Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería Mecánica Industrial

DISEÑO DE INVESTIGACIÓN PARA EL DESARROLLO DE UN ANÁLISIS DE SERIES TEMPORALES Y ESTIMACIÓN DE CORRELACIÓN ENTRE LA POBLACIÓN ESTUDIANTIL Y LA CANTIDAD DE CENTROS EDUCATIVOS PÚBLICOS PARA CUANTIFICAR LA SOBREPOBLACIÓN EN LOS NIVELES DE EDUCACIÓN BÁSICA DEL MUNICIPIO DE VILLA NUEVA, DEPARTAMENTO DE GUATEMALA

Luis Fernando Tojin Us
Asesorado por Msc. Ing. Javier Fidelino Garcia Tetzaquic

Guatemala, enero de 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

DISEÑO DE INVESTIGACIÓN PARA EL DESARROLLO DE UN ANÁLISIS DE SERIES TEMPORALES Y ESTIMACIÓN DE CORRELACIÓN ENTRE LA POBLACIÓN ESTUDIANTIL Y LA CANTIDAD DE CENTROS EDUCATIVOS PÚBLICOS PARA CUANTIFICAR LA SOBREPOBLACIÓN EN LOS NIVELES DE EDUCACIÓN BÁSICA DEL MUNICIPIO DE VILLA NUEVA, DEPARTAMENTO DE GUATEMALA

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

LUIS FERNANDO TOJIN US
ASESORADO POR MSC. ING. JAVIER FIDELINO GARCIA TETZQUIC

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO INDUSTRIAL

GUATEMALA, ENERO DE 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martinez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Inga. Nora Leonor Garcia Tobar
EXAMINADOR	Inga. Priscila Yohana Sandoval Barrios
SECRETARIO	Ing. Hugo Humerto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

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

DISEÑO DE INVESTIGACIÓN PARA EL DESARROLLO DE UN ANÁLISIS DE SERIES TEMPORALES Y ESTIMACIÓN DE CORRELACIÓN ENTRE LA POBLACIÓN ESTUDIANTIL Y LA CANTIDAD DE CENTROS EDUCATIVOS PÚBLICOS PARA CUANTIFICAR LA SOBREPOBLACIÓN EN LOS NIVELES DE EDUCACIÓN BÁSICA DEL MUNICIPIO DE VILLA NUEVA, DEPARTAMENTO DE GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería Mecánica Industrial, con fecha 10 de noviembre de 2022.

Luis Fernando Tojin Us

EEPFI-PP-1806-2022
Guatemala, 10 de noviembre de 2022

Director
César Ernesto Urquizú Rodas
Escuela Ingenieria Mecanica Industrial
Presente.

Estimado Ing. Urquizú

Reciba un cordial saludo de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería.

El propósito de la presente es para informarle que se ha revisado y aprobado el Diseño de Investigación titulado: **ANÁLISIS DE SERIES TEMPORALES Y ESTIMACIÓN DE CORRELACIÓN ENTRE LA POBLACIÓN ESTUDIANTIL Y LA CANTIDAD DE CENTROS EDUCATIVOS PÚBLICOS PARA CUANTIFICAR LA SOBRE POBLACIÓN EN LOS NIVELES DE EDUCACIÓN BÁSICA DEL MUNICIPIO DE VILLA NUEVA**, el cual se enmarca en la línea de investigación: **Todas las áreas - Pronósticos**, presentado por el estudiante **Luis Fernando Tojin Us** carné número **200915339**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la Maestría en ARTES en Estadistica Aplicada.

Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

Atentamente,

"Id y Enseñad a Todos"

Mtro. Javier Fidelino García Tetzaguic
Asesor(a)

Mtro. Edwin Adalberto Bracamonte Orozco
Coordinador(a) de Maestría

UNIVERSIDAD DE SAN CARLOS
ESCUELA DE POSTGRADO
FACULTAD DE INGENIERIA
DE GUATEMALA

Javier Fidelino García Tetzaguic
Ingeniero Mecánico Industrial
Colegiado No. 14190

Mtro. Edgar Dario Alvaréz Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERIA
DIRECCIÓN
ESCUELA DE POSTGRADO

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

FACULTAD DE INGENIERÍA

EEP-EIMI-1456-2022

El Director de la Escuela Ingenieria Mecanica Industrial de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **ANÁLISIS DE SERIES TEMPORALES Y ESTIMACIÓN DE CORRELACIÓN ENTRE LA POBLACIÓN ESTUDIANTIL Y LA CANTIDAD DE CENTROS EDUCATIVOS PÚBLICOS PARA CUANTIFICAR LA SOBRE POBLACIÓN EN LOS NIVELES DE EDUCACIÓN BÁSICA DEL MUNICIPIO DE VILLA NUEVA**, presentado por el estudiante universitario **Luis Fernando Tojin Us**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería en esta modalidad.

ID Y ENSEÑAD A TODOS

UNIVERSIDAD DE SAN CARLOS
DIRECCION
Escuela de Ingeniería Mecánica Industrial
FACULTAD DE INGENIERIA

Ing. César Ernesto Urquizú Rodas
Director
Escuela Ingenieria Mecanica Industrial

Guatemala, noviembre de 2022

Decanato
Facultad de Ingeniería
24189101- 24189102
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.078.2023

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería Mecánica Industrial, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN PARA EL DESARROLLO DE UN ANÁLISIS DE SERIES TEMPORALES Y ESTIMACIÓN DE CORRELACIÓN ENTRE LA POBLACIÓN ESTUDIANTIL Y LA CANTIDAD DE CENTROS EDUCATIVOS PÚBLICOS PARA CUANTIFICAR LA SOBREPOBLACIÓN EN LOS NIVELES DE EDUCACIÓN BÁSICA DEL MUNICIPIO DE VILLA NUEVA, DEPARTAMENTO DE GUATEMALA,** presentado por: **Luis Fernando Tojin Us,** después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DECANA
FACULTAD DE INGENIERÍA

Inga. Aurelia Anabela Cordova Estrada
Decana

Guatemala, enero de 2023

AACE/gaoc

ACTO QUE DEDICO A:

Dios	Por darme la sabiduría necesaria para cumplir las metas propuestas.
Mis padres	Diego Tojin y Rosario Us (q.d.e.p), por su apoyo, ejemplo y amor incondicional.
Mis hermanas	Silvia y Sandra Tojin, por motivarme para cumplir los objetivos.
Mis hermanos	Manuel, Marco, Segio Tojin, por estar estar siempre a mi lado animándome.
Mis sobrinos	Diego, Erick, Paulina, Jimena y Jeremy Tojin, por ser parte del éxito.
Maestra	Maria Calderón, por su paciencia al brindarme los estudios a mi temprana edad.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por brindarme la oportunidad de obtener la formación a nivel superior para lograr ser un profesional.
Facultad de Ingeniería	Por brindarme todos los conocimientos profesionales.
Mis amigos y compañeros	Por compartir cada momento y experiencia durante la carrera.
Catedráticos	Por su dedicación y alto profesionalismo en cada curso.
Ing. Javier Garcia	Por brindarme sus consejos y asesoria profesional.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
r_{xy}	Coeficiente de Correlacion de Pearson
cov	Covarianza
MAE	Error Absoluto Medio
RMSE	Error Cuadrático Medio
p_t	Error de Porcentaje
mi_{T+h}	Error de Pronóstico
FADEP	Familia Desarrollo Población
ACF	Función de Autocorrelacion
Qq-plot	Gráfico de Cuantiles
INE	Instituto Nacional de Estadística
K-S	Kolmogorov-Smirnov
MINEDUC	Ministerio de Educación
ARIMA	Modelo Autorregresivo Integrado Promedio Movil
nc	Número de Clases
$\overline{y}$	Promedio de la Variable Dependiente
$\overline{x}$	Promedio de la Variable Independiente
q	Prueba Box-Pierce
q^*	Prueba de Ljung-Box
ε_t	Ruido Blanco
y_i	Variable Aleatoria Dependiente
x_i	Vairable Aleateria Independiente
s^2	Varianza

GLOSARIO

Cuartiles Son los tres elementos de un conjunto de datos ordenados que dividen el conjunto en cuatro partes iguales.

Curva normal También llamada curva de Gauss, es una función de probabilidad continua, simétrica, cuyo máximo coincide con la media y que tiene dos puntos de inflexión situados a ambos lados de la media.

Regla de Sturger Es una regla que sirve para calcular el número, clases o intervalos idóneo en los que se debe dividir un conjunto de datos.

Test de normalidad Conjunto de pruebas que evalúan si los datos analizados se adaptan a distribuciones normales o conocidas.

RESUMEN

En el presente trabajo se describe y se detalla la metodología para el desarrollo de un diseño de investigación, que tiene como fin determinar la correlacion entre la capacidad de los centros educativos habilitados en el municipio de Villa Nueva y la población estudiantil y de como será la situación en los próximos años a través de un análisis de series temporales.

En la primera sección están los antecedentes, que son investigaciones previas relacionadas al tema y que brindan fundamentos para el planteamiento del problema, que a su vez permiten la definición de objetivos. En la justificación se detallan el tipo de investigación y las entidades que están involucradas.

En la siguiente sección del trabajo se encuentran las necesidades a cubrir y el esquema de solución, lo que permite describir la problemática y los procedimientos estadísticos a utilizar, basados en los tópicos descritos en el marco teórico. Finalmente se encuentra la metodología, que describe la naturaleza del estudio y las variables a utilizar; asimismo se detallan las fases de la investigación.

1. INTRODUCCIÓN

El presente estudio de investigación es una sistematización del proceso de la relación de población estudiantil del nivel básico y los centros educativos públicos habilitados para la cobertura de dicha demanda estudiantil.

Algunos centros educativos públicos del municipio de Villa Nueva reciben más estudiantes de lo que están aptas sus instalaciones lo que se convierte en una problemática que año con año experimentan, no se tiene la relación de la cantidad de centros educativos que deben operar en la actualidad en función a la población estudiantil.

La importancia de realizar un análisis del nivel de sobrepoblación es fundamental para crear bases que ayuden a las instituciones correspondientes a generar nuevos estudios e implementar nuevas medidas que contribuyan a la educación del sector. Poder medir y evaluar la situación, contribuye a generar las bases de nuevas políticas y una mejor planificación.

La metodología de la investigación utilizada en este estudio es de enfoque cuantitativo con un diseño no experimental u observacional y un alcance descriptivo correlacional.

Los aportes esperados en este estudio beneficiarán a los estudiantes que estén próximos a iniciar la etapa de educación básica, los maestros de dicho nivel, tendrán una cantidad idónea para trabajar.

Este estudio tendrá la fase de revisión de literatura, en la que se revisarán tópicos de pruebas de normalidad, correlación lineal, análisis de series temporales; fase de recolección y análisis de información; en esta fase se hará la recolección, análisis e interpretación de datos, una vez teniendo la información se interpretará la naturaleza sometiendo a diversas pruebas estadísticas y, por último, la elaboración de informe final.

Es un estudio factible, porque cuenta con los recursos humanos, financieros y tecnológicos, necesarios para su ejecución.

En el Capítulo I, se hará el Marco Referencial en el que se revisarán estudios previos sobre temas relacionados a la investigación, que brindarán fundamentos para su desarrollo. En el Capítulo II, se desarrollará el marco teórico, que describirá los tópicos de la estadística en general; posteriormente conceptos de análisis de normalidad y series temporales. En el Capítulo III, se expondrán los resultados obtenidos sobre el análisis de normalidad, correlación y series temporales. Finalmente en el Capítulo IV, se hará la discusión de resultados de cada análisis realizado para elaborar las conclusiones y recomendaciones respectivas.

2. ANTECEDENTES

En diversas regiones y niveles en la educación pública del país se experimenta una demanda que supera la capacidad que ofrece el sistema educativo público, esto afecta de manera directa la calidad educativa y el desempeño académico del estudiante.

A nivel de educación superior, la Universidad de San Carlos siendo la única pública en el país, experimenta gran demanda por parte de la sociedad guatemalteca, Solórzano (2017), expone las estadísticas de la cantidad de estudiantes que se atienden por aula y cómo las clases no personalizadas dificultan el aprendizaje, tanto para el estudiante como para el docente, en la que sin duda esta situación exige redoblar esfuerzos para lograr el objetivo de transmitir conocimientos.

En el estudio realizado en la Facultad de Humanidades en la Universidad de San Carlos, Solórzano (2017), analiza el impacto que tiene la sobrepoblación en el estudiantes y docente. En el estudio selecciona un estrado de la población y extrae una muestra con el método aleatorio simple de estudiantes y docentes quienes son encuestados y entrevistados, respectivamente, y define las variables a medir, posteriormente plantea las hipótesis nulas y alternativas para concluir los hallazgos del estudio, que afirman la deficiencia en la calidad del aprendizaje.

Ahora bien, al estar en el nivel de educación superior e ir los a niveles básico y primario, esta situación dificulta aún más el proceso de aprendizaje,

puesto que el estudiante a este nivel necesita una atención más personalizada, López (2016) indica: “Las aulas repletas de estudiantes, lejos de ser una satisfacción para maestros y motivación para los escolares, se convierte en una barrera para el adecuado aprendizaje”. Su estudio se enfoca en manifestar las dificultades que atraviesan las estudiantes derivadas de la falta de atención de los docentes.

En 2019 el Centro de Investigación Económicas Nacionales realizó un informe sobre el diagnóstico del sistema educativo y enmarcó un panorama muy amplio en el que expone el deficiente nivel de aprendizaje de distintas materias y marcos legales; el objetivo del artículo es ser una guía que contribuyan a generar nuevas políticas para la mejora en a la educación a nivel nacional.

El MINEDUC a partir del año 2014 ha generado una base de datos con cifras e indicadores del sistema nacional de educación de los niveles pre-primario a diversificado. La dinámica consiste en que cada centro educativo privado y público; urbano y rural cargue cierta información, la última fue en mayo de 2022. El objetivo de esta data es tener una fuente de información al alcance de la sociedad guatemalteca o internacional, que decida realizar estudios estadísticos y necesite información histórica y del último año cerrado. Este anuario es básicamente una plataforma que permite descargar información en formatos fáciles de moldear con análisis estadístico.

De manera paralela a los datos que muestra el MINEDUC, existe una organización formada por un grupo de empresarios que buscan lograr un cambio en la educación del país. En su portal *web* dan a conocer sus proyectos, metas y objetivos, pero la parte que resalta son los indicadores, que son informes que han elaborado por medio de auditorías sociales en trabajos de campo, que muestran datos del sistema nacional desde otro punto de vista. Esta data permite

hacer un comparativo de la información oficial del Estado contra la información social.

El fondo de Población de las Naciones Unidas a través del censo realizado en 2020 a nivel nacional, exponen variedad de datos estadísticos relacionados con la juventud guatemalteca; es un informe completo que, entre varios temas, manifiestan los rangos de escolaridad de distintos niveles y razones de deserción. El informe proporciona un panorama completo sobre datos en el ámbito de realidad educativa.

Con los notables índices de deserción escolar en los jóvenes guatemaltecos, la Agencia de Estados Unidos para el Desarrollo Internacional, ha fomentado en institutos por cooperativa de nivel básico, programas para incentivar al estudiante. Por medio de la recopilación de datos a través de encuestas en ciertos sectores rurales han implementados programas que se adecúan a la necesidad de aprendizaje de la población que también los motiva a continuar sus estudios.

Un factor importante para considerar en el ámbito de educación es la relación docente-estudiante, es decir, la cantidad de estudiantes que es capaz de atender un docente. En el estudio realizado por la Universidad Nacional de Educación de Ecuador (2020), por medio la comparación de parámetros utilizados en América Latina y varios países de Europa y África, se busca una relación matemática y estadística que brinde un punto de equilibrio que tome en cuenta calidad educativa y optimice recursos. Asimismo, utilizan un sistema de proyección para calcular la cantidad de docentes hacia los siguientes 10 años. El estudio concluye que un docente debe atender a no más de 15 estudiantes, para garantizar una educación de calidad y que serán necesarios 221,887 docentes para atender la demanda de población estudiantil.

FADEP (2019), explica los datos estadísticos del INE y de la UNESCO a través de diagramas de dispersión con datos de los últimos 30 años muestran la tendencia de la deserción escolar en el país. Explican las barreras que cada nivel educativo conlleva.

Zhuji World (2022), muestra una serie de estadísticos y clasificación de la población del municipio de Villa Nueva, realiza una previsión de la población que tendrá la localidad en los siguientes años hasta el año 2100. En el portal de Datos Mundial se muestra un análisis similar, pero tomando en cuenta la tasa de natalidad y mortalidad.

En los estudios realizados, se busca expresar los obstáculos que conlleva la educación pública en el país a todos niveles. Los artículos tienen un punto en común, que para el proceso de aprendizaje, se necesita un ambiente y políticas adecuadas para el estudiante que sea acorde en calidad, recursos y espacios.

Los antecedentes ayudarán a tener un panorama claro de lo que ha sido, es y será esta problemática, Es necesario analizar la problemática desde el punto de vista estadístico, para conocer la correlación entre la población estudiantil y la capacidad de los centros educativos.

3. PLANTEAMIENTO DEL PROBLEMA

3.1. Contexto general

En diversos centros educativos del país, la capacidad disponible es insuficiente para atender estudiantes, lo que se conoce como sobrepoblación estudiantil. Esta problemática ha surgido por diversos factores como la tasa de natalidad del país, proyectos mal planificados por los gobiernos en turno, falta de proyección, entre otras razones.

3.2. Descripción del problema

En el municipio de Villa Nueva los establecimientos educativos de educación básica y primaria experimentan sobrepoblación de estudiantes. No se conoce la cantidad de establecimientos que deben operar para cumplir la demanda de alumnos, asimismo debido a la falta de planificación por parte de las autoridades educativas de cada gobierno no hay una proyección que permita definir la cantidad estimada de centros educativos que deban habilitar en función al crecimiento poblacional, tasa de natalidad, entre otros factores.

3.3. Formulación del problema

Pregunta central

¿Cuál es el nivel de correlación que tienen la población estudiantil y la capacidad de centros educativos, y cuál es el modelo de series temporales

adecuado para pronosticar la situación en el futuro y proyectar la cantidad optima de centros educativos?

Preguntas auxiliares

- ¿Cuál es el nivel de correlación que existe entre la población actual de estudiantes y la capacidad de los centros educativos públicos?

- ¿Qué pronostica el análisis de series temporales sobre la sobrepoblación estudiantil en los siguientes años?

- ¿Cuál es el pronóstico para la cantidad de centros educativos idóneo para la demanda de alumnos en el futuro?

4. JUSTIFICACIÓN

La línea de investigación en que se desarrollará este estudio es de Pronósticos ya que se analizará la demanda de estudiantes antes los centros educativos públicos de nivel básico a futuro, con base al histórico de datos de los últimos cinco años.

Es importante que se determine la relación entre las variables, población estudiantil y centros educativos, para tener una cantidad idónea de estudiantes por centros educativos.

Brindar un estudio con bases estadística que genere las bases de nuevas políticas de planeación a las autoridades correspondientes para que en el futuro el sistema educativo tenga un balance entre cantidad de estudiantes y centros educativos.

Los beneficiados con este estudio serán los alumnos, maestros, padres de familia, autoridades educativas en turno; puesto que con ello se busca no exceder sobrepasar los límites de las instalaciones de los centros educativos y la capacidad de enseñanza de los docentes.

A nivel social, es satisfactorio que los alumnos cuenten con un ambiente idóneo para su educación. Por lo que es necesario tener un análisis que permita pronosticar la dimensión de la población en los próximos años para fundamentar la apertura de nuevos centros educativos.

5. OBJETIVOS

5.1. General

Desarrollar un análisis de series temporales y estimación de correlación entre la población estudiantil y la cantidad de centros educativos públicos para cuantificar la sobrepoblación en los niveles de educación básica del municipio de Villa Nueva.

5.2. Específicos

- Construir un modelo estadístico de regresión lineal para evaluar la correlación entre las variables (independiente) la población estudiantil actual y (dependiente) la capacidad de los centros educativos públicos habilitados, para explicar la sobrepoblación estudiantil.

- Realizar un análisis de series temporales que evalué la tasa de crecimiento poblacional y el número de centros educativos que habilita el Mineduc anualmente para medir la sobrepoblación en los siguientes años.

- Estimar por medio de series temporales el pronóstico de centros educativos que el Mineduc deberá habilitar para cubrir la demanda de estudiantes y así mitigar la sobrepoblación de los centros educativos.

6. NECESIDADES POR CUBRIR Y ESQUEMA DE SOLUCIÓN

A través del estudio a realizar se proporcionará un informe que cubre la necesidad de exponer la problemática de sobrepoblación que sufren algunos centros educativos públicos de nivel básico, lo que ha sido una situación que experimenta la sociedad guatemalteca en cada inicio de ciclo escolar, y que si no se toman las políticas y medidas necesarias persistirán y podrían empeorar en el futuro.

Para cubrir la necesidad se utilizan conceptos estadísticos para lo cual el análisis de la información se hará a través de variables de tipo cualitativo de escala de razón, siendo la variable independiente la población estudiantil y la variable dependiente la capacidad de estudiantes que puede cubrir los centros educativos, los datos serán extraídos de fuentes oficiales como lo son INE y MINEDUC, se realizar pruebas de normalidad, como Histograma, Grafico de comparación de cuantiles (qqplot) y Kolmogorov-Smirnov, porque no se conoce la naturaleza la información.

Una vez realizadas estas pruebas y si los datos son normales se estimará el coeficiente de Pearson; en caso de no cumplir la normalidad de los datos se estimará el coeficiente de Tau de Kendall. Lo que se desea calcular, es el grado de asociación entre la población estudiantil y la capacidad de los centros educativos, definir si es acorde el sistema educativo en cubrir la demanda estudiantil.

Basados en la información histórica, se realizará un modelo de series temporales específicamente el cálculo de Tendencia determinista, el crecimiento

poblacional estudiantil es la variable que necesita ser estudiada, pues marca la pauta de cómo preparar el sistema educativo, se definirá la tendencia, el comportamiento estacional, componente cíclico y el error.

7. MARCO TEÓRICO

7.1. Estadística

Es un grupo de técnicas que se utilizan para recopilar, analizar, interpretar y presentar datos masivos, Contento (2019), la cual se divide en estadística descriptiva e inferencial.

La estadística descriptiva se encarga de recolectar datos y utiliza métodos para describir de manera correcta las variables en la que se está interesado. Mientras que la estadística inferencial utiliza métodos que apoyen a la emisión de conclusiones sobre las características de una población en particular por medio de juicios de carácter probabilístico.

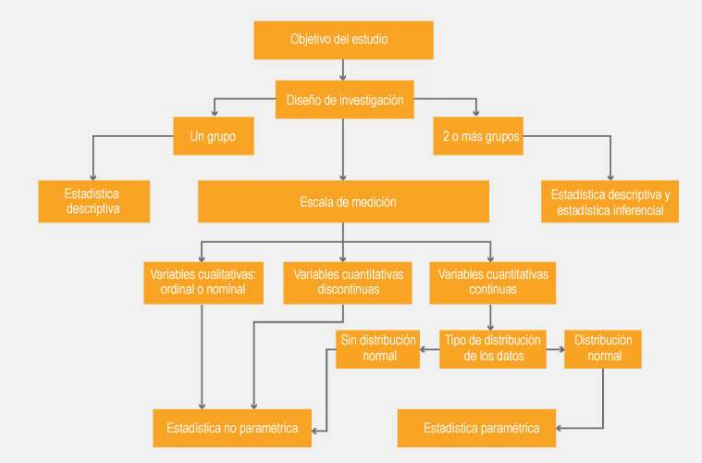
7.1.1. Estadística paramétrica y no paramétrica

Estadística paramétrica, es la rama de la estadística inferencial en la que los datos del fenómeno estudiado tienen un comportamiento de una distribución normal o conocida. Como lo indica López (2019) es la parte de la estadística que utiliza pruebas estadísticas y criterios de toma de decisiones basados en distribuciones conocidas.

Estadística no paramétrica, es una rama de la estadística inferencial, menciona López (2019), en la cual los cálculos y procedimientos son de distribuciones desconocidas o distribuciones libres. Ambas ramas de la estadística son complementarias y una vez identificado el tipo de distribución que tengan los datos de la muestra extraída, así serán los cálculos estadísticos para

trabajar. Según el tipo de variable que tenga el fenómeno estudiado, así será el análisis.

Figura 1. **Division de la estadística**



Fuente: Universidad Autónoma del Estado de Hidalgo (2017)

Es necesario realizar pruebas de normalidad a los datos, para tener la certeza de trabajar con determinadas distribuciones (paramétrico o no paramétrico), para ello se tiene:

7.1.2. Pruebas de normalidad

Son contrastes que permiten analizar la información previa a cualquier análisis estadístico, estas también son llamadas pruebas de bondad de ajuste. Estas pueden ser gráficas, analíticas y pruebas de hipótesis.

7.1.2.1. Histograma

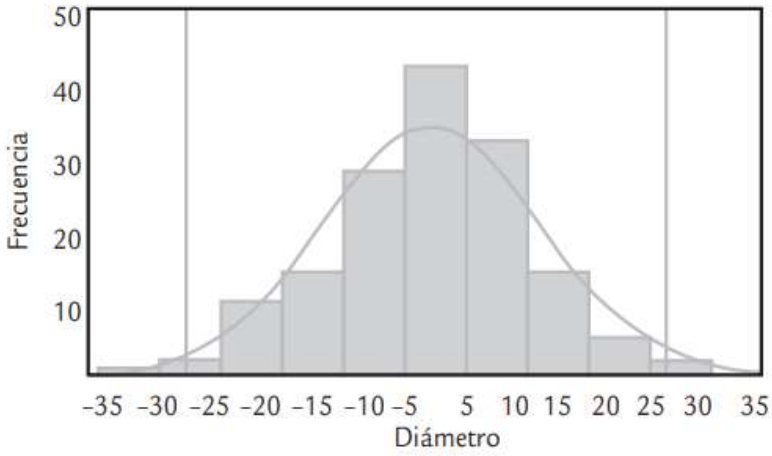
Es una prueba gráfica y nos muestra un primer panorama del comportamiento de los datos extraídos de la población. De manera visual nos

permite analizar si los datos se ajustan a una distribución normal, la sencillez de este método lo hace de los más utilizados, el inconveniente es que es subjetivo, pues su interpretación depende del observador. Saldaña (2016).

“Representación gráfica de la distribución de un conjunto de datos o de una variable, donde los datos se clasifican por su magnitud en cierto número de clases. Permite visualizar la tendencia central, la dispersión y la forma de la distribución.” (Gutiérrez 2013).

Al histograma es conveniente agregar la curva de normalidad ajustada a la serie de datos para tener un mejor panorama, tal como se ve en la gráfica.

Figura 2. **Histograma con curva normal teórica**



Fuente: Gutiérrez, (2016)

Para crear un histograma, se trabaja previamente una tabla de frecuencias, conformada por intervalos o clases que abarcan el rango de datos, posteriormente se verifican cuantos datos entran en cada uno. Para definir la cantidad de clases se puede utilizar la regla de Sturges:

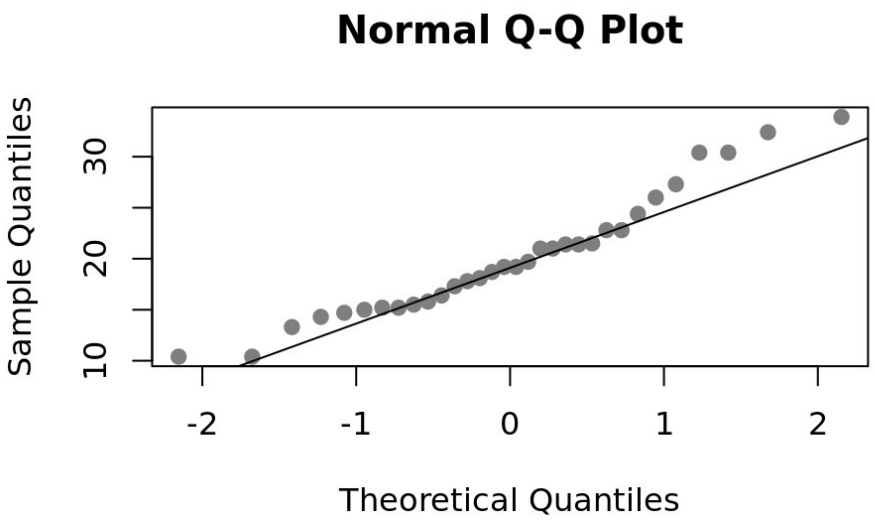
$$NC = 1 + 3.3 * log_{10}(numero\ de\ datos) \qquad \text{(Ec. 01)}$$

7.1.2.2. Gráfico de cuartiles

Este análisis gráfico realiza una comparación de los cuantiles de una distribución de datos y la distribución ideal (normal), utilizando la misma media y desviación estándar. Kaminsky (2008).

Estos gráficos de probabilidad buscan demostrar gráficamente la normalidad de los datos estudiados, consiste en que los valores observados vayan lo más apegados a su pareja con valor ideal (normal). Los valores ideales, se muestran en una línea recta y los observados, deben apegarse a esta línea recta. Saldaña (2016).

Figura 3. **Gráfico de cuartiles**



Fuente: Amat, (2016)

7.1.2.3. Test de Kolmogorov-Smirnov

Esta prueba se utiliza para variables continuas y compara la distribución acumulada de los datos con una distribución teórica determinada, como la Distribución normal, Poisson, Exponencial, entre otras. Rodó (2020).

Estadístico de prueba:

$$KS = \max |F_1(x) - F_2(x)| \quad \text{(Ec. 02)}$$

Donde:

x es el i-esimo valor observado en la muestra.

F_1 (Frecuencia acumulada observada) estimador de la probabilidad de observar valores distintos a x.

F_2 (Frecuencia acumulada teórica) es la probabilidad de observar valores menores o iguales que x.

El contraste de hipótesis de la prueba K-S:

$$Ho: Los\ datos\ siguen\ una\ distribucion\ normal$$

$$F_1(x) = F_2(x) \quad \text{(Ec. 03)}$$

Para todo $x(-\infty, +\infty)$ y $x \sim N(\mu, \sigma^2)$

$$Ha: Los\ datos\ no\ siguen\ una\ distritucion\ normal$$

$$F_1(x) \neq F_2(x) \quad \text{(Ec. 04)}$$

Para al menos una x.

7.1.2.4. Correlación

Es un método estadístico que analiza la relación entre dos variables, cuyo objetivo es cuantificar su nivel de asociación. Asimismo, es posible conocer si las variables tienen una relación directa o inversa, es decir, si mientras una crece la otra de igual manera lo hace, o si mientras una crece la otra decrece. El objetivo es establecer la dirección y nivel de asociación. Fallas (2012).

Hay una forma gráfica para visualizar la relación entre dos variables el cual es el Diagrama de Dispersión. Asimismo, hay una medida que nos respalda la observación que es Coeficiente de correlación Lineal de Pearson.

La correlación es definida en términos de la varianza s^2de las variables (x, y) asimismo de la covarianza cov de las mismas variables.

7.1.2.4.1. Varianza (s^2)

La varianza expresa el valor medio de la desviación de los datos tomando con referencia a la media.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{N - 1} \qquad \text{(Ec. 05)}$$

7.1.2.4.2. Covarianza $(cov\ x, y)$

La varianza expresa el valor medio de la desviación de los datos tomando con referencia a la media.

$$cov = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N - 1} \qquad \text{(Ec. 06)}$$

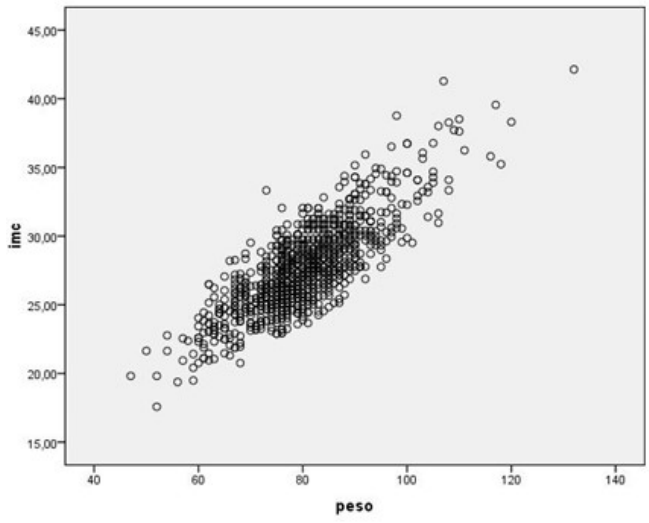
7.1.2.5. Diagrama de dispersión

Es la representación gráfica de dos variables graficadas en un plano cartesiano (x, y). Donde x es la variable independiente y y es la variable dependiente. En este se grafican los datos extraídos del fenómeno estudiado.

Este es el primer acercamiento que se tiene para ver el comportamiento de los datos, que pueden ser lineales, exponenciales, logarítmicas y otros.

Brinda un análisis preliminar y con ello se tiene un adelanto de que se puede esperar en el coeficiente de correlación de Pearson.

Figura 4. **Diagrama de Correlación**



Fuente: Instituto Aragonés de Ciencias de la Salud

7.1.2.6. Coeficiente de correlación lineal de Pearson

Este coeficiente se establece en términos de la covarianza de las variables x y y, el cual establece si los puntos tienen una tendencia en línea recta, el valor

que pude tomar esta entre -1 y 1. Para ello el investigador debe estar seguro de la normalidad de los datos, funcional para pruebas con datos paramétricos. Fallas (2012).

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad \text{(Ec. 07)}$$

Donde:

x_i Variable aleatoria, independiente

y_i Variable aleatoria, dependiente

$\bar{x}$ Promedio variable independiente

$\bar{y}$ Promedio de variable dependiente

7.1.2.6.1. Características del coeficiente (r)

El coeficiente r tiene características según su magnitud y su signo, a continuación, se detallará los rangos y dirección de este.

Según se la magnitud del coeficiente este tiene una interpretación, como se ve en la Tabla I:

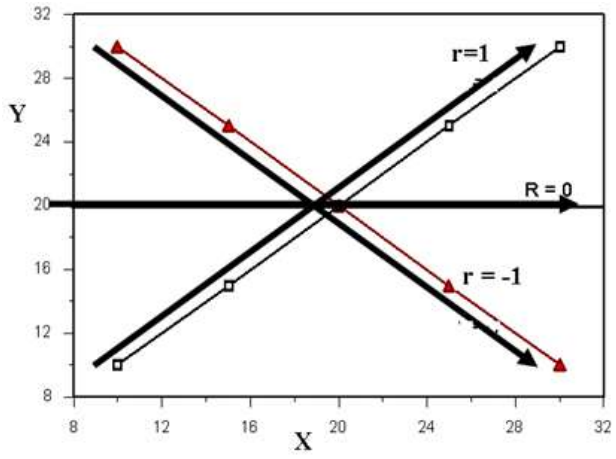
Tabla I. **Rangos coeficiente de correlación de Pearson**

Rango r_{xy}	Interpretación
$0.0 \leq \lvert r_{xy} \rvert < 0.1$	Correlación Nula
$0.1 \leq \lvert r_{xy} \rvert < 0.3$	Correlación Débil
$0.3 \leq \lvert r_{xy} \rvert < 0.5$	Correlación Moderada
$0.5 \leq \lvert r_{xy} \rvert < 1.0$	Correlación Fuerte

Fuente: Elaboración propia

Según el signo del coeficiente muestra la dirección de asociación de las variables, por ejemplo, si r es positivo la correlación es directa, esto quiere decir que si los valores de x aumentan de igual manera lo harán los valores de y. Ahora bien, si los valores de x aumentan y los valores de y disminuyen, o viceversa, el signo de r es negativo y la correlación inversa.

Figura 5. **Signo de Coeficiente r**



Fuente: Fallas (2012)

7.1.3. Pronósticos

Los métodos de pronósticos toman en cuenta diversos factores como el horizonte tiempo, los patrones en los datos entre otros. Existen métodos de pronósticos simples, también llamados ingenuos, además existen métodos complejos. En algunos casos no habrá datos históricos para consultar, esto origina los pronósticos de juicio. Algo que debe quedar claro es que la elección del método dependerá mucho de los datos con que se cuenta y de la previsibilidad que se desea pronosticar.

7.1.4. Métodos de pronósticos cualitativos

Este método surge cuando no se tiene datos o lo que se tiene son poco o nada relevantes para el fenómeno estudiado, y no es que sean conjeturas o algo similar, ya que cuentan con una estructura de análisis que los validan.

7.1.5. Métodos de pronósticos cuantitativos

Para este tipo de pronóstico el fenómeno debe cumplir lo siguiente:

- Existen datos históricos del fenómeno.
- Es lógico que el fenómeno vuelva a ocurrir en el futuro, es decir que vuelva a pasar.

Este tipo de pronóstico puede usar datos de series temporales o datos transversales, los cuales han sido recopilados en un tiempo corto.

Para realizar una labor de pronóstico hay cinco pasos básicos a seguir:

1. Definición del fenómeno o problema
2. Recopilación de datos
3. Análisis exploratorio
4. Elección y ajuste de modelos
5. Evaluación del modelo de pronóstico

El pronóstico puede ser tomado como una variable aleatoria, entonces obtenido que al ser calculado se está determinando un rango de posibles valores que podría tomar dicha variable, llamado Intervalo de Predicción. Existen también

más concretas que son puntos medios de intervalos, que se definen como Pronósticos Puntuales. Ahora bien, al grupo de valores junto con las probabilidades relativas, se le define como distribución de probabilidades o distribución de pronósticos.

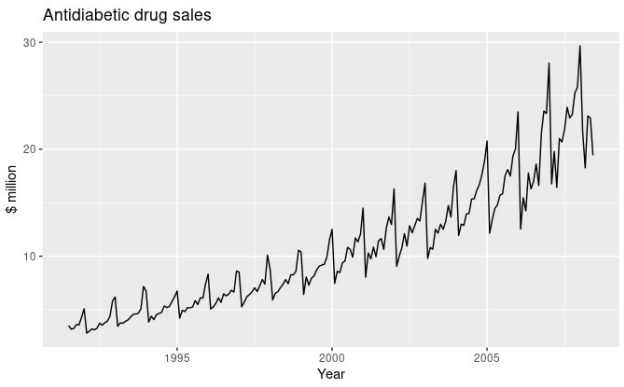
Las gráficas proporcionan una vista de los datos que se están trabajando, relación de variables y patrones o comportamientos. Con ello se puede tener una idea del método a trabajar.

7.1.6. Series de tiempo

Este tipo de datos puede tener tendencia y/o estacionalidad, y es importante distinguir una de la otra

- Tendencia: Existe cuando hay una disminución o aumento en tiempo prolongado en los datos. Es importante considerar que para considerar tendencia esta no debe ser lineal.

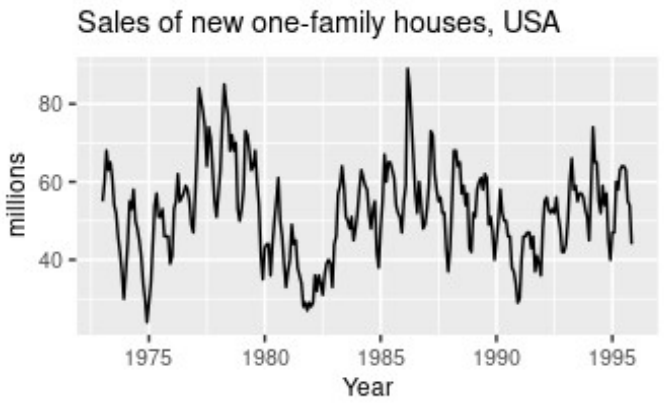
Figura 6. **Venta de medicamentos anti-diabetes**



Fuente: Hyndman (2018)

- Estacionalidad: surge cuando los datos de serie de tiempo se ven afectados por factores temporales como los días de la semana o los meses del año, o inclusive durante los años. La frecuencia dependerá del fenómeno y esta debe ser conocida y estable.

Figura 7. **Venta mensual de viviendas**



Fuente: Hyndman (2018)

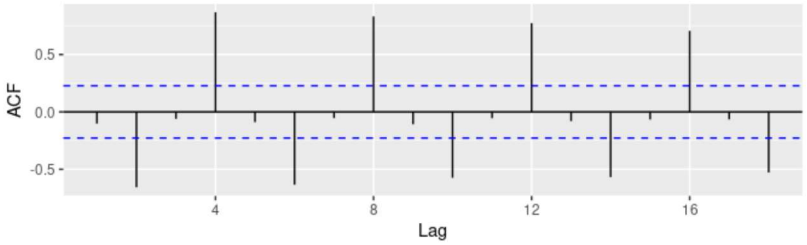
7.1.7. Autocorrelación

Mide la relación lineal existente entre los valores rezagados de una serie de tiempo la cual está definida por medio del coeficiente de correlación r_k:

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \qquad \text{(Ec. 08)}$$

Los coeficientes encontrados de una serie de tiempo se pueden graficar para indicar la función de autocorrelación (ACF) y poder crear un Correlograma.

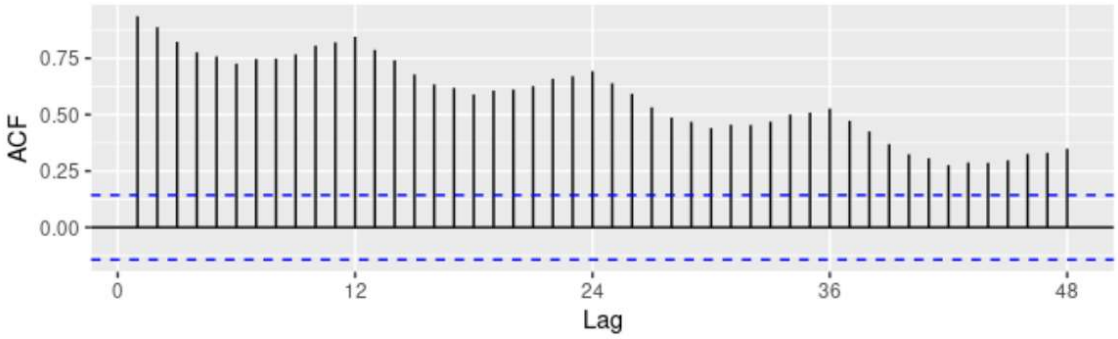
Figura 8. **Correlograma**



Fuente: Hyndman (2018)

Cuando la serie de datos tiene tendencia, las auto correlaciones son grandes y positivas ante retrasos pequeños. Se puede decir que los ACF de datos con tendencia tienden a mostrar resultados positivos que lentamente disminuyen conforme aumentan los retrasos. En la figura 9 se muestra ACF de la serie de datos de la serie de datos mostrados en la figura 8.

Figura 9. **Correlograma venta de medicamentos anti-diabetes**

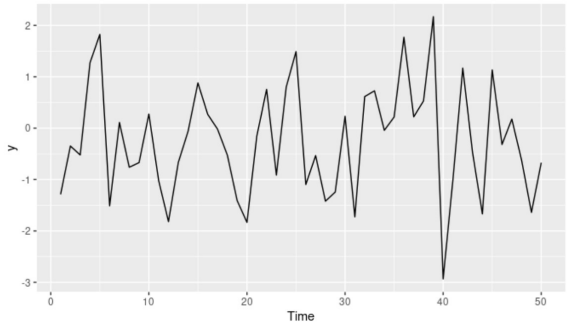


Fuente: Hyndman (2018)

7.1.8. Ruido blanco

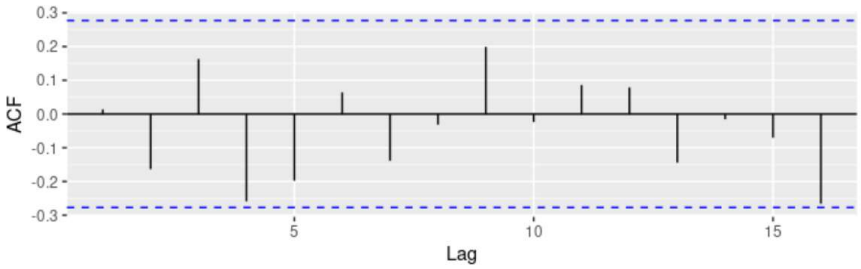
Las series con ruido blanco tiene la característica que las auto correlaciones sean muy cercanas a cero en su mayoría o al menos hasta en un 95 % de los picos.

Figura 10. **Ruido Blanco**



Fuente: Hyndman (2018)

Figura 11. **Correlograma ruido blanco**



Fuente: Hyndman (2018)

7.1.9. Transformaciones y ajustes

El fin de estas transformaciones es simplificar los patrones de los datos o hacer que sean más consistentes en la serie de datos. Es importante considerar que los datos más sencillos conllevan a pronósticos más exactos. Los ajustes para realizar pueden ser:

- Calendario: Debido a la variedad de cantidad de días por meses.
- Población: Considerar a las poblaciones en miles o millones o datos más concretos.

- Inflación: Para estos ajustes utilizar índices de precios.
- Transformaciones matemáticas: Las transformaciones más comunes son las logarítmicas puesto que son cambios relativos y de mejor interpretación. Existen otras transformaciones de potencias en la que están las raíces cuadradas y cubicas, que no son tan comunes ya que no son tan interpretables. Hay un grupo de transformaciones en las que se incluyen tanto de logaritmo y de potencias, y son las transformaciones Box-Cox las cuales dependen de λ:

$$w_t = \begin{cases} Iniciar\ sesión\ (y_t)\ si\ \lambda = 0; de\ lo\ contrario \\ (y_t^{\lambda} - 1/\lambda) \end{cases} \qquad \text{(Ec. 09)}$$

El logaritmo que utiliza la prueba Box-Cox es natural bajo la condición de $\lambda = 0$, en caso contrario utilizará una trasformación de potencias.

7.1.10. Diagnósticos residuales

Las observaciones de una serie de tiempo es posible pronosticarla usando todas las observaciones del pasado, lo que se denomina valores ajustados y se definen por $\hat{y}_{t|t-1}$, lo que indica que la previsión de y_t está basado en observaciones $y_1, \dots, y_{t-1}$.

Los valores ajustados regularmente no son pronósticos certeros ya que los parámetros usados en el método de pronóstico se calculan haciendo uso de todas las observaciones a disposición en la serie de tiempo, incluyendo las observaciones futuras.

Los residuales de un modelo de series de tiempo son aquellos que permanecen posterior a ajustar un modelo. En la mayoría de los modelos de series temporales, los residuos se definen por la diferencia entre las observaciones y los valores ajustados correspondientes, así.

$$mi_t = y_t - \hat{y}_t \qquad \text{(Ec. 10)}$$

Los residuos tienen como objetivo comprobar si un modelo ha retenido adecuadamente la información de los datos. Para que un método se defina como bueno los residuos deberán cumplir los siguientes puntos:

- No deben estar correlacionados, ya que sí la hay quedo información en los residuos que servirá para el pronóstico.
- Deben tener una media con valor cero, en caso contrario los pronósticos están sesgados.
- Deben tener varianza constante.
- Se distribuyen normalmente.

Además de tener la gráfica ACF es posible realizar pruebas con mayor formalidad:

- Prueba Box-Pierce:

$$q = T * \sum_{k=1}^{h} r_k^2 \qquad \text{(Ec. 11)}$$

Donde:

h: Retardo máximo

T: Numero de observaciones

- Prueba de Ljung-Box:

$$q^* = T(T+2) * \sum_{k=1}^{h} (T-k)^{-1} * r_k^2 \qquad \text{(Ec. 12)}$$

7.1.11. Evaluación de exactitud de pronósticos

Los errores de pronóstico están definidos entre la diferencia entre el valor observado y su pronóstico. Importante hay que considerar que el concepto de error no es precisamente error sino la fracción de imprecisión de una observación la cual se define:

$$mi_{T+h} = y_{T+h} - \hat{y}_{T+h|T} \qquad \text{(Ec. 13)}$$

Donde los datos de entrenamiento están definidos por $\{y_1, \ldots y_T\}$ y los datos de prueba están definidos por $\{y_{T+1}, y_{T+2}, \ldots\}$.

Es necesario tener presente que los errores y los residuos no son iguales, ya que los residuos se estiman en el conjunto de entrenamiento, y los errores se estiman con el conjunto de prueba. Otra consideración es que los residuos se fundamenta pronósticos de un paso y los errores pueden implicar pronósticos de varios pasos.

Los errores dependiendo de las escalas, están en la misma escala que la serie de datos, estas medidas de precisión se basan en mi y por ende son dependientes de la escala y no son útiles para comprar entre series. Las medidas dependientes de la escala más usadas están fundamentadas en:

- Error absoluto medio:

$$MAE = media(|mi_t|) \quad \text{(Ec. 14)}$$

- Error cuadrático medio:

$$RMSE = \sqrt{media(mi_t^2)} \quad \text{(Ec. 15)}$$

Los errores de porcentajes (p_t) tienen la virtud de no unidad por lo que son usados para desempeño de pronósticos entre conjunto de datos.

$$p_t = \frac{100mi_t}{y_t} \quad \text{(Ec. 16)}$$

$$Error\ porcentual\ absoluto\ medio: MAPE = media\ (|p_t|) \quad \text{(Ec. 17)}$$

7.1.12. Pronósticos basados en juicios

Surge cuando hay falta de datos históricos del fenómeno estudiado, o cuando se está trabajando con algo innovador de lo que no hay información previa. También surge cuando la información está incompleta o cuando tiene cierto rezago.

7.1.12.1. Método Delphi

Tiene como objetivo elaborar un pronóstico a través del consenso de un grupo de expertos de manera coordinada, en este hay un moderador para gestionar todo el proceso

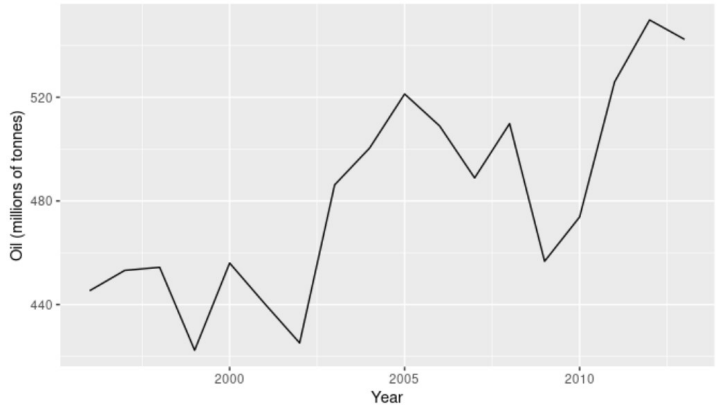
Para este método se tiene las siguientes etapas:

- ✓ Reunión del grupo de expertos
- ✓ Se definen los desafíos del pronóstico y se comparten con los expertos.
- ✓ Los expertos realizan sus pronósticos prematuros con sus justificaciones.
- ✓ Se retroalimenta a los expertos
- ✓ Se realizan los pronósticos definitivos compilando la información.

7.1.13. Suavizado exponencial simple

Este método es adecuado para pronosticar cuando los datos no tienen tendencia o estacionalidad, podemos ver la figura 12 como ejemplo:

Figura 12. **Serie de tiempo sin tendencia y estacionalidad**



Fuente: Fuente: Hyndman (2018)

Y, la ecuación que define este método:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots \quad \text{(Ec. 18)}$$

Donde:

$0 \leq \alpha \leq 1$ es el parámetro suavizado

7.1.14. Método de tendencia lineal de Holt

Es básicamente una ampliación del método suavizado exponencial, con la variante que este si permite pronosticar datos con una tendencia. La ecuación de pronóstico está definida por:

$$\hat{y}_{t+h|t} = l_t + hb_t \qquad \text{(Ec. 19)}$$

Este método tiene dos ecuaciones, una para el nivel y otra para la tendencia.

- Ecuación de nivel:

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \qquad \text{(Ec. 20)}$$

- Ecuación de tendencia:

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \qquad \text{(Ec. 21)}$$

Donde $0 \leq \alpha \leq 1$ y $0 \leq \beta \leq 1$

7.1.15. Modelo ARIMA

Junto con el modelo de Suavizado Exponencial, el modelo ARIMA son los más utilizados. Para este modelo es necesario aclarar estacionariedad lo cual es una propiedad que no depende del tiempo en el que se está observando la serie, por ende, es necesario tomar en cuenta, que tendencia y estacionalidad, no es lo mismo que estacionariedad. Por otro lado, la diferenciación aporta estabilidad a la media de una serie temporal, lo elimina en los niveles de una serie temporal.

El modelo ARIMA realiza un pronóstico de la variable estudiada haciendo uso de combinaciones lineales de valores pasados de la variable

$$y_t = C + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \qquad \text{(Ec. 22)}$$

Donde ε_t es ruido blanco. Los que ha sido definido en la ecuación 22 es como una regresión múltiple con valores rezagados de y_t. Ha esto se refiere como un Modelo $AR(p)$ es decir un modelo autorregresivo de orden p.

Para un modelo $AR(1)$ cuando:

- $\emptyset_1 = 0, y_t$ es igual a ruido blanco.
- $\emptyset_1 = 1\ y\ C = 0$ es equivalente a un paseo aleatorio.
- $\emptyset_1 = 1\ y\ C \neq 0$ es equivalente a un paseo aleatorio con deriva.
- $\emptyset_1 < 1\ y_t$ tiende a oscilar alrededor de la media.

Regularmente los modelos auto regresivos están restringidos a datos estacionarios en el que se requieren ciertas limitaciones en los valores de los parámetros. Así:

Para un modelo AR(1): $-1 < \emptyset_1 < 1$

Para un modelo AR(2): $-1 < \emptyset_2 < 1, \emptyset_1 + \emptyset_2 < 1, \emptyset_2 - \emptyset_1 < 1$

Y para $p \geq 3$ son más complejas

7.2. Población estudiantil

Individuos que debido a su edad son incluidos en un determinado grado de escolaridad dentro del sistema educación. Es el grupo de niños en un rango

de edades clasificado quienes son apropiados para exigir los servicios educativos acorde a su edad.

En Guatemala se tiene la siguiente escala:

- Preprimaria, de 5 a 6 años.
- Primaria, de 7 a 12 años.
- Ciclo Básico, de 13 a 15 años.

7.2.1. Centro educativo

Es un establecimiento cuyo fin es la enseñanza. Pueden ser de diversos tipos y características.

7.2.1.1. Público

Son establecimientos administrados y financiados por el Estado y brindan sin excepción el servicio educativo a su población, en función al rango de edad.

7.2.1.2. Privado

Básicamente son empresas financiadas por los padres de familia de los alumnos que asisten. Tiene la libertad de decidir a quien le permiten el acceso. Cabe destacar que se rigen bajo las leyes educativas, cada establecimiento fundamenta sus políticas apegadas a ella.

7.2.1.3. Promedio alumno-docente

Es la relación entre la cantidad de estudiantes que en promedio atiende un docente en cada aula.

8. PROPUESTA DE ÍNDICE DE CONTENIDOS

9. METODOLOGÍA

9.1. Características del estudio

El enfoque del estudio propuesto es cuantitativo ya que se utilizarán variables numéricas para el análisis el crecimiento de población estudiantil y la capacidad de los centros educativos públicos.

El diseño adoptado será no experimental u observacional, porque se analizarán bases de datos, y no se manipularán variables.

Asimismo, será retrospectivo ya que se utilizarán base de datos con información histórica.

El alcance del estudio es descriptivo y correlacional, pues se buscará describir la situación de relación entre población estudiantil y centros educativos.

9.2. Unidades de análisis

La población en estudio estará conformada por estudiantes de nivel básico del sector educativo público del municipio de Villa Nueva con información de los últimos cinco años.

9.3. Variables

Tabla II. **Operativización de variables**

Variable	Definición teórica	Definición operativa
Población estudiantil de nivel básico del sector público (x)	Cantidad de niños que se encuentra en un rango de edades preestablecido como la edad apropiada para demandar los servicios de un nivel.	Variable independiente Se medirá y definirá a través de una base datos histórica. Número natural a escala de razón. Variable numérica discreta.
Centros educativos públicos activos (y)	Cantidad de establecimientos que administra y financia el Estado para ofrecer sin discriminación, el servicio educacional a los habitantes del país, de acuerdo con las edades correspondientes a cada nivel y tipo de escuela, normados por el reglamento específico.	Variable dependiente. Se analizará por medio de bases de datos. Número natural a escala de razón.

Fuente: Elaboración propia.

9.4. Fases del estudio

Para el presente estudio se han definido las siguientes fases:

- Fase 1: Revisión de literatura

Se buscará literatura que respalde los temas y definiciones en que se fundamentará este estudio. Los tópicos como pruebas paramétricas de normalidad, correlación lineal, series temporales y pronósticos.

- Fase 2: Gestión o recolección de la información

Se analizará la información que el Ministerio de Educación ha recolectado de cada establecimiento a través de su portal *web*. Por lo que se tiene a disposición una base de datos muy completa sobre la cantidad de estudiantes por grado de escolaridad.

- Fase 3: Análisis de información

Primero se realizará un diagnóstico de la tasa de crecimiento poblacional estudiantil del nivel básico y se estudiará su comportamiento a lo largo de los años. Asimismo, se realizará el mismo análisis sobre la cantidad de centros educativos que el MINEDUC ha ido habilitando a través del tiempo.

Se hará un análisis de correlación entre ambas variables para identificar su comportamiento en conjunto. Se realizará un análisis de series temporales sobre los datos históricos y se hará un pronóstico para evaluar la situación en los próximos años.

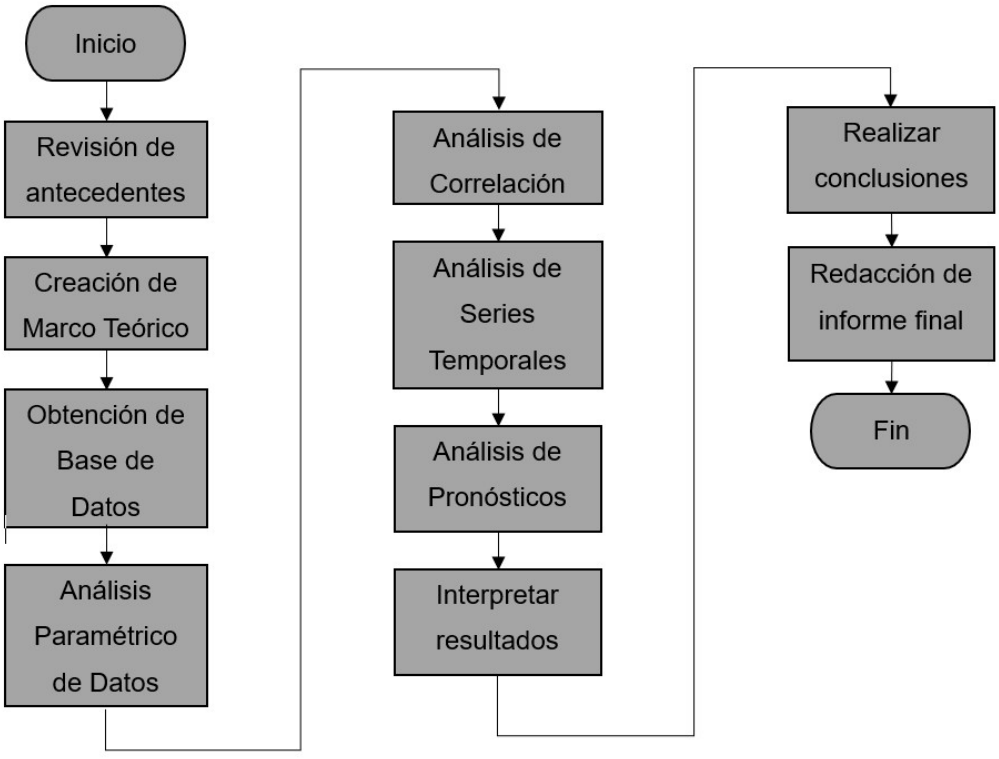
- Fase 4: Interpretación de información

Se analizarán los resultados de la correlación para definir el grado de sobrepoblación, asimismo de las series temporales y pronósticos, con el fin de ver los patrones, historias y evaluar el método idóneo de pronósticos

- Fase 5: Elaboración del informe final

Todos los análisis y cálculos numéricos y gráficos será interpretados sobre el estudio en mención, tales como correlación, pruebas de normalidad, series temporales y pronósticos.

Figura 13. **Flujograma del proceso de investigación**



Fuente: Elaboración propia.

10. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN

En este apartado se detallarán las técnicas que se utilizarán para realizar el estudio.

Los datos para analizar serán la poblacional estudiantil del nivel básico que debe cubrir los centros de educación básica del sector público en el municipio de Villa Nueva y con ello analizar la sobrepoblación que tienen algunos centros educativos. La información se encuentra en la base de datos del Ministerio de Educación por lo que la observación es directa para el estudio.

10.1. Kolmogorov-Smirnov

Una de las pruebas de normalidad que se realizaran será Kolmogorov-Smirnov, de ser rechazados se realizarán transformaciones por anamorfosis que contribuyan al supuesto de normalidad, esto con el propósito poder realizar pruebas paramétricas sobre los datos.

10.2. Q-Q plot e Histograma

Para continuar con las pruebas de normalidad de los datos, se realizarán gráficos de los residuos de análisis de varianza que serán contrastados con la prueba de Kolmogorov-Smirnov. Este grafico dará una primera vista sobre el comportamiento de los datos de manera visual.

Se realizará una segunda prueba grafica para contrastar las pruebas anteriores con el propósito de realizar una tercera prueba de normalidad para que aporte más información sobre la naturaleza de los datos.

10.3. Análisis de correlación

Se realizará un análisis de correlación sobre las variables independientes y dependientes para determinar la relación y la fuerza o debilidad que existe entre ellas, por medio del coeficiente de Pearson, en caso exista normalidad en los datos. De lo contrario se realizará un análisis sobre el coeficiente de Tau de Kendall.

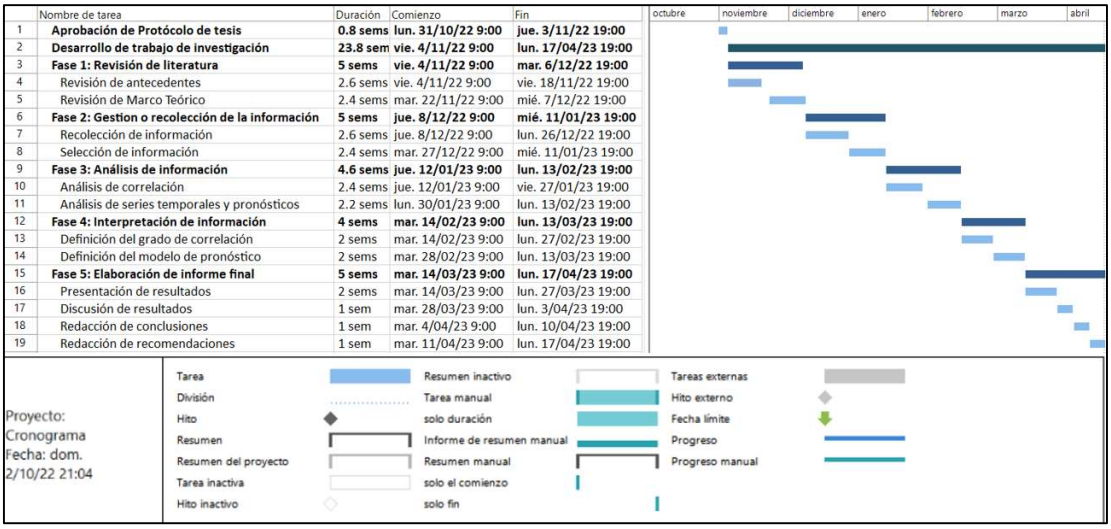
10.4. Evaluación de series temporales

Se realizará un análisis de series temporales para predecir la población estudiantil y centros educativos a futuro bajo las condiciones históricas de ambas variables.

11. CRONOGRAMA

A continuación, se presenta la distribución por semana de la metodología propuesta para el desarrollo del proyecto:

Figura 14. **Cronograma I**



	Nombre de tarea	Duración	Comienzo	Fin
1	**Aprobación de Protócolo de tesis**	**0.8 sems**	**lun. 31/10/22 9:00**	**jue. 3/11/22 19:00**
2	**Desarrollo de trabajo de investigación**	**23.8 sem**	**vie. 4/11/22 9:00**	**lun. 17/04/23 19:00**
3	**Fase 1: Revisión de literatura**	**5 sems**	**vie. 4/11/22 9:00**	**mar. 6/12/22 19:00**
4	Revisión de antecedentes	2.6 sems	vie. 4/11/22 9:00	vie. 18/11/22 19:00
5	Revisión de Marco Teórico	2.4 sems	mar. 22/11/22 9:00	mié. 7/12/22 19:00
6	**Fase 2: Gestion o recolección de la información**	**5 sems**	**jue. 8/12/22 9:00**	**mié. 11/01/23 19:00**
7	Recolección de información	2.6 sems	jue. 8/12/22 9:00	lun. 26/12/22 19:00
8	Selección de información	2.4 sems	mar. 27/12/22 9:00	mié. 11/01/23 19:00
9	**Fase 3: Análisis de información**	**4.6 sems**	**jue. 12/01/23 9:00**	**lun. 13/02/23 19:00**
10	Análisis de correlación	2.4 sems	jue. 12/01/23 9:00	vie. 27/01/23 19:00
11	Análisis de series temporales y pronósticos	2.2 sems	lun. 30/01/23 9:00	lun. 13/02/23 19:00
12	**Fase 4: Interpretación de información**	**4 sems**	**mar. 14/02/23 9:00**	**lun. 13/03/23 19:00**
13	Definición del grado de correlación	2 sems	mar. 14/02/23 9:00	lun. 27/02/23 19:00
14	Definición del modelo de pronóstico	2 sems	mar. 28/02/23 9:00	lun. 13/03/23 19:00
15	**Fase 5: Elaboración de informe final**	**5 sems**	**mar. 14/03/23 9:00**	**lun. 17/04/23 19:00**
16	Presentación de resultados	2 sems	mar. 14/03/23 9:00	lun. 27/03/23 19:00
17	Discusión de resultados	1 sem	mar. 28/03/23 9:00	lun. 3/04/23 19:00
18	Redacción de conclusiones	1 sem	mar. 4/04/23 9:00	lun. 10/04/23 19:00
19	Redacción de recomendaciones	1 sem	mar. 11/04/23 9:00	lun. 17/04/23 19:00

Fuente: Elaboración propia.

12. FACTIBILIDAD DEL ESTUDIO

La investigación es factible por contar con los recursos necesarios para llevarla a cabo. A continuación, se detallan:

12.1. Recurso humano

Es la persona que se dedicará a realizar todas las atribuciones para llevar a cabo la investigación.

12.2. Recursos financieros

Son los costos económicos que se utilizar la ejecución del estudio, el investigador cubrirá la totalidad de ellos. En la siguiente tabla se detallan todos los rubros necesarios.

Tabla III. **Costos de investigación**

Elemento	Unidad	Costo Unitario/(Q.)	Cantidad necesaria	Costo /(Q.)
FASE 1-5: En general				
Investigador	Persona	10,000	1	10,000
Internet	Paquete/mensual	250	4	1,000
Alimentos	Ración	40	4	160
Viáticos	Gasolina-parqueos	40	4	160
Equipo de computo	Dispositivo	4,000	1	4,000
FASE 3: Análisis de información				
Software	Programa	0	2	0
FASE 5: Elaboración de informe final				
Hojas de papel bond	Resma	75	1	75
Impresora	Dispositivo	400	1	400
TOTAL				**15,795**

Fuente: Elaboración propia.

12.3. Recursos tecnológicos

Es el software que será utilizado para la ejecución del estudio, con que se hará el análisis de la información, pruebas y gráficas necesarias; se utilizará Microsoft Excel y R Studio.

12.4. Acceso a información y permisos

Son las bases de datos donde se extraerá la información que está en los repositorios del Ministerio de Educación y están a libre disposición

12.5. Equipo e infraestructura

Son los dispositivos electrónicos como equipo de cómputo e impresora que estarán a disposición del investigador.

13. REFERENCIAS

1. Amat, Joaquín (2016), *Análisis de Normalidad: Gráficos y contraste de hipótesis.* *Recuperado de* https://www.cienciadedatos.net/documentos/8_analisis_normalidad#:~:text=Los%20an%C3%A1lisis%20de%20normalidad%2C%20tambi%C3%A9n,misma%20media%20y%20desviaci%C3%B3n%20t%C3%ADpica.

2. Centro de Investigación Económicas Nacionales (2019) *El Sistema Educativo en Guatemala*. Recuperado de https://cien.org.gt/wp-content/uploads/2019/05/Educacio%CC%81n-y-Tecnologi%CC%81a-documento-final.pdf

3. Contento, Manuel, (2020), *Estadística con aplicaciones en R*, Colombia Editorial UTADEO

4. Datos Mundial, (2021), *Desarrollo demográfico en Guatemala desde 1960.* *Recuperado de* https://www.datosmundial.com/america/guatemala/crecimiento-poblacional.php

5. Empresarios por la educación (2022), *Indicadores, Cobertura, eficiencia, calidad, inversión pública*. Recuperado de http://www.empresariosporlaeducacion.org/content/indicadores

6. FADEP, (2019), *Cifras Educativas de Guatemala*. Recuperado de https://fadep.org/principal/desarrollo/cifras-educativas-de-guatemala/

7. Fallas, Jorge, (2012), *Correlación Lineal, Midiendo la relación entre dos variables.* Recuperado de www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/correlacion_lineal_2012.pdf

8. Fondo de Población de las Naciones Unidas, (2020), *Juventudes en Guatemala. Recuperado de* https://conjuve.gob.gt/wp-content/uploads/2021/05/UNFPA.pdf

9. Gutiérrez, Humberto (2013), *Control estadístico de la calidad y seis sigmas*, México McGraw-Hill

10. Hyndman, Rob, (2018), *Forecasting: Principles y Practice*. Recuperado de https://otexts.com/fpp2/determining-what-to-forecast.html

11. López, Hércules (2016) *Ensayo Sobrepoblación Escolar en el Nivel Primario*. Recuperado de https://ley.exam-10.com/pravo/26127/index.html

12. Lopez, Jose (2019), *Diferencia entre estadística paramétrica y no paramétrica*. Recuperado de https://economipedia.com/definiciones/diferencia-entre-estadistica-parametrica-y-no-parametrica.html#:~:text=La%20estad%C3%ADstica%20param%

C3%A9trica%20utiliza%20c%C3%A1lculos,utilizar%20t%C3%A9cnicas%20de%20estad%C3%ADstica%20param%C3%A9trica

13. Ministerio de educación de Guatemala (2022) *Anuario estadístico de la educación en Guatemala*. Recuperado de http://estadistica.mineduc.gob.gt/Anuario/home.html#

14. Ministerio de educación de Guatemala (2022) *Anexos información adicional de interés general.* Recuperado de https://www.mineduc.gob.gt/estadistica/2012/data/index_anexo.html#:~:text=Letra%20P-,Poblaci%C3%B3n%20en%20edad%20escolar,los%20servicios%20de%20un%20nivel.

15. Perez, Harold, (2008), *Estadística para las ciencias sociales, del comportamiento y de la salud*, México Cengage Learning

16. Ritchey, Ferris, (2008), *Estadística para las ciencias sociales*. México, McGraw-Hill

17. Rodó, Paula (2020), *Prueba de Kolmogorov – Smirnoff (K-S).* Recuperado de https://economipedia.com/definiciones/prueba-de-kolmogorov-smirnoff-k-s.html

18. Saldaña, Manuel (2016), *Metodología de la investigación: Pruebas de bondad de ajuste a una distribución normal*. Recuperado de https://dialnet.unirioja.es/servlet/articulo?codigo=5633043

a.

19. Ruiz, Eric, (2017), *El protocolo de investigación VI: como elegir la prueba estadística adecuada. Estadística inferencial*. Recuperado de www.scielo.org.mx/pdf/ram/v64n3/2448-9190-ram-64-03-0364.pdf

20. Solorzano, Jose (2017) *Dificultades académicas que enfrentan los docentes del plan sábado por sobrepoblación estudiantil de la Facultad de Humanidades [Tesis Maestría*, Universidad de San Carlos de Guatemala] http://biblioteca.usac.edu.gt/tesis/07/07_2364.pdf

21. Solorzano, Jose (2017) *La sobrepoblación estudiantil en la USAC*, Recuperado de https://docplayer.es/66830055-La-sobrepoblacion-estudiantil-en-la-usac-y-la-calidad-educativa.html

22. Universidad Nacional de Educación y Universidad Central del Ecuador, (2020), *La ratio dorado Estudiante-Profesor y el número de Docentes que necesita Ecuador. Recuperado de* https://unae.edu.ec/wp-content/uploads/2020/03/RATIO-DORADO-FINAL.pdf

23. USAID (2019) *Educación: Una oportunidad de desarrollo para Guatemala*. Recuperado de https://www.thedialogue.org/blogs/2019/04/educacion-una-oportunidade-de-desarrollo-para-guatemala/

24. Vinuesa, Pablo, (2016), UNAM, *Tema 8 - Correlación: Teórica y práctica*. Recuperado de www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.pdf

25. Zhuji World, (2022), *Villa Nueva, Guatemala – Estadística*. Recuperado de https://fadep.org/principal/desarrollo/cifras-educativas-de-guatemala/

14. APÉNDICES

Apéndice 1. **Matriz de coherencia**

ELEMENTOS	PROBLEMA ESTADISTICO	PREGUNTAS DE INVESTIGACIÓN	OBJETIVOS	SOLUCIÓN PROPUESTA	FUNDAMENTO	METODOLOGIA
GENERAL O CENTRAL	El nivel de correlación de la población estudiantil y la capacidad de los centros educativos habilitados actual, el análisis de series temporales para evaluar la situación a futuro y el pronóstico que permita definir la cantidad optima de centros educativos.	¿Qué tan optimo es el nivel de correlación que tienen la población estudiantil y la capacidad de centros educativos, y cual es el modelo de series temporales adecuado para pronosticar la situación en el futuro para proyectar la cantidad optima de centros educativos?	Desarrollar un modelo de regresión lineal que permita medir el grado de correlación lineal entre la población estudiantil actual y la capacidad de los centros educativos habilitados. Asimismo, crear un modelo estadístico de series temporales que permita proyectar la situación a futuro que además permita pronosticar la cantidad de centros educativos idóneo en función a la población estudiantil.	Realizar un análisis de Correlación a los datos, asimismo un pronóstico por medio del análisis de series temporales	Modelo de regresión lineal Coeficiente de correlación Gráficos y series temporales	El enfoque del estudio propuesto es cuantitativo ya que se analizarán el crecimiento de población estudiantil y la capacidad de los centros educativos públicos. El alcance será descriptivo y correlacional; descriptivo ya que indicará el tamaño de la población estudiantil y los centros educativos públicos activos y correlacional porque se definirá la relación entre la población estudiantil y la capacidad de los centros educativos. El diseño adoptado será no experimental-longitudinal, pues la información del comportamiento de las variables se analizar con bases de datos de años anteriores
PECÍFICOS O AUXILIARES	El nivel de correlación entre la población de estudiantes actual y la capacidad centros educativos públicos, por medio de una correlación lineal	¿Cuál será el nivel de correlación que existe entre la población actual de estudiantes y la capacidad de los centros educativos públicos?	01. Crear un modelo estadístico de regresión lineal para evaluar la correlación entre la población estudiantil actual y la capacidad de los centros educativos públicos habilitados, para indicar la sobrepoblación estudiantil.	Se realizará un análisis previo de normalidad de los datos, se realizarán histogramas, gráfico de cuartiles y por último test de Kolmogorov-Smirnov. Posterior a ello se realizará el análisis de correlación, por medio de un diagrama de dispersión y con el coeficiente de correlación de Pearson.	Modelos matemáticos Gráficos	
	El pronóstico a atreves de series temporales indica una situación de sobrepoblación igual o peor.	¿Qué pronostica el análisis de series temporales sobre la sobrepoblación estudiantil en los siguientes años?	02. Definir un análisis estadístico de pronósticos a través de series temporales tomando en cuenta la tasa de crecimiento poblacional y el número de centros educativos que habilita el Mineduc anualmente, para medir la sobrepoblación en los siguientes años.	Por medio del análisis de series temporales se evaluará una proyección de la situación en los próximos años.	Gráficos Pronósticos	
	El pronóstico por medio de series temporales que indica la cantidad de centros educativos idóneo para la demanda de estudiantes pronosticados.	¿Cuál es el pronóstico por medio de series temporales que indique la cantidad de centros educativos idóneo para la demanda de alumnos en el futuro?	03. Definir por medio de series temporales el pronóstico de centro centros educativos que el Mineduc deberá aperturar para cubrir la demanda de estudiantes y mitigar la sobrepoblación de los centros educativos.	Se analizará una proyección de la cantidad de centros educativos que deben operar en función a la población estudiantil en los próximos años.	Gráficos Pronósticos	

Fuente: Elaboración propia.