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

DISEÑO DE INVESTIGACIÓN DE UN MODELO ESTADÍSTICO DE PREDICCIÓN DE LA DIMENSIÓN EDUCATIVA DEL IDH EN FUNCIÓN DE LA INVERSIÓN PORCENTUAL DEL PIB Y TASA DE COBERTURA EDUCATIVA NETA EN LOS DIFERENTES NIVELES EN GUATEMALA

Ricardo Alberto Meza Santos
Asesorado por el Mtro. Luis Carlos Bolaños Mendez

Guatemala, octubre 2021

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

DISEÑO DE INVESTIGACIÓN DE UN MODELO ESTADÍSTICO DE PREDICCIÓN DE LA DIMENSIÓN EDUCATIVA DEL IDH EN FUNCIÓN DE LA INVERSIÓN PORCENTUAL DEL PIB Y TASA DE COBERTURA EDUCATIVA NETA EN LOS DIFERENTES NIVELES EN GUATEMALA

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

RICARDO ALBERTO MEZA SANTOS
ASESORADO POR EL MTRO. LUIS CARLOS BOLAÑOS MENDEZ

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO INDUSTRIAL

GUATEMALA, OCTUBRE DE 2021

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

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

NÓMINA DEL TRIBUNAL EXAMINADOR

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Alberto Eulalio Hernández García
EXAMINADOR	Ing. Ismael Homero Jerez González
EXAMINADOR	Ing. Byron Gerardo Chocooj Barrientos
SECRETARIO	Ing. Hugo Humberto 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 DE UN MODELO ESTADÍSTICO DE PREDICCIÓN DE LA DIMENSIÓN EDUCATIVA DEL IDH EN FUNCIÓN DE LA INVERSIÓN PORCENTUAL DEL PIB Y TASA DE COBERTURA EDUCATIVA NETA EN LOS DIFERENTES NIVELES EN GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 27 de julio 2021.

Ricardo Alberto Meza Santos

Escuela de Estudios de Postgrado, Edificio S-11
Teléfono: 2418-9142 / 24188000 ext. 1382
WhatsApp: 5746-9323
Email: informacion_eep@ing.usac.edu.gt

Ref. EEPFI-0919-2021
Guatemala, 27 de julio de 2021

Director
César Ernesto Urquizú Rodas
Escuela de Ingeniería Mecánica Industrial
Presente.

Estimado Ing. Urquizú:

Reciba un cordial saludo de la Escuela de Estudios de Postgrado. El propósito de la presente es para informarle que se ha revisado y aprobado el **DISEÑO DE INVESTIGACIÓN: MODELO ESTADÍSTICO DE PREDICCIÓN DE LA DIMENSIÓN EDUCATIVA DEL IDH EN FUNCIÓN DE LA INVERSIÓN PORCENTUAL DEL PIB Y TASA DE COBERTURA EDUCATIVA NETA EN LOS DIFERENTES NIVELES EN GUATEMALA**, presentado por el estudiante **Ricardo Alberto Meza Santos** carné número **9312016**, 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 Estadística 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"

Luis Carlos Bolaños Méndez
Ing. electrónico
Col. 7653

Mtro. Luis Carlos Leonardo Bolaños Méndez
Asesor

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

Mtro. Edwin Adalberto Bracamonte Orozco
Coordinador de Maestría
Estadística Aplicada

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

Mtro. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

FACULTAD DE INGENIERÍA

EEP-EIMI-052-2021

El Director de la Escuela de Ingeniería Mecánica 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: **MODELO ESTADÍSTICO DE PREDICCIÓN DE LA DIMENSIÓN EDUCATIVA DEL IDH EN FUNCIÓN DE LA INVERSIÓN PORCENTUAL DEL PIB Y TASA DE COBERTURA EDUCATIVA NETA EN LOS DIFERENTES NIVELES EN GUATEMALA**, presentado por el estudiante universitario **Ricardo Alberto Meza Santos**, 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

DIRECCION
Escuela de Ingeniería Mecánica Industrial
FACULTAD DE INGENIERIA

Firmada digitalmente por Cesar Ernesto Urquizu Rodas
Motivo: Ingeniero Industrial
Ubicación: Facultad de Ingenieria, Escuela de Ingenieria Mecanica Industrial, USAC
Colegiado 4,272

Ing. César Ernesto Urquizú Rodas
Director
Escuela de Ingeniería Mecánica Industrial

Guatemala, julio de 2021

USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

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

DTG. 552.2021

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 DE UN MODELO ESTADÍSTICO DE PREDICCIÓN DE LA DIMENSIÓN EDUCATIVA DEL IDH EN FUNCIÓN DE LA INVERSIÓN PORCENTUAL DEL PIB Y TASA DE COBERTURA EDUCATIVA NETA EN LOS DIFERENTES NIVELES EN GUATEMALA,** presentado por el estudiante universitario: **Ricardo Alberto Meza Santos,** y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Inga. Anabela Cordova Estrada
Decana

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DECANO
FACULTAD DE INGENIERIA

Guatemala, octubre de 2021

AACE/cc

ACTO QUE DEDICO A:

Dios	Por ser la fuente de fortaleza en mi vida y otorgarme la bendición de concluir este proceso.
Mis padres	Juan Francisco Meza y Aura Margarita Santos, por su apoyo y guía a lo largo de toda mi vida.
Mi hermana	Karen Janeth Meza, por ser parte de mi vida comprometiéndome a ser ejemplo continuo.
Mi sobrino	Liam Joaquín Zacarias Meza, por ser una brisa de aire fresco y ser un recordatorio constante de lo fácil que es ser feliz.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser mi casa de estudios, y la fuente de formación profesional de que he alcanzado.
Facultad de Ingeniería	Por albergarme en sus pasillos y formarme en sus aulas, regalándome además muchos de mis amigos más entrañables.
Mis amigos de la maestría	Por ser los compañeros de desvelos y fines de semana interminables, que fueron fuente de apoyo para no claudicar, gracias, Julio Monterroso, Erick Marroquín, Fabiola Ramírez y Andrea Corado.
Mi amigo	Luis Fernando Alvarado, quien además de su amista, compartió conmigo la aventura de cursar la maestría.
Mi asesor	Luis Carlos Bolaños Méndez por compartir sus conocimientos y brindarme su orientación.
Mi primo	El ingeniero Luis Manuel Ávila, a quien considero un hermano y amigo, que siempre me impulso para alcanzar esta y muchas metas.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
gl	Grados de libertad
S^2	Varianza muestral
μ	Media poblacional
σ	Varianza poblacional

GLOSARIO

AICc	Criterio de la información Akaike
Aleatoriedad	Carencia de una tendencia o patrones reconocibles en una serie de datos.
BIC	Criterio de la información Bayesiano
Correlación	Relación o correspondencia entre dos o más variables.
Hipótesis nula	Supuesto que se desea validar o invalidar.
Homocedasticidad	Uniformidad de la varianza.
Homogeneidad	Igualdad en el comportamiento de todos los elementos que conforman un conjunto determinado de datos.
IDH	Índice de Desarrollo Humano
Independencia	Falta de relación entre variables, es decir, el comportamiento de una de ellas no se ve afectado por los valores que toma la otra.
Indicador	Valor estadístico de referencia al tamaño de parámetros o atributos.

PIB	Producto Interno Bruto

RESUMEN

En el presente trabajo se detalla la metodología empleada para construir un modelo estadístico de predicción con el máximo nivel de ajuste, de la dimensión educativa del IDH en función de la inversión porcentual del PIB y tasa de cobertura educativa neta en los diferentes niveles en Guatemala, y con ello se ofrecer una herramienta útil para la elaboración de políticas educativas de desarrollo.

Se inicia presentando los antecedentes que permiten contextualizar la importancia de índice del desarrollo humano, particularmente la dimensión de educación y la utilización de modelos de regresión basados en indicadores, dando con ello sustento al desarrollo y planteamiento del problema. Seguidamente se explica la importancia del estudio por medio de la justificación y presenta los objetivos con los cuales se pretende dar respuesta al problema.

La investigación continua con el análisis de las necesidades por cubrir y el esquema de solución propuesto, que está directamente relacionado con el marco teórico y las fuentes allí plasmadas, que describen las ecuaciones que se aplicaran y el fundamento teórico para realizarlo. Por último, se describe a detalle la metodología y el cronograma, siendo estos los que brindan la pauta para el desarrollo de estudio de forma estructurada, permitiendo su factibilidad.

1. INTRODUCCIÓN

La investigación propuesta va orientada al análisis del desarrollo humano en Guatemala, toma como base el índice de desarrollo humano (IDH), del Programa de las Naciones Unidas para el Desarrollo (PNUD).

Se propone el análisis de la dimensión de educación en función de la inversión porcentual del PIB en dicha área, así como otras variables que tengan incidencia directa, este análisis de inversión en el área de educación en Guatemala incluye los últimos 10 años.

Se analizarán modelos de inversión en materia de educación, de países latinoamericanos con contextos similares al de Guatemala en lo referente la composición de su población, entre los cuales se encuentran Bolivia, México y Perú, con el fin de establecer parámetros que ayuden a validar y construir un modelo propio que sea representativo de la realidad nacional, usando para ello datos del 2004 al 2019.

Se propondrá la construcción de un modelo de regresión, que permita predecir el comportamiento de la dimensión de educación, parte del índice de desarrollo humano (IDH), en relación con la inversión realizada en materia de educación como un porcentaje asignado del producto interno bruto (PIB) en esta área con su respectivo análisis de regresión que valide la precisión de este.

En el capítulo uno: se hará una descripción de los antecedentes utilizados como referencia de trabajos anteriores relacionados con el tema del desarrollo

humano, las técnicas y estrategias aplicadas sobre los datos y la construcción de modelos sobre este tema.

En el capítulo dos: se presentará el marco teórico organizado en dos secciones, la primera de ellas incluyendo todos los fundamentos estadísticos que se utilizaran para la construcción de los modelos de pronóstico de la dimensión de educación, del índice de desarrollo humano; la segunda sección describe los conceptos básicos relacionados con el desarrollo humano.

En el capítulo tres: se presentarán los resultados obtenidos a partir del análisis estadístico aplicado y se muestra la información de forma numérica y gráfica.

En el capítulo cuatro: se analizarán los resultados obtenidos y a partir de ellos se ofrecen conclusiones y recomendaciones.

2. ANTECEDENTES

La medida del desarrollo humano y de las variables que intervienen en su análisis se han reducido a índices o escalas porcentuales que, si bien son resultados continuos se ven influenciados por los efectos de tener cotas superiores e inferiores que tienden a frenar o amortiguar los modelos de regresión al aproximarse a ellos, si no se les da a los datos el tratamiento adecuado, dado que los porcentajes no son exclusivos de las áreas sociales, a continuación encontraremos diferentes áreas donde se plantearon estrategias diferentes en cuanto al manejo de la información que serán tomadas como base en la resolución del problema estadístico.

Elacqua y Martínez (BID, 2018) describen como en las últimas dos décadas en Latinoamérica, ha crecido la inversión en educación de 3.6 % a 5.3 % del PIB, mientras que en Guatemala en el mismo año se destinó el 2.93 % de su PIB para ello, destacando que en los países con mejores resultados se invierten anual mente 8000 US$ por estudiante en comparación con Latinoamérica que invierte 2000 US$ por estudiante, y sumado a ello la poco eficiente gestión de los recursos.

Liscano y Ortiz (2018) recomiendan modelos de regresión lineal y mixto cuando se está trabajando con datos composicionales, este tipo de dato se define como multivariado ya que el resultado de la suma es contante, y se utiliza por conveniencia, esta clase de datos se presentan como porcentajes, proporciones o concentraciones, se aplican por lo general en áreas de estudio como geología, arqueología, economía, ciencias políticas y ciencias forenses. El que los datos no puedan ser negativos y la suma sea constante implica que las técnicas

multivariantes habitualmente utilizadas no son adecuadas para su análisis y modelización, en este caso fue utilizado para proyectar la intención de voto en unas elecciones distritales en Colombia.

Lopéz-Roldan y Fachelli (2016) los modelos multivariados con datos composicionales habitualmente son empleados en estudios de investigación social, en estos casos es habitual el uso de transformaciones log-cociente, para la aplicación de regresión logística se trata de predecir una variable cualitativa o categórica, con la ventaja, frente al modelo de regresión clásico, de no tener que establecer la serie de condiciones de aplicación que dificultan su utilización y sus posibilidades, en particular, en el contexto de estudios por encuesta, en este caso relaciona el voto con el nivel educativo.

Enke, Graue y Mehdiyev (2011) presentan un sistema de predicción en tres etapas. Etapa 1, se realiza el análisis de regresión múltiple para definir las variables que tienen una fuerte relación con la salida. En la segunda fase se implementa la evolución diferencial con base en el Fuzzy Clustering (algoritmo de agrupamiento) tipo 2 para crear un modelo de predicción. En la tercera fase se utiliza una red Fuzzy neural (clasificación por peso) tipo-2 para realizar el razonamiento de la predicción. Este modelo ha superado a los modelos clásicos de predicción financiera basados en índices.

Montero (2016) cuando la variable dependiente es un porcentaje, se puede estimar un modelo de probabilidad lineal, sobre todo si el modelo solo tiene valores intermedios. Sin embargo, cuando los porcentajes están muy próximos a los extremos superior o inferior ya no se comportan igual cuando están en mitad de la tabla porque se “frenan”, se “tuercen” acotados los límites superior e inferior. Entonces hay que estimar un modelo de respuesta fraccional, que es similar a un logit o probit. Hay dos versiones similares francreg y betareg que se pueden

estimar como logit, probit log-log, entre otros. La única diferencia entre ambas es que la última no se puede ajustarse cuando se tienen 0 y 1 exactos en la variable dependiente.

Trucco (CEPAL, 2014) en el documento educación y desigualdad en América latina de la serie políticas sociales, destaca el papel fundamental de la educación en el desarrollo social, no limitando su importancia únicamente a los años de escolaridad, señalando que la calidad de esta potencializa en el aumento en el IDH en todas sus dimensiones, reafirmando el papel destacado de la educación y justificando su importancia.

Abuín (2007) indica que al momento elegir las variables que se consideren como generadoras, estas deben poseer un sentido numérico, se debe evitar variables que se repitan, deben cumplir con un sentido teórico asociado al problema y que entre la variable dependiente y las generadoras debe darse una relación de proporcionalidad.

Llaugel (2011) sugiere emplear el método de mínimos cuadrados para construir el modelo de regresión múltiple, permite un mejor ajuste entre variables dependiente y regresoras. Utilizar el coeficiente de determinación para la calidad del ajuste que ofrece la ecuación de regresión.

Walpole Myers y Myers, (2012) resaltan que el coeficiente de correlación únicamente mide la fuerza existente en un modelo lineal, mientras el coeficiente de determinación puede ser usado en relaciones no lineales y en relaciones con dos o más variables generadoras. Teniendo, el coeficiente de determinación una mayor utilidad.

“Cualesquiera conclusiones acerca de una relación causa y efecto deben basarse en los conocimientos de los especialistas en la aplicación de que se trate.” (Anderson, Sweeney y Williams, 2001, p. 565). Esto significa que la experiencia del analista y la profundidad de sus conocimientos sobre el tema son más importantes que propiamente el análisis de regresión que se realice.

En general se pueden destacar los siguientes elementos, una forma objetiva aceptada de poder comparar resultados es en forma de índices o porcentajes, indistintamente del tamaño de las poblaciones, si bien los modelos de regresión en consenso pueden ser aplicados, la limitante que da propiamente el resultado de la variable dependiente y de los propios índices utilizados para su predicción sufren un efecto de amortiguamiento o frenado, que determina la necesidad de un tratamiento especial, asociado a los datos composicionales, y modelos de regresión múltiple o multivariado, y en algunos aplicando métodos posteriores a la regresión que le den robustez a las predicciones finales.

3. PLANTEAMIENTO DEL PROBLEMA

3.1. Contexto general

El índice de desarrollo humano (IDH) es una forma aceptada, generalmente, de medir la calidad de vida de los habitantes de un país o región, este índice se construye tomando en cuenta tres dimensiones, las cuales son salud, riqueza y educación, siendo esta última, fundamental para alcanzar las primeras dos; esta información es recogida y presentada por el Programa de las Naciones Unidas para el Desarrollo conocido por sus siglas PNUD, y que según informes presentados por este programa actualmente le otorgan a Guatemala un índice que lo sitúa por debajo de la posición 128 de 192 países, ubicando a la nación en un nivel medio mundial, pero que es el más bajo de América Latina y de Centro América, aunque ha mostrado una tendencia a conservar dichas posiciones e índices en los últimos años.

El hecho de que Guatemala se haya estancado en su Índice de Desarrollo Económico es una muestra del estancamiento en materia de las políticas que favorecen el desarrollo de la sociedad como tal, a pesar de que el PNUD existe desde 1965 (con presencia en Guatemala desde 1975) y aporta esta información de carácter objetivo, se evidencia la poca eficiencia de las políticas aplicadas en la búsqueda del desarrollo integral de la sociedad

3.2. Descripción del problema

Guatemala no ha sido capaz de elevar sus índices de desarrollo humano en los últimos 5 años, y se sospecha que es debido a políticas poco eficientes en el

área de la educación, una de las 3 dimensiones para la construcción del índice, pero también como un elemento indirecto que afecta a las otras dos dimensiones, en el caso de salud donde no se dispone de personal capacitado para prestar estos servicios y en el caso del producto interno bruto (PIB) donde el aporte de su contribución individual es escaso en comparación con otros países donde se dispone de mano de obra especializada debido a su formación.

Se desconoce si existe algún grado significativo de correlación entre la proporción del IDH correspondiente a la dimensional de educación y el porcentaje de inversión, del PIB, que el estado hace al rubro de educación.

Se requiere un modelo estadístico que permita pronosticar la dimensión de educación en función de las variables generadoras como el porcentaje del PIB destinado a educación y que sean estadísticamente más significativas, con la finalidad de alcanzar los estándares educativos mínimos mundiales para Guatemala.

3.3. Formulación del problema

El planteamiento de los objeticos se desarrollará por medio de un análisis crítico a partir de una serie de preguntas con una clara orientación estadística que permitirá dar respuestas al problema.

3.3.1. Pregunta central

¿Cuál es el modelo con máximo ajuste que permita predecir el valor de la dimensión de educación del índice de desarrollo humano (IDH) en función de la inversión en educación como un porcentaje del PIB y el nivel de cobertura neta educativa en los diferentes niveles?

3.3.2. Preguntas auxiliares

- ¿Cuál es el nivel de correlación entre la dimensión educativa del índice de desarrollo humano (IDH) y la inversión en educación como un porcentaje del PIB y tasa de cobertura educativa neta en los diferentes niveles?

- ¿Cuál es el modelo estadístico optimo que describe mejor el comportamiento de la dimensión educativa del índice de desarrollo humano (IDH) cuando hay variaciones del presupuesto de educación en su forma porcentual y en las tasas de cobertura neta educativa de los diferentes niveles?

- ¿Qué diferencias existen entre los modelos de inversión de educación de los países con valores en la dimensión de educación superiores al nacional y que tienen un contexto similar?

3.4. Delimitación del problema

El problema se analizará con los datos históricos de inversión en educación en Guatemala, y los resultados de la dimensión de educación asociados a ellos de manera anual, en el periodo comprendido del 2004 al 2019, se contemplará la tasa de cobertura educativa neta de educativa.

La información puede ser recopilada del Ministerio De Educación (MINEDUC), Instituto Nacional de Estadística (INE), Ministerio de Finanzas Públicas (MINFIN) y del Programa de las Naciones Unidas para el Desarrollo (PNUD).

4. JUSTIFICACIÓN

La línea de investigación utilizada es la de pronósticos, ya que se busca la construcción de un modelo de regresión, que permita ofrecer una estimación a la dimensión de desarrollo a partir de datos históricos a nivel nacional en educación, y tomando como referencia modelos construidos, de la misma manera, en las sociedades latinoamericanas con los índices más altos.

Estudiar los efectos que influyen en la educación como un índice de desarrollo humano brindará las herramientas de conocimiento que permitan plantear políticas de desarrollo más efectivas en materia de educación, y que además de ser una clara muestra de la mejora en la calidad de vida de los ciudadanos, también sirve como incentivo entre otros factores para la atracción de inversión extranjera, creación de fuentes de trabajo con mano de obra especializada y mucho mejor remunerada, que influye de forma directa en el incremento del PIB, que también forma parte de las dimensiones usadas en el cálculo del índice de desarrollo humano.

La educación como tal también es en sí una herramienta que apoya procesos como investigación y desarrollo, que es otra forma de generar emprendimientos y fuentes de trabajo no dependientes de la inversión extranjera.

La mejora educativa también permitir una mayor diversidad de profesiones, entre ellas las relacionadas con la rama de la medicina, ayudando a incrementar otro indicador de desarrollo humano como lo es la salud.

5. OBJETIVOS

5.1. General

Construir un modelo estadístico con nivel máximo de ajuste, por medio del análisis de regresión, que ofrezca una herramienta para la predicción de la dimensión educativa del índice de desarrollo humano (IDH) en función de la inversión en educación como un porcentaje del PIB y de las tasas de cobertura educativa neta en los diferentes niveles.

5.2. Específicos

- Estimar el nivel de correlación entre la dimensión de educación del índice de desarrollo humano (IDH) en función de la inversión en educación como un porcentaje del PIB y de las tasas de cobertura educativa neta en los diferentes niveles, mediante el análisis de regresión, para aplicarlo como criterio de selección del modelo óptimo.

- Seleccionar el mejor modelo estadístico, que ayude a predecir con mayor precisión la dimensión de educación, en función de la inversión realizada en forma porcentual del PIB, evaluando los diferentes modelos desarrollados, utilizando los criterios de información Akaike (AICc) y Bayesiano (BIC).

- Comparar la inversión otorgada a educación en Guatemala con respecto a Bolivia, México y Perú, países con contextos similares al guatemalteco,

por medio de modelos estadísticos de regresión que proyecten el comportamiento de la dimensión de educación en cada uno de ellos.

6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN

Con la construcción del modelo de predicción, se busca ofrecer argumentos con fundamento estadístico que permitan medir la importancia de la relación entre la dimensión educación y el porcentaje del PIB asignado a educación, nivel de cobertura educativa neta en los diferentes niveles en Guatemala, y se utilice como una herramienta confiable en la predicción en políticas de desarrollo.

La información será obtenida de las siguientes fuentes, Instituto Nacional De Estadística INE y sus similares en los países de Bolivia, México y Perú, otra fuente de información serán los programas de las naciones unidas para el desarrollo PNUD establecidos en cada país, gracias a que es información de dominio público.

El proceso da inicio analizando las variables más importantes de los modelos de los países con los que se realizara la comparación y que hayan demostrado tener mayor significancia, estas variables se incluirán en el modelo.

Los modelos propuestos serán evaluados rigurosamente sometiéndolos a diferentes pruebas entre las que destacaremos: de significancia de coeficientes, coeficiente de determinación ajusto, pruebas de normalidad de residuos y prueba de homocedasticidad residuos, buscando con ello determinar el modelo que proporciones el mejor ajuste.

7. MARCO TEÓRICO

7.1. Modelos de regresión y predicciones

Llevar a cabo la construcción de un modelo de regresión y predicción confiable comienza con la calidad de los datos, en algunos casos simplemente son tomados y utilizados en la forma en la que se encuentran, y en otros casos requiere un tratamiento o transformación previa, que permita utilizar dicha información de la manera más adecuada, en este sentido la información debe ser sometida a ciertas consideraciones que deben tomarse en cuenta.

Construir un modelo estadístico que permita predecir resultados con un grado de confiabilidad aceptable entre al menos dos variables, una variable objetivo y una o más variables generadoras, requiere de una buena cantidad de herramientas que pueden ayudar con la operatoria pero que no pueden sustituir el análisis de un profesional de la información.

Dentro de los aspectos a considerar, previo a iniciar la construcción del modelo deseado se debe cumplir con un análisis de la información que habrá de ser considerada en un espacio de tiempo determinado, empleando la técnica de análisis de correlación la cual mide el nivel de asociación entre variables. Cuando los resultados del análisis de correlación confirman la viabilidad de los datos, se puede construir el modelo que tenga como fin predecir a futuro los valores en función de las variables evaluadas previamente, este proceso lleva por nombre análisis de regresión, que medirá el nivel de confiabilidad de nuestro modelo.

7.1.1. Supuestos estadísticos de los datos

La calidad del modelo de regresión está en función de la calidad de la información o datos que se utilicen, en el caso de la construcción de un modelo es imprescindible garantizar la calidad de los datos y por consiguiente del modelo que se construya a partir de ellos, existe una serie de supuestos sobre el comportamiento de los datos y sus residuos, que deben ser cumplidos para este cometido.

La gran mayoría de autores destacan que los supuestos estadísticos que deben ser verificados previo a la construcción de un modelo son los siguiente: homocedasticidad, homogeneidad, independencia, linealidad y normalidad.

7.1.1.1. Homocedasticidad

Según Hair, Anderson, Tatham y Black (1999) los datos presentan homocedasticidad cuando se analiza la dispersión de la varianza de ellos y se puede identificar un comportamiento constante, la uniformidad de la varianza de los datos nos lleva a concluir que existe homocedasticidad en caso contrario heterocedasticidad.

Guàrdia, Freixa, Peró, y Turbany (2008) indican que cuando se está seguro la muestra presentan un comportamiento normal, pueden ser aplicada la prueba F de Snedecor para determinar la igualdad de varianzas, que ofrece una gran potencia, pero es muy sensible a las desviaciones en la distribución normal, una segunda alternativa la prueba de *Bartlet*.

7.1.1.2. Homogeneidad

Gómez, Aparicio y Patiño (2010) indican que el supuesto de homogeneidad es observable cuando la media o la varianza de los datos es susceptible a los cambios, muchas veces presentado por un proceso inadecuado en la toma de datos o por la presencia de datos atípicos, sugiriendo para su comprobación la aplicación de la prueba *t-student* para evaluación de la homogeneidad.

7.1.1.3. Independencia

Díaz (2006) hace énfasis en que la independencia de los datos está fuertemente asociada a la aleatoriedad de estos, por lo que sugiere analizar el comportamiento del cambio signos en los residuos en busca de series largas de signos sin cambio ya sea positivos o negativos, de ser así, recomienda la aplicación de una prueba de rachas para determinar la independencia de los datos.

7.1.1.4. Linealidad

Hair et al. (1999) destacan que el supuesto de linealidad tiene implicación directa al realizar un análisis de regresión, por la dependencia que tiene de la correlación dada entre las variables que sean analizadas, a pesar de ello, no es posible representar los efectos no lineales. Si en caso se diera la no linealidad se hace necesario el análisis de todas las relaciones para determinar de qué manera influyen en la correlación. Con todo esto lo que se pretende es determinar si los datos se logran ajustar a una recta de regresión, para que el modelo pueda ser considerado válido.

Lind, D., Marchal, W. y Wathen, S. (2012) recomiendan que se verifique el cumplimiento del supuesto de linealidad, evaluando la correlación entre las variables, para demostrar la existencia de relación entre las mismas, por medio de la prueba de significancia del coeficiente de correlación.

La prueba de significancia compara un valor t calculado versus un valor t teórico donde un valor calculado mayor a un valor t teórico confirma la significancia del coeficiente de correlación y por consiguiente del supuesto de linealidad.

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

7.1.1.5. Normalidad

Hair et al. (1999) destaca la importancia de la prueba del supuesto de normalidad, debido a que la mayor parte de las pruebas estadísticas paramétricas que se pueden realizar sobre los datos, son válidas cuando los datos se ajustan a una distribución normal.

Para evaluar el supuesto de normalidad se dispone de diversas pruebas y la adecuada selección de ellas, esto permite identificar cuando la distribución de los datos se ajusta a una distribución normal, como se puede apreciar a continuación.

Una forma visual de identificar la normalidad en el comportamiento de los datos o de sus residuos es por medio de un gráfico *Q-Q plot*, el cual muestra los datos observados versus los datos esperados, donde se espera que la dispersión de los puntos se encuentre en su gran mayoría alrededor de una recta a 45°

desde el origen, esto nos da indicios de que el supuesto se cumple, aunque vale resaltar que no es una prueba definitiva.

Para la comprobación del supuesto de normalidad, Guisande (2006) sugiere la utilización de la prueba de Shapiro-Wilks que por lo general se recomienda para muestras no mayores a 30 datos, donde se calcula un p-valor que debe ser comparado con el nivel de significancia deseado, que en caso de ser mayor se concluye que los datos o sus residuos tiene un comportamiento que se ajusta a una distribución normal, una prueba de normalidad alternativa a Shapiro-Wilks es la de Ryan-Joiner, ya que ambas se basan regresión y correlación.

$$R_p = \frac{\Sigma Y_i b_i}{\sqrt{s^2(n-1)\Sigma b_i^2}} \qquad \text{(Ec.2)}$$

Donde:

Y_i: Observaciones ordenadas de los datos

b_i: Ponderación normal de los datos ordenados

S^2: Varianza muestras.

Según Marques (2001) cuando una serie o muestra posee más de 50 datos la prueba más recomendable para la comprobación del supuesto de normalidad es Kolmogórov–Smirnov aunque también es posible utilizar Anderson-Darling ya que ambas basadas en la distribución empírica. Donde el objetivo es encontrar un p-valor que sea mayor al nivel de significancia que se pretende, para concluir si se cumple con el supuesto.

Kolmogorov-Smirnov

$$D = \max\{D^+, D^-\}$$

Donde:

D^+: $max_i\left\{\frac{i}{n} - Z_{(i)}\right\}$

D^-: $max_i\left\{\left(Z_{(i)} - \frac{i-1}{n}\right)\right\}$

Z : $F(X_{(i)})$

$F_{(i)}$: $Función\ de\ distribución\ normal$

$x_{(i)}$: $Estadístico\ del\ i^{esimo}\ orden\ de\ una\ muestra\ aleatoria\ 1 \leq i \leq n$

n : $Tamaño\ de\ la\ muestra$ (Ec.3)

Anderson-Darling

$$A^2 = -N - \frac{1}{N}\sum(2i-1)(lnF(Y_i) + \ln(1 - F(Y_{N+1})))$$

$F(Y_i)$: $Función\ de\ distribución\ normal\ acumulada$

Y_i : $Observaciones\ ordenas\ de\ los\ datos$ (Ec.4)

7.1.2. Análisis de correlación

Lind et al. (2012) recomienda como paso inicial para evaluar la relación entre dos variables la construcción de diagramas de dispersión, con el fin de ofrecer una representación visual del comportamiento de estas, aunque al tratarse de una inspección visual no puede considerarse una prueba definitiva para identificar algún tipo de tendencia, pero sí como un indicio.

Después de elaborado el diagrama de dispersión, y haber identificado algún patrón o tendencia asociado a un modelo particular, se procede, de ser necesario, a realizar las transformaciones que ayuden a linealizar el comportamiento de los

datos, que se puede observar en un nuevo diagrama de dispersión, repitiendo este paso hasta que se esté satisfecho con el resultado.

Cuando los datos ya presentan esta tendencia lineal se puede proceder con un análisis cuantitativo de la relación entre los datos observados, para ello calcularemos el coeficiente de correlación tal y como se indica en la Ec.5.

Navidi (2006), ve el cálculo del coeficiente de correlación como una medida para cuantificar de manera porcentual que tan fuerte es la relación entre las variables, siendo un valor absoluto del resultado próximo a la unidad un indicador de una relación muy fuerte y lo opuesto cuando el resultado está próximo a cero.

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{S_x}\right)\left(\frac{y_i - \bar{y}}{S_y}\right) \qquad \text{(Ec.5)}$$

Para el cálculo del coeficiente de correlación se utilizarán los valores transformados de haberse dado el caso, con ellos se calculará las medias de cada uno y sus respectivas desviaciones, con estos valores ser alimenta la ecuación y calcular el valor r, para su interpretación.

7.1.3. Análisis de regresión

Walpole, Myers, Myers y Ye (2012), señalan que en todo modelo de regresión vamos a encontrar una variable dependiente o de respuesta y al menos una variable que será la variable independiente o regresora que tradicionalmente se representan como “Y” y “X” respectivamente, aunque esto no quiere decir que no se puedan representar bajo otra simbología.

Lind et al. (2012) simplifica la definición a una ecuación formada por al menos dos variables, una de ellas dependiente y el resto independientes, que refleja la relación gráfica y cuantificada entre las variables y que puede ser utilizada para realizar predicciones futuras.

La mayoría de los autores consideran el modelo básico de regresión, el que se presenta a continuación:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{(Ec.6)}$$

Donde β_0 y β_1 representan el intercepto en el eje "y" y la pendiente respectivamente, por otra parte "ε_i" constituye el término aleatorio. Si el valor exacto del término aleatorio fuera conocido se podría conocer el valor de Y; pero, como solo puede estimarse, el modelo de regresión queda simplificado de la siguiente forma:

$$y = \beta_0 + \beta_1 x \quad \text{(Ec.7)}$$

Donde β_0 y β_1 se obtiene a partir de las siguientes ecuaciones:

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{(Ec.8)}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad \text{(Ec.9)}$$

Webster (2000) señala que el objetivo del análisis de regresión es determinar la ecuación de la recta que mejor se ajuste a los datos reales

obtenidos a partir de una muestra. Para encontrar la ecuación de la recta se utiliza el método matemático conocido como mínimos cuadrados ordinario, en donde el resultado obtenido será los coeficientes β_0 y β_1 de la recta, y que esta se ajuste lo mejor posible a todos los valores de la muestra representativa.

Posada y Noguera (2007) destacan que el criterio de la información de Akaike (AIC) considera cambios en la bondad de ajuste y el número de parámetros existentes en el modelo que al ser evaluados en la ecuación AIC ofrecen un parámetro de la calidad del modelo y cuando se tiene varios modelos diseñados para realizar la misma predicción, el mejor modelo se determinara por el que ofrezca el menor valor.

Existentes dos variantes para el cálculo del criterio de la información Akaike, AIC, utilizado para muestras con más de 50 datos y AICc cuando el conjunto de datos es pequeño.

$$AIC = 2k - 2 \times \ln(L)$$
$$k: \ Número\ de\ parametros\ del\ modelo.$$
$$l\,n(L): \ Función\ de\ log - verosimilitud.$$

(Ec.10)

$$AICc = AIC + \frac{2k(k+1)}{N-k-1}$$
$$N: \quad El\ tamaño\ de\ la\ muestra\ de\ datos$$

(Ec.11)

Posada y Noguera (2007) indican la importancia del calcular el criterio de información Bayesiano (BIC), para los diferentes modelos propuestos como un indicador de la bondad de ajuste de la función de verosimilitud, la cantidad de

datos de la muestra y el número de parámetros con que cuenta el modelo, siendo la mejor opción el modelo con el BIC de menor valor.

$$BIC = -2 \times \ln(L) + \ln(N) \times k$$
$$k: Número\ de\ parametros\ del\ modelo$$
$$N: El\ tamaño\ de\ la\ muestra\ de\ datos$$
$$\ln(L): Función\ de\ log-verosimilitud$$

(Ec.12)

7.1.4. Series temporales

González (2009) señala que cuando se identifica una serie temporal, en primer lugar, existe un orden cronológico, no hay independencia entre las observaciones, y que incluso muchas de ellas manifestarán dependencia entre ellas, algo que debe ser tomado en cuenta en el proceso de análisis.

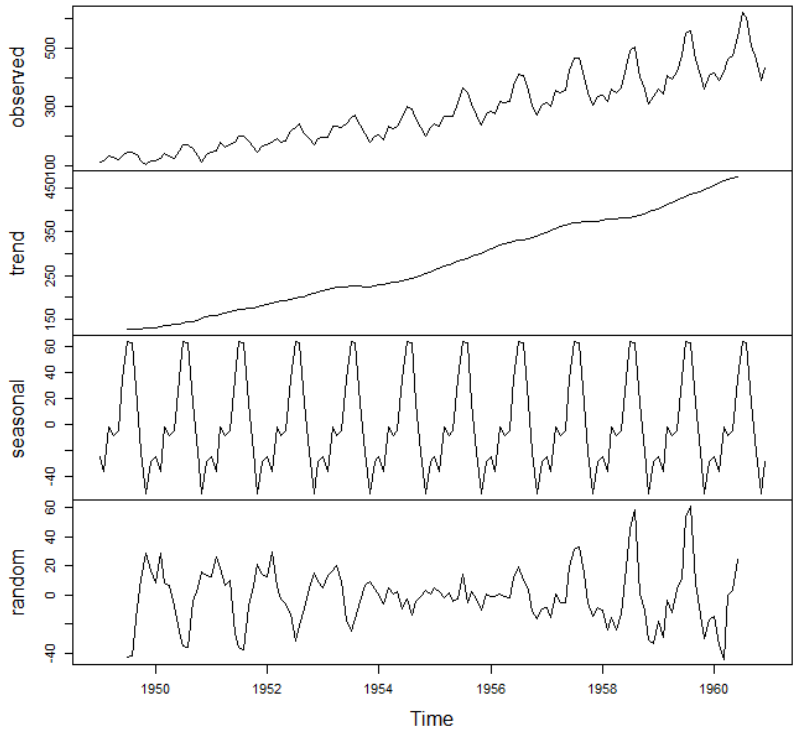
Debido a la naturaleza de las series temporales, los modelos desarrollados a partir de ellas pueden aprovechar la dependencia entre observaciones y a partir de ello proyectar el comportamiento futuro relativamente inmediato. Las técnicas desarrolladas para el análisis de datos o registros influenciados de manera significativa por el tiempo son un área de estudio a la cual se le conoce como análisis de series temporales.

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

Lind et al. (2012), amplia la definición de series temporales, como un conjunto de registros ordenados de forma cronológica, y la periodicidad de su registro se puede realizar de forma anual, mensual, semanal, diaria o de cualquier manera en la que el tiempo pueda ser medido.

Dadas las condiciones que presentan las series de tiempo es necesario analizar otros aspectos como la estacionalidad, la tendencia, la aleatoriedad y el ajuste.

Figura 1. **Descomposición de una serie de tiempo aditiva en: observaciones, tendencia, estacionalidad y aleatoriedad**



Fuente: Escutia, I. (2019) *Descomposición de series de tiempo*. Consultado el 9 de noviembre de 2020. Recuperado de RPubs by RStudio: https://rpubs.com/ItzelEscutia/546278

Antúnez (2011) señala que la estacionalidad se puede entender como un factor que se repite en espacios de tiempo similares y que por lo general no suele ser superior a 12 meses, sugiriendo aplicar la prueba de Dickey Fuller, para su determinación.

La tendencia, según González (2009), se puede presentar tanto lineal como no lineal, en el comportamiento general de los datos y muestra cómo se comporta la media a lo largo del tiempo.

Por último, Gómez et al. (2010) se refiere al ajuste de los datos con respecto a una función de distribución de probabilidad.

7.2. Índice de desarrollo humano

> El índice de desarrollo humano (IDH) se creó para hacer hincapié en que la ampliación de las oportunidades de las personas debería ser el criterio más importante para evaluar los resultados en materia de desarrollo. El crecimiento económico es un medio que contribuye a ese proceso, pero no es un objetivo en sí mismo. (PNUD, 2016, pár. 1)

Dicho de forma sencilla el índice de desarrollo humano es un indicador porcentual que refleja que tanto se satisfacen los estándares básicos mundiales en lo referente a la calidad de vida en una sociedad como la guatemalteca.

7.2.1. Programa de las Naciones Unidas para el Desarrollo

El Programa de las Naciones Unidas para el Desarrollo (PNUD) fue establecido en 1965, en el 2010 el índice de desarrollo humano (IDH) sufrió un ajuste en el método utilizado para su cálculo y es el método que actualmente se

utiliza, este programa está presente en muchos países del mundo y en particular en Guatemala desde el año de 1975, para aportar información que oriente al desarrollo del país.

> El Programa de las Naciones Unidas para el Desarrollo (PNUD) es la red mundial de la ONU para el desarrollo, que propugna el cambio y hace que los países tengan acceso al conocimiento, a la experiencia y a los recursos necesarios para ayudar a que las personas se labren un futuro mejor. El programa está presente en 177 países y territorios, y colabora con gobiernos y ciudadanos para que den con sus propias soluciones frente a los desafíos que plantea el desarrollo nacional y mundial. De este modo, a medida que desarrollan su capacidad local, los países se benefician del personal del PNUD y de su amplia variedad de asociados para obtener resultados. (ONUSIDA, 2015, página 2)

7.2.2. Dimensiones de desarrollo

Como se menciona en el Informe nacional de desarrollo humano Guatemala, (s.f.) este se descompone en tres dimensiones básicas las cuales son disfrutar de una vida larga y saludable (salud), acceso a educación (educación) y nivel de vida digno (nivel de vida).

7.2.3. Dimensión de educación

Como se indica en el informe nacional de desarrollo humano Guatemala, (s.f.) la dimensión de educación es un indicador compuesto que toma en cuenta aspectos como la esperanza educativa en niños y la escolaridad obtenida por los adultos mayores de 25 años.

Este indicador mide el porcentaje de alcance en educación obtenido por un país con respecto a un estándar básico a nivel mundial, y esta dimensión aporta el 33 % del peso del índice de desarrollo humano que puede ser alcanzado por una sociedad o país.

7.2.4. Producto interno bruto

“Indicador económico que refleja el valor monetario de todos los bienes y servicios finales producidos por un país o región en un determinado periodo de tiempo, normalmente un año. Se utiliza para medir la riqueza que genera un país.” (Sevilla, 2021, pár 1). Es la definición técnica que normalmente es acepta para referirse al producto interno bruto.

8. PROPUESTA DE ÍNDICE DE CONTENIDOS

9. METODOLOGÍA

9.1. Características del estudio

El enfoque del estudio propuesto es cuantitativo, ya que pretende determinar un valor porcentual, el cual es la dimensión de educación, una de las tres dimensiones que conforman el índice de desarrollo humano (IDH), que estará en función del porcentaje de producto interno bruto (PIB) asignado a dicho rubro y de las tasas de cobertura educativa neta de los diferentes niveles.

El alcance es correlacional, dado que se desea inferir el efecto de las variables explicativas: porcentaje de producto interno bruto (PIB) asignado a dicho rubro y las tasas de cobertura educativa neta en los diferentes niveles.

El diseño adoptado será observacional, pues busca construir un modelo de regresión estadístico tomando parte de la información socioeconómica de Guatemala y otros países latinoamericanos de contexto similar, específicamente en el área de educación, se analizará en su estado original sin ninguna manipulación; además será transversal pues se estudiarán los datos de manera independiente para cada país involucrado en el estudio. Sera longitudinal, pues se analizará el comportamiento de los mismos países a lo largo del tiempo con el propósito de la construcción de un modelo predictivo, funcional a mediano plazo.

9.2. Unidades de análisis

El objeto de estudio es la dimensión de educación, del periodo comprendido del año 2004 al 2019 y datos socioeconómicos relacionados directamente con ella, la cual se encuentra dividida en subpoblaciones dadas por los espacios geográficos específicos ordenados por países, entre ellos a Guatemala, Bolivia, México y Perú, de la cual se extraerán los datos anuales del periodo ya mencionado en cada país, que serán estudiadas en su totalidad.

9.3. Variables

A continuación, se presentan las variables que serán incluidas en el análisis estadístico con la nomenclatura propuesta y una breve descripción de sus características más importantes.

Tabla I. **Variables del estudio**

Variable	Definición teórica	Definición operativa	Escala
Dimensión Educación (Y)	Nivel de alcance logrado según los estándares educativos, del Programa de las Naciones Unidas para el Desarrollo.	Porcentual, adimensional, rango de 0 a 1	De razón
Porcentaje del PIB invertido en educación (X_1)	Dato histórico de la inversión porcentual que ha asignado cada gobierno en el rubro de educación en los diferentes países del estudio.	Porcentual, adimensional, rango de 0 a 1	De razón
Tasa de cobertura educativa neta en preprimaria (X_2)	Datos recopilados del Ministerio de Educación e institutos de estadística de cada país sobre cantidad relativa de personas en edad nivel preprimaria.	Porcentual, adimensional, rango de 0 a 1	De razón
Tasa de cobertura educativa neta en primaria (X_3)	Datos recopilados de Ministerio de Educación e institutos de estadística de cada país sobre cantidad relativa de personas en edad nivel primaria.	Porcentual, adimensional, rango de 0 a 1	De razón
Tasa de cobertura educativa neta en secundaria (X_4)	Datos recopilados de Ministerio de Educación e institutos de estadística de cada país sobre cantidad relativa de personas en edad nivel secundaria.	Porcentual, adimensional, rango de 0 a 1	De razón

Fuente: elaboración propia.

9.4. Fases del estudio

A continuación, se describen dichas fases.

- Fase 1: revisión bibliográfica

El análisis bibliográfico se realizará sobre los libros, revistas, tesis y artículos, consultados para elaborar la investigación. Son documentos con no

más de 10 años de antigüedad, especializados en estadística, socioeconomía, y en el caso de las tesis solo de nivel de postgrado.

- Fase 2: gestión y recolección de la información

Los datos utilizados corresponden al total de la población, que proviene del informe anual generado por cada país en el Programa de las Naciones Unidas para el Desarrollo y de instituciones estatales oficiales.

- Fase 3: análisis preliminar de la información

Los datos serán analizados sometiéndolos a pruebas de aleatoriedad, normalidad, homogeneidad, entre otras, para determinar su comportamiento y como criterio para confirmar la viabilidad del tratamiento de estos en la construcción del modelo de regresión.

- Fase 4: modelización

Construcción de modelos a partir de los datos diagramas de dispersión que ayuden a visualizar la relaciona de cada variable generadora, y se considerarán las posibles transformaciones que ayuden a la construcción de un modelo funcional, incluidos los modelos multivariados y con datos composicionales para cada país.

- Fase 5**:** validación y selección

Evaluación de cada modelo bajo los criterios de las siguientes pruebas de regresión de significancia de coeficientes, coeficientes de determinación ajustado, homocedasticidad y de normalidad de residuos y con base en los resultados la selección del mejor modelo estadístico en base al nivel máximo de ajuste, al nivel de correlación.

- Fase 6: redacción de informe final

El informe final detalla el proceso empleado para responder a las preguntas que se depreden del problema y que se convierten en los objetivos del estudio, describe las diferentes herramientas estadísticas empleadas, y las conclusiones obtenidas a partir de ellas.

10. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN

Análisis de datos: los datos serán analizados mediante gráficos de dispersión para determinar si existe relación significativa entre variables y se completara con análisis de regresión, y a partir de los resultados, seleccionar las variables de mayor significancia en el modelo.

Análisis de modelos: se realizará la prueba de R2 ajustado para determinar el nivel de correlación de cada modelo, también se estimará la calidad del modelo utilizando las medidas de AIC y el BIC, y a partir de toda esta información determinar cuál es el modelo que ofrece el máximo ajuste.

Pruebas de bondad de ajuste: se realizarán prueba de bondad de ajuste utilizando para ello los residuos y a partir de ellos identifica, independencia, homocedasticidad y normalidad.

11. CRONOGRAMA

En la tabla siguiente se presenta el tiempo propuesto para el estudio planteado, descompuesto en fases por semana.

Tabla II. **Cronograma del estudio**

FASES	**TIEMPO (Semanas)**															
Semana	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
FASE 1: Revisión literaria	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
FASE 2: Gestión y recolección de la información	■	■	■	■	■	■	■									
FASE 3: Análisis preliminar de la información					■	■	■	■	■	■	■					
FASE 4: Modelización								■	■	■	■	■	■	■		
FASE 5: Informe Final													■	■	■	■
TIEMPO TOTAL:	16 semanas															

Fuente: elaboración propia.

12. FACTIBILIDAD DEL ESTUDIO

12.1. Recurso humano

Asesor de tesis ad honorem, secretaria contratada a destajo, revisor lingüista externo contratado por servicios profesionales, su contribución será requerida en momento específicos, para la elaboración del informe.

12.2. Recursos financieros

En la tabla siguiente se observa una estimación de los gastos requeridos para la elaboración del trabajo de investigación, durante todo su desarrollo.

Tabla III. **Presupuesto asignado al estudio**

Elemento	Unidad	Costo Unitario (Q.)	Cantidad necesaria	Costo (Q.)
Papel	resma	35.00	4	140.00
Tablet	Tablet	2300.00	1	2300.00
Internet	Mes	510.00	4	2040.00
Consumo de energía	Mes	25.00	4	100.00
Gastos de asesoramiento	Comida	200.00	4	800.00
Servicios secretariales	horas	100.00	8	800.00
Servicios de lingüística	servicio	500	1	500.00
Computadora portátil	-	0.00	1	0.00
Software estadístico	-	0.00	1	0.00
Transporte	Galones	25.00	10	250.00
Imprevistos	-	-	-	1000.00
Total				7930.00

Fuente: elaboración propia.

12.3. Recursos tecnológicos

Hardware: se dispone de computadora portátil, equipo de impresión, periféricos, *tablet* y celular.

Software: se cuenta con plataformas Windows 10 y Android para *tablet* y teléfono, además se dispone de diferentes softwares de análisis estadístico.

12.4. Acceso a información y permisos

La información es de acceso público y se obtiene de los diferentes institutos de estadística, de programa de las naciones unidad para el desarrollo, y ministerios de finanzas de los diferentes países involucrados en el estudio.

12.5. Equipo e infraestructura

Se dispone de un espacio destinado como estudio, para trabajar durante todo el proceso de investigación y desarrollo del estudio.

13. REFERENCIAS

1. Anderson, D., Sweeney, D., Williams, T. (2012). *Estadística para Negocios y Economía.* México: Cengage Learning.

2. Barros, V., Gallegos, D. y Pavón, C., (2018). Muestreo para el levantamiento de datos acerca de la enseñanza de física experimental en Guayaquil*. Revista Lasallista de Investigación 15*(2), 223-231. Recuperado de http://repository.lasallista.edu.co :8080/ojs/index.php/rldi/article/view/ 1859/210210317

3. Enke, D., Grauer, M. y Mehdiyev, N. (2011). Stock Market Prediction with Multiple Regression,Fuzzy Type-2 Clustering and Neural Networks*. Procedia Computer Science 6*(2011), 201-206. Recuperado de https://doi.org/10.1016/j.procs.2011.08.038

4. Escutia, I. (9 de 11 de 2019). *Descomposición de series de tiempo*. Recuperado de RPubs by RStudio. https://rpubs.com/ItzelEscutia/546278

5. Frías-Navarro, D. (2011). *Técnica estadística y diseño de investigación.* España: Palmero ediciones.

6. Guisande, C. (2006). *Tratamiento de datos.* España: Universidad de Vigo.

7. Gutiérrez, H., y De la Vara, R., (2012) *Análisis y diseño de experimentos.* México: McGraw-Hill.

8. Lind, D., Marchal, W. y Wathen, S. (2012). *Estadística Aplicada a los Negocios y Economía* México: McGraw-Hill Interamericana.

9. Montgomey, D., Hines, W. (1996)*. Probabilidad y estadística para ingeniería y administración.* México: Compañía Editorial Continental.

10. Morduchowicz, A. (2006) *Los indicadores educativos y las dimensiones que los integran*. Buenos Aires, Argentina: IIPE – UNESCO.

11. Navidi, W. *(2006). Estadística para ingenieros.* México: McGraw-Hill Interamericana.

12. Pawlowsky-Glahn, V. & Buccianti, A. (2011), *Compositional data analysis: Theory and applications*, United Kingsdom: John Wiley & Sons.

13. Programa de las Naciones Unidas para el Desarrollo. (2016) *Informe nacional de desarrollo humano.* Guatemala: PNUD. Recuperado de http://desarrollohumano.org.gt/desarrollohumano/calculo-de-idh/

14. Posada, S., y Rosero, R. (2007). Comparación de modelos matemáticos: una aplicación en la evaluación de alimentos para animales. *Revista Colombiana de Ciencias Pecuarias 2*(2), 141-148.

Recuperado de https://www.redalyc.org/articulo.oa?id= 295023034006

15. ONUSIDA. (2015). *Programa de las Naciones Unidas para el Desarrollo.* Recuperado de https://www.unaids.org/sites/default/files/media_asset/ PNUD_es.pdf

16. Torrents, D., Fachelli, S. (2015). El efecto del origen social con el paso del tiempo: la inserción laboral de los graduados universitarios españoles durante la democracia. *Revista Complutense de Educación 26* (2) 331-349. Recuperado de https://revistas.ucm.es/index.php/RCED/article/view/43070/45512

17. Sevilla, A. (2021) *Producto interior bruto (PIB).* Recuperado de Economipedia https://economipedia.com/definiciones/producto-interior-bruto-pib.html.

18. *Triola, M. (2018). Estadística.* México: Pearson Educación.

19. Walpole, R., Myers, R., Myers, S., Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias.* México: Pearson Educación.

20. Webster, A. (2000). *Estadística aplicada a los negocios y la economía.* Colombia: McGraw-Hill.

21. Wooldridge, J. (2010). *Introducción a la econometría: un enfoque moderno*. México: Thomsom Learning.