ser usado para estimar el consumo de energía eléctrica en el departamento del Vichada para el periodo 2019-2039 es el poblacional logístico o modelo de Verhulst, puesto que los datos estimados por medio de este modelo representan un crecimiento del consumo de energía eléctrica consistente con los datos histórico del SUI.

Se logró estimar el consumo de energía eléctrica mediante el uso del modelo Poblacional Logístico para el departamento del vichada durante el periodo 2019-2039 estableciendo tres escenarios futuros a saber, el pesimista, que corresponde a los datos estimados directamente del modelo lo que representa un promedio del 45,7 % de la población con acceso al servicio de energía eléctrica, el optimista, que corresponde al consumo que tendría el departamento del Vichada si toda la población (100 %) tuviese acceso al servicio de energía eléctrica y el transicional, que corresponde a un aumento gradual del porcentaje de la población con acceso al servicio de energía eléctrica, partiendo de 45,7 % hasta llegar al 100 % para el 2039, como se muestra en la tabla 19.

5. RECOMENDACIONES

Se recomienda, que a medida que surjan más datos de entidades públicas como el DANE y el SUI, se ajuste el modelo poblacional logístico para que los consumos estimados se aproximen más a la realidad del contexto en el periodo estudiado.

Frente a la formulación de política pública de energización rural sostenible en el departamento de Vichada en el periodo 2019-2039 se debe tener en cuenta lo siguiente: que en la actualidad hay un déficit de prestación de servicio de energía eléctrica de 54,3 % de acuerdo a los datos obtenidos del DANE y del SUI; que según la población, los usuarios y el año, hay un aumento de demanda para el periodo 2019-2039 del consumo mínimo de energía eléctrica estimado para el departamento que se encontrará entre 2,2 millones y 3,1 millones de kWh al mes; y el consumo de energía eléctrica máximo estimado estará entre 4,8 millones y 6,8 millones de kWh al mes.

Por lo anterior, es importante que, frente a las ofertas energéticas del territorio de Vichada, el departamento este apostando por la creación de energías alternativas para suplir el consumo potencial de energía eléctrica en el departamento del vichada para el periodo 2019-2039

BIBLIOGRAFÍA

- Andeson, D., Sweeney, D., & Williams, T. *Estadística para administración y economía Estadística para administración y economía* (10 ed.). México D.F.: Cengage Learning Editores,S.A. 2008
- Barrientos Marín, J., & Toro Martínez, M. Análisis de los fundamentales del precio de la energía eléctrica: evidencia empírica para Colombia. *revista de economía del caribe*(19), 33-59. 2017
- Barrientos, A. F., Olaya, J., & González, V. M. Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, 30(2), 187-202. 2007
- Camarero Rioja, L., Almazán Llorente, A., & Mañas Ramírez, B. (s.f.). Regresión Logística: Fundamentos y aplicación a la investigación sociológica. 2. Recuperado el 2019, de <https://www2.uned.es/socioestadistica/Multivariante/TextoMultivarianteOddRatioV1mayo2017.pdf>. 2019
- Cardona Madariaga, D. F., González Rodríguez, J. L., Rivera Lozano, M., & Cárdenas Vallejo, E. *Inferencia estadística. Módulo de regresión lineal simple. Documento de investigación No. 147*. Bogotá D.C. 2013
- Costa Cor, T., Boj del Val, E., & Fortiana Gregori, J. Bondad de ajuste y elección del punto de corte en regresión logística basada en distancias. Aplicación al problema de "credit scoring". *Dialet*(18), 19-40. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=4084989>. 2012
- Cuadras, C. *Nuevos métodos de análisis multivariante*. Barcelona: CMC Editions. 2014
- DANE. *Censo general 2005*. 2008
- DANE. *Estimados población censo 2005 y proyecciones a 2020*. Bogotá. 2016
- De la Puente Viedm, C. *Esadística descriptiva e inferencial y una introducción al método cinético* (1 ed.). Madrid: Complutense S. A. 2019
- Días Monrroy , L. G. *Análisis estadístico de datos multivariados* (1 ed.). Bogotá D.C. 2012
- Equipo PERS Orinoquia. *Diagnostico de la generación electrica en las Zonas No Interconectadas de la Orinoquia*. Universidad de los Llanos, Villavicencio, Meta. 2019

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. *Metodología de la investigación*. México D.F.: McGraw-Hill Interamericana. 2004
- IPSE. *Informe mensual de telemetría Diciembre 01- 31 de 2018*. Bogotá D.C. 2019
- IPSE. (s.f.). *Planes de energización rural sostenibles*. Bogotá D.C. Obtenido de http://www.ipse.gov.co/pages/ipse/Informe_PERS_Direcci%C3%B3n1.pdf
- Llinás, H. J. Precisiones en la teoría de los modelos logísticos. *Revista Colombiana de Estadística*, 29(2), 239-265. 2006
- Martínez Bencardino, C. *Estadística y muestreo* (13 ed.). Bogotá D.C.: Ecoe ediciones. 2012
- Medina Hurtado, S., & García Aguado, J. Predicción de demanda de energía en Colombia mediante un sistema de inferencia difuso neuronal. *Energética. Revista Energética*(33), 15-24. Recuperado el 12 de Febrero de 2019, de <http://www.redalyc.org/articulo.oa?id=147019387002>. 2005
- Montgomery, D. C., Peck, E. A., & Geoffrey, V. G. *Introducción al análisis de regresión lineal* (3 ed.). México D.F.: Compañía editorial continental. 2006
- Ocaña-Riola, R. Modelos de Markov aplicados a la investigación en Ciencias de la Salud. *Revista Interciencia*, 34(3), 157-162. 2009
- Orellana Romero, J. L. Modelación y Pronóstico de la Demanda de Energía Eléctrica de Mediano Plazo de El Salvador. Ciudad Universitaria, El salvador. 2012
- PERS_Orinoquía. *Planes de energización rural sostenibles – pers, región orinoquia (arauca, casanare, meta vichada). Energías renovables: una opción para el desarrollo rural productivo*. Universidad de los Llanos, Villavicencio. 2016
- PERS_Orinoquía. Diagnostico de la generación electrica en las Zonas No Interconectadas de la Orinoquia. Villavicencio, Meta, Colombia. 2019
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. *Probabilidad y estadística para ingeniería y ciencias* (8 ed.). México D.F.: Pearson Educación. 2007
- Zetina Mogueli, C. E., & Ríos Lara, G. V. Estimación del tamaño de la población de langostas *Panulirus Argus* en las costas de Yucatan, usando diferentes modelos de evaluación. 1998

ANEXOS

ANEXO A: Demostración de la ecuación (20) a partir de la ecuación (19)

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right) \quad (19)$$

Se empieza a aplicar el método de variables separables

$$\frac{dP}{dt} = rP \left(\frac{K - P}{K} \right)$$

$$K \frac{dP}{P(K - P)} = r dt$$

$$\int_{P_0}^P \frac{K}{P(K - P)} dP = r \int_0^t dt$$

Se aplica fracciones parciales

$$\frac{K}{P(K - P)} = \frac{A}{P} + \frac{B}{K - P}$$

$$K = A(K - P) + BP$$

Con $P = K$

$$K = A(K - K) + BK$$

$$K = A(0) + BK$$

$$K = BK$$

$$B = 1$$

Con $P = 0$

$$K = A(K - 0) + B * 0$$

$$K = AK$$

$$A = 1$$

Se continua con la integral

$$\int_{P_0}^P \frac{K}{P(K-P)} dP = r \int_0^t dt$$

$$\int_{P_0}^P \left(\frac{1}{P} + \frac{1}{K-P} \right) dP = r \int_0^t dt$$

$$\ln(P) - \ln(K-P) - (\ln(P_0) - \ln(K-P_0)) = rt$$

$$\ln(P) - \ln(K-P) - \ln(P_0) + \ln(K-P_0) = rt$$

$$\ln(P) + \ln(K-P_0) - (\ln(K-P) + \ln(P_0)) = rt$$

$$\ln(P(K-P_0)) - \ln(P_0(K-P)) = rt$$

$$\ln\left(\frac{P(K-P_0)}{P_0(K-P)}\right) = rt$$

Se despeja a P

$$\frac{P(K-P_0)}{P_0(K-P)} = e^{rt}$$

$$P(K-P_0) = e^{rt} P_0(K-P)$$

$$P(K-P_0) = Ke^{rt} P_0 - Pe^{rt} P_0$$

$$P(K-P_0) + Pe^{rt} P_0 = Ke^{rt} P_0$$

$$P(K-P_0 + P_0 e^{rt}) = Ke^{rt} P_0$$

$$P(K - P_0(1 - e^{rt})) = Ke^{rt} P_0$$

$$P = \frac{Ke^{rt} P_0}{K - P_0(1 - e^{rt})}$$

Tenemos que la única variable, aparte de la población, es el tiempo, por lo que se llega a la ecuación (20)

$$P(t) = \frac{Ke^{rt} P_0}{K - P_0(1 - e^{rt})}$$

ANEXO B: Script en R, Consumo promedio de energía eléctrica al mes por año.

```
setwd("D:/Tesis/Datos/Vichada sui")
library (xlsx)
for(k in 2003:2016){
  for(j in 1:12){
    dir<-paste("VICHADA_",k,"_",j,".csv", sep="")
    d1<-read.table(dir,header=T,sep=";",dec=".",stringsAsFactors =
FALSE)
    df1<-data.frame(d1)
    for(i in 1:5){
      df1[1]<-NULL
    }
    for(i in 1:4){
      df1[9]<-NULL
    }
    for(i in 1:2){
      df1[length(df1)]<-NULL
    }
    min<-nrow(df1)-3
    max<-nrow(df1)
    df1<-df1[-c(min:max),]
    df1[df1=='ND']<-'0'
    archivo<-paste("Consumo energía_", k,"_",j,".csv", sep="")
    write.csv(df1, file=archivo)
    d1<-read.table(archivo,header=T,sep=";",dec=".",stringsAsFactors =
FALSE)
    df1<-data.frame(d1)
    df2<-0
    names(df2)<-c("Residencial")
    for(i in 1:6){
      df2[1]<-df2+df1[i]
    }
    Resultado<-cbind(df2,df1)
    for(i in 1:7){
      Resultado[2]<-NULL
    }
    archivo<-paste("VICHADA_",k,".xlsx", sep="")
    Hoja<-paste("Mes", j, sep="")
    write.xlsx(Resultado,file=archivo, sheetName=Hoja, append=TRUE)
  }
}
```

ANEXO C: Script en R, Modelo lineal simple aplicado a la población de los censos del DANE.

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Pobl<-read_excel("Base de datos Vichada general.xlsx", sheet='Poblacion')
Pobl2<-data.frame(Pobl)
P.t<-function(i) {
  round(coef[1]+coef[2]*i, 0)
}
PRESS<-function(i) {
  residuales<-resid(i)
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
names(Pobl2)<-c('t', 'P')
regresion<-lm(Pobl2$P ~ Pobl2$t)
coef=coefficients(regresion)
summary(regresion)
anova(regresion)
PRESS(regresion)
x11()
par(mfrow=c(2,2))
plot(regresion,1:4)
x11()
plot(Pobl2$P ~ Pobl2$t, main='Regresión Crecimiento lineal de la Población\ndel
departamentento del Vichada', xlab='tiempo en años', ylab='Población', lwd=3,
pch=19)
lines(seq(1985, 2018, 1),P.t(seq(1985, 2018, 1))), col='blue', lwd=3)
polygon(c(1996,1997), c(25000,25000), border='blue', lwd=3)
text(2008,25000,substitute("P(t)"==paste(-3655679.548+1850.436*t)),cex=1.5)
a<-2003:2016
b<-P.t(2019:2039)
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Población')
Proyeccion
```

ANEXO D: Script en R, Modelo exponencial aplicado a la población de los censos del DANE.

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Pobl<-read_excel("Base de datos Vichada general.xlsx", sheet='Poblacion')
Pobl2<-data.frame(Pobl)
Ln<-function(i){
  log(i,exp(1))
}
P.t<-function(p0,t,r){
  round(p0*exp(r*(t-1985)))
}
PRESS<-function(i) {
  residuales<-resid(i)
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
x11()
plot(Pobl2$P ~ Pobl2$t, main='Regresión Crecimiento exponencial de la
Población\ndel departamentento del Vichada',lwd=5, xlab='tiempo en años
contados a partir de 1985', ylab='Población', pch=19)
lines(seq(0, 33, 1),coef2[2]*P.t(Pobl2$P[1],seq(1985, 2018, 1),r)+coef2[1],
col='orange', lwd=3)
a<-2019:2039
b<-P.t(Pit,a,r.est)
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Población')
Proyeccion
```

ANEXO E: Script en R, Modelo parabólico aplicado a la población de los censos del DANE.

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Pobl<-read_excel("Base de datos Vichada general.xlsx", sheet='Poblacion')

Pobl2<-data.frame(Pobl)
Pobl2[3]<-(Pobl2[1])^2
names(Pobl2)<-c('t', 'P', 't2')
View(Pobl2)
Regresion<-lm(Pobl2$P ~ Pobl2$t+Pobl2$t2)
Regresion
coef=coefficients(Regresion)
coef
P.t<-function(i){
  round(coef[1]+coef[2]*(i)+coef[3]*(i)^2, 0)
}
PRESS<-function(i) {
  residuales<-resid(i)
  residuales
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
summary(Regresion)
anova(Regresion)
PRESS(Regresion)
par(mfrow=c(2,2))
plot(Regresion,1:4)
x11()
plot(Pobl2$P ~ Pobl2$t, main='Regresión Crecimiento parabolico de la
Población\ndel departamentento del Vichada',lwd=5, xlab='Año', ylab='Población',
pch=19)
lines(seq(1985, 2018, 1),P.t(seq(1985, 2018, 1)), col='orange', lwd=2)
text(2005,25000,substitute("P(t)"==paste(-90970422+89096*t'-
22*t"^2")),cex=1.5)
a<-2019:2039
b<-P.t(2019:2039)
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Población')
Proyeccion
```


ANEXO F: Script en R, Modelo poblacional logístico aplicado a la población de los censos del DANE.

```

setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Pobl<-read_excel("Base de datos Vichada general.xlsx", sheet='Poblacion')
Ln<-function(i){
  log(i,exp(1))
}
k<-function(p0,p1,p2) {
  (2*p0*p1*p2-(p0+p2)*p1^2)/(p0*p2-p1^2)
}
A<-function(k,p0) {
  (k-p0)/p0
}
P.t<-function(p0,r,k,t,A){
  Ak<-(k-p0)/p0
  round(k/(1+A*exp(r*(t-1985))))
}
PRESS<-function(i) {
  residuales<-resid(i)
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
Pobl2[3]<-Pobl2[1]-1985
names(Pobl2)<-c('año', 'P', 't')
P01<-Pobl2$P[1]
P11<-Pobl2$P[3]
P21<-Pobl2$P[4]
k1<-k(P01,P11,P21)
n1<-Pobl2$t[4]/2
r1<-(1/n1)*Ln((P01*(k1-P11))/(P11*(k1-P01)))
A1<-A(k1, P01)
Pobl2[4]<-P.t(P01,r1,k1,Pobl2[1],A1)
names(Pobl2)<-c('Año', 'P', 't', 'P.est1')
regresion1<-lm(Pobl2$P~Pobl2$P.est1)
regresion1
coef=coefficients(regresion1)
summary(regresion1)
anova(regresion1)
PRESS(regresion1)
x11()
par(mfrow=c(2,2))
plot(regresion1,1:4)
x11()

```

```

plot(Pobl2$P ~ Pobl2$t, main='Regresión Crecimiento logístico de la Población\ndel
departamentento del Vichada',lwd=5, xlab='tiempo en años contados a partir de
1985', ylab='Población', pch=19)
lines(seq(0, 33, 1),0.9263*P.t(P01, r1, k1, seq(1985, 2018, 1),A1)+2707, col='blue',
lwd=3)
text(25,30000,substitute("P(t)"==paste(frac(73657,
paste(1+4.78*"e"^0.146t"))+2707)),cex=1)
a<-2019:2039
b<-round(0.9263*P.t(P01, r1, k1, seq(2019, 2039, 1),A1)+2707, 0)
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Población')
Proyeccion

```

ANEXO G: Script en R, Gráfico de dispersión y matriz de correlación

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Consumo<-read_excel("Base de datos Vichada general.xlsx", sheet='Consumo')
Consumo2<-data.frame(Consumo)
plot(Consumo2, main='Matriz de dispersión del Año, usuarios
(Viviendas),\nPoblación y Consumo del departamento del Vichada')
cor(Consumo2)
for(i in 1:2){
  Consumo2[2]<-NULL
}

names(Consumo2)<-c('t', 'C')
x11()
plot(Consumo2$C ~Consumo2$t, main='Consumo de energía eléctrica
del\ndepartamento del Vichada según SUI',lwd=5, xlab='Tiempo en años',
ylab='Consumo (kWh/mes)',pch=19)
Consumo2<-Consumo2[-2,]
x11()
plot(Consumo2$C ~Consumo2$t, main='Consumo de energía eléctrica
del\ndepartamento del Vichada según SUI',lwd=5, xlab='Tiempo en años',
ylab='Consumo (kWh/mes)',pch=19)
```

ANEXO H: Script en R, Modelo Lineal simple del consumo de energía eléctrica

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Consumo<-read_excel("Base de datos Vichada general.xlsx", sheet='Consumo')
Consumo2<-data.frame(Consumo)
for(i in 1:2){
  Consumo2[2]<-NULL
}
names(Consumo2)<-c('Año', 'C')
Consumo2<-Consumo2[-2,]
C.t<-function(i) {
  round(coef[1]+coef[2]*i, 2)
}
PRESS<-function(i) {
  residuales<-resid(i)
  Cr<-residuales/(1-lm.influence(i)$hat)
  sum(Cr^2)
}
names(Consumo2)<-c('t', 'C')
regresion<-lm(Consumo2$C ~ Consumo2$t)
regresion
coef=coefficients(regresion)
summary(regresion)
anova(regresion)

PRESS(regresion)
x11()
par(mfrow=c(2,2))
plot(regresion,1:4)
x11()
plot(Consumo2$C ~ Consumo2$t, main='Regresión Crecimiento lineal del
Consumo de energía\n eléctrica del departamentento del Vichada', xlab='tiempo en
años', ylab='Consumo (kWh/mes)', lwd=3, pch=19)
lines(seq(2003, 2016, 1),C.t(seq(2003, 2016, 1)), col='blue', lwd=3)
text(2012,900000,substitute("C(t)"==paste(-182335758+91346*t)),cex=1.5)
a<-2019:2039
b<-C.t(2019:2039)
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Consumo')
Proyeccion
```

ANEXO I: Script en R, Modelo parabólico del consumo de energía eléctrica

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Consumo<-read_excel("Base de datos Vichada general.xlsx", sheet='Consumo')
Consumo2<-data.frame(Consumo)
for(i in 1:2){
  Consumo2[2]<-NULL
}
names(Consumo2)<-c('Año', 'C')
Consumo2<-Consumo2[-2,]
Consumo2[3]<-(Consumo2[1])^2
names(Consumo2)<-c('t', 'C', 't2')
View(Consumo2)
Regresion<-lm(Consumo2$C ~ Consumo2$t+Consumo2$t2)
Regresion
coef=coefficients(Regresion)
coef
C.t<-function(i){
  round(coef[1]+coef[2]*(i)+coef[3]*(i)^2, 2)
}
PRESS<-function(i) {
  residuales<-resid(i)
  residuales
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
summary(Regresion)
anova(Regresion)
PRESS(Regresion)
par(mfrow=c(2,2))
plot(Regresion,1:4)
x11()
plot(Consumo2$C ~ Consumo2$t, main='Regresión Crecimiento parabólico del
consumo de energía\n eléctrica del departamentento del Vichada',lwd=5,
xlab='Año', ylab='Consumo (kWh/mes)', pch=19)
lines(seq(2003, 2016, 1),C.t(seq(2003, 2016, 1)), col='orange', lwd=2)
text(2010.5,850000,substitute("C(t)"==paste(31747264640-
31684061*t+7906*t^2)),cex=1.5)
a<-2019:2039
b<-C.t(2019:2039)
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Consumo')
Proyeccion
```

ANEXO J: Script en R, Modelo exponencial del consumo de energía eléctrica

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Consumo<-read_excel("Base de datos Vichada general.xlsx", sheet='Consumo')
Consumo2<-data.frame(Consumo)
for(i in 1:2){
  Consumo2[2]<-NULL
}
names(Consumo2)<-c('Año', 'C')
Consumo2<-Consumo2[-2,]
Ln<-function(i){
  log(i,exp(1))
}
C.t<-function(p0,t,r){
  round(p0*exp(r*(t-2003)))
}
PRESS<-function(i) {
  residuales<-resid(i)
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
Consumo2[3]<-Consumo2[1]-2003
names(Consumo2)<-c('año', 'C', 't')
n<-Consumo2$t[13]
r<-(1/n)*Ln(Consumo2$C[13]/Consumo2$C[1])
Consumo2[4]<-C.t(Consumo2$C[1], Consumo2$año, r)
names(Consumo2)<-c('año', 'C', 't', 'C.est')
regresion2<-lm(Consumo2$C ~ Consumo2$C.est)
regresion2
coef2=coefficients(regresion2)
summary(regresion2)
anova(regresion2)
Residuales PRESS
PRESS(regresion2)
x11()
par(mfrow=c(2,2))
plot(regresion2,1:4)
x11()
plot(Consumo2$C ~ Consumo2$t, main='Regresión Crecimiento exponencial del
consumo de energía\neléctrica del departamentento del Vichada',lwd=5,
xlab='tiempo en años contados a partir de 2004', ylab='Consumo (kWh/mes)',
pch=19)
```

```
lines(seq(0, 13, 1),1.055*C.t(Consumo2$C[1],seq(2003, 2016, 1),r)-168775.356,  
col='blue', lwd=3)  
text(9,900000,substitute("P(t)"==paste(1.055*"e"^(0.067t"-168775.356))),cex=1.5)  
a<-2019:2039  
b<-1.055*C.t(Consumo2$C[1],a,r)-168775.356  
Proyeccion<-data.frame(cbind(a,b))  
names(Proyeccion)<-c('Año','Consumo')  
Proyeccion
```

ANEXO K: Script en R, Modelo exponencial del consumo de energía eléctrica

```
setwd("D:/Tesis/Datos/Datos y script")
library(readxl)
Consumo<-read_excel("Base de datos Vichada general.xlsx", sheet='Consumo')
Consumo2<-data.frame(Consumo)
for(i in 1:2){
  Consumo2[2]<-NULL
}
names(Consumo2)<-c('Año', 'C')
Consumo2<-Consumo2[-1,]
Ln<-function(i){
  log(i,exp(1))
}
k<-function(p0,p1,p2) {
  abs((2*p0*p1*p2-(p0+p2)*p1^2)/(p0*p2-p1^2))
}
A<-function(k,c0) {
  (k-c0)/c0
}
C.t<-function(c0,r,k,t,A){
  Ak<-(k-c0)/c0
  round(k/(1+Ak*exp(-r*(t-2004))),1)
}
PRESS<-function(i) {
  residuales<-resid(i)
  Pr<-residuales/(1-lm.influence(i)$hat)
  sum(Pr^2)
}
Consumo2[3]<-Consumo2[1]-2004
names(Consumo2)<-c('año', 'C', 't')
C01<-Consumo2$C[1]
C11<-Consumo2$C[6]
C21<-Consumo2$C[13]
k1<-k(C01,C11,C21)
n1<-Consumo2$t[6]
r1<-abs((1/n1)*Ln((C01*(k1-C11)/(C11*(k1-C01))))))
A1<-A(k1, C01)
Consumo2[4]<-C.t(C01,r1,k1,Consumo2[1],A1)
names(Consumo2)<-c('Año', 'C', 't', 'C.est1')
x11()
plot(Consumo2$C.est1~ Consumo2$t, main='gráfica de dispersión del consumo de
energía eléctrica en kWh/mes del departamentento del Vichada',lwd=5,
xlab='tiempo en años contados a partir de 2004', ylab='Consumo', pch=19)
regresion1<-lm(Consumo2$C~Consumo2$C.est1)
```



```

regresion1
coef=coefficients(regresion1)
summary(regresion1)
anova(regresion1)
PRESS(regresion1)
x11()
par(mfrow=c(2,2))
plot(regresion1,1:4)
x11()
plot(Consumo2$C ~ Consumo2$t, main='Regresión Crecimiento logístico del
consumo de energía\neléctrica del departamentento del Vichada',lwd=5,
xlab='tiempo en años contados a partir de 2004', ylab='Consumo (kWh/mes)',
pch=19)
lines(seq(0, 12, 1),1.002*C.t(C01, r1, k1, seq(2004, 2016, 1),A1)+45472.338,
col='blue', lwd=3)
text(8,800000,substitute("C(t)"==paste(frac(paste("3151910"), paste(1+3.97*"e"^-
0.144t")))+45472.34),cex=1.5)
a<-2019:2039
b<-C.t(C01, r1, 3151910, seq(2019, 2039, 1),A1)+45472.34
Proyeccion<-data.frame(cbind(a,b))
names(Proyeccion)<-c('Año','Consumo')
Proyeccion

```

RESUMEN ANALITICO ESPECIALIZADO

A. TIPO DE DOCUMENTO OPCIÓN DE GRADO	Trabajo de grado modalidad EPI como requisito para optar por el título de Licenciado(a) en Matemáticas y Física
B. ACCESO AL DOCUMENTO	Universidad de los Llanos, Biblioteca Central.
1. TÍTULO DEL DOCUMENTO	Estimación del consumo potencial de energía eléctrica mediante el uso de un modelo matemático para el departamento del Vichada durante el periodo 2019-2039.
2. AUTORES	Avilan Trujillo, Karen; González, Fernando Aleisy
3. LUGAR Y AÑO DE PUBLICACIÓN	Villavicencio, 2019.
4. UNIDAD PATROCINANTE	Universidad de los Llanos
5. PALABRAS CLAVES	ANOVA, Consumo, Energía, Estimación, Exponencial, Hipótesis, Lineal, Logístico, Modelo, Parabólico, PERS, PRESS, Regresión, residuales, Vichada.
6. DESCRIPCIÓN	El Plan de Energización Rural Sostenibles de la Orinoquia (PERS Orinoquía) es un plan que nace de reconocimiento de la problemática asociada a la falta de expansión de la cobertura de la energía eléctrica en las zonas rurales apartadas y que, a partir de los elementos regionales más relevantes en términos de condiciones socioeconómicas, de recurso, de oferta y demanda energética, busca identificar las fuentes locales aprovechables para el suministro de energía mediante la estructuración y elaboración de proyectos integrales y sostenibles en el corto plazo y en un horizonte de mínimo 15 años, que además de generar energía, apoyen el crecimiento y el

	<p>desarrollo de las comunidades rurales de las regiones objetivos.</p> <p>PERS Orinoquía será la formulación de una política pública energética que vaya en sintonía con el entorno y la visión de desarrollo regional con emprendimientos y productividad local, a partir de la generación de energía eléctrica que apoye el crecimiento y el mejoramiento de las condiciones de las comunidades locales en los departamentos de Arauca, Casanare, Meta y Vichada.</p> <p>Con este trabajo se busca aportar a PERS Orinoquía puesto que la estimación del consumo de energía eléctrica entre los años 2019 y 2039 permitirá que se establezca, entre la política pública, la formulación de proyectos de energización en donde se haga uso de energías alternativas a partir de la realidad del departamento del Vichada. Estos proyectos permitirán que la población del departamento del Vichada tenga acceso a los fondos de apoyo financiero, que brinda el gobierno.</p> <p>El departamento del Vichada presenta gran interés para PERS Orinoquía puesto que no se encuentra en el sistema interconectado nacional (SIN), se estableció que, a partir de los datos del DANE y del SUI, que el departamento tiene un déficit en prestación del servicio de 54,3 %. Los departamentos de la periferia son los más afectados principalmente por lo costoso que sería llevar la infraestructura eléctrica. Según la IPSE, al 31 de diciembre de 2018, el 94% de la energía consumida en Puerto Carreño correspondía a la energía suministrada por la interconexión con Venezuela, el 6 % restante se genera desde la central DIESEL de Puerto Carreño.</p> <p>El departamento del Vichada está situado en el extremo oriental de Colombia y se encuentra</p>
--	---

	<p>conformado por los municipios de Cumaribo, La primavera, Puerto Carreño y Santa Rosalía</p> <p>Para obtener información energética en el Vichada, se realizó una investigación con enfoque cuantitativo, en el que se estableció el tipo de correlación entre distintas variables, tales como, la población total, el número de viviendas con energía eléctrica, con la variable consumo en kWh/mes en el periodo 2005-2018. Lo anterior permitió identificar el modelo adecuado para estimar el consumo de energía eléctrica para el departamento en el periodo 2019-2039 en diversos escenarios.</p> <p>Como estrategia de investigación se usó el estudio de caso. Se analizó la información existente sobre la población y el consumo de energía eléctrica del departamento de Vichada en las bases del Sistema único de información (SUI), y DANE, junto con el consumo de energía eléctrica del 2018 del Departamento de Vichada según los resultados obtenidos en el proyecto PERS Orinoquía.</p> <p>Luego de tener la información recopilada, se diseñó y se aplicó un script de R Project por cada modelo de regresión, a los datos y se obtuvo modelos ajustados con los que se estimó los consumos de energía eléctrica en el periodo del 2005 al 2018 en el departamento. Los valores estimados se usaron para los análisis de varianza (ANOVA), cálculo del coeficiente de determinación o determinación ajustado y de residuales. Luego de determinar el modelo más adecuado, se procedió a estimar los consumos de energía eléctrica para el periodo del 2019-2039 del departamento, con el modelo que mejor se ajustó a los datos recolectados.</p> <p>En este proyecto se encontró que el consumo de energía eléctrica para el periodo 2019-2039 está entre 2201982 kWh/mes y 3117186 kWh/mes; pero para que toda la población tenga acceso a la energía eléctrica se requiere una</p>
--	--

	<p>producción de energía de 4818341.36 kWh/mes para el año 2019 y 6820975.93 kWh/mes para el año 2039.</p>
7. FUENTES	<p>Cuadras, C. Nuevos métodos de análisis multivariante. Barcelona: CMC Editions. 2014</p> <p>Equipo PERS Orinoquia. Diagnostico de la generación eléctrica en las Zonas No Interconectadas de la Orinoquia. Universidad de los Llanos, Villavicencio, Meta. 2019</p> <p>Llinás, H. J. Precisiones en la teoría de los modelos logísticos. Revista Colombiana de Estadística, 29(2), 239-265. 2006</p> <p>Martínez Bencardino, C. Estadística y muestreo (13 ed.). Bogotá D.C.: Ecoe ediciones. 2012</p> <p>Montgomery, D. C., Peck, E. A., & Geoffrey, V. G. Introducción al análisis de regresión lineal (3 ed.). México D.F.: Compañía editorial continental. 2006</p> <p>PERS_Orinoquía. Planes de energización rural sostenibles – pers, región orinoquia (arauca, casanare, meta vichada). Energías renovables: una opción para el desarrollo rural productivo. Universidad de los Llanos, Villavicencio. 2016</p>
8. CONTENIDOS	<p>INTRODUCCIÓN</p> <ol style="list-style-type: none"> 1. MARCO REFERENCIAL 2. MATERIALES Y MÉTODOS 3. RESULTADOS Y ANALISIS DE RESULTADOS 4. CONCLUSIONES 5. RECOMENDACIONES <p>BIBLIOGRAFÍA</p>

	ANEXOS
<p>9. METODOLOGÍA</p>	<p>ENFOQUE METODOLÓGICO</p> <p>El trabajo se sustentó bajo un enfoque cuantitativo, ya que buscó establecer el tipo de correlación entre distintas variables, tales como, la población total, el número de hogares que tienen acceso a energía eléctrica y el tiempo en el que tienen acceso a esa energía con la variable consumo de energía eléctrica en el periodo 2005-2018.</p> <p>ESTRATEGIA DE INVESTIGACIÓN</p> <p>Se usó el estudio de caso como estrategia de investigación. Se analizó la información existente sobre la población y el consumo de energía eléctrica del departamento de Vichada en las bases de datos del, DANE y del Sistema único de información (SUI), junto con el consumo de energía eléctrica del 2018 del Departamento de Vichada establecido por PERS Orinoquía.</p> <p>El departamento de Vichada está situado en el extremo oriental de Colombia, localizado entre los 06°19'34" y 02°53'58" de latitud norte y 67°25'1" y 71°7'10" de longitud oeste. Cuenta con una superficie de 98.970 km² lo que representa el 8.6 % del territorio nacional. Limita por el Norte con el río Meta que lo separa de los departamentos de Casanare, Arauca y la República de Venezuela; por el Este con el río Orinoco que lo separa de la República de Venezuela, por el Sur con el río Guaviare que lo separa de los departamentos de Guainía y Guaviare y por el Oeste con los departamentos de Meta y Casanare. Este departamento se encuentra conformado por los municipios de Cumaribo al sur del departamento, La primavera al norte, Puerto Carreño al noreste y Santa Rosalía al norte.</p>

	<p>El alcance de este estudio fue correlacional ya que se tuvo como hipótesis las existencias de correlación entre cada una de las siguientes variables a saber, la población total, el número de hogares que tienen acceso a energía eléctrica y el tiempo con el que cuentan con energía eléctrica estos hogares con la variable el consumo de energía eléctrica, pero se buscó establecer qué tipo de correlación existe entre estas variable ya establecida con el consumo de energía eléctrica, si esta responde a un modelo lineal, exponencial, logístico, entre otros o una combinación de algunos de ellos.</p> <p>Como herramientas estadísticas se utilizó los modelos de regresión, como el lineal simple, el múltiple, el exponencial y el poblacional logístico, el ANOVA, el cálculo del coeficiente de determinación y determinación ajustado y el análisis de residuales.</p> <p>Como herramienta tecnológica se usó el software de uso libre R Project. Se creó un script en un archivo “.R” para cada modelo de regresión y sus respectivos ANOVA, cálculo del coeficiente de determinación o determinación ajustado, según fue el caso y el análisis de residuales.</p> <p>TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN.</p> <p>Para establecer qué tipo de correlación existente entre la población total, se recopiló información de documentación de mediciones del DANE, SUI y PERS Orinoquia de la siguientes forma, la información correspondiente al número de habitantes, la cantidad de hogares o personas, en promedio, por vivienda en la base de datos del DANE, el número de viviendas que tienen acceso a energía eléctrica, el tiempo en el que tienen acceso a esa energía y el consumo de energía eléctrica en las bases de datos del SUI, junto con el consumo de energía eléctrica del 2018</p>
--	---

del Departamento de Vichada establecido por PERS Orinoquía.

ANÁLISIS Y SISTEMATIZACIÓN

Luego de tener la información recopilada, se creó y aplicó un script de R Project, por cada modelo a los datos de la población de los censos 1985, 1993, 2005 y 2018 y con estos modelos ajustados se estimó la población correspondiente a cada uno de los 4 censos ya nombrados. Estos valores estimados se usaron para realizar el análisis de varianza (ANOVA), el cálculo del coeficiente de determinación o determinación ajustado y el análisis de residuales. A partir de los resultados de estos análisis se determinó el modelo que mejor se ajusta a los datos de los censos del DANE. Con el modelo seleccionado se estimó la población del departamento en el periodo 2003-2016.

Teniendo la información completa de cada una de las variables, se creó un script de R Project por cada modelo. Con los modelos de regresión se estimó los consumos de energía eléctrica en el periodo del 2003 al 2018 en el departamento. Estos valores estimados se usaron para los análisis de varianza (ANOVA), del coeficiente de determinación o determinación ajustado y de residuales. Estos análisis se utilizaron para determinar el modelo que mejor se ajusta a los datos del SUI.

Luego se determinó el modelo más adecuado y se procedió a estimar los consumos de energía eléctrica para 3 escenarios futuros (pesimista, optimista y transicional), para el periodo 2019-2039 del departamento del Vichada. El escenario pesimista corresponde al consumo que tendría el departamento si se mantiene estable el porcentaje de la población que tiene acceso al servicio de energía eléctrica. El escenario optimista corresponde al consumo que tendría el departamento si el 100 % de la población llega a tener acceso al servicio de

	<p>energía en los primeros años del periodo 2019-2039. El escenario transicional corresponde al consumo si el porcentaje de la población que tiene acceso al servicio de energía eléctrica va incrementando gradualmente hasta llegar al 100 % para el año 2039.</p>
<p>10. CONCLUSIONES</p>	<p>Se determinó la población del departamento de Vichada en los años 1985, 1993, 2005 y 2018 a partir de los datos del DANE 2005 y 2019 y por medio de un modelo lineal simple estimó la población para el periodo 2003-2016.</p> <p>Se obtuvo el consumo de energía eléctrica por mes desde el año 2003 hasta el año 2017 de la base de datos del SUI, pero no se tuvo en cuenta el consumo mensual promedio del año 2017 ya que la información correspondiente a este año estaba incompleta. También, del SUI, se estableció la cantidad de viviendas que tienen acceso a la energía eléctrica en promedio en el periodo 2003-2017.</p> <p>A partir de la matriz de correlación, se estableció que los parámetros que influyen en el consumo de energía eléctrica son la población, la cantidad de viviendas que tienen acceso al servicio de energía eléctrica y el año; con la matriz de dispersión se encontró que estas variables también están correlacionadas entre sí, en consecuencia, se escogió solo el año para la aplicación de los modelos, al consumo de energía eléctrica para evitar problema de multicolinealidad.</p> <p>A partir de los análisis realizados, se determinó que el modelo que debe ser usado para estimar el consumo de energía eléctrica en el departamento del Vichada para el periodo 2019-2039 es el poblacional logístico o modelo de Verhulst, puesto que los datos estimados por medio de este modelo representan un crecimiento del consumo de energía eléctrica consistente con los datos histórico del SUI.</p>

	<p>Se logró estimar el consumo de energía eléctrica mediante el uso del modelo Poblacional Logístico para el departamento del vichada durante el periodo 2019-2039 estableciendo tres escenarios futuros a saber, el pesimista, que corresponde a los datos estimados directamente del modelo lo que representa un promedio del 45,7 % de la población con acceso al servicio de energía eléctrica, el optimista, que corresponde al consumo que tendría el departamento del Vichada si toda la población (100 %) tuviese acceso al servicio de energía eléctrica y el transicional, que corresponde a un aumento gradual del porcentaje de la población con acceso al servicio de energía eléctrica, partiendo de 45,7 % hasta llegar al 100 % para el 2039.</p>
--	---

