



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA DE INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

“ANÁLISIS DE SERIES DE TIEMPO Y REGRESIÓN LOGIT DEL DESEMPLEO JUVENIL EN EL ECUADOR, PERIODO 2013 - 2018”

Trabajo de Titulación

Tipo: Proyecto De Investigación

Presentado para obtener el grado académico de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

AUTORA: TANIA FERNANDA TIERRA VILEMA

DIRECTOR: Dr. JORGE WASHINGTON CONGACHA AUSHAY

Riobamba – Ecuador
2020

©2020, Tania Fernanda Tierra Vilema

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo cita bibliográfica del documento, siempre y cuando se reconozca el Derecho del Autor.

Yo, Tania Fernanda Tierra Vilema, declaro que el presente trabajo de titulación es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor/autora asumo la responsabilidad legal y académica de los contenidos de este trabajo de titulación; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 21 de agosto de 2020

Tania Fernanda Tierra Vilema
CI: 060446490-9

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA DE INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

El Tribunal del Trabajo de Titulación certifica que: El trabajo de titulación: Tipo: Proyecto de Investigación, **ANÁLISIS DE SERIES DE TIEMPO Y REGRESIÓN LOGIT DEL DESEMPLEO JUVENIL EN EL ECUADOR, PERIODO 2013 - 2018**, realizado por la señorita **Tania Fernanda Tierra Vilema**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, el mismo que cumple con los requisitos científicos, técnicos y legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Ing. Johanna Enith Aguilar Reyes PRESIDENTE DEL TRIBUNAL	_____	2020 – 08 - 21
Dr. Jorge Washington Congacha Aushay DIRECTOR DE TRABAJO DE TITULACIÓN	_____	2020 – 08 - 21
Ing. Nancy Elizabeth Chariguaman Maurisaca MIEMBRO DEL TRIBUNAL	_____	2020 – 08 - 21

DEDICATORIA

A Dios: por haberme permitido llegar hasta este punto de mi vida, darme salud y sabiduría para lograr mis objetivos profesionales.

A mis padres: por ser ejemplos de perseverancia y constancia, sus consejos y motivación constante me han permitido ser una persona de bien.

A mi abuelita y madrina Paula: quien es mi motivación y ejemplo de lucha diaria para una vida solvente.

Abuelito Leonidas mi papito Vilema, con gratitud te lo dedico especialmente a ti, que te me adelantaste durante la realización de este trabajo, ahora desde el cielo junto a **Mamá María** son mis angelitos y estarán en primera fila viéndome cumplir mi meta profesional.

A mis hermanos Alex y Magaly: por ser un buen ejemplo de hermanos mayores, que no importa el orden en que se cumplan nuestras metas lo importante es llegar a cumplirlas.

A mis cuñados Aixa y Carlos: quienes de una u otra manera me han motivado a culminar mi carrera universitaria, he aprendido mucho de ustedes.

A mis sobrinas Dayanara y Emily: quienes en tiempos difíciles han sido mi motor e inspiración para no desistir y ser cada día la mejor mamá Tana.

Tania

AGRADECIMIENTO

A Dios por ser mi guía, apoyo y fortaleza en los momentos de dificultad y debilidad, permitiendo cumplir con éxito mi sueño más anhelado.

A la Escuela Superior Politécnica de Chimborazo, por darme la oportunidad de formarme profesionalmente y poder compartir con la sociedad los conocimientos adquiridos.

A los docentes de la carrera de Estadística Informática, por haber compartido sus conocimientos durante toda mi formación académica, de manera especial al Dr. Jorge Washington Congacha Aushay director del trabajo de titulación y la Ing. Nancy Chariguaman, quienes con sus conocimientos, tiempo, paciencia y motivación permitieron culminar con éxito el presente trabajo de investigación.

A mis padres por ser mi principal apoyo y sustento económico y moral.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	x
ÍNDICE DE GRÁFICOS.....	xi
ÍNDICE DE ANEXOS	xii
INDICE DE ABREVIATURAS	xiii
RESUMEN	xiv
SUMMARY	xv
INTRODUCCIÓN	1
CAPÍTULO I	
1 MARCO TEÓRICO REFERENCIAL	9
1.1 Encuesta Nacional de Empleo, Subempleo y Desempleo (ENEMDU).....	9
1.1.1 Clasificación de la población	9
1.1.2 Desempleo.....	10
1.2 Desempleo juvenil.....	12
1.2.1 Resumen del desempleo juvenil en Ecuador.....	14
1.2.2 Fundamentos del desempleo juvenil	14
1.3 Estudios de serie de tiempo en Encuestas de Hogares	14
1.3.1 Determinantes del desempleo juvenil en Encuestas de Hogares	15
1.4 Bases teóricas de estadística	17
1.4.1 Regresión logit (RL).....	22
1.4.2 Análisis de Series de tiempo	27
1.4.3 Metodología Box-Jenkins	28
1.4.4 Modelo ARIMA para datos estacionales.....	28
1.4.5 Método bootstrap.....	34
CAPÍTULO II	
2 MARCO METODOLÓGICO	36
2.1 Tipo y Diseño de Investigación	36
2.2 Población de estudio.....	37
2.3 Tamaño de la muestra.....	37

2.4	Método de muestreo	37
2.5	Recolección de información	38
2.6	Operacionalización de las variables	38
2.7	Alcance de la investigación	39
2.8	Análisis de datos	39
CAPÍTULO III		
3	ANÁLISIS Y DISCUSIÓN DE RESULTADOS	41
3.1	Tasa de desempleo juvenil por provincias	41
3.2	Análisis exploratorio de datos	42
3.3	Modelos de series de tiempo de la tasa de desempleo juvenil nacional.....	52
3.3.1	<i>AED del desempleo juvenil nacional y por provincias</i>	53
3.3.2	<i>Modelo ARIMA del desempleo juvenil en el periodo 2013 - 2018</i>	56
3.3.3	<i>Pronósticos</i>	60
3.4	Análisis de regresión logit	63
3.4.1	<i>Modelo para el desempleo juvenil</i>	63
CONCLUSIONES		66
RECOMENDACIONES		67
BIBLIOGRAFÍA		
ANEXOS		

ÍNDICE DE TABLAS

Tabla 1-1: Tasa de desempleo juvenil de Ecuador de DIC 2007 – DIC 2018	14
Tabla 2-1: Combinaciones más comunes de las variables respuestas y variables explicativas ..	21
Tabla 1-2: Tamaño de la muestra, por años y trimestres	37
Tabla 2-2: Operacionalización de las variables	38
Tabla 1-3: Tasas del desempleo juvenil en Ecuador por provincia y nacional 2013-2018	41
Tabla 2-3: Frecuencias de la variable desempleo en el período 2013 - 2018.....	42
Tabla 3-3: Frecuencias de la variable sexo.....	43
Tabla 4-3: Frecuencias de la variable edad	44
Tabla 5-3: Frecuencias de la variable relación de parentesco	44
Tabla 6-3: Frecuencias de la variable estado civil	45
Tabla 7-3: Frecuencias de la variable asiste a clases	46
Tabla 8-3: Frecuencias de la variable nivel de instrucción	47
Tabla 9-3: Frecuencias de la variable ingreso laboral	48
Tabla 10-3: Frecuencias de la variable rama de actividad	49
Tabla 11-3: Frecuencias de la variable grupo de ocupación	50
Tabla 12-3: Frecuencias de la variable recibe seguro social.....	52
Tabla 13-3: Test Dickey-Fuller para las 24 provincias.....	54
Tabla 14-3: Test Dickey-Fuller para las series diferenciadas	56
Tabla 15-3: Modelos <i>ARIMA</i> obtenidos para las 24 provincias del Ecuador	57
Tabla 16-3: Pronósticos de las provincias más grandes del Ecuador.....	60
Tabla 17-3: Pronósticos de la tasa nacional de desempleo juvenil	62
Tabla 18-3: Codificación de las variables independientes.....	63
Tabla 19-3: Resultados de la aplicación del modelo de regresión logit.....	65
Tabla 20-3: Tasas del desempleo juvenil nacional y provincial de Ecuador 2013-2018.....	73

ÍNDICE DE FIGURAS

Figura 1-1: Diagrama de Cajas y Bigotes	21
Figura 2-2: División geográfica de Ecuador	36

ÍNDICE DE GRÁFICOS

Gráfico 1-1: Clasificación de la Población Económicamente Activa (PEA) Ecuador	10
Gráfico 2-1: Tasa de desempleo juvenil nacional de Ecuador de DIC 2007–DIC 2018	13
Gráfico 1-3: Gráfico de frecuencias de la variable desempleo en el período 2013	42
Gráfico 2 -3: Gráfico de frecuencias de la variable sexo	43
Gráfico 3-3: Gráfico de frecuencias de la variable edad.....	44
Gráfico 4 -3: Gráfico de frecuencias de la variable relación de parentesco	44
Gráfico 5 -3: Gráfico de frecuencias de la variable estado civil	45
Gráfico 6 -3: Gráfico de frecuencias de la variable asiste a clases	47
Gráfico 7 -3: Gráfico de frecuencias de la variable nivel de instrucción	48
Gráfico 8 -3: Gráfico de frecuencias de la variable ingreso laboral.....	49
Gráfico 9 -3: Gráfico de frecuencias de la variable rama de actividad	50
Gráfico 10 -3: Gráfico de frecuencias de la variable grupo de ocupación	51
Gráfico 11 -3: Gráfico de frecuencias de la variable recibe seguro social	52
Gráfico 12 -3: Series de tiempo de la tasa de desempleo juvenil provincial y nacional 2013-2018	53
Gráfico 13 -3: Función de autocorrelación simple y parcial para la tasa de desempleo juvenil provincial y nacional del Ecuador 2013-2018	55
Gráfico 14 -3: Función de autocorrelación simple y parcial diferenciada para la tasa de desempleo juvenil provincial y nacional del Ecuador 2013-2018	56
Gráfico 15 -3: Pronósticos de la tasa de desempleo juvenil provincial 2019-2020.....	61
Gráfico 16 -3: Pronósticos de la tasa de desempleo juvenil nacional 2019-2020	62

ÍNDICE DE ANEXOS

ANEXO A: AVAL DE LA INVESTIGACIÓN

ANEXO B: TASAS DEL DESEMPLEO JUVENIL EN ECUADOR NACIONAL Y
PROVINCIAL 2013-2018

ANEXO C: CÓDIGO R

INDICE DE ABREVIATURAS

INEC:	Instituto Nacional de Estadística y Censos
OIT:	Organización Internacional del Trabajo
TNDJ:	Tasa nacional de desempleo juvenil
TD:	Tasa de desempleo
SBU:	Salario básico unificado
AED:	Análisis exploratorio de datos
FAS:	Función de autocorrelación simple
FAP:	Función de autocorrelación parcial
RL:	Regresión logit
MGL:	Modelos lineales generalizados
AIC:	Criterio de Información Akaike

RESUMEN

El presente trabajo de titulación tuvo como objetivo analizar mediante las técnicas de series de tiempo y regresión logit el desempleo juvenil en el Ecuador, período 2013-2018, para lo cual la información necesaria fue tomada del Instituto Nacional de Estadística y Censos (INEC), recolectada mediante la Encuesta Nacional de Empleo y Desempleo (ENEMDU). La técnica de series de tiempo permitió analizar la información de las 24 provincias del Ecuador en relación al tiempo, enfocándose en identificar cual es la magnitud del desempleo juvenil en el país y los cambios que presenta a lo largo del tiempo, además se aplicó la metodología Box-Jenkins en la cual se detectó que las series de tiempo presentaban estacionariedad, identificando modelos *SARIMA* y el cumplimiento parcial de supuestos teóricos, el mejor modelo estimado fué la tasa de desempleo juvenil nacional *SARIMA*(1,0,0)(1,0,1)₄ con un AIC= 70.67. Se seleccionó como mejor modelo probabilístico al modelo logit del año 2013, el cual muestra que las variables seleccionadas para el estudio fueron significativas, además indica que los factores determinantes del desempleo juvenil de los jóvenes ecuatorianos se relacionan con las variables sociodemográficas nivel de instrucción y la edad que mayor peso tienen en el modelo. Conjuntamente se determinó que si un joven ecuatoriano, cumple con las características de: ser mujer, tener 26 años, no ser jefe de hogar, estar sin pareja, asista a clases, tenga un nivel académico superior, sus ingresos sean menores al Salario Básico Unificado de \$357.50, no especifique el sector económico y grupo de ocupación laboral y no cuente con un seguro social, tendrá una probabilidad del 94.7% de encontrarse desempleado. En general se debe socializar con entes gubernamentales la evolución y los cambios a futuro de la tasa desempleo juvenil estudiada, para la toma de decisiones en medidas de contingencia que eviten el aumento de la misma.

Palabras clave: <ESTADÍSTICA>, <DESEMPLEO JUVENIL>, <SERIES DE TIEMPO>, <METODOLOGÍA BOX-JENKINS>, <REGESIÓN LOGIT>.



0273-DBRAI-UPT-2020

SUMMARY

The objective of this paper was to analyze youth unemployment in Ecuador, using time series and logit regression techniques, for the period 2013-2018. The necessary information was taken from the National Institute of Statistics and Census (INEC), collected through the National Survey of Employment and Unemployment (ENEMDU). The time series technique allowed for the analysis of information from Ecuador's 24 provinces concerning time, focusing on identifying the magnitude of youth unemployment in the country and the changes that it presents over time. Besides, the Box-Jenkins methodology was applied, in which it was detected that the time series were stationary, identifying SARIMA models and the partial fulfillment of theoretical assumptions, the best-estimated model was the national youth unemployment rate SARIMA (1,0,0)(1,0,1)₄ with an AIC= 70.67. The logit model of 2013 was selected as the best probabilistic model, which shows that the variables selected for the study were important. It also indicates that the determinants of youth unemployment among Ecuadorian youth are related to the sociodemographic variables of educational level and age that have the greatest weight in the model. It was jointly determined that if an Ecuadorian youth meets the characteristics of being a woman, being 26 years old, not being the head of household, being without a partner, attending classes, having a higher academic level, the income is less than the Unified Basic Salary of \$357.50, do not specify the economic sector and labor occupation group and do not have social security, it will have a 94.7% probability of being unemployed. In general, the evolution and future changes of the youth unemployment rate studied should be discussed with government agencies to make decisions on contingency measures to avoid an increase in the rate.

Keywords: <STATISTICS>, <YOUTH UNEMPLOYMENT>, <TIME SERIES>, <BOX-JENKINS METHODOLOGY>, <LOGIT REGRESSION>.

INTRODUCCIÓN

Desde el ámbito social el desempleo juvenil, es una de las principales preocupaciones de los gobiernos, ya que obstruye a la mejora del mercado laboral a nivel mundial, especialmente a los países en desarrollo, dado que la población joven en edad de trabajar presenta una tasa de desempleo (TD) mucho más alta que la del resto de la población, este fenómeno usualmente genera altas tasas de informalidad entre los jóvenes y es debido a muchos factores, como el nivel de instrucción y la falta de experiencia, que inducen a un círculo de sin experiencia no hay trabajo y sin trabajo no hay experiencia. (Fawcett, 2002).

A nivel mundial según cifras de la Organización Internacional de Trabajo (OIT, 2017), presentado en el informe *Panorama Laboral América Latina y el Caribe* la tasa de desempleo juvenil se redujo en un punto porcentual pasando de 13,7% en el 2016 a 12,7% en el 2017, despertando una mayor preocupación por la proporción de jóvenes, que viven en pobreza extrema o moderada, en particular en los países emergentes y en desarrollo. Para el año 2018 la tasa de desempleo juvenil fue de aproximadamente el 20%, tres veces superior a la de los adultos, del 6.5% (OIT, 2018). La inclusión laboral de los jóvenes es deficiente, puesto que uno de cada cinco jóvenes busca trabajo y lo encuentran, ocasionando así que ni estudien ni trabajen, este segmento de la población son denominados ninis¹ (OIT, 2013).

En el contexto ecuatoriano según las estadísticas del Instituto Nacional de Estadística y Censos (INEC), el desempleo se enfoca más al grupo de los jóvenes, dado que la mayoría por su edad tienen poca o insuficiente experiencia laboral, años de educación, el mismo hecho de ser mujer, entre otras variables impiden el ingreso al mercado y a lo largo del tiempo imponen un costo económico y social. La tasa de desempleo juvenil del 2018 fue de 8,25% ubicándose por debajo de países caracterizados como económicamente débiles como Perú, Bolivia y Paraguay, que presentaron tasas de desempleo juvenil de 18%, 8.5% y 15% respectivamente.

Lo anteriormente expuesto se vio reflejado en el estudio realizado por (Mullo y Marcatoma, 2018, pp.4-5) donde se muestra que Ecuador en el primer semestre del 2017 en todo el país existía 366.003 personas desocupadas, de este total 197.019 tenían entre 18 a 29 años, es decir el 59,32% de los desocupados ecuatorianos eran jóvenes. La tasa de desempleo juvenil evidencia que los jóvenes (9,26%) poseen una tasa de desempleo mayor en comparación al de los no jóvenes (4,49%). El problema es aún más grave para las mujeres que presentan el 12.48% de desempleo, mientras que en los hombres es del 6.99%. El origen de este problema es debido a muchos factores, entre ellos el nivel de instrucción y la falta de experiencia, que induce a un círculo de sin experiencia no hay

¹ En España este término se conoce como NENT (no estudian ni trabajan) o NINI.

trabajo y sin trabajo no hay experiencia. Por tal motivo, es importante analizar y observar los cambios que tiene este fenómeno para así tener una visión más clara del desarrollo de la sociedad.

El desempleo juvenil se relaciona con el género, ya que en los países de la Unión Europea el desempleo es significativamente mayor en mujeres que en hombres, lo que conlleva a la discriminación de género y por último está ligado a los fenómenos de desempleo y trabajo informal, en especial de los países en desarrollo (Ramírez, 2002).

De esta forma, esta investigación tiene como objetivo analizar mediante la utilización de técnicas de series de tiempo y regresión logit el desempleo juvenil en el Ecuador periodo 2013-2018. Este trabajo se dividirá en tres capítulos, los mismos que se presentan a continuación.

En el capítulo I se desarrolla la base teórica y conceptual necesaria para el desarrollo del proyecto de investigación, donde se presenta una breve descripción de las técnicas estadísticas de series de tiempo y regresión logit, análisis descriptivo, detección de atípicos e imputación de datos faltantes. Además, se presenta información del desempleo juvenil en el Ecuador.

En el capítulo II se plantea la metodología de a investigación, el tipo de investigación llevado a cabo, la población y muestra, e identificación y operacionalización de las variables. Para ello se utiliza la base de datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo del Ecuador (ENEMDU), de acceso público perteneciente a las estadísticas de trabajo del INEC².

En el capítulo III se expone los resultados alcanzados para la investigación, que se sustentaron en tablas de frecuencia, gráficos y análisis e interpretaciones, en función de cada uno de los objetivos específicos previamente planteados. Además se analiza las estadísticas de nuestro país, mediante dos tipos de análisis: 1) el análisis de series de tiempo del tipo ARIMA (Autoregressive Integrated Moving Average) para identificar el comportamiento real y próximo de la tasa de desempleo juvenil de Ecuador, la selección del mejor modelo predictivo está basado en los criterios de información del AIC (Criterio de Información de Akaike), con estimaciones para un horizonte de predicción de 3 meses³ y así visualizar la evolución del desempleo juvenil a lo largo del tiempo. 2) modelos de probabilidad no lineal como lo es el caso de la regresión logit, que explica cuáles son los factores sociodemográficos y sociolaborales influyentes del desempleo juvenil, estimando así la probabilidad de que un joven ecuatoriano se encuentre o no desempleado, al mismo tiempo que se pueda aportar con información e identificar cuáles son las variables que inciden en el desempleo juvenil. Todo se lo realiza utilizando el software estadístico R versión 3.5.2.

² Instituto Nacional de Estadística y Censos: fuente oficial de estadísticas de Ecuador.

³ Los 3 meses corresponden al período de cada trimestre, considerando que el cierre estadístico del presenta trabajo fue diciembre de 2018

Antecedentes

El desempleo juvenil es uno de los problemas sociales que ha venido trascendiendo desde años atrás, a nivel nacional e internacional, especialmente cuando se lo vincula a situaciones de pobreza y de exclusión social, convirtiéndose así en un tema prioritario para las autoridades gubernamentales de cada país, ya que la tasa de desempleo juvenil presenta un comportamiento fluctuante en el tiempo y varía en cada país, por diversos factores tales como económicos, sociales, políticos, entre otros, además a nivel mundial se evidencia que ciertos grupos en edad joven son los más vulnerables a formar parte de este fenómeno (Ramírez, 2002).

En *Determinants of youth unemployment in grece with an emphasis on tertiary education graduates*, (Mitrakos et al., 2010, pp. 21-62) consideran que el desempleo juvenil es el resultado de la falta de educación, ya que un bajo nivel de escolaridad aumenta el tiempo de que un individuo joven permanezca desempleado antes de encontrar un empleo estable, lo cual implica que con el transcurso del tiempo la tasa de desempleo sea más alta.

A nivel mundial, la Organización Internacional del Trabajo (OIT) en la edición 2017 de la serie *“Tendencias Mundiales del Empleo Juvenil”* publicó que la tasa mundial de desempleo juvenil en el 2017 fue del 13.1%. A nivel local la tasa de desempleo juvenil más alta se encuentra en los Estados árabes, donde alcanzó el 30% para el 2017, mientras que la tasa de desempleo nacional actual para el 2018 a nivel mundial es del 11.8%. En Ecuador, el director Regional de la OIT, José Manuel Salazar, indicó que el panorama de los mercados laborales en América Latina y Ecuador para el 2018, la tasa de desempleo nacional fue del 8.4% frente a la tasa de desempleo juvenil fue casi del 20%; es decir la mayoría de la población desempleada pertenece al segmento de los jóvenes que buscan trabajo y no lo consiguen o se encuentran trabajando en la informalidad. Además, la informalidad es casi la mitad de la fuerza de trabajo, en promedio en América Latina el 47% y en ciertos países llega el 70% (Rodríguez, 2018).

En el estudio *“Un análisis del Desempleo Juvenil en el Perú”*, analiza el desempleo juvenil utilizando datos estadísticos oficiales a partir del Instituto Nacional de Estadística e Informática INEI, con una metodología tipo descriptivo y gráfico, se realizó comparaciones de los resultados laborales, el cual reveló que Perú tiene tanto la tasa de desempleo juvenil y general más bajas en comparación a los países que se encuentran a su entorno concretamente Bolivia, Brasil, Chile, Ecuador, Paraguay, Uruguay y Venezuela. No obstante, el desempleo juvenil constituye un problema en la medida que triplica el correspondiente a la población adulta; en el 2017 la tasa de desempleo juvenil alcanzó el 13.7% reduciéndose solo 0.4 puntos porcentuales respecto al año anterior 14.1% (2016). Además su análisis muestra que las mujeres presentan más problemas de acceder a un empleo; una de las principales causas de este problema es la educación, dado que

los jóvenes con estudios superiores presentan una menor tasa de desempleo y los que tienen menor experiencia laboral se encuentran desocupadas (Zegarra, 2015).

En cuanto, España constituye la principal fuente subexplotada de beneficio y empleo de calidad, este país se caracteriza por tener ciertos problemas estructurales en el mercado de trabajo que otras economías europeas y mundiales no poseen, tales como menor cantidad de contratos temporales y mayor cantidad de contratos indefinidos (Manzano, 2015). De esta forma en la publicación *“Desempleo juvenil en España: qué determina su duración”*, usando los datos de la Encuesta de Condiciones de Vida y Trabajo (ECTV) aplican un modelo univariante de supervivencia Kaplan-Meier para estimar la duración del desempleo juvenil mediante el cálculo de la tasa de riesgo muestral, a partir de la cual es posible comparar las tasas de supervivencia entre las variables sociodemográficas tales como el género, el nivel educativo y el estado civil, obteniendo como resultado que en los jóvenes hombres no existe diferencias significativas entre casados y no casados, pero en el caso de las mujeres, las casadas tienen un 58% menor de probabilidad de encontrar empleo que las mujeres solteras, en cuanto a la educación los hombres no muestran evidencias significativas, mientras que las mujeres con educación universitaria tienen una probabilidad del 52% y con educación secundaria el 43% de encontrar empleo (Ahn y Ugidos-Olazabal, 2009).

En este sentido en Colombia, se realizó la investigación de *“Los determinantes del desempleo juvenil durante el periodo del 2015”*, usando como fuente de información la Gran Encuesta Integrada de Hogares (GEIH) del IV trimestre de 2015, mediante la aplicación de un modelo logit para determinar las causas que afectan al desempleo juvenil y estimar la probabilidad de que un joven este desempleado, concluyó que la variables escolaridad al cuadrado y el número de hijos en el hogar son estadísticamente no significativas por lo que representa una reducción de la probabilidad de que un joven permanezca desempleado además que el incremento de la edad en un año de vida incrementa la probabilidad de ser desempleado. Mientras que el nivel de educación, reflejó ser estadísticamente significativa, describiendo una mayor incidencia de desempleo en los jóvenes que están o cursaron la educación media y superior (Martínez Porras, 2018).

En el Ecuador, si bien resulta evidente que el estudio del desempleo juvenil ha sido uno de los temas a estudiar por diversos autores a nivel internacional, en el ámbito local existen pocos estudios que sirven como punto de referencia para seguir profundizando acerca de los aspectos más relevantes que aumentan la probabilidad de que un joven se encuentre desempleado. Entre los trabajos existentes se destacan los trabajos de (Cacuango y Lucero, 2014), (Espinoza, 2016) y (Mullo y Marcatoma, 2018, pp. 4-5), usando las encuestas trimestrales de la Encuesta Nacional de Empleo y Desempleo (ENEMDU).

En la investigación “Desempleo juvenil: estimación de la probabilidad de que un joven trabaje”, se determinó que las variables más importantes de la teoría del capital humano (Educación, Salud, Experiencia) resultaron significativas como es el sexo si el joven es hombre aumenta un 24.79% de que trabaje, si el joven estudia aumento un 55.58%, con relación a la experiencia existe una relación directa con la variable dependiente (trabaja) en un 8% y de acuerdo al nivel de instrucción un 98.72% si el joven tuvo una educación media. Ajustándose así un modelo de probabilidad de que un joven azuayo trabaje si se encuentra entre las edades de 18 y 24 años, cumpliendo las condiciones de ser mestizo de sexo masculino, no soltero, viva en el área rural, no se encuentre estudiando, haya alcanzado al menos la educación media y tenga cierto grado de experiencia laboral entre 0 y 3 años; con una efectividad de modelo 76.96% (Cacuango y Lucero, 2014).

A nivel nacional en el 2016, se realizó la investigación centrada en la “Inserción laboral de los jóvenes con énfasis en la formación dual a nivel nacional. Periodo 2010-2015” el cual determinó que el desempleo afecta en mayor proporción al segmento comprendido entre 15 y 24 años de edad, y en mayor magnitud a las mujeres, a pesar de que existe un 32% más mujeres que hombres jóvenes con título universitario. Tomando en cuenta estas generalidades ajustó un modelo econométrico logit, determinando como principales variables socioeconómicas a la edad, sexo, estado civil, auto identificación y nivel de instrucción que influyen en la probabilidad de que un joven se inserte laboralmente, donde se identificó que por el hecho de ser hombre, estar soltero, separado, no asistir a clases y poseer cierto nivel de instrucción, existe menor probabilidad de que un joven esté desempleado (Espinoza, 2016).

Además, en la investigación “Desempleo juvenil Ecuador, 2017: Magnitud y construcción de perfiles sociodemográficos”, se indica que en Ecuador en el primer semestre del 2017 existía 366.003 personas desocupadas, de este total 197.019 tenían entre 18 a 29 años, es decir el 59,32% de los desempleados ecuatorianos eran jóvenes en relación con los adultos (40.66%). La tasa de desempleo juvenil evidencia que los jóvenes (9,26%) poseen una tasa de desempleo mayor en comparación al de los no jóvenes (4,49%). El problema es aún más grave para las mujeres que presentan el 12.48% de desempleo, mientras que en los hombres es del 6.99%, mismas que muestran la gravedad del problema, en función de variables sociodemográficas el perfil planteado engloba la mayor problemática a las mujeres que son fejes de hogar, con nivel de instrucción secundaria o superior y que tuvieron alguna relación formal o son solteras por el 73.3% (Mullo y Marcatoma, 2018, pp. 4-5).

La presente investigación pretende analizar los factores que influyen en el desempleo juvenil nacional utilizando los datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), esta encuesta presenta datos de panel que resultan ser más adecuados para estudiar

la dinámica del cambio en una variable determinada, en especial los conjuntos de datos respecto al desempleo, la tendencia en el trabajo y la inestabilidad laboral (Gujarati y Porter, 2010).

Planteamiento del problema

Ecuador, ha experimentado problemas puntuales, tales como económicos, políticos, sociales y naturales; algunos por causas externas como la crisis mundial del 2008 o como el terremoto que se dio en abril del 2016, mismos que han sido de una u otra manera determinante de los incrementos en las tasas de desempleo general y juvenil. En la última Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) para diciembre del 2018, se evidencia que la tasa de desempleo fue de 8.25% para los ciudadanos entre 18 y 29 años, evidenciándose así una disminución en relación a diciembre del 2017 cuando el porcentaje era 9.61%, sin embargo, se mantienen con cantidades altas en relación a de los adultos 3,7% y 4,6% respectivamente convirtiéndose así el grupo de los jóvenes en el que mayor cantidad de desempleados tiene el país.

La preocupación que surge tras este fenómeno son los efectos negativos que tendrá la población joven en nuestro país, por lo que es imprescindible dar respuesta a la siguiente pregunta:

¿El análisis de series de tiempo y de región logística determina el comportamiento y probabilidad de que un joven ecuatoriano se encuentre desempleado, en base a la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) 2013 - 2018?

Justificación

Ecuador a partir del año 1999 ha pasado por severas crisis económicas, que han causado un sin número de afectaciones en el ámbito social, político y económico, debilitando el sistema financiero y llevando a las pequeñas y medianas empresas al recorte en gran medida de personal joven que son los que cuentan con menor experiencia laboral y reciben salarios bajos. Por este motivo dificulta a los jóvenes insertarse en el mercado laboral.

Dentro del Plan Nacional Toda una Vida aplicable para los años 2017 al 2021, en el eje 1 entre sus objetivos consta que se debe garantizar una vida digna con iguales oportunidades para todas las personas, cuya meta va dirigida a reducir del 10.4% al 7.6% la tasa de desempleo juvenil de 18 a 29 años y es por eso que de acuerdo con la “Ley de la Juventud”, Art. 1 se reconoce a todas las personas comprendidas entre 18 y 29 años de edad como jóvenes (Congreso Nacional, 2001), permitiendo al Gobierno ecuatoriano identificar el segmento de la población que puede gozar de la protección que el Estado otorga en calidad de subsidios y programas implantados para las personas jóvenes tales como: Mi Primer Empleo, Red Socio Empleo, Educación Dual, entre otros.

El presente estudio está enfocado específicamente en los jóvenes ecuatorianos, que tengan entre 18 y 29 años de edad, tiene como propósito indicar cuál es la probabilidad de que un joven ecuatoriano se encuentre o no desempleado, analizando la encuesta ENEMDU en el período 2013-2018 ya que ésta recoge variables categóricas de carácter sociodemográficas y sociolaborales.

Por lo antes mencionado se propone analizar la evolución del desempleo juvenil mediante un análisis estadístico que presenta una serie de técnicas estadísticas para detectar patrones de comportamiento ocultos, basados en los mismos para establecer predicciones y posibles relaciones entre variables, además es conocido que los métodos estadísticos clásicos incorporan distintos supuesto teóricos como normalidad e independencia, la violación de los supuestos incurre sobre la estimación de la probabilidad asociada al estadístico, por lo que para solucionar este problema, existen los métodos no paramétricos que son más apropiados cuando no se tiene un conocimiento previo de la relación entre variables que forman parte de un estudio.

Los métodos no paramétricos tienen múltiples ventajas tales como: facilidad de aplicación, manipulación de datos categóricos, etc., así como también se lo puede aplicar cuando dos series de observaciones provienen de distintas poblaciones, dentro de estas técnicas para la clasificación de variables se tiene la técnica de regresión logit debido a su importancia por sus diferentes aplicaciones en áreas socioeconómicas, por lo que su estudio cada vez es más frecuente. El objetivo de esta técnica estadística es la estimación en relación de las variables independientes con la variable dependiente, utilizada en su mayoría en estudios causales que consiste en investigar factores de una determinada característica de la población y estudiar qué factores modifican la probabilidad en la aparición de un suceso determinado; o también predictivo que consiste en predecir lo mejor posible la variable dependiente a través de las independientes (Solano et al., 2014).

En la actualidad, el análisis de las series de tiempo juega un papel muy importante dentro del estudio económico, pues a través de este se puede obtener información de variables en el tiempo, reduciendo así la incertidumbre de su comportamiento en el futuro y con esto se facilitarían la toma de decisiones, los datos son representados mediante un modelo estructurado que visualmente es una curva que evoluciona a lo largo del tiempo y mediante la realización de inferencias se podrá efectuar predicciones aplicando el método Bootstrap, el cual permite eludir las limitaciones del enfoque estadístico clásico, que en muchas ocasiones prácticas resultan poco realistas cumplirlas, este método utiliza la información de la muestra original, para realizar remuestras con reemplazamiento que generan nuevas muestras, mismas que serán usadas para ajustar un modelo de predicción, en este sentido, el método Bootstrap es considerado como una metodología de inferencia estadística (Araya, 2010). Este proceder ha sido considerado por la American Statistical Association como “el único gran descubrimiento en estadística desde 1970” (Kotz y Johnson, 1997).

La regresión logit permitirá explicar el comportamiento de los factores que influyen directa o indirectamente en el desempleo de jóvenes ecuatorianos, con el análisis de series de tiempo se podrá evidenciar la evolución de los patrones que sigue la serie del desempleo juvenil a lo largo del periodo en estudio y con la aplicación del método Bootstrap se ajustará un modelo de predicción adecuado sin violar los supuestos determinados del método clásico de predicción.

La herramienta estadística a utilizar para la obtención de resultados óptimos dentro de esta investigación será el software estadístico libre R, ya que permite autonomía tecnológica de la información y el ahorro de recursos logrando así la innovación que fortalece la inclusión digital, que gracias a su versatilidad en la aplicación de diversas estrategias analíticas y de gran potencialidad gráfica, se puede realizar un análisis exploratorio de los datos, la implementación de técnicas de estimación y series temporales facilitando así el manejo de las bases de datos de una forma sencilla (Santana y Farfán, 2014).

Objetivos

Objetivo General

Analizar mediante técnicas de series de tiempo y regresión logit el desempleo juvenil en el Ecuador, periodo 2013 – 2018.

Objetivos Específicos

- Determinar la magnitud del desempleo juvenil en Ecuador a nivel nacional y provincial.
- Encontrar modelos empíricos de pronóstico de series de tiempo que se ajusten a los datos del desempleo juvenil por provincias para el periodo 2013 – 2018.
- Establecer los factores sociodemográficos y sociolaborales que influyen en la probabilidad de ser un joven desempleado en Ecuador durante el periodo 2013-2018.

CAPÍTULO I

1 MARCO TEÓRICO REFERENCIAL

1.1 Encuesta Nacional de Empleo, Subempleo y Desempleo (ENEMDU)

La ENEMDU es una encuesta por muestro probabilístico y bietápico; la Unidad Primaria de Muestro (UPM) es el conglomerado y la Unidad Secundaria de Muestreo (USM) son las viviendas ocupadas; dicha encuesta es realizada y publicada en Ecuador por el Instituto Nacional de Estadística y Censos (INEC) desde el año 1987; se realizada de forma periódica y está orientada a los hogares ecuatorianos con el objetivo de determinar la estructura de la población en los ámbitos sociales, económicos y demográficos. Mediante la aplicación de esta encuesta se calculan los indicadores de mercado laboral, relacionados con el empleo, subempleo y desempleo, determinando así la estructura de la Población Económicamente Activa (PEA), además la información recolectada contribuye para la formación de las bases de datos de manera mensual, trimestral y anual que permiten realizar análisis, seguimiento y evaluación de las políticas públicas de empleo y desempleo.

El universo para la muestra de la ENENDU según el INEC se conforma de todas las viviendas particulares dentro del territorio del Ecuador, excluyendo la Región Insular, la población en viviendas colectivas, viviendas flotantes y sectores con población indigente; sin embargo, se incluye las zonas no delimitadas.

El marco muestral es en base a resultados definitivos y la cartografía del VII Censo de Población y Vivienda del 2010 (CPV-2010), el mismo que será renovado con el nuevo censo en noviembre del 2020.

La muestra de la ENEMDU proviene de la Muestra Maestra del Sistema Integrado de Encuestas de Hogares (SIEH), la cual está basada en un esquema de rotación de viviendas que permite calcular los cambios del empleo, subempleo y desempleo; y se efectúa utilizando métodos de muestreo con bases sólidas en la teoría estadística y de probabilidades, por lo cual toda información obtenida de esta encuesta son estimaciones confiables de los valores reales que se quiere conocer (Encuesta Nacional de Empleo Manual del Encuestador, 2018).

1.1.1 Clasificación de la población

Las siguientes definiciones están basados en los conceptos publicados por el INEC (Metodología para la medición del empleo en Ecuador, 2016).

- **Condición de actividad**

Permite clasificar a la población en edad de trabajar (personas de 10 años y más), en población económicamente activa (PEA) y población económicamente inactiva (PEI).

- **Población económicamente activa (PEA)**

Personas de 10 años y más que trabajaron al menos una hora en la semana de referencia o, aunque no trabajaron, tuvieron trabajo (ocupados), o bien aquellas personas que no tenían empleo, pero estaban disponibles para trabajar y buscan empleo (desempleados).

- **Población económicamente inactiva (PEI)**

Se consideran personas inactivas todas las personas de 10 años y más, no clasificadas como personas con empleo (ocupadas) o desempleadas (desocupados) durante la semana de referencia, como: rentistas, jubilados, pensionistas, estudiantes, amas de casa, entre otros (INEC, 2018, p.5).

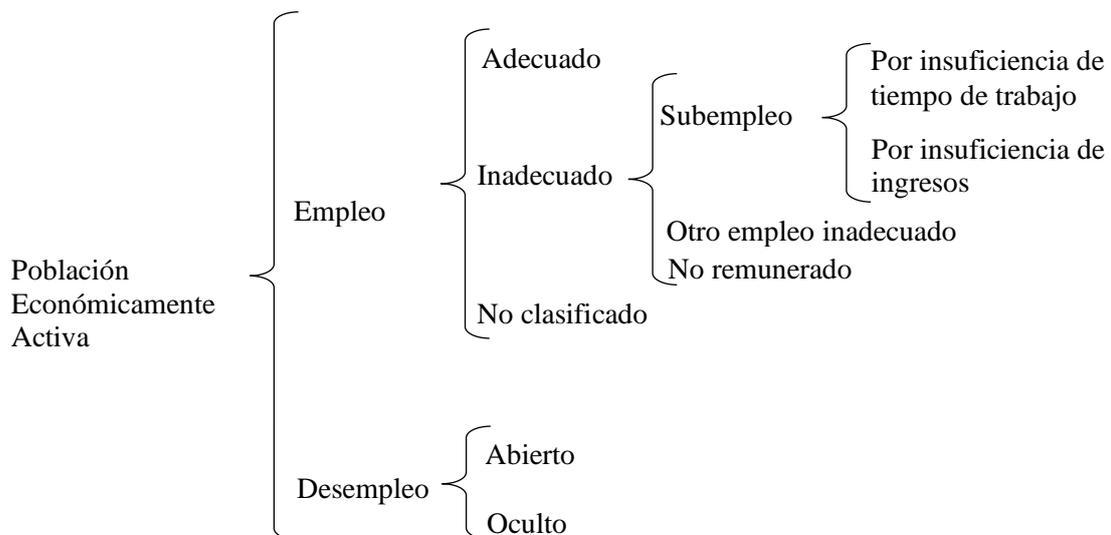


Gráfico 1-1: Clasificación de la Población Económicamente Activa (PEA) Ecuador
 Realizado por: Tierra, Tania. 2020

1.1.2 Desempleo

Según la OIT una persona desempleada es “toda aquella persona que tenga más de cierta edad específica, que se encuentren sin empleo, actualmente está disponible para trabajar y esté buscando trabajo durante el período de referencia”.

En Ecuador según el Instituto de Estadísticas y Censos (INEC) el desempleo lo conforman las personas de edad de 10 años y más, que en periodo de referencia presentan las siguientes características: 1) No tuvieron empleo, no estuvieron empleados la semana pasada y están disponibles para trabajar; 2) buscaron trabajo o realizaron tareas concretas para conseguir empleo o para levantar algún negocio en las cuatro semanas anteriores. Se distinguen dos tipos de desempleo: abierto y oculto.

1.1.2.1 Desempleo abierto

Personas sin empleo de 10 años y más que se encontraban sin empleo en la semana pasada y buscaron trabajo e hicieron gestiones concretas para conseguir empleo o para establecer algún negocio en las cuatro semanas anteriores a la entrevista.

1.1.2.2 Desempleo oculto

Personas sin empleo de 10 años y más, sin empleo durante la semana pasada, que no buscaron trabajo y que no hicieron gestiones para conseguir empleo o para establecer algún negocio en las cuatro semanas por alguna de las siguientes razones: tienen un trabajo ocasional, tienen un trabajo para empezar inmediatamente, esperan respuesta de un empleador, se cansaron de buscar trabajo, espera cosecha o temporada de trabajo entre otras.

1.1.2.3 Tasa de desempleo

La tasa de desempleo es el cociente entre el número de personas desempleadas y el de activos (ocupados o buscando empleo), expresado como porcentaje (Larroulet y Mochón, 1995).

$$\text{Tasa de desempleo} = \frac{\text{Núm. de desempleados}}{\text{Población activa}} \times 100 \quad (1.1)$$

1.1.2.4 Tipos de desempleo

Los tipos de desempleo según Cuadrado, Mancha, Villena, Casares, González, Marín y Peinado (2010), son los siguientes:

- **Estructural:** es aquel que persiste en el tiempo en ciertas regiones o ramas de actividad, como consecuencia de la inadecuada estructura económica. Este tipo de desempleo es causado principalmente por la existencia de actividades productivas en retroceso, es decir debido a los cambios producidos por las estructuras industriales, ocupacionales y demográficas de la economía, se ve reducida la demanda de trabajo, por lo que les es difícil a las personas cesantes⁴ encontrar un empleo (Cuadrado et al., 2010, pp.147-148).
- **Friccional:** la evolución tecnológica y los cambios en la demanda provocan una continua rotación en el empleo y existe siempre una masa flotante de personas que abandonaron su puesto de trabajo o a su vez fueron despedidos. En contexto este volumen de personas representa determinados porcentajes de la población activa. En algunos países más

⁴ Las personas cesantes son aquellas que en la semana de referencia se encontraban desempleados, pero han trabajado anteriormente.

desarrollados la elevación de los niveles educativos y algunos cambios sociales han provocado que cada día sea más difícil que una persona acabe su vida activa en el mismo empleo en que la inició (Cuadrado et al., 2010, p.149).

- **Estacional:** es un trabajo con actividades que dado sus características de producción se requiere la mano de obra en determinadas épocas del año. Un ejemplo claro de esto es la estacionalidad agrícola y la demanda de trabajadores, en épocas de siembra y de cosecha, donde una vez cubiertas las dos fases cesa el trabajo, al igual que en épocas de mantenimiento de cultivos y cría de ganado se reduce la demanda de empleo (Cuadrado et al., 2010, p.149).
- **Cíclico:** consiste en la falta de trabajo durante un momento de crisis económica, por lo general de periodos no demasiado extensos en el tiempo y que se revierten junto a la reactivación de la Economía. Durante esos periodos, los trabajadores que se pierden su trabajo tienden acogerse a los subsidios de desempleo y esperar una fase más propicia del ciclo económico (Cuadrado et al., 2010, pp.149-150).

1.2 Desempleo juvenil

El desempleo juvenil es un fenómeno generalizado a nivel mundial que afecta a todas las economías del mundo, ya que los jóvenes son los más vulnerables, debido a múltiples condiciones entre ellas principalmente el nivel de educación y experiencia laboral. Según el censo de 2010 en Ecuador eran 3'043.513 personas jóvenes (18 a 29 años) lo que representaba el 21.4% de la población y el 9.90% se encontraba desempleada, en el 2018 fueron 3'106.687 jóvenes siendo el 18.04% de la población total y el 9.08% sin empleo.

A nivel mundial, tanto hombres como mujeres se enfrentan a incontables problemas para poder encontrar un trabajo adecuado. En los últimos años el número de jóvenes desempleados ha triplicado su cifra, y es cinco veces mayor que la cifra de desempleados adultos (Oficina Internacional del Trabajo, 2012). Una posible causa de las alarmantes cifras es que en tiempos de crisis económicas los jóvenes son los más vulnerables en ser contratados y despedidos de un puesto de trabajo.

En la mayoría de países a nivel global la tasa de desempleo juvenil supera a la de los adultos con una variación considerable, Ecuador no es la excepción dado a los últimos acontecimientos naturales y económicos que ha pasado el país, el desempleo ha aumentado considerablemente creando una gran brecha en el mercado laboral, como se observa en el siguiente gráfico:

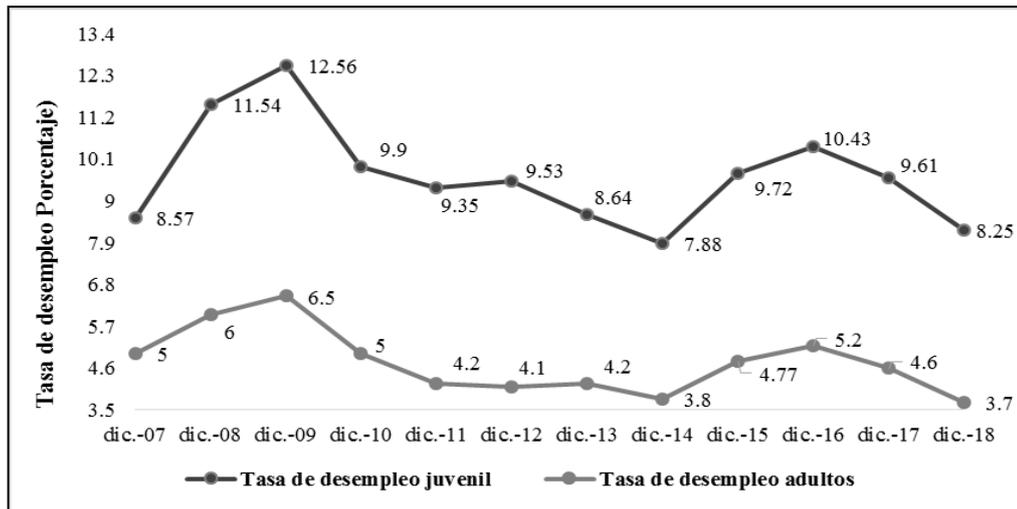


Gráfico 2-1: Tasa de desempleo juvenil nacional de Ecuador de DIC 2007–DIC 2018
 Realizado por: Tierra, Tania. 2020

Según el gráfico 2-1, se evidencia que los jóvenes alcanzan tasas cinco veces más altas de estar desempleados en relación a la tasa de desempleo de los adultos. Por esta razón, la tasa de desempleo juvenil tiene más sensibilidad ante una crisis económica. Un dato que refleja este hecho es que el desempleo juvenil incrementó en 2.18% en comparación al 1.5% que incrementó el desempleo adulto para el periodo 2016-2018.

El desempleo juvenil en el tiempo muestra en general una tendencia decreciente, pero con tasas elevadas en comparación al desempleo de los adultos, causado por muchos problemas especialmente por las malas decisiones de los gobernantes del país.

En el año 2000 el país atravesó una de las peores crisis económicas vividas en los últimos tiempos, pues los indicadores económicos mostraban niveles muy bajos, no había confianza en las instituciones del Estado y en las instituciones financieras, por lo que para evitar que el sistema financiero colapsara se tomaron una serie de medidas drásticas como declarar un feriado bancario, en el que se congelaron cuentas bancarias para evitar la fuga de capital y posterior a ello se dolarizó oficialmente la moneda, correspondiendo un equivalente de 25000 sucres un dólar.

Para comprender la magnitud del desempleo juvenil en Ecuador es importante indicar que el país ha tenido un alto grado de desigualdad social, inestabilidad política, escaso desarrollo del capital humano y la falta de transparencia en los programas de empleo creadas por las diferentes instituciones del Estado.

1.2.1 Resumen del desempleo juvenil en Ecuador

Tabla 1-1: Tasa de desempleo juvenil de Ecuador de DIC
2007 – DIC 2018

Mes-Año	18 a 29 años	15 años y mas
	TNDJ	TND
Dic-07	8,57	5,00
Dic-08	11,54	6,00
Dic-09	12,56	6,50
Dic-10	9,90	5,00
Dic-11	9,35	4,20
Dic-12	9,53	4,10
Dic-13	8,64	4,20
Dic-14	7,88	3,80
Dic-15	9,72	4,77
Dic-16	10,43	5,20
Dic-17	9,61	4,6
Dic-18	8,25	3,7

Fuente: Banco de datos abiertos del INEC-ENEMDU

Realizado por: Tierra, Tania. 2020

El cuadro anterior muestra el comportamiento del desempleo juvenil en Ecuador, en el cual se evidencia que los más jóvenes son quienes tienen tasas más altas de desempleo. La tasa de desempleo juvenil en el 2018 fue 9.08%, tres veces mayor a la tasa de desempleo nacional promedio 3.7%.

1.2.2 Fundamentos del desempleo juvenil

(Chen et al., 1990) señala que existe una serie de variables, algunas cuantificables que pueden explicar la presencia de una tasa de desempleo elevada, entre las principales se tiene:

- **Seguro de desempleo y afines:** el seguro de desempleo es una medida de protección social, que contribuye de manera negativa, dado que los beneficiarios mantienen durante un tiempo prolongado la búsqueda de trabajo y no tienen el deseo de aceptar uno nuevo.
- **Salario mínimo:** es una de las principales causas por la que los jóvenes no aceptan su primer empleo o provocan que abandonen sus trabajos voluntariamente en búsqueda de una mejor oportunidad.

1.3 Estudios de serie de tiempo en Encuestas de Hogares

Cárdenas, 2011 utiliza el análisis de series de tiempo para predecir el desempleo en el Ecuador a largo plazo, en base a la información contenida en la variable es el “número de semanas que busca trabajo por más de 12 semanas” de la encuesta ENEMDU. Al modelar los datos de la variable únicamente en función de su propia evolución pasada, la serie presenta un patrón de tendencia decreciente, la cual se podría dar por las múltiples políticas públicas de empleo que se han

implementado en los últimos años. En el 2003 se tenía un 4.5% de individuos de la población económicamente activa que buscaba trabajo por más de 12 semanas, sin embargo, en el 2010 esta tasa bajo a 2.1%, es decir, que una persona puede encontrar trabajo en menos de tres meses. Además, presento un modelo *ARIMA* (1,1,1), el pronóstico calculado por este modelo para los siguientes tres periodos fue una tasa de alrededor 1.9%, lo que indica que existe una pequeña baja en la tasa de desempleo a largo plazo, esto se podría explicar que es debido a la ausencia de cambios significativos en los determinantes del desempleo a largo plazo. De acuerdo con los resultados encontrados, se determina que las predicciones pueden ser empleadas como herramientas que producen señales asociadas a la necesidad de aplicar políticas correctivas además de servir como indicadores referidos a la eficiencia de la política pública de ajuste en el corto plazo (Cárdenas, 2011).

1.3.1 Determinantes del desempleo juvenil en Encuestas de Hogares

1.3.1.1 Determinantes sociodemográficos

- **Edad joven:** según (Verdera ,2010; citado en Vicente, 2018), en varios países de Centroamérica a la población joven se la define según diferentes grupos de edad, tales como la comprendida entre 10 y 24 años, en Costa Rica y México entre 12 y 24 años, dentro de la política nacional para la juventud de varios países delimitan a la edad entre los 15 y 29 años, como es el caso de la mayoría de los países andinos, mientras que el 17 de octubre de 2001 la Ley de la Juventud en Ecuador, en el artículo 1, reconoció a todas las personas comprendidas entre 18 y 29 años como jóvenes (Ley de la juventud, 2001).
- **Género:** muestra las diferencias psicológicas, sociales y culturales entre los hombres y mujeres (Giddens, 2010; citado en Vicente, 2018).
- **Estado civil:** es la legalización de su inscripción del estado conyugal actual de las personas en el Registro Civil, tanto el estado civil legal como las relaciones civiles de hecho. Para este estudio se lo seccionará a jóvenes con pareja es decir casados o en unión libre, mientras que los jóvenes sin pareja estable serán los separados, divorciados, viudos y solteros.
- **Relación de parentesco:** el parentesco es un vínculo o lazo de unión existente entre dos personas pertenecientes a un mismo hogar; estos lazos pueden ser de consanguinidad, conyugal, afinidad, legal o de costumbre. Se clasificará como jefe de hogar y en la opción otros se asociarán a los jóvenes que son yerno/nuera, nieto, padres, suegros, hijos, etc. del jefe de hogar.
- **Asiste a clases:** las personas tienen del derecho de recibir una formación académica que responda a unos estándares de calidad y así tener la oportunidad de llegar a una vida adulta con éxito sin distinciones de edad y nivel social.

- **Nivel de educación:** Para el (Ministerio de Educación, 2013, p. 9), de acuerdo a la Ley Orgánica de Educación Intercultural (LOEI) y el Reglamento a la LOEI, la clasificación de los niveles educativos corresponde a educación inicial, educación general básica y bachillerato. Al mismo tiempo, la Ley Orgánica de Educación Superior (LOES), en su Art. 118 considera que los niveles de formación de la educación superior con: nivel técnico o tecnológico superior, tercer nivel o universitario y cuarto nivel o de posgrado (LOES, 2018, p. 22).

1.3.1.2 Determinantes sociolaborales

- **Ingreso laboral:** cantidad monetaria o en especie que recibe la persona por parte de la empresa o entidad para la que trabaja, en forma habitual y en intervalos regulares de tiempo (diario, semanal, quincenal, mensual).
- **Rama de la actividad:** actividad económica que permite clasificar al establecimiento donde labora la persona dentro del sector de la economía, según la clase de bienes o servicios que produce. Esta determinante se clasifica en 21 secciones de acuerdo a la Revisión 4 de la Clasificación Internacional Industrial Uniforme (CIIU 4.0). En el presente estudio las actividades se clasifican en tres grandes sectores: primario formado por las actividades relacionadas con la obtención de los recursos de la naturaleza, agrícola, ganadero, forestal y pesquero; el sector secundario conforman las actividades que implica la transformación de materia prima y alimentos a través de variados procesos productivos, reservas de petróleo, industria de la energía eléctrica e industria textil; sector terciario se encarga de las actividades que no producen bienes materiales de forma directa, sino servicios que satisfacen las necesidades de la población (Gomezjurado Jarrin 2016, p.268).
 - ✓ **Sector primario:** agricultura, silvicultura, caza y pesca.
 - ✓ **Sector secundario:** explotación de minas y canteras, industrias manufactureras y construcción.
 - ✓ **Sector terciario:** suministros de electricidad, gas, aire acondicionado, distribución de agua, alcantarillado, comercio, reparación de vehículos, transporte y almacenamiento, hoteles y restaurantes, información y comunicación, actividades financieras y de seguros, actividades inmobiliarias, actividades profesionales, científicas y técnicas, actividades y servicios administrativos y de apoyo, administración pública, defensa y seguridad social, enseñanza, actividades, servicios sociales y de salud, artes, entretenimiento y recreación, otras actividades de servicios, actividades en hogares privados con servicio doméstico, actividades de organizaciones extraterritoriales.
- **Seguro social:** es la protección social o cobertura de las necesidades reconocidas socialmente, como la salud, la vejez o las discapacidades que reciben las personas durante el contrato

laboral.

- **Grupo de ocupación:** tarea o actividad específica que desarrolla el empleado dentro del establecimiento en que labora, básicamente trata de las características de los puestos de trabajo, se clasifica en 10 secciones a un dígito de acuerdo a la Clasificación Internacional Uniforme de Ocupaciones (CIUO -08) (Instituto Nacional de Estadística y Censos (INEC), 2012).

Esta variable fue recodificada como **ocupación no manual:** personal direct. /admin. Publica y empresas, profesionales científicos e intelectuales, técnicos y profesionales de nivel medio, empleados de oficina; mientras que la **ocupación manual** contiene las siguientes categorías trabajadores de los servicios y comerciantes, trabajadores calificados agropecuarios y pesqueros, oficiales operarios y artesanos, operadores de instalaciones de máquinas y montadoras, trabajadores no calificados, ocupaciones elementales y fuerzas armadas.

1.4 Bases teóricas de estadística

Dentro de la rama de la estadística existe una serie de técnicas aplicables para el análisis de datos, que buscan dar respuesta a una multitud de preguntas relacionadas con los datos. El principal inconveniente que se presenta a la hora de realizar un análisis es el desconocimiento de cual técnica aplicar y como utilizarla, dado que las variables que se analizan en cada técnica tienen diferentes características y pueden ser cualitativas o cuantitativas.

Antes de realizar un análisis estadístico, se debe tener claro el problema que se va a estudiar y cuales son objetivos a cumplir, fijando las clasificaciones que se van a realizar, las variables que se van a observar y como medirlas, además de las gráficas que mejor representen la información.

Análisis exploratorio de datos (AED)

El AED tiene como finalidad examinar los datos de manera descriptiva, gráfica y analítica, para así dar una explicación clara de los datos y las relaciones existentes entre las variables analizadas, también ayuda a encontrar si existen datos atípicos.

Además el AED se caracteriza por el empleo de procedimientos analíticos y descriptivos de carácter gráfico y/o numérico de un conjunto de datos univariantes o multivariantes, para organizar y preparar los datos, detectar fallos en el diseño y recogida de datos, tratamiento y evaluación de datos faltantes y atípicos, a la vez ayudan a comprobar los supuestos subyacentes en la mayor parte de las técnicas multivariantes (Batanero y Godino, 1991, pp. 25-31).

De este modo, es posible obtener una estructura explicativa de los datos a través de la combinación de herramientas básicas de la estadística (descriptivos, correlaciones, tablas de frecuencias), con un análisis multivariante avanzado, diseñado especialmente para identificar perfiles en grandes

bases de datos (análisis clúster, escalas multidimensionales, análisis logit, correlación canónica, análisis de correspondencias, etc.) (Chasco, 2003).

Imputación de datos faltantes

La falta de información en las investigaciones representa un problema constante que se presenta al momento de estudiar un conjunto de datos sobre el cual se busque realizar un análisis estadístico, convirtiéndose en un obstáculo que no solamente afecta a las estadísticas oficiales (encuestas de hogares, datos censales, etc.) en la cual la mayoría se limita a dar respuesta a las variables relacionadas al ingreso económico, sino también a los registros administrativos de empresas u organismos. Por tal motivo contar con métodos que puedan dar solución al problema de datos faltantes ha sido y continua siendo una necesidad en el análisis estadístico moderno, ya que los paquetes estadísticos asumen que se trabaja con una matriz de datos completos (Rosati, 2017, p.93).

Cuando se aplica una encuesta, todos los investigadores se han encontrado con el problema de datos cuantitativos faltantes en algún momento del registro o análisis de la información, esto se debe a que las personas informantes pueden haberse rehusado u olvidado de responder a una pregunta o tal vez se hayan perdido los archivos, los datos no hayan sido grabados apropiadamente (Little y Rubin, 2002; citados en Calafati, 2017).

A pesar del avance de los softwares estadísticos, la situación se complica más cuando se presenta una matriz de datos formada por diferentes tipos de variables sobre la cual se realiza análisis multivariantes, haciéndose necesario la aplicación de métodos que convenientemente imputen conjuntamente los datos.

La proporción de datos faltantes en un registro puede variar; dependiendo del estudio, la dificultad de llenar el cuestionario o de la medición de una unidad, sin embargo, dependerá del investigador, considerar el registro como pérdida parcial o total si el número de faltantes es considerable. De igual forma el investigador debe determinar el porcentaje de pérdida de datos si se puede tratar mediante imputación o considerarlo como una mala recolección de información y por lo tanto la base de datos no debe ser utilizada; en la práctica se habla de pérdidas máximas entre 1 y 20% de la información, dependiendo de la exactitud del estudio y del área de investigación entre otros factores. Por ejemplo, en las ciencias de la salud, la precisión es un factor muy importante en la obtención de resultados por lo que no se admite la imputación de un gran número de valores que no serán reales, mientras que en las ciencias sociales se permite porcentajes de datos faltantes más altos, la decisión siempre estará en el investigador (Useche y Mesa, 2006, pp.129-130).

De esta manera los datos faltantes no se los puede ignorar, la manera más adecuada de tratarlos es llenar los espacios faltantes con valores aceptables; existe una gran variedad de métodos para analizar datos faltantes que requieren suposiciones acerca de la naturaleza de los datos y también tener en cuenta las razones por las cuales las observaciones están incompletas.

Varios estudios de (Goicoechea, 2002), (Platek, 1986) y (Government Statistical Service, 1996), indican que las técnicas de imputación se pueden clasificar de la siguiente manera:

Técnicas fundamentales en información externa y técnicas determinísticas; dentro de esta última se encuentran la imputación de la media o moda, imputación de media de clases e imputación por regresión. En este estudio la imputación será mediante el método de la moda.

Método de imputación por la moda

Este método se lo utiliza específicamente para variables categóricas, el cual llena el vacío del dato faltante de cada variable con la moda de los registros. Su ventaja es la facilidad de la aplicación del método. Se utiliza la imputación por la moda para este tipo de variables en vez de la media muestral o la mediana, para evitar que el número de clases no afecte en la variable.

Dada una variable X_i que presenta valores perdidos, mediante este método se reemplaza cada uno de ellos por un valor calculado, es decir la moda de los valores observados de X_i (Useche y Mesa, 2006).

Análisis de datos atípicos

Los datos pueden presentar diferentes tipos de problemas a la hora de ser analizados como duplicados, datos faltantes o atípicos, entre otros y para cada tipo existen técnicas diferentes para detectarlos y corregirlos. La calidad de la limpieza lograda sobre los datos, depende de la técnica aplicada y la elección de la técnica debe tener en cuenta la naturaleza y comportamiento de los datos específicos sobre los que se está trabajando.

Un dato atípico es una observación claramente diferente al resto de los datos, es decir es una observación externa, este tipo de casos no pueden ser tomados como favorables o problemáticos, sino que deben ser verificados en el contexto del análisis y debe evaluarse el tipo de información que puede proporcionar. Dentro del alcance estadístico existen varios métodos para detectar valores atípicos: gráficos como los de normalidad, diagrama de cajas o métodos basados en distribuciones. Dichos métodos asumen que los datos provienen de una distribución normal. Un método común para la detección de valores atípicos es mediante el rango intercuartílico, en el cual una observación se considera atípica si esta fuera de los límites y; k se fija normalmente a 1.5 o 3. para muestras pequeñas (Abellana y Farran, 2015, pp. 191-192).

El principal problema radica en que pueden ser datos no representativos de la población pudiendo distorsionar seriamente el comportamiento de los contrastes estadísticos, aunque en algunos casos pueden ser diferentes en mayor parte a la muestra de la población, llegando a ser indicativos de las características de un segmento válido de la población y por consiguiente una señal de la falta de representatividad de la muestra. Los casos atípicos pueden clasificarse en 4 categorías (Ocaña, 2019):

- **Error de procedimiento:** surgen al momento de ingresar los datos o un error de codificación, estos casos atípicos deberían corregirse en el filtrado de los datos, caso contrario deberían eliminarse del análisis o recodificarse como datos ausentes.
- **Observación que se produce de un acontecimiento inusual:** en este caso la observación atípica no representa ningún segmento valido de la población y puede ser eliminado del análisis.
- **Observaciones que se ubican dentro del rango de las variables observadas pero que son únicas en la combinación de sus valores:** estas observaciones suelen ser retenidas en el análisis, pero se debe estudiar la influencia que realiza en los procesos de estimación de los modelos considerados.
- **Observaciones inusuales cuyo comportamiento no tiene explicación:** en este caso lo mejor que se puede hacer es replicar el análisis con y sin dichas observaciones, con el fin de analizar su influencia sobre los resultados, si dichas observaciones son influyentes el investigador debería explicarlo en sus conclusiones e investigar el comportamiento de dichas observaciones.

Diagrama de cajas y bigotes

Conocido como Boxplot se emplea como representación gráfica de variables cuantitativas, permite resumir, describir y analizar aspectos generales y particulares de la variable en estudio, representando así los datos centrales, adyacentes, atípicos y extremos, puede ser graficada de manera vertical u horizontal (Ballesteros, 2013, p.89).

Creado por Tukey en 1977, en el cual se representa cinco medidas descriptivas: mediana, primer cuartil (Q_1), tercer cuartil (Q_3), valor máximo y valor mínimo. Está compuesto por un rectángulo que forma una caja la cual se construye con la ayuda del primer y tercer cuartil representando el 50% de los datos que particularmente están ubicados en la zona central de la distribución, la mediana es la línea que atraviesa la caja y dos bigotes que son las líneas que se extienden desde la caja hasta los valores más altos y más bajos (Palomino, 2004). Al mismo tiempo presenta información sobre la tendencia central, dispersión y simetría de los datos en estudio.

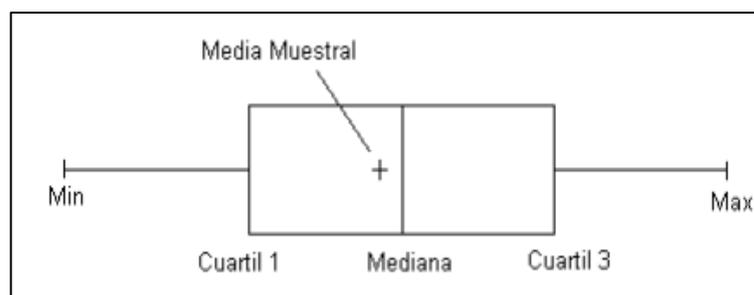


Figura 1-1: Diagrama de Cajas y Bigotes
Fuente: (Palomino, 2004)

Además, permite identificar con claridad y de forma individual, observaciones que se alejan de manera poco usual del resto de los datos, es decir permite detectar los valores atípicos, misma que por su facilidad de construcción e interpretación, también permite comparar a la vez varios grupos de datos sin perder información.

Para la detección de valores atípicos, la longitud máxima de cada uno de los bigotes es de $k = 1.5$ veces el rango intercuartil (IQR) es decir $1.5 \times (Q_3 - Q_1)$ por encima y por debajo de los cuartiles. Las observaciones fuera de los bigotes son dibujadas separadamente y etiquetadas como valores atípicos (Amón, 2010, pp. 84-86).

Modelos de regresión

En estadística existe una serie de técnicas para el análisis de datos, las mismas que expresan la estructura de dependencia que mejor explique el comportamiento de la variable dependiente o respuesta (y) a través de un conjunto de variables independientes o explicativas (x_1, x_2, \dots, x_p), con las que se supone está relacionada; como lo es la regresión, cuya ecuación a ajustar puede ser lineal o no lineal, que en ambos casos el objetivo es el mismo: encontrar las mejores estimaciones de los parámetros y cuantificar la precisión de los mismos (Olaechea, 2006, p. 55).

Tabla 2-1: Combinaciones más comunes de las variables respuestas y variables explicativas

Tipo de análisis	Variable respuesta	Variable explicativa	Función de vínculo	Distribución de errores
Regresión	Continua	Continua	Identidad	Normal
ANOVA	Continua	Factor	Identidad	Normal
Regresión	Continua	Continua	Recíproca	Gamma
Regresión	Conteo	Continua	Logarítmica	Poisson
Tabla de contingencia	Conteo	Factor	Logarítmica	Poisson
Proporciones	Proporción	Continua	Logit	Binomial
<i>Regresión logística</i>	<i>Binaria</i>	<i>Continua</i>	<i>Logarítmica</i>	<i>Binomial</i>
Análisis de supervivencia	Tiempo	Continua	Recíproca	Exponencial

Fuente: (Cayuela, 2010)

El modelo de regresión utilizada para esta investigación es la *Regresión Logit*, dado que la variable respuesta es categórica.

1.4.1 Regresión logit (RL)

La técnica de regresión logit se origina en la década de los 60 con el trabajo de Cornfield, Gordon y Smith (1961), siendo a partir de los 80 que se generaliza su uso con la ayuda de la informática aplicada. La regresión logit puede considerarse un caso especial del análisis de regresión en donde la variable dependiente es dicotómica (« Si » [1] o « No » [0]), por tanto, trata de estimar la probabilidad de que ocurra un suceso en función de la dependencia de otras variables. En este caso cuando la variable dependiente Y tome el valor 1 dependerá de un conjunto de p covariables X_1, X_2, \dots, X_p que inciden de forma diferente según un coeficiente β_p , siendo β_0 la constante del modelo (López y Fachelli, 2016, pp. 5, 11-22).

Para encontrar la probabilidad de ocurrencia de un suceso se utiliza la función logística la cual, es la encargada de encontrar para cada individuo según los valores de una serie de variables independientes (X_p), la probabilidad (P) de que presente el efecto en estudiado. Mediante una transformación logarítmica de dicha ecuación, a la que se le llama *logit*, la probabilidad (P) se convertirá en odds, que indica cuanto más probable es el éxito que el fracaso. De aquí surge la ecuación de la regresión logit, que es parecida a la ecuación de la regresión lineal múltiple (Fiuza y Rodríguez, 2000).

El objetivo del modelo de regresión logit es calcular la probabilidad de que una de las opciones de la variable dicotómica dependiente (Y) ocurra en función de una serie de variables independientes (X_i) que pueden estar en diferentes escalas de medida. Cuando se tiene una variable de respuesta binaria obtenida a partir de una variable independiente se denomina *regresión logit simple*, mientras que cuando la probabilidad del modelo se obtiene a partir de 2 o más variables independientes se denomina *regresión logit múltiple*.

Si la probabilidad de que suceda un evento es (P), la probabilidad de que no suceda es igual a 1 menos la probabilidad (P):

$$\begin{aligned}P_r(y = 1) &= P \\P_r(y = 0) &= 1 - P\end{aligned}$$

Cuando se tiene una sola variable independiente, el modelo es:

$$Y = \ln\left(\frac{P}{1-P}\right) = \alpha + \beta X \quad (2.1)$$

En cambio, cuando exista p variables independientes el modelo puede representarse de la siguiente forma:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (3.1)$$

Para el caso con más de una variable independiente o covariable, la expresión de la ecuación de regresión logit es:

$$Pr(Y = 1|X_1, X_2, \dots, X_p) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)}} \quad (4.1)$$

Otra forma de presentar esta relación es aplicando una transformación logarítmica a la ecuación (4.1). Esta transformación se denomina, transformación logit:

$$\text{logit}\left(Pr(Y = 1|X_p)\right) = \ln\left(\frac{Pr(Y = 1|X_p)}{1 - Pr(Y = 1|X_p)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (5.1)$$

Donde:

- $Pr(Y = 1|X_1, X_2, \dots, X_p)$ es la probabilidad condicionada de que ocurra (Y).
- X_1, X_2, \dots, X_p es el conjunto de p covariables que forman parte del modelo pueden ser numéricas o categóricas.
- β_0 es la constante del modelo o termino independiente.
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes de las covariables.

De este modo se obtiene un modelo lineal para el logaritmo de la probabilidad $P(X_i)$, así pues, este tipo de modelos se consideran un caso particular de los modelos lineales generalizados (GLM) (McCullagh y Nelder, 1989).

1.4.1.1 Odds (ratio del riesgo)

La transformación logit (ecuación 5.1) surge de considerar la relación o el cociente de probabilidades en dos sucesos, llamada *ventaja* o *razón (odds)*. La razón de un suceso es el cociente entre la probabilidad de que este suceda y la probabilidad de que no suceda.

$$\text{logit}(Pr(Y = 1|X_i)) = \ln\left(\frac{Pr(Y = 1|X_1, \dots, X_p)}{1 - Pr(Y = 1|X_1, \dots, X_p)}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (6.1)$$

$$\text{con } \alpha = \beta_0$$

Consideremos que el estimador del parámetro β es la variación del logaritmo natural del cociente de probabilidades, originada por una variación unitaria de dos individuos con valores X_1, X_2 de la variable X (suponiendo constante el resto de variables explicativas), para el primer individuo tendremos: $\ln\left(\frac{P_1}{1-P_1}\right) = \alpha + \beta X_1$ y, para el segundo $\ln\left(\frac{P_2}{1-P_2}\right) = \alpha + \beta X_2$, donde P_1 es la

probabilidad de que un individuo con valor X_1 de la variable X presente la característica de interés, mientras que P_2 representa la misma probabilidad para un individuo con valor X_2 . Restando estas dos igualdades se tiene:

$$\ln \frac{P_1}{1 - P_1} - \ln \frac{P_2}{1 - P_2} = (\alpha + \beta X_1) - (\alpha + \beta X_2) = \beta X_1 - \beta X_2 = \beta(X_1 - X_2) \quad (7.1)$$

Al aplicar la diferencia entre los dos logaritmos se tiene el cociente de logaritmos, de la siguiente forma:

$$\ln \left[\frac{\frac{P_1}{1 - P_1}}{\frac{P_2}{1 - P_2}} \right] = \beta(X_1 - X_2) \quad (8.1)$$

Al aplicar el logaritmo al cociente entre dos odds asociados (el obtenido al realizar el incremento y el anterior al mismo), aparece el concepto de **odds-ratio**, que se lo representa mediante el símbolo OR :

$$OR = \frac{\frac{P_1}{1 - P_1}}{\frac{P_2}{1 - P_2}} = \frac{P_1(1 - P_2)}{P_2(1 - P_1)} \quad (9.1)$$

Entonces a la expresión anterior se puede escribir, de la siguiente forma:

$$\ln (OR) = \beta(X_1 - X_2) \quad (10.1)$$

Considerando el caso, si los valores de X para los individuos se diferencien solo en una unidad tendremos: Aplicando la función exponencial a la ecuación (10.1) se tiene la forma:

$$\ln (OR) = \beta\{(X_1 - (X_1 - 1))\} = \beta \quad (11.1)$$

De modo que β se pueda interpretar como el algoritmo de la razón de *odds* para los individuos que se diferencian en una unidad respecto a la variable independiente.

Finalmente, aplicando la función exponencial al odds ratio para el modelo general logístico múltiple queda expresado así:

$$OR = \exp^{\beta} \quad (12.1)$$

De donde se deduce que, un coeficiente β cercano a cero, es decir, un odds-ratio próximo a 1, indicará que los cambios en la variable explicativa X_p en cuestión no tendrán efecto alguno sobre la probabilidad de ocurrencia del suceso (variable dependiente Y).

Análisis de riesgo

Dentro del análisis de riesgo de debe tener en cuenta que:

Si el $OR > 1$ y su intervalo de confianza no contiene al 1 se denomina a esa variable como un *factor de riesgo*.

Si el $OR < 1$ y el intervalo de confianza no contiene al 1 la variable será un *factor de protección*.

Si los intervalos de confianza contienen al 1 entonces las variables no serán significativas para el planteamiento de un modelo de regresión logístico solo serán consideradas como variables independientes aquellas que hayan sido definidas como factores de riesgo.

1.4.1.2 Análisis estadístico de variables confusoras

El test estadístico más usado para comprobar la posible confusión que puede generar una tercera variable en la relación entre dos variables es el *Test de Cochran-Mantel-Haenzel*.

Dicho test es un contraste de hipótesis que contrasta la igualdad del Odds ratio entre k tablas de contingencia de 2×2 examina la asociación ponderada de un conjunto de 2×2 tablas. Por consiguiente, es una forma estadística que permite evaluar la posible influencia que pueda tener sobre la relación entre las variables cualitativas dicotómicas y una tercera variable también cualitativa con k posibles valores.

Contraste de la Prueba de Cochran-Mantel-Haenzel

1. Planteamiento de hipótesis

$$H_0: OR_1 = \dots = OR_k$$

$$H_1: \exists_i tqOR_i \neq 1$$

2. Nivel de significancia

$$\alpha = 0.05$$

3. Estadístico

$$\widehat{OR} = \frac{\left(\sum_{i=1}^k \left(a_i - \frac{(a_i + b_i)(a_i + c_i)}{n_i} \right) \right)^2}{\sum_{i=1}^k \frac{(a_i + b_i)(a_i + c_i)(b_i + d_i)(c_i + d_i)}{n_i^3 - n_i^2}} \cong \chi_1^2 \quad (13.1)$$

χ_1^2 = es una variable chi cuadrada con un grado de libertad

Donde tenemos k tablas de 2×2

$$\begin{array}{cc} \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} & a_1 + b_1 + c_1 + d_1 = n_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix} & a_k + b_k + c_k + d_k = n_k \end{array}$$

4. Región de rechazo

Si $\widehat{OR} \geq 1$ No Rechazo H_0

5. Conclusión

Si se mantiene la hipótesis nula (H_0), se afirma que las Odds ratio son iguales en todas las tablas de contingencia y la tercera variable no será considerada como una variable confusora.

Si se rechaza la hipótesis nula, significa que la tercera variable a través de la cual se construyó las diferentes tablas de contingencia, está influyendo en la relación entre las dos primeras variables por lo tanto será considerada como una variable confusora (Llopis, 2013, p.35).

Test de significancia de los coeficientes

El estadístico de Wald sigue una distribución normal estándar y evalúa el coeficiente estimado en la población y se define como un cociente entre el coeficiente y el error estándar del coeficiente en la hipótesis (López y Fachelli, 2016, p. 25).

1. Planteamiento de hipótesis

$H_0: \beta_i = 0$ (la probabilidad de éxito no depende de las variables independientes x_i)

$H_1: \beta_i \neq 0$

2. Nivel de significancia

$\alpha = 0.05$

3. Estadístico

$$W = \left(\frac{\beta_i}{SE_{\beta_i}} \right)^2 \sim \chi_{1,1-\alpha}^2 \quad (14.1)$$

4. Región de rechazo

Si $p - value > 0.05$
Si $W < \chi_{1,1-\alpha}^2$ } No se rechaza H_0

5. Conclusión

Si el p-value es menor que el nivel de significancia establecido ($\alpha = 0.05$), se concluye que hay una relación entre el predictor de X_i y la probabilidad de éxito.

Si el estadístico W es inferior a $\chi_{1,1-\alpha}^2$ se concluye que la variable independiente no influye en la probabilidad de las características de la variable dependiente. Si la variable independiente es cualitativa los grados de libertad es igual al número de categoría menos 1.

1.4.2 Análisis de Series de tiempo

Es el estudio de datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares tales períodos pueden ser diarios, semanales, semestrales, trimestrales, anuales o de cualquier otro intervalo temporal; pueden estar referidas a una determinada característica (serie univariante) o a varias características (serie multivariante), las series son representadas por medio de una gráfica de líneas en su eje horizontal se representan los períodos y en el eje vertical los valores de la serie de tiempo; en el cual se busca construir un modelo para explicar la estructura y prever la evolución de los datos de una variable que se observa a lo largo del tiempo. Para analizar una serie de tiempo primero se debe identificar los componentes de la serie, aplicar las técnicas estadísticas para su análisis y realizar las proyecciones o pronósticos de eventos futuros (Hanke y Wichem, 2010).

El estudio de las series de tiempo tiene dos objetivos principales:

- Explicar el comportamiento pasado de la variable en estudio y efectuar análisis comparativos entre 2 o más series, ya sea de lugares o tiempos diferentes.
- Predecir el valor de la variable en el futuro, teniendo en cuenta que dicho comportamiento futuro se considera como una clase de extensión del pasado, que difícilmente será exactamente igual.

Matemáticamente, una serie de tiempo se define por los valores Y_1, Y_2, \dots, Y_T , de una variable Y (por ejemplo, la tasa de desempleo) en los momentos t_1, t_2, \dots, t_T . Así Y es una función de t y se simboliza por $Y = F(t)$. En cuanto al conjunto de observaciones historias de la variable Y son derivadas de intervalos regulares de tiempo y constituyen la serie temporal de datos Y_t con $t = 1, 2, \dots, T$.

1.4.3 Metodología Box-Jenkins

La metodología Box-Jenkins trata sobre el análisis de series de tiempo en cuanto a predicción y control, la cual tiene una serie de herramientas de predicción para el análisis univariante, en la cual la estimación y predicción de valores futuros de una serie de tiempo se basa únicamente en los valores anteriores de la misma serie y no en los de otras variables independientes. Las predicciones en este tipo de análisis se generan mediante un modelo algebraico que relaciona el valor actual de la serie con valores anteriores de la misma y con ciertas perturbaciones, que se ajusten a los datos de la serie. Este tipo de modelos se conocen con el nombre de Modelos Autorregresivos Integrados de Medias Móviles (ARIMA). Como primer paso para identificar un posible modelo ARIMA, es que si la serie no presenta estacionalidad, a través de diferenciarla d veces, se debe determinar los parámetros del modelo $ARIMA(p, d, q)$ y si la serie presenta estacionalidad de periodo k , los parámetros del modelo serán de tipo multiplicativo $ARIMA(p, d, q)(P, D, Q)_k$ (Bowerman et al., 2007).

Hanke y Wichern, 2010 define la metodología Box-Jenkins como “un método para generar pronósticos diferente a la mayoría de los métodos porque no supone ningún patrón en particular en los datos históricos de las series que se van a pronosticar, sino que se basa en un enfoque iterativo para identificar un modelo posible a partir de una clase general de modelos, que mejor describa la serie; teniendo en cuenta que el modelo está bien ajustado si, los residuos son generalmente pequeños, están distribuidos aleatoriamente y no contienen información útil; caso contrario el proceso se repite usando un nuevo modelo hasta encontrar uno útil para pronosticar” (Hanke y Wichern, 2010).

En síntesis, esta metodología describe un conjunto de procedimientos para identificar, ajustar y verificar modelos *ARIMA* con los datos de una serie de tiempo. Los pronósticos resultan directamente de un modelo ajustado de una serie de tiempo.

1.4.4 Modelo ARIMA para datos estacionales

$$ARIMA(p, d, q) \times (P, D, Q)_s$$

El primer modelo (p, d, q) corresponde a la parte regular y el segundo $(P, D, Q)_s$ a la parte estacional.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_{p+p+D+d} Y_{t-p-p-D-d} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \dots - \omega_q \varepsilon_{t-q} \quad (15.1)$$

$$+ \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} - \Omega_1 \varepsilon_{t-1} - \dots - \Omega_{Q+q} \varepsilon_{t-Q-q}$$

Donde:

- **p:** términos autorregresivos ordinales

- **d**: diferencias ordinales ($Y_t - Y_{t-1}$)
- **q**: términos de promedio móvil ordinales
- **P**: términos autorregresivos estacionales de orden **k**
- **D**: diferencias estacionales ($Y_t - Y_{t-k}$)
- **Q**: términos de promedio móvil estacionales
- **s**: grado estacional (**4, 12**)

Proceso de modelación Box-Jenkins

Para identificar qué tipo de modelo sigue la serie que se está estudiando y conocer los valores de los parámetros p , d y q correspondientes. Para dar solución a este problema Box-Jenkins propone un proceso de cuatro fases para encontrar el mejor modelo que permita estimar la serie de tiempo en estudio: (Rosales et al., 2008).

Fase 1: Identificación

Consiste en reconocer el posible modelo ARIMA que sigue la serie, lo que requiere determinar si la serie de tiempo es estacionaria y si hay alguna estacionalidad significativa que necesite ser modelada.

Análisis exploratorio de datos

Dado que al momento de selección del mejor modelo se presentan una serie de modelos *ARIMA*, se debe realizar como primer punto un análisis exploratorio de la serie en estudio.

Primero se grafica la serie original para identificar si es o no estacionaria mediante FAC. Si no lo es se procede a diferenciarla (d) tantas veces como fuese necesario para convertirla en estacionaria. Una vez diferenciada la serie, una buena estrategia consiste en comparar los correlogramas de la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP).

Una vez determinados los parámetros d y D y en el análisis exploratorio, el siguiente paso será identificar los parámetros p y q y, si la serie presentara el caso de estacionalidad, los parámetros P y Q .

Prueba de estacionariedad

Para comprobar la existencia de estacionariedad se aplica el contraste aumentado de Dickey-Fuller (ADF), que permite identificar la presencia de raíces unitarias, considera un modelo autorregresivo de primer orden (Guerrero, 2003):

$$Y_t = \mu + \rho Y_{t-1} + \varepsilon_t \quad (16.1)$$

Donde μ y ρ son parámetros a estimar y ε_t es un término del error que se supone cumple la propiedad de ruido blanco. Dado ρ es un coeficiente de autocorrelación que, toma valores $-1 < \rho < 1$, si $\rho = 1$, la serie Y_t es no estacionaria, la varianza decrece en el tiempo. Si el valor de $\rho > 1$, entonces se dice que la serie es explosiva.

De esta forma, la hipótesis de estacionariedad de una serie puede ser evaluada analizando si el valor absoluto de ρ es estrictamente menor que uno. Este test plantea como hipótesis nula la existencia de una raíz unitaria es decir la serie no es estacionaria.

Contraste:

1. Planteamiento de hipótesis

$$H_0: \rho = 1$$

$$H_1: \rho > 1$$

2. Nivel de significancia

$$\alpha = 0.05$$

3. Estadístico

$$DF_r = \frac{\hat{\rho}}{\sigma_{\hat{\rho}}} \sim t_{\alpha/2, n-2} \quad (17.1)$$

Donde:

- $\hat{\rho}$: es el coeficiente estimado
- $\sigma_{\hat{\rho}}$: es el error estándar en el coeficiente estimado

4. Región de rechazo

Si $p - value > 0.05$
 Si $t_{\alpha/2, n-2} > DF_r$ } No se rechaza H_0

5. Conclusión

Al no rechazar la hipótesis nula se concluye que la serie no es estacionaria en la componente regular, es decir existe un problema de raíces unitarias que amerita una primera diferenciación para resolver el problema de no estacionariedad.

Identificación de posibles modelos ARIMA

No resulta fácil seleccionar el mejor modelo *ARIMA*, por lo que es importante indicar que la selección del modelo no representa una decisión irreversible, sino como el inicio de un proceso de selección que puede ser revisado en las siguientes etapas; por lo que se recomienda realizar la prueba estadística **Q de Ljung-Box**

1. Planteamiento de hipótesis

H_0 : $\rho_1 = \rho_2 = \dots = \rho_m$ los datos se distribuyen de forma independiente

H_1 : Al menos un $\rho_k \neq 0 \quad \forall k = 1, 2, \dots, m$

2. Nivel de significancia

$\alpha = 0.05$

3. Estadístico

$$Q = n(n+2) \sum_{k=1}^m \frac{\rho_k^2}{n-k} \sim \chi^2_{1-\alpha, m-p-q-1} \quad (18.1)$$

Donde:

- n = numero de observaciones
- k = retraso de tiempo
- m = numero de retrasos de tiempo que se van a probar
- ρ_k = la función de autocorrelación muestral en el retraso k

4. Región de rechazo

Si $Q < \chi^2_{1-\alpha, (m-p-q-1)}$
Si $p\text{-value} > 0.05$ } No se rechaza H_0

5. Conclusión

Al No rechazarse la hipótesis nula, se puede afirmar que los errores del modelo siguen un proceso de ruido blanco.

Fase 2: Estimación

Una vez identificado el posible modelo estacionario se procede a estimar los parámetros, que pueden ser obtenidos mediante paquetes estadísticos, en este trabajo se utilizó el software R Studio la función *auto.arima* del paquete *forecast*.

Se procede a comprobar si los coeficientes β_i son significativamente distintos de cero.

Contraste de significancia de los coeficientes

1. Planteamiento de hipótesis

$H_0: \beta_i = 0$ (No hay relación lineal)

$H_1: \beta_i \neq 0$ (Relación lineal con pendiente $\neq 0$)

2. Nivel de significancia

$\alpha = 0.05$

3. Estadístico

$$\hat{\beta}_i = \frac{\beta_i - 1}{SE(\beta_i)} \sim t_{\frac{\alpha}{2}, n-k-1} \quad (19.1)$$

4. Región de rechazo

Si $\hat{\beta}_i < t_\alpha$
Si $p\text{-value} > 0.05$ } No se rechaza H_0

5. Conclusión

Se concluye que el parámetro no es significativo y no puede ser considerado en el modelo. El hecho de que el modelo identificado no sea adecuado no se debe descartar la posibilidad de que exista otro modelo más complejo que pueda ajustarse mejor a la serie observada.

Fase 3: Verificación del modelo

En esta fase se busca evaluar si el modelo estimado se ajusta a los datos en forma razonablemente buena, ya que es posible que exista otro modelo. La validación o verificación incluye el análisis de los coeficientes o parámetros del modelo, la evaluación de la bondad de ajuste y análisis de los residuos.

Análisis de los coeficientes

El modelo estimado debe cumplir con las condiciones de estacionariedad y debe existir significancia estadística en los retrasos. Teniendo en cuenta las propiedades asintóticas de la estimación por mínimos cuadrados ordinarios, los estadísticos t de student pueden utilizarse para probar significancia individual de cada uno de los coeficientes (ecuación 19.1).

Bondad de ajuste

Debido a que en la fase de identificación se plantea más de un modelo tentativo, en esta fase de verificación es importante identificar cuál de los modelos presenta una mejor bondad de ajuste.

Una herramienta estadística útil para elegir el número adecuado de retrasos de p y q , es el criterio de información Akaike o AIC, la cual se basa en la suma de los cuadrados de los errores, buscando minimizar a partir de diversas combinaciones de p y q (Hanke y Wichern, 2010).

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n}r \quad (20.1)$$

Donde:

\ln = Logaritmo natural

$\hat{\sigma}^2$ = Suma residual de cuadrados dividida entre el número de observaciones

n = Número de observaciones

r = Número total de parámetros (incluyendo el término constante)

Análisis de los residuos

El supuesto de que los errores del modelo son aleatorios (media cero, varianza constante y no correlación serial) debe verificarse.

Independencia

Se lo puede identificar mediante la inspección del comportamiento del correlograma a medida que aumenta el número de retrasos y a través del estadístico **Q de Ljung_Box** (ecuación 18.1).

Normalidad

Se lo puede identificar gráficamente mediante la construcción de un histograma de frecuencias y analíticamente se efectúa con el estadístico Jarque Bera.

Test

1. Planteamiento de hipótesis

H_0 : Los residuos siguen una distribución normal

H_1 : Los residuos siguen una distribución normal

2. Nivel de significancia

$\alpha = 0.05$

3. Estadístico

$$JB = n \left(\frac{SK^2}{6} + \frac{(KU - 3)^2}{24} \right) \sim \chi_2^2 \quad (21.1)$$

Donde:

$$SK = \frac{\hat{\mu}_3}{\hat{\mu}_3^{3/2}} = \frac{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^3}{\left(\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2 \right)^{3/2}} \text{ Coeficiente muestral de asimetría}$$

$$KU = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} = \frac{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^4}{\left(\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2 \right)^2} \text{ Coeficiente muestral de curtosis}$$

n = Número de residuos

4. Región de rechazo

$$\left. \begin{array}{l} JB < \chi_2^2 \\ \text{Si } p\text{-value} > 0.05 \end{array} \right\} \text{ No se rechaza } \mathbf{H}_0$$

5. Conclusión

Al no rechazarse la hipótesis nula, se puede afirmar que existe linealidad o los errores se distribuyen normalmente.

Fase 4: Pronóstico

Para el paso final de pronosticar un periodo futuro a partir del mejor modelo seleccionado, es importante considerar si la variable original fue diferenciada. Además, que los residuos deben cumplir rigurosamente con los supuestos de normalidad e independencia caso contrario los valores pronosticados a futuro no serán óptimos y confiables. Es decir, el modelo seleccionado no es adecuado para realizar predicciones, por lo que se debe utilizar otra técnica de predicción, por ejemplo, el método Bootstrap, que a continuación definimos.

1.4.5 Método bootstrap

Bootstrap fue propuesta por Efron en el año 1979, con el propósito de hallar intervalos de confianza para parámetros desconocidos en circunstancias donde es imposible encontrar analíticamente la distribución estadística de interés; la aplicación de este tipo de remuestreo computacional (González y Fabio, 2008).

Este método se basa en que la distribución de la muestra es la mejor forma de representar la distribución de la población y que remuestrear la misma permite distinguir lo que sucedería si se remuestrea la población. Los datos observados son remuestreados con reemplazamiento para generar nuevas muestras que ajustarán un modelo adecuado. En términos generales este método tiene una muestra aleatoria de tamaño n , en donde $x = (x_1, x_2, \dots, x_n)$ son los valores observados de dicha muestra. Muestreando n veces con reemplazamiento de la muestra original (x_1, x_2, \dots, x_n) se obtienen las remuestras Bootstrap $x_b^* = (x_{1b}, x_{2b}, \dots, x_{nb})$ para $b = 1, 2, \dots, B$.

De esta manera, a partir de la muestra aleatoria se define la función de distribución empírica \hat{F} como la distribución discreta que asigna una probabilidad $1/n$ a cada elemento de la muestra original.

En el caso de contraste de hipótesis, este método desarrolla el siguiente proceso (Araya, 2010, pp.28-29):

1. Dado una muestra de tamaño n , calcular el estadístico de interés ($\hat{\theta}$).
2. Generar una muestra Bootstrap de tamaño n mediante un muestreo con reemplazo de la muestra original, asignando a cada observación una probabilidad de selección de $1/n$, denominada “remuestra” y representada por $x_b^* = (x_{1b}, x_{2b}, \dots, x_{nb})$.
3. Se calcula el valor del estadístico $\hat{\theta}_b^*$ que se utiliza como estimador del parámetro poblacional θ , para la remuestra.
4. Se repite los pasos 2 y 3 un número grande de veces (N) hasta obtener B estimaciones $\hat{\theta}_b^*$.
5. Con la secuencia de estimadores se construye la distribución Bootstrap del estadístico asignando una probabilidad de $1/B$ a cada punto $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$, que representan una aproximación de la función de probabilidad para cada estadístico.

Finalmente, una de las ventajas y la más importante del método Bootstrap con respecto a otros métodos de estimación es que, consiste en mantener la estructura de los datos observados, sin recurrir a estudios paramétricos, es decir, no es necesario asumir que los datos se ajustan a un determinado modelo estadístico.

CAPÍTULO II

2 MARCO METODOLÓGICO

2.1 Tipo y Diseño de Investigación

La presente investigación es de tipo mixta, dado que la variable dependiente es una mudable estadística y las series temporales explican el comportamiento del desempleo juvenil; según el objetivo de estudio será investigación aplicada ya que se busca determinar los factores sociodemográficos y sociolaborales que explican el comportamiento de la tasa de desempleo juvenil; respecto al nivel de profundización en el objeto de estudio es considerada investigación explicativa, puesto que una vez obtenida y procesada la información, permitirá identificar el comportamiento de las variables del modelo logit, las variables que se van a manipular son no experimental ya que la matriz de información proviene de una fuente de información secundaria, dado el tipo de inferencia es investigación deductiva ya que busca determinar cuáles son los factores influyentes en el desempleo juvenil a través de la información presentada por la encuesta ENEMDU, además es considerada según el periodo investigación longitudinal (análisis de series de tiempo) e investigación transversal (regresión logit) ya que se trabajará con datos pertenecientes al periodo 2013 – 2018 (Hernández et al, 2014, p. 154).

Descripción del área de estudio



Figura 2-2: División geográfica de Ecuador
Fuente: («Provincias del Ecuador y sus Capitales», 2020)

La presente investigación se llevó a cabo a nivel nacional con información del Instituto de Estadística y Censos (INEC) con sede en la ciudad de Quito – Ecuador. La república del Ecuador se localiza en la costa noroccidental de América del Sur; limita al norte con Colombia, sur y este con Perú y al oeste el Océano Pacífico, el cual separa de las Islas Galápagos; tiene una extensión territorial de 283.561 km², es el cuarto país más pequeño del subcontinente y el décimo país más

poblado de América con más de 17 millones de habitantes; a pesar de sus pequeñas dimensiones, Ecuador es uno de los países con mayor diversidad geográfica del mundo, ostenta de cuatro regiones geográficas: Costa (Litoral), Sierra (Interandina), Oriente (Amazonia) y la Región Insular (Islas Galápagos), dividida políticamente en 24 provincias, las mismas que se dividen en cantones y estos en parroquias urbanas y rurales gobernadas por alcaldes, jefes políticos y tenientes políticos, según se trate de provincia, cantón o parroquia respectivamente.

2.2 Población de estudio

La encuesta de la ENEMDU está enfocada a hogares y personas que en ellas habitan, dentro de viviendas particulares ubicadas en el territorio ecuatoriano, excluyendo la Región Insular, también está excluida la población en viviendas colectivas, viviendas flotantes y sectores con población indigente, sin embargo, se incluye las zonas no delimitadas. Para el análisis estadístico de esta investigación, la población objeto de estudio son todas las personas informantes que se encuentran en una edad comprendida entre 18 y 29 años, presentes en la base de datos de la encuesta (ENEMDU) del periodo 2013 – 2018.

2.3 Tamaño de la muestra

Basándonos en los objetivos de esta investigación, se decidió filtrar la información presentada por la ENEMDU de la población económicamente activa (PEA) enfocándose específicamente en los jóvenes que se encuentran dentro de la edad de 18 a 29 años y residen en el Ecuador, obteniendo los siguientes tamaños de muestra:

Tabla 1-2: Tamaño de la muestra, por años y trimestres

TRIMESTRE	2013	2014	2015	2016	2017	2018
MARZO	2957	6883	7414	7497	7512	7383
JUNIO	9069	13544	14075	7472	7361	7255
SEPTIEMBRE	2932	7187	7422	8054	7314	7560
DICIEMBRE	9180	13896	13886	14129	13951	7110
TOTAL	24138	41510	42797	37152	36138	29308

Fuente: Banco de datos abiertos del INEC-ENEMDU

Realizado por: Tierra, Tania. 2020

2.4 Método de muestreo

Se trabaja con información secundaria proveniente de la encuesta ENEMDU la cual utiliza un tipo de muestreo probabilístico en dos etapas, bietápico, con estratificación geográfica por dominios de estudio y área urbana - rural

2.5 Recolección de información

La recolección de datos se realizó con la encuesta ENEMDU aplicada por el INEC, misma que tiene cobertura geográfica a nivel nacional urbano de 5 ciudades mayores (Quito, Cuenca, Guayaquil, Machala, Ambato), resto sierra, costa y región Amazónica, para las encuestas de hogares rurales, los dominios son: costa, sierra y amazonia. Estos dominios cubren todo el territorio del Ecuador continental, dicha recopilación de la información es de carácter social, demográfico y económico. La encuesta ENEMDU, se realiza trimestralmente para ser criticada, codificada, digitada, validada e ingresada al sistema y para luego ser publicada de forma trimestral.

2.6 Operacionalización de las variables

Siguiendo la estructura de la encuesta ENEMDU y basado en la necesidad de esta investigación para alcanzar los objetivos planteados se trabajó con las siguientes variables:

Tabla 2-2: Operacionalización de las variables

Ítem	VARIABLE	DESCRIPCION	TIPO DE VARIABLE	ESCALA DE MEDIDA	CATEGORIAS
1	Desempleo	Indica el estado laboral del joven de 18 a 29 años.	Cualitativa: Dicotómica	Nominal	0: No desempleo 1: Desempleo
2	Sexo	Determinado por naturaleza al nacer, puede ser hombre o mujer.	Cualitativa: Dicotómica	Nominal	1: Hombre 2: Mujer
3	Edad	Tiempo en años que una persona ha vivido desde su nacimiento.	Cuantitativa	Ordinal	1: De 18 a 21 años 2: De 22 a 25 años 3: De 26 a 29 años
4	Parentesco	Es el lazo consanguíneo y de afinidad que los miembros del hogar tienen con respecto al jefe.	Cualitativa: Dicotómica	Ordinal	1: Jefe 0: Otros
5	Estado Civil	Indica el estado conyugal del joven.	Cualitativa: Dicotómica	Nominal	1: Con pareja 0: Sin pareja estable
6	Asiste a clases	Indica si el joven asiste o no a clases.	Cualitativa: Dicotómica	Nominal	1: Si 2: No
7	Nivel de instrucción	Indica la instrucción académica del joven de 18-29 años.	Cualitativa: Politómica	Ordinal	0: Ninguno 1: Primaria 2: Secundaria 3: Superior y más
8	Ingreso laboral	Indica la cantidad monetaria que recibe el joven de 18-29 años en intervalos regulares (diario, semanal, quincenal, mensual) Se toma como base el promedio del período en estudio del Salario Básico Unificado (SBU) de \$356,50.	Cuantitativa	Ordinal	1: Ingreso mayor al SBU 0: Ingreso menor al SBU

9	Rama de la actividad	Indica la clasificación de las empresas de trabajo dentro de un sector de la economía.	Cualitativa: Politómica	Nominal	1: Sector Primario 2: Sector Secundario 3: Sector Terciario 4: No especificado
10	Grupo de ocupación	Indica la clasificación de la tarea o actividad que desarrolla el trabajador dentro del establecimiento.	Cualitativa: Politómica	Nominal	1: Ocupación no manual 2: Ocupación manual 3: No especificado
11	Recibe Seguro social	Indica si el joven se encuentra dentro de los sistemas de seguros de salud público o privado.	Cualitativa: Dicotómica	Nominal	1: Si 2: No

Fuente: Banco de datos abiertos del INEC-ENEMDU

Realizado por: Tierra, Tania. 2020

2.7 Alcance de la investigación

Teniendo presente el problema planteado, la investigación tiene alcances descriptivos y correlacional. Descriptiva porque se puede visualizar el comportamiento del desempleo juvenil a nivel nacional y provincial en Ecuador que se reduce, pero mantiene tasas significativamente; y correlacional porque describe la interdependencia entre las variables relacionadas al desempleo juvenil.

2.8 Análisis de datos

En esta investigación se aplicó la técnica de Series de Tiempo Box- Jenkins, la cual permitió analizar la información en relación al tiempo, enfocándose en describir los patrones que presenta la serie, e identificar cual es la magnitud del desempleo juvenil en el país y los cambios que presentan a lo largo del tiempo. Se aplicó la Regresión Logit, cuyo objetivo es modelar como influye en la probabilidad de aparición de un suceso, la presencia o no de diversos factores y el valor o nivel de los mismos, que explica el comportamiento de cómo los factores influyentes inciden en la probabilidad de que un joven ecuatoriano se encuentre desempleado o no, en el periodo de marzo 2013 a diciembre 2018. Como análisis adicional a las series de tiempo al no cumplir el supuesto de normalidad se ajusta un modelo de predicción mediante el método Bootstrap realizando remuestras de la serie original.

Todo lo antes mencionado se lo realizó utilizando el software estadístico R versión 3.5.2 para las series de tiempo, método bootstrap y estimación del modelo de regresión logit y PSPP versión 0.10.4 reemplazo libre de SPSS para el cálculo de las tasas de desempleo.

En el siguiente apartado se considerará las variables: desempleo, sexo, edad, estado civil, relación de parentesco, nivel de instrucción, asiste a clases, ingreso laboral, rama de la actividad, seguro social y grupo de ocupación para la regresión logit y en relación a las series de tiempo se trabajará con las tasas de desempleo nacional y provincial por trimestres de marzo de 2013 a diciembre de

2018. Así mismo el análisis exploratorio de datos se realizará con todas las variables, mismas que fueron recodificadas, por medio de discretización, para un mejor entendimiento y explicación de los cambios más relevantes dentro de la investigación.

Para una mejor presentación se utiliza las abreviaturas de las provincias del Ecuador subdivisiones enumeradas en el estándar ISO 3166-2 publicada por la ISO 3166 Maintenance Agency (ISO 3166/MA).

EC-A Azuay
EC-B Bolívar
EC-F Cañar
EC-C Carchi
EC-X Cotopaxi
EC-H Chimborazo
EC-O El Oro
EC-E Esmeraldas
EC-G Guayas
EC-I Imbabura
EC-L Loja
EC-R Los Ríos
EC-M Manabí
EC-S Morona Santiago
EC-N Napo
EC-Y Pastaza
EC-P Pichincha
EC-T Tungurahua
EC-Z Zamora Chinchipe
EC-U Sucumbíos
EC-D Orellana
EC-SD Santo Domingo
EC-SE Santa Elena
EC-ZND Zonas no delimitadas

CAPÍTULO III

3 ANÁLISIS Y DISCUSIÓN DE RESULTADOS

3.1 Tasa de desempleo juvenil por provincias

Tabla 1-3: Tasas del desempleo juvenil en Ecuador por provincia y nacional 2013-2018

Año	Trimestre	Provincia					TNDJ
		Azuay	El Oro	Guayas	Pichincha	Tungurahua	
2013	Marzo	3.96	10.37	12.59	9.32	9.48	10.57
	Junio	6.81	8.03	12.93	7.77	4.92	9.11
	Septiembre	9.39	9.24	11.61	6.05	5.03	9.24
	Diciembre	5.59	7.43	12.18	5.99	4.86	8.64
2014	Marzo	6.04	10.07	13.80	9.21	6.45	10.71
	Junio	6.82	8.71	13.95	9.58	4.99	10.08
	Septiembre	4.63	4.45	9.43	8.51	6.16	8.46
	Diciembre	2.74	5.93	10.54	6.18	4.58	7.88
2015	Marzo	3.22	6.22	8.64	10.31	8.12	8.50
	Junio	4.04	5.17	11.04	8.20	3.82	8.70
	Septiembre	2.88	6.53	11.05	6.43	4.22	8.70
	Diciembre	4.00	8.09	12.18	8.91	6.92	9.72
2016	Marzo	5.01	8.73	16.39	13.19	10.03	11.73
	Junio	4.50	5.51	9.41	11.77	9.64	10.01
	Septiembre	6.97	7.52	11.72	13.31	10.47	10.75
	Diciembre	5.45	8.67	13.60	12.79	6.27	10.43
2017	Marzo	5.13	6.84	11.40	15.83	4.83	9.12
	Junio	5.45	7.21	11.28	13.68	4.08	9.26
	Septiembre	4.55	9.83	9.04	13.94	4.94	8.47
	Diciembre	7.75	8.51	10.58	15.59	8.40	9.61
2018	Marzo	7.75	7.04	10.20	12.12	9.91	9.31
	Junio	7.74	7.41	8.98	14.97	6.29	8.88
	Septiembre	8.04	10.31	7.24	15.28	6.75	9.22
	Diciembre	5.56	9.42	7.30	13.70	4.10	8.25

Fuente: Banco de datos abiertos del INEC-ENEMDU

Realizado por: Tierra, Tania. 2020

Ecuador es un país en vías de desarrollo, que durante los últimos años ha venido preocupándose del desempleo, en el cual los jóvenes son los más afectados debido a que por diferentes factores no han encontrado trabajo. En la tabla anterior se muestra que en los últimos tres años los niveles de desempleo juvenil nacional han bajado, no obstante si hacemos una comparación entre las tasas de desempleo de las cinco provincias pertenecientes a las ciudades más grandes del país donde se centra la mayor parte de la actividad económica ciudades como Quito, Guayaquil, Cuenca, Ambato y Machala⁵; este indicador presenta tasas de desempleo juvenil altas para las provincias de El Oro con 8.73%, Tungurahua con 10.03% y Pichincha con 13.19%, Guayas es la provincia que registra la mayor tasa de desempleo juvenil a marzo del 2016 con 16.39% y Azuay registra

⁵ Según la nueva metodología del ENEMDU vigente desde el año 2007, esta se la realiza en las cinco ciudades más importantes del país y en 127 centros poblados, de forma trimestral y anual.

la tasa de desempleo juvenil más baja con 5.01%. Si se observa la tabla del ANEXO B se puede apreciar notablemente que la provincia de Bolívar es una de las provincias que menor tasa de desempleo juvenil tiene con un 3.41%.

3.2 Análisis exploratorio de datos

Variable: Desempleo

Tabla 2-3: Frecuencias de la variable desempleo en el período 2013 - 2018

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	Desempleo	312	10.55	768	8.47	267	9.11	749	8.16
	No desempleo	2645	89.45	8301	91.53	2665	90.89	8431	91.84
2014	Desempleo	683	9.92	1188	8.77	582	8.1	1067	7.68
	No desempleo	6200	90.08	12356	91.23	6605	91.9	12829	92.32
2015	Desempleo	610	8.23	1164	8.27	627	8.45	1296	9.33
	No desempleo	6804	91.77	12911	91.73	6795	91.55	12590	90.67
2016	Desempleo	806	10.75	682	9.13	779	9.67	1325	9.38
	No desempleo	6691	89.25	6790	90.87	7275	90.33	12804	90.62
2017	Desempleo	632	8.41	637	8.65	627	8.57	1234	8.85
	No desempleo	6880	91.59	6724	91.35	6687	91.43	12717	91.15
2018	Desempleo	677	9.17	623	8.59	685	9.06	579	8.14
	No desempleo	6706	90.83	6632	91.41	6875	90.94	6531	91.86

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

En el año 2016 para el 1^{er} trimestre se tenía un 10.75% de jóvenes desempleados. Respondiendo así a la incógnita de la magnitud del desempleo juvenil, mismo que presenta porcentajes bajos en el 4^{to} trimestre, siendo así que el período en estudio cierra con una tasa de desempleo juvenil de 8.14%, en el cual se puede evidenciar que para el año 2018 la tasa de desempleo juvenil tuvo una evolución decreciente, pero con porcentajes altos.

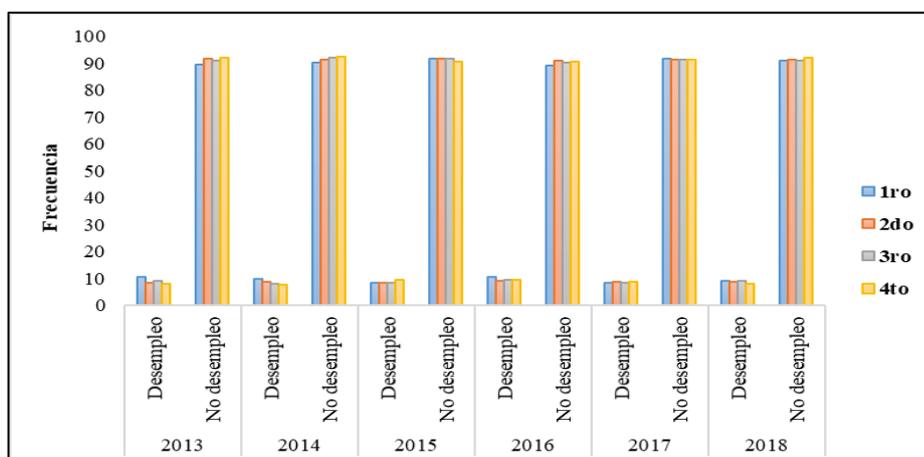


Gráfico 1-3: Gráfico de frecuencias de la variable desempleo en el período 2013-2018

Realizado por: Tierra, Tania. 2020

Variable: Sexo

Tabla 3-3: Frecuencias de la variable sexo

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		f_i	%	f_i	%	f_i	%	f_i	%
2013	Hombre	1790	60.53	5665	62.47	1743	59.45	5795	63.13
	Mujer	1167	39.47	3404	37.53	1189	40.55	3385	36.87
2014	Hombre	4184	60.79	8403	62.04	4403	61.26	8456	60.85
	Mujer	2699	39.21	5141	37.96	2784	38.74	5440	39.15
2015	Hombre	4473	60.33	8419	59.82	4333	58.38	8264	59.51
	Mujer	2941	39.67	5656	40.18	3089	41.62	5622	40.49
2016	Hombre	4263	56.86	4329	57.94	4640	57.61	8268	58.52
	Mujer	3234	43.14	3143	42.06	3414	42.39	5861	41.48
2017	Hombre	4338	57.75	4304	58.47	4272	58.41	8153	58.44
	Mujer	3174	42.25	3057	41.53	3042	41.59	5798	41.56
2018	Hombre	4398	59.57	4379	60.36	4510	59.66	4392	61.77
	Mujer	2985	40.43	2876	39.64	3050	40.34	2718	38.23

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

En la tabla se observa que en el 1^{er} trimestre del 2016 el 56.86% de los jóvenes son hombres lo que indica que este segmento de jóvenes es más propenso a estar desempleado, frente al 43.14% de mujeres, esto se puede observar en la siguiente gráfica, en la cual se nota claramente que las mujeres tienen menor porcentaje de encontrarse desempleadas a diferencia de los hombres.

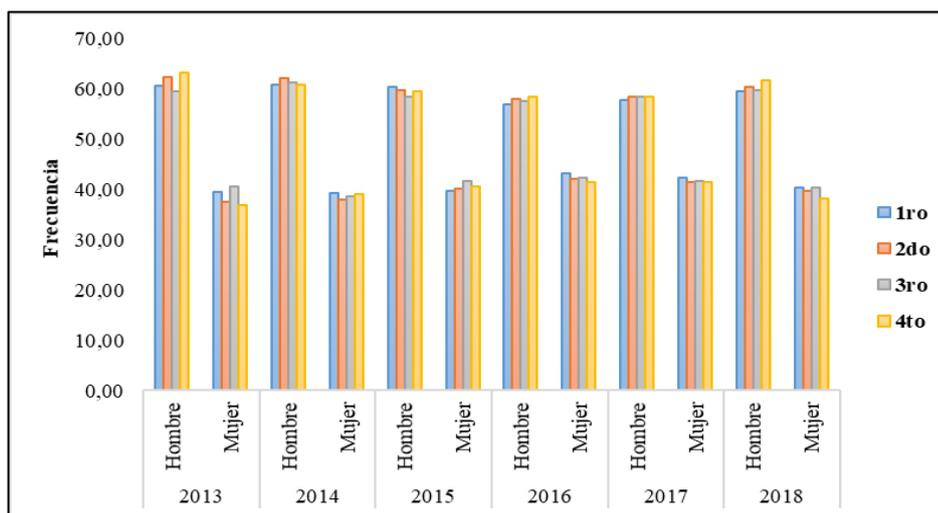


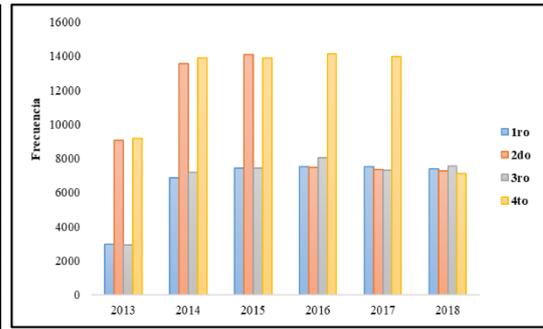
Gráfico 2 -3: Gráfico de frecuencias de la variable sexo

Realizado por: Tierra, Tania. 2020

Variable: Edad 18-29 años

Tabla 4-3: Frecuencias de la variable edad

Año	Trimestre							
	1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	2957	100	9069	100	2932	100	9180	100
2014	6883	100	13544	100	7187	100	13896	100
2015	7414	100	14075	100	7422	100	13886	100
2016	7497	100	7472	100	8054	100	14129	100
2017	7512	100	7361	100	7314	100	13951	100
2018	7383	100	7255	100	7560	100	7110	100



Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

Gráfico 3-3: Gráfico de frecuencias de la variable edad

Realizado por: Tierra, Tania. 2020

En el 1^{er} trimestre del 2016 existe 7497 jóvenes de entre 18 a 29 años, mientras que para el 4^{to} existe 14129 jóvenes que se encuentran impredecibles en el campo laboral.

Variable: Relación de parentesco

Tabla 5-3: Frecuencias de la variable relación de parentesco

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	Jefe	478	16.17	1211	13.35	556	18.96	1792	19.52
	Otros	2479	83.83	7858	86.65	2376	81.04	7388	80.48
2014	Jefe	1366	19.85	3661	27.03	1657	23.06	2852	20.52
	Otros	5517	80.15	9883	72.97	5530	76.94	11044	79.48
2015	Jefe	1392	18.78	3168	22.51	1657	22.33	3148	22.67
	Otros	6022	81.22	10907	77.49	5765	77.67	10738	77.33
2016	Jefe	1621	21.62	1563	20.92	1465	18.19	2661	18.83
	Otros	5876	78.38	5909	79.08	6589	81.81	11468	81.17
2017	Jefe	1349	17.96	1306	17.74	1352	18.49	2666	19.11
	Otros	6163	82.04	6055	82.26	5962	81.51	11285	80.89
2018	Jefe	1233	16.70	1189	16.39	1212	16.03	1099	15.46
	Otros	6150	83.30	6066	83.61	6348	83.97	6011	84.54

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

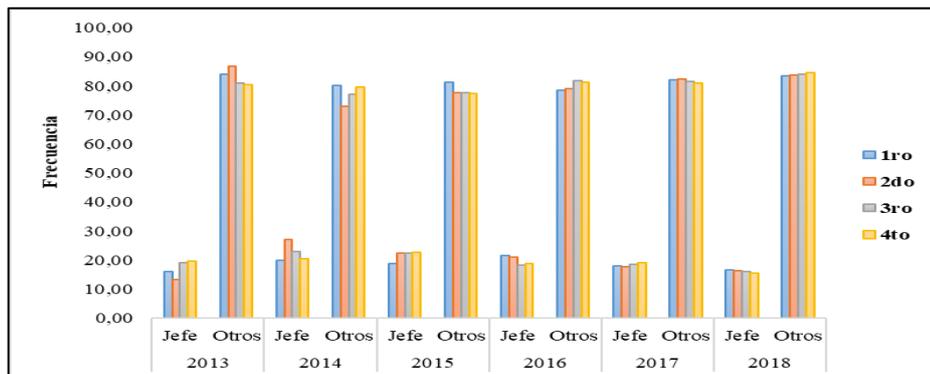


Gráfico 4-3: Gráfico de frecuencias de la variable relación de parentesco

Realizado por: Tierra, Tania. 2020

Con esta variable se identifica si el joven es el encargado del sustento del hogar y de sus miembros debido a diversas circunstancias sean estas económicas o sociales, reflejándose así que en el 1^{er} trimestre del 2016 el 21.62% de los jóvenes asumieron el papel primordial de ser jefe de hogar. En la gráfica se puede observar que la categoría otros tienen mayor porcentaje de participación, lo que indica que la función de jefe de hogar lo lleva a cargo el padre, madre, abuelos entre otros y no el joven.

Variable: Estado civil

Tabla 6-3: Frecuencias de la variable estado civil

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	Con pareja	904	30.57	2684	29.60	1038	35.40	3303	35.98
	Sin pareja estable	2053	69.43	6385	70.40	1894	64.60	5877	64.02
2014	Con pareja	2625	38.14	5932	43.80	3004	41.80	5563	40.03
	Sin pareja estable	4258	61.86	7612	56.20	4183	58.20	8333	59.97
2015	Con pareja	2827	38.13	5784	41.09	3218	43.36	5977	43.04
	Sin pareja estable	4587	61.87	8291	58.91	4204	56.64	7909	56.96
2016	Con pareja	3191	42.56	3135	41.96	3240	40.23	5691	40.28
	Sin pareja estable	4306	57.44	4337	58.04	4814	59.77	8438	59.72
2017	Con pareja	2898	38.58	2836	38.53	2767	37.83	5380	38.56
	Sin pareja estable	4614	61.42	4525	61.47	4547	62.17	8571	61.44
2018	Con pareja	2673	36.20	2567	35.38	2595	34.33	2348	33.02
	Sin pareja estable	4710	63.80	4688	64.62	4965	65.67	4762	66.98

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

Podemos apreciar que en el 1^{er} trimestre del 2016 la mayoría de los jóvenes se encontraban sin pareja estable es decir el 57.44% de los jóvenes estaban solteros, a diferencia del 38.58% de los jóvenes se encontraban con pareja ya sea en estado conyugal o de unión libre.

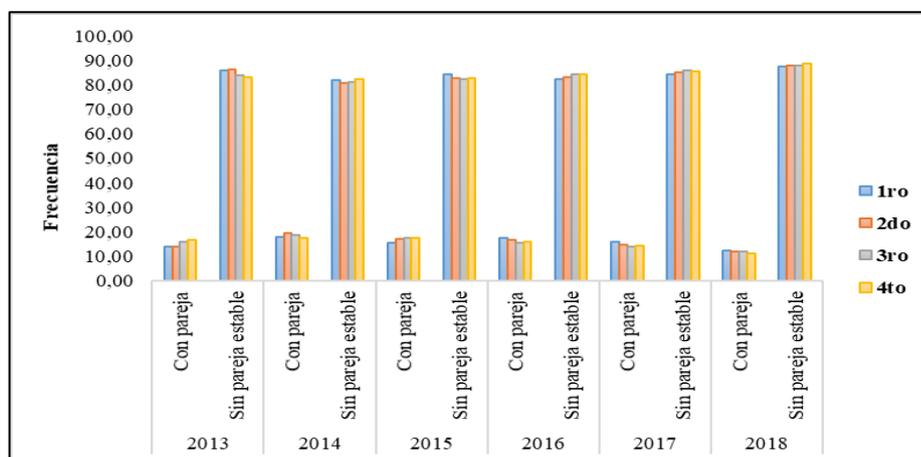


Gráfico 5-3: Gráfico de frecuencias de la variable estado civil

Realizado por: Tierra, Tania. 2020

Es necesario analizar si el joven tiene o no una relación de convivencia, ya que esta variable podría ser una razón por la cual el joven este desempleado o a su vez no se sienta presionado a conseguir empleo de inmediato dado que no tendría responsabilidad alguna, evidenciándose así en el gráfico anterior que la mayoría de jóvenes no tienen pareja estable, es decir se encuentran solteros, divorciados o viudos.

Variable: Asiste a clases

Tabla 7-3: Frecuencias de la variable asiste a clases

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	No	2348	79.40	7626	84.09	2398	81.79	8094	88.17
	Si	609	20.60	1443	15.91	534	18.21	1086	11.83
2014	No	5911	85.88	12050	88.97	6261	87.12	12180	87.65
	Si	972	14.12	1494	11.03	926	12.88	1716	12.35
2015	No	6336	85.46	12331	87.61	6434	86.69	12229	88.07
	Si	1078	14.54	1744	12.39	988	13.31	1657	11.93
2016	No	6417	85.59	6439	86.18	6875	85.36	12370	87.55
	Si	1080	14.41	1033	13.82	1179	14.64	1759	12.45
2017	No	6429	85.58	6254	84.96	6272	85.75	12131	86.95
	Si	1083	14.42	1107	15.04	1042	14.25	1820	13.05
2018	No	6334	85.79	6204	85.51	6406	84.74	6094	85.71
	Si	1049	14.21	1051	14.49	1154	15.26	1016	14.29

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

Esta variable nos permite investigar la cantidad de jóvenes que se encuentran asistiendo a clases además de conocer el porcentaje de jóvenes que abandonan sus estudios, en la cual en el 1^{er} trimestre del 2016 el 85.59% de los jóvenes no asistieron a clases, lo cual provoca que la probabilidad de desempleo se incremente, no obstante se aprecia dentro del mismo período un menor número de jóvenes que están asistiendo a clases con un 14.41%., como se observa en el gráfico siguiente existe un mayor número de jóvenes que actualmente no asisten a clases, las causas posibles pueden ser no aprobaron a una carrera universitaria o a su vez están en busca de empleo y no lo consiguen, motivos por lo cual dejan sus estudios paralizados, perdiendo la oportunidad de conseguir un título universitario.

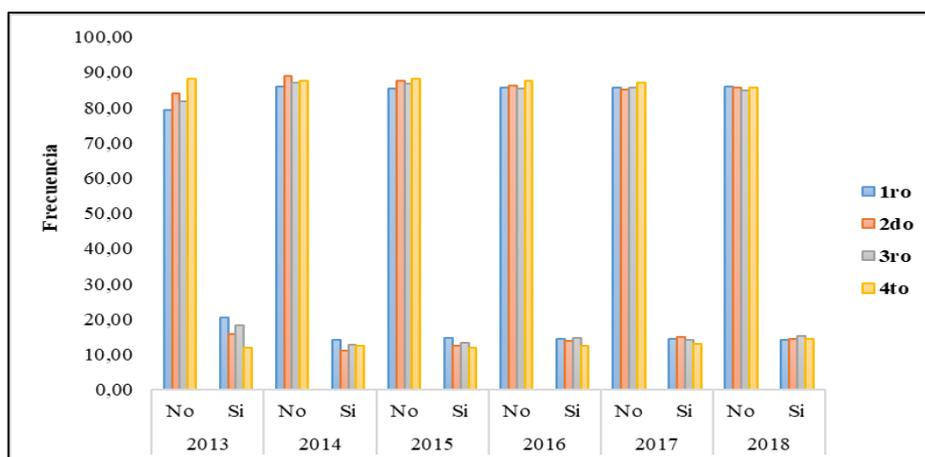


Gráfico 6 -3: Gráfico de frecuencias de la variable asiste a clases

Realizado por: Tierra, Tania. 2020

Variable: Nivel de instrucción

Tabla 8-3: Frecuencias de la variable nivel de instrucción

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	Ninguno	7	0.24	64	0.71	10	0.34	87	0.95
	Primaria	384	12.99	2511	27.69	422	14.39	2710	29.52
	Secundaria	1488	50.32	4267	47.05	1439	49.08	4488	48.89
	Superior y más	1078	36.46	2227	24.56	1061	36.19	1895	20.64
2014	Ninguno	31	0.45	86	0.63	28	0.39	65	0.47
	Primaria	1672	24.29	3798	28.04	1838	25.57	3676	26.45
	Secundaria	3485	50.63	6874	50.75	3662	50.95	7321	52.68
	Superior y más	1695	24.63	2786	20.57	1659	23.08	2834	20.39
2015	Ninguno	29	0.39	83	0.59	27	0.36	72	0.52
	Primaria	1723	23.24	3913	27.80	2016	27.16	3448	24.83
	Secundaria	3936	53.09	7213	51.25	3738	50.36	7510	54.08
	Superior y más	1726	23.28	2866	20.36	1641	22.11	2856	20.57
2016	Ninguno	36	0.48	28	0.37	30	0.37	73	0.52
	Primaria	1738	23.18	1712	22.91	1819	22.59	3252	23.02
	Secundaria	3958	52.79	4016	53.75	4383	54.42	7867	55.68
	Superior y más	1765	23.54	1716	22.97	1822	22.62	2937	20.79
2017	Ninguno	30	0.40	27	0.37	25	0.34	62	0.44
	Primaria	1644	21.88	1573	21.37	1486	20.32	2965	21.25
	Secundaria	4089	54.43	4081	55.44	4145	56.67	7932	56.86
	Superior y más	1749	23.28	1680	22.82	1658	22.67	2992	21.45
2018	Ninguno	26	0.35	24	0.33	15	0.20	19	0.27
	Primaria	1382	18.72	1322	18.22	1339	17.71	1266	17.81
	Secundaria	4195	56.82	4146	57.15	4322	57.17	4112	57.83
	Superior o más	1780	24.11	1763	24.30	1884	24.92	1713	24.09

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

La variable nivel de instrucción académica es la que definirá la profesión futura del joven, ya que de esta dependerá el cargo que vaya a ocupar cuando desee conseguir un empleo. Analizando las características de esta variable revelan que en 1^{er} trimestre del 2016 el 23.54% corresponde a los

jóvenes que tienen educación superior o más, sin embargo, el porcentaje más significativo pertenece a los jóvenes que poseen educación secundaria con el 52.79%, dicha información se corrobora con el gráfico 7-3 en el cual se puede observar que el nivel de instrucción más alcanzado por los jóvenes es la secundaria seguido por la primaria además se puede evidenciar que son pocos los jóvenes que llegan a tener un educación superior o de tercer nivel.

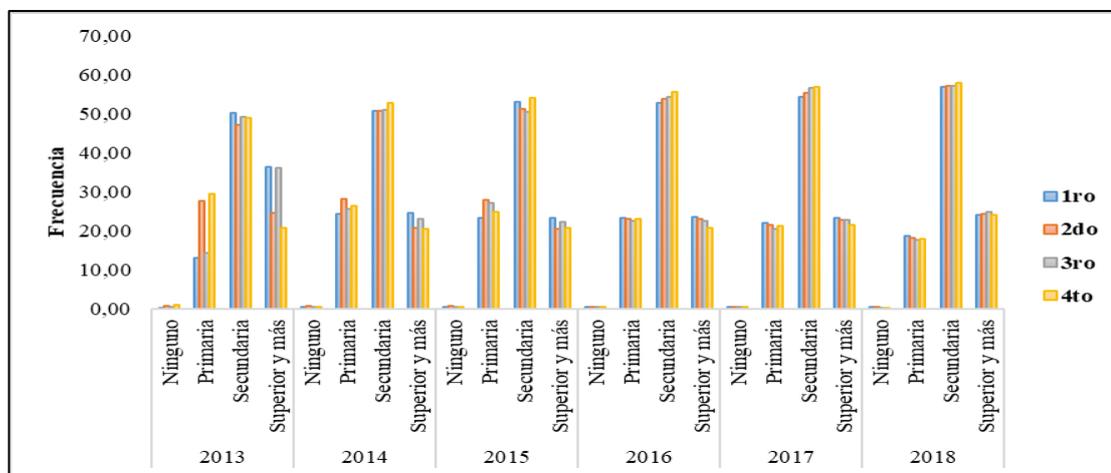


Gráfico 7 -3: Gráfico de frecuencias de la variable nivel de instrucción

Realizado por: Tierra, Tania. 2020

Variable: Ingreso laboral

Tabla 9-3: Frecuencias de la variable ingreso laboral

Año	Categoría	TRIMESTRE							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		f_i	%	f_i	%	f_i	%	f_i	%
2013	Mayor o igual a un SBU	920	35.95	1930	26.73	1057	40.28	2157	28.48
	Menor a un SBU	1639	64.05	5290	73.27	1567	59.72	5417	71.52
2014	Mayor o igual a un SBU	2022	35.47	4133	36.43	2453	40.64	4292	37.76
	Menor a un SBU	3679	64.53	7213	63.57	3583	59.36	7076	62.24
2015	Mayor o igual a un SBU	2310	39.29	4568	40.41	2554	42.90	4358	39.70
	Menor a un SBU	3570	60.71	6735	59.59	3400	57.10	6619	60.30
2016	Mayor o igual a un SBU	2263	38.94	2260	38.85	2388	39.05	4294	42.93
	Menor a un SBU	3548	61.06	3557	61.15	3728	60.95	5709	57.07
2017	Mayor o igual a un SBU	2244	42.78	2187	42.58	2308	43.96	4361	43.07
	Menor a un SBU	3001	57.22	2949	57.42	2942	56.04	5764	56.93
2018	Mayor o igual a un SBU	2461	45.93	2464	45.71	2537	46.53	2468	46.12
	Menor a un SBU	2897	54.07	2926	54.29	2915	53.47	2883	53.88

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

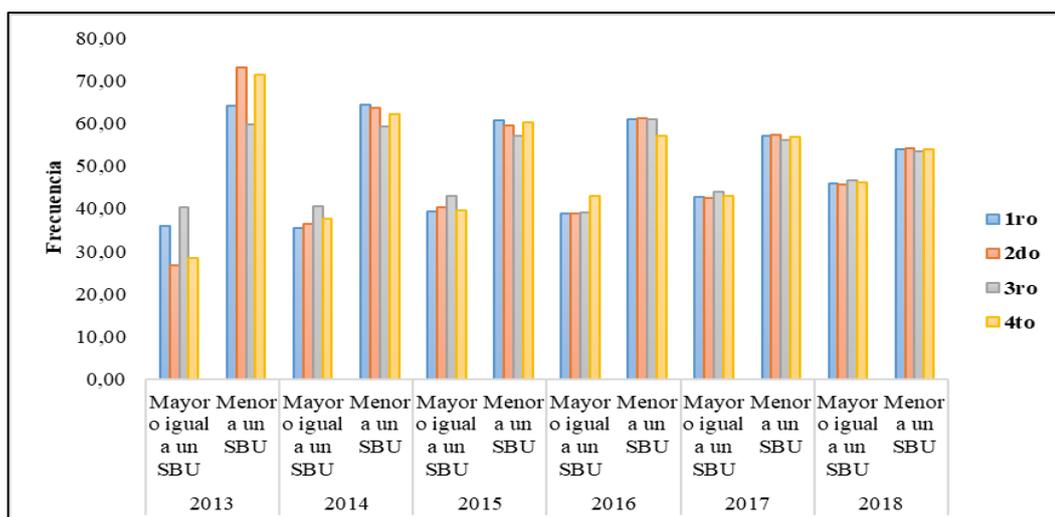


Gráfico 8 -3: Gráfico de frecuencias de la variable ingreso laboral

Realizado por: Tierra, Tania. 2020

Los datos de la encuesta muestran que en el 1^{er} trimestre del 2016 los jóvenes que recibieron un ingreso mayor o igual promedio del salario básico unificado dentro del periodo de estudio (\$356.50) fue de 38.94% y por consiguiente el resto percibe un ingreso menor el cual representa el 61.06%, esta diferencia se puede observar en el gráfico anterior, la mayoría de los jóvenes reciben un sueldo menor a un salario básico, esto se puede deber al horario de trabajo, años de experiencia, nivel de educación entre otros ya que por ser jóvenes y no contar con la suficiente experiencia no se les da la oportunidad de ser contratados u ocupar cargos superiores.

Variable: Rama de actividad

Tabla 10-3: Frecuencias de la variable rama de actividad

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	No especificado	99	3.35	356	3.93	190	6.48	749	8.16
	Primario	183	6.19	2652	29.24	144	4.91	2635	28.70
	Secundario	707	23.91	1752	19.32	629	21.45	1799	19.60
	Terciario	1968	66.55	4309	47.51	1969	67.16	3997	43.54
2014	No especificado	683	9.92	1188	8.77	582	8.10	1067	7.68
	Primario	1574	22.87	3525	26.03	1760	24.49	3690	26.55
	Secundario	1448	21.04	2739	20.22	1495	20.80	2803	20.17
	Terciario	3178	46.17	6092	44.98	3350	46.61	6336	45.60
2015	No especificado	610	8.23	1167	8.29	627	8.45	1296	9.33
	Primario	1911	25.78	4011	28.50	2001	26.96	3894	28.04
	Secundario	1478	19.94	2739	19.46	1422	19.16	2495	17.97
	Terciario	3415	46.06	6158	43.75	3372	45.43	6201	44.66
2016	No especificado	806	10.75	682	9.13	779	9.67	1325	9.38
	Primario	2094	27.93	2098	28.08	2230	27.69	4144	29.33
	Secundario	1244	16.59	1272	17.02	1381	17.15	2438	17.26
	Terciario	3353	44.72	3420	45.77	3664	45.49	6222	44.04
2017	No especificado	632	8.41	637	8.65	627	8.57	1234	8.85
	Primario	2209	29.41	2088	28.37	1975	27.00	3968	28.44
	Secundario	1293	17.21	1264	17.17	1222	16.71	2301	16.49

	Terciario	3378	44.97	3372	45.81	3490	47.72	6448	46.22
2018	No especificado	677	9.17	625	8.61	685	9.06	579	8.14
	Primario	1944	26.33	1841	25.38	1960	25.93	1753	24.66
	Secundario	1319	17.87	1323	18.24	1324	17.51	1293	18.19
	Terciario	3443	46.63	3466	47.77	3591	47.50	3485	49.02

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

La variable rama de actividad permite identificar a que sector pertenece el establecimiento donde los jóvenes laboran, lo cual indica que en el 1^{er} trimestre del 2016 el 44.72% de jóvenes laboraron en el sector terciario es decir prestaron sus servicios en empresas públicas o privadas ya sea en actividades relacionadas al comercio o cargos administrativos, a comparación de aquellos jóvenes que no especificaron exactamente el establecimiento donde laboraron siendo este el 10.75%, este segmento de jóvenes pueden haber realizados alguna actividad que les permitió obtener ingresos tales como chaucha o cachuelo, comprar y revender cosas, pasantías por pago, etc.; al observar el gráfico 9-3 se puede confirmar que existe un gran número de jóvenes que prestan sus servicios al sector terciario seguido del primario y secundario.

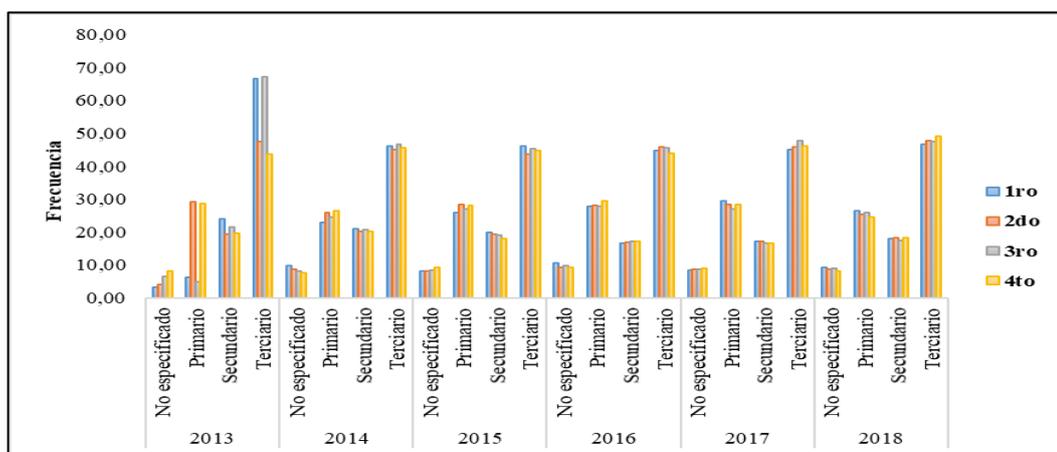


Gráfico 9 -3: Gráfico de frecuencias de la variable rama de actividad

Realizado por: Tierra, Tania. 2020

Variable: Grupo de ocupación

Tabla 11-3: Frecuencias de la variable grupo de ocupación

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	No especificado	99	3.35	356	3.93	117	3.99	749	8.16
	Ocupación manual	1949	65.91	7121	78.52	1934	65.96	7033	76.61
	Ocupación no manual	909	30.74	1592	17.55	881	30.05	1398	15.23
2014	No especificado	683	9.92	1188	8.77	582	8.10	1067	7.68
	Ocupación manual	4957	72.02	10155	74.98	5311	73.90	10612	76.37
	Ocupación no manual	1243	18.06	2201	16.25	1294	18.00	2217	15.95
2015	No especificado	610	8.23	1164	8.27	627	8.45	4133	29.76
	Ocupación manual	5468	73.75	10660	75.74	5492	74.00	7563	54.46
	Ocupación no manual	1336	18.02	2251	15.99	1303	17.56	2190	15.77
2016	No especificado	806	10.75	682	9.13	779	9.67	1325	9.38

	Ocupación manual	5435	72.50	5567	74.50	6009	74.61	10825	76.62
	Ocupación no manual	1256	16.75	1223	16.37	1266	15.72	1979	14.01
2017	No especificado	632	8.41	637	8.65	627	8.57	1234	8.85
	Ocupación manual	5666	75.43	5613	76.25	5579	76.28	10683	76.58
	Ocupación no manual	1214	16.16	1111	15.09	1108	15.15	2034	14.58
2018	No especificado	677	9.17	2018	27.82	685	9.06	579	8.14
	Ocupación manual	5522	74.79	4121	56.80	5667	74.96	5347	75.20
	Ocupación no manual	1184	16.04	1116	15.38	1208	15.98	1184	16.65

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

Esta variable permite definir el conjunto de funciones, obligaciones y tareas que desempeña habitualmente el joven en su puesto de trabajo, el cual indica que en el 1^{er} trimestre del 2016 el 72.50% de jóvenes se desempeñaron en ocupaciones manuales, es decir realizaron actividades de comercio, agentes de venta, operadores de maquinaria, etc., mientras que el 16.75% se desempeñaron en actividades no manuales como funcionarios, directores, jefes y demás actividades administrativas, dentro de empresas públicas o privadas.

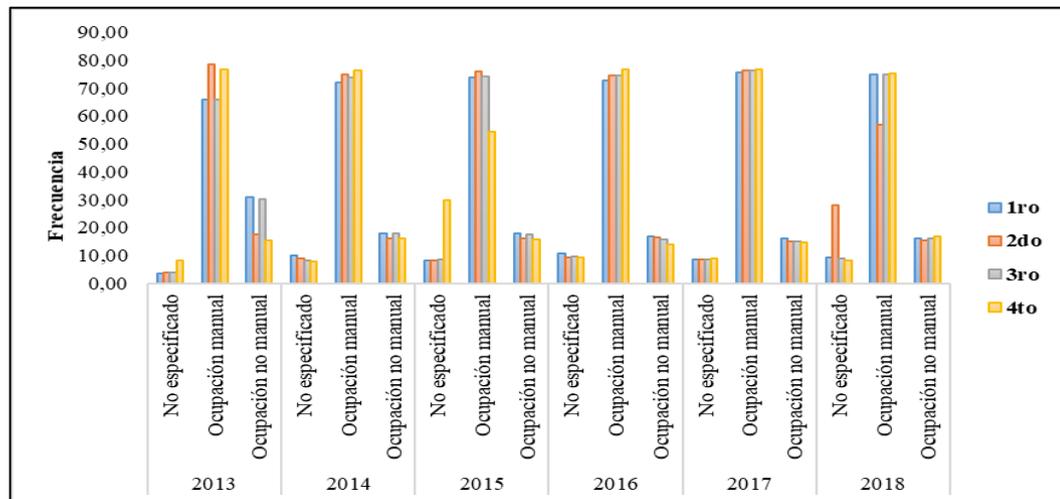


Gráfico 10 -3: Gráfico de frecuencias de la variable grupo de ocupación

Realizado por: Tierra, Tania. 2020

Se debe tener en cuenta que al no especificar la rama de actividad que desempeña en la empresa o institución donde labora, es imposible definir las funciones que desempeña la persona dentro del grupo de ocupación, el cual se puede observar en el gráfico 10-3 que son pocos los jóvenes que pertenecen a este segmento frente a los que se desenvuelven en ocupaciones manuales que es el grupo más representativo.

Variable: Recibe seguro social

Tabla 12-3: Frecuencias de la variable recibe seguro social

Año	Categoría	Trimestre							
		1 ^{ro}		2 ^{do}		3 ^{ro}		4 ^{to}	
		<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%	<i>f_i</i>	%
2013	No	1612	54.51	5598	61.73	1560	53.21	5879	64.04
	Si	1345	45.49	3471	38.27	1372	46.79	3301	35.96
2014	No	4242	61.63	8597	63.47	4403	61.26	8758	63.03
	Si	2641	38.37	4947	36.53	2784	38.74	5138	36.97
2015	No	4580	61.78	9060	64.37	4670	62.92	9108	65.59
	Si	2834	38.22	5015	35.63	2752	37.08	4778	34.41
2016	No	4932	65.79	4904	65.63	5420	67.30	9618	68.07
	Si	2565	34.21	2568	34.37	2634	32.70	4511	31.93
2017	No	5111	68.04	4975	67.59	4974	68.01	9517	68.22
	Si	2401	31.96	2386	32.41	2340	31.99	4434	31.78
2018	No	4869	65.95	4708	64.89	5009	66.26	4700	66.10
	Si	2514	34.05	2547	35.11	2551	33.74	2410	33.90

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

Finalmente, el resultado de esta variable nos permite conocer si el joven recibe un seguro sea este público o privado, lo cual indica que en el 1^{er} trimestre del 2016 el 65.79% de los jóvenes no recibieron ningún beneficio de salud por sus servicios prestados, no obstante, se aprecia que el 34.21% si lo recibieron, además en el gráfico 11-3 se puede observar que en su mayoría los jóvenes no reciben un seguro de salud.

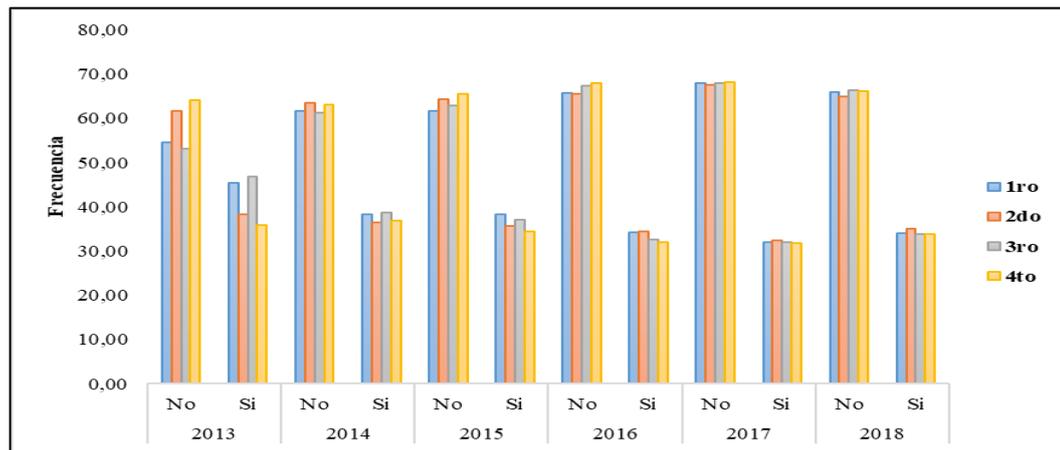


Gráfico 11 -3: Gráfico de frecuencias de la variable recibe seguro social

Realizado por: Tierra, Tania. 2020

3.3 Modelos de series de tiempo de la tasa de desempleo juvenil nacional

Como primer paso se realizó el cálculo de las tasas de desempleo juvenil provincial y nacional (ANEXO B) en el software PSPP versión 0.10.4 reemplazo libre de SPSS, una vez creadas las series trimestrales ($s=4$) del 2013 al 2018, ahora se procede al respectivo análisis de la serie y modelación aplicando la metodología Box Jenkins.

3.3.1 AED del desempleo juvenil nacional y por provincias

El desempleo es un fenómeno que afecta con mayor impacto a los jóvenes y de acuerdo al gráfico 12-3 de las series de tiempo para las 24 provincias del Ecuador en el período marzo 2013 a diciembre 2018, se puede apreciar el comportamiento y variabilidad de las series originales, en la cual se puede apreciar que las provincias de la costa son las más afectadas en cuanto a desempleo y pobreza y una de las provincias con mayores tasas de desempleo es Esmeraldas (EC_E), según el censo del INEC del 2010, dicha provincia muestra elevados porcentajes de desempleo, siendo la primera provincia con mayor tasa de desempleo entre el año 2013 – 2015 en cual presenta mayor rebrote a comparación de las otras series de cada provincia, seguida de Sucumbíos (EC_U). Esto se corrobora también con la publicación realizada por Bonilla, 2018 en la cual indica que en Esmeraldas la principal actividad económica es la agrícola, especialmente la siembra de palma africana y en Sucumbíos la actividad petrolera es la más importante, pero las empresas requieren de mano de obra que no siempre hay en la zona y los jóvenes se reusan a ir a trabajar en los cultivos o de obreros, por lo que, emigran a otras provincias en busca de mejores puestos de trabajo, aumentando así la brecha de desempleo en su provincia natal (Bonilla, 2018). Es así como se concibe que las provincias de la costa

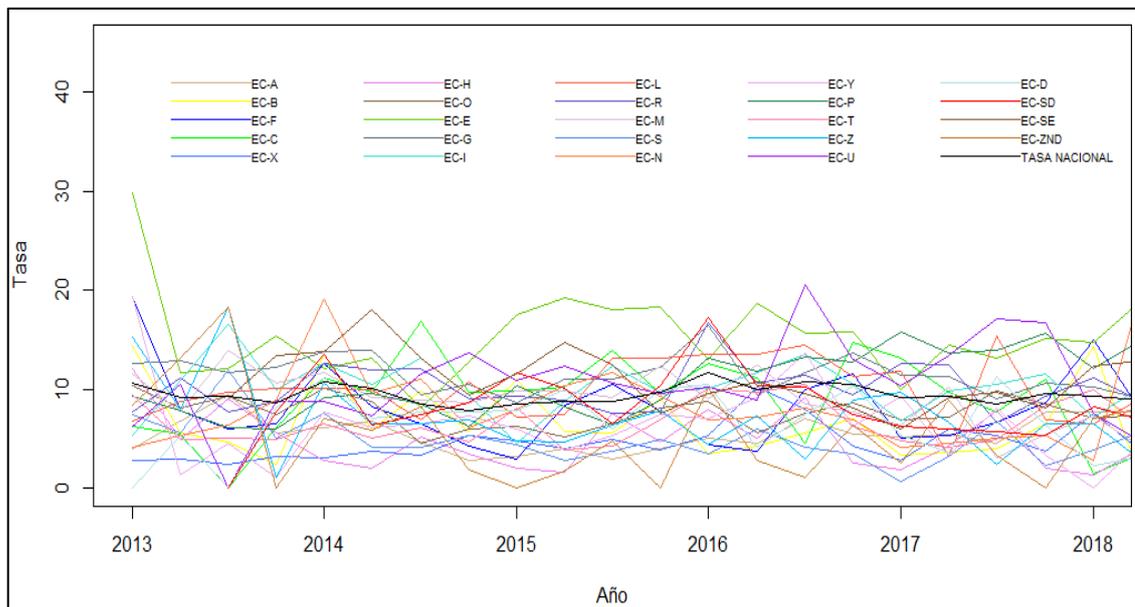


Gráfico 12 -3: Series de tiempo de la tasa de desempleo juvenil provincial y nacional 2013-2018
Realizado por: Tierra, Tania. 2020

La tasa nacional de desempleo juvenil, se puede apreciar que se mantiene en un rango de aproximadamente 8% y 12% a lo largo del período en estudio, además de no presentar variabilidad significativa en comparación con las tasas desagregadas por provincias, siendo así Azuay (EC_A) la provincia que presenta menor tasa de desempleo juvenil en comparación a las demás provincias.

Análisis de estacionariedad para las series de tiempo del desempleo juvenil

Se aplicó la metodología Box-Jenkins para construir el modelo *ARIMA* apropiado, que consta de las fases de identificación, estimación, verificación del modelo y pronóstico.

Fase de identificación

Para verificar si las series bajo estudio cumplen la condición de estacionariedad se aplica el test de Dickey-Fuller que estudia la siguiente hipótesis:

$$H_0: \rho = 1$$

$$H_1: \rho < 1$$

Tabla 13-3: Test Dickey-Fuller para las 24 provincias

Provincia	Estadístico DF_t	p-valor	Nivel De Significancia	Región de rechazo	Decisión
Azuay	-2.63	0.33	0.05	No se rechaza H_0	Serie no estacionaria
Bolívar	-6.36	0.01	0.05	Se rechaza H_0	Serie estacionaria
Cañar	-5.13	0.01	0.05	Se rechaza H_0	Serie estacionaria
Carchi	-3.39	0.08	0.05	No se rechaza H_0	Serie no estacionaria
Cotopaxi	-3.32	0.09	0.05	No se rechaza H_0	Serie no estacionaria
Chimborazo	-5.18	0.01	0.05	Se rechaza H_0	Serie estacionaria
El Oro	-3.62	0.04	0.05	No se rechaza H_0	Serie no estacionaria
Esmeraldas	-6.26	0.01	0.05	Se rechaza H_0	Serie estacionaria
Guayas	-3.49	0.04	0.05	No se rechaza H_0	Serie estacionaria
Imbabura	-4.93	0.01	0.05	Se rechaza H_0	Serie estacionaria
Loja	-4.13	0.02	0.05	Se rechaza H_0	Serie estacionaria
Los Ríos	-4.39	0.01	0.05	Se rechaza H_0	Serie estacionaria
Manabí	-5.09	0.01	0.05	Se rechaza H_0	Serie estacionaria
Morona Santiago	-4.84	0.01	0.05	Se rechaza H_0	Serie estacionaria
Napo	-4.78	0.01	0.05	Se rechaza H_0	Serie estacionaria
Pastaza	-6.21	0.01	0.05	Se rechaza H_0	Serie estacionaria
Pichincha	-3.99	0.02	0.05	Se rechaza H_0	Serie estacionaria
Tungurahua	-3.37	0.04	0.05	Se rechaza H_0	Serie estacionaria
Zamora Chinchipe	-5.07	0.01	0.05	Se rechaza H_0	Serie estacionaria
Sucumbíos	-4.16	0.02	0.05	Se rechaza H_0	Serie estacionaria
Orellana	-4.46	0.01	0.05	Se rechaza H_0	Serie estacionaria
Santo Domingo	-4.35	0.01	0.05	Se rechaza H_0	Serie estacionaria
Santa Elena	-3.71	0.04	0.05	Se rechaza H_0	Serie estacionaria
Zonas no delimitadas	-4.58	0.01	0.05	Se rechaza H_0	Serie estacionaria
Tasa Nacional	-3.02	0.18	0.05	No se rechaza H_0	Serie no estacionaria

Realizado por: Tierra, Tania. 2020

Al aplicar el test de estacionariedad analíticamente, las series de tiempo del desempleo juvenil, a un nivel de confianza del 95% y observando los p-valores, se obtiene como resultado que las provincias de Azuay, Carchi, Cotopaxi, El Oro y la tasa nacional no son series estacionarias, para corroborar los resultados obtenidos de forma gráfica se procede a realizar las funciones de autocorrelación simple y parcial de dichas provincias.

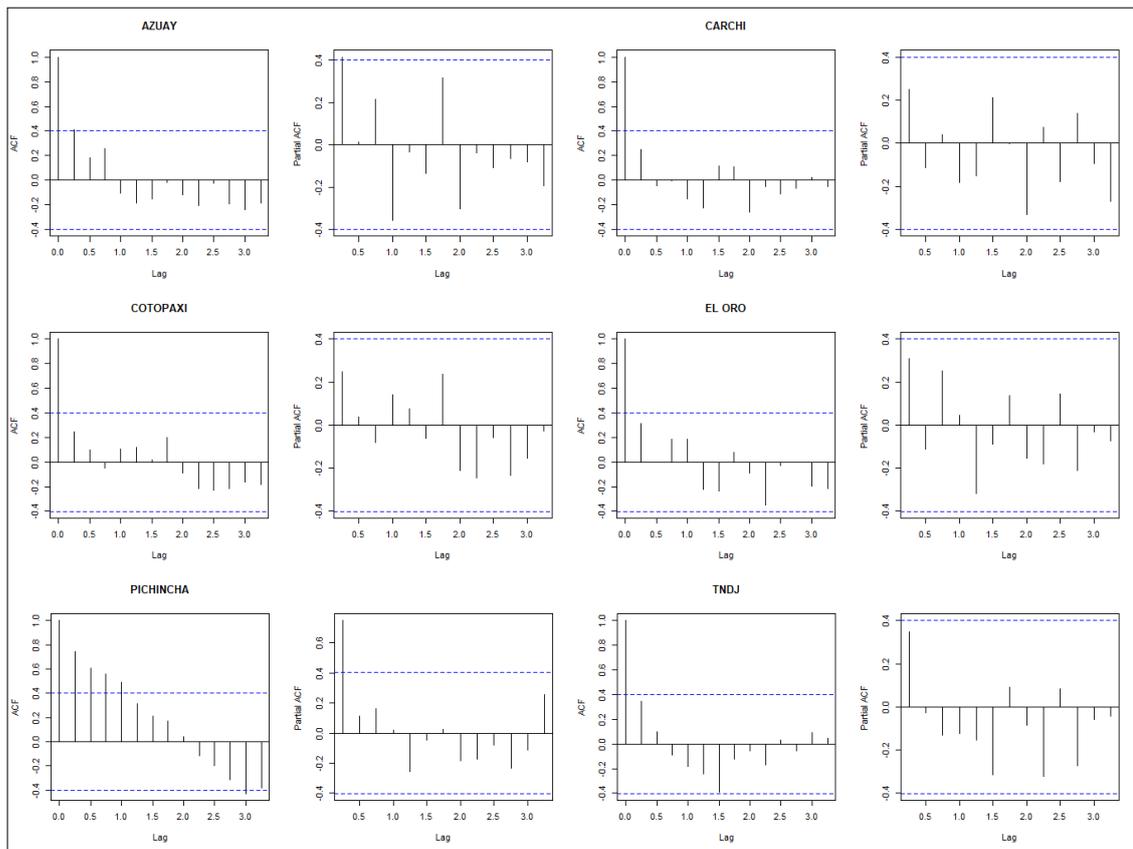


Gráfico 13 -3: Función de autocorrelación simple y parcial para la tasa de desempleo juvenil provincial y nacional del Ecuador 2013-2018

Realizado por: Tierra, Tania. 2020

En el gráfico anterior se observa que las funciones de autocorrelación simple y parcial para las series de desempleo juvenil, no siguen un patrón de estacionario, para dar solución a este problema se debe realizar una diferenciación de primer orden.

Al realizar las gráficas de FAC y FAP para las series diferenciadas gráfico 14-3 se observa que el primer coeficiente es significativamente diferente de cero y el resto no son significativos, identificando así el patrón estacionario, lo cual se argumenta analíticamente con la tabla 14-3, con un p-valor inferior a 0.05 que las series de las tasas de desempleo diferenciadas de las provincias analizadas son estacionarias, corrigiendo así el problema de estacionariedad realizando una diferenciación de primer orden para las series de las provincias de Azuay, Carchi, Cotopaxi, El Oro, Pichincha y la tasa de desempleo nacional.

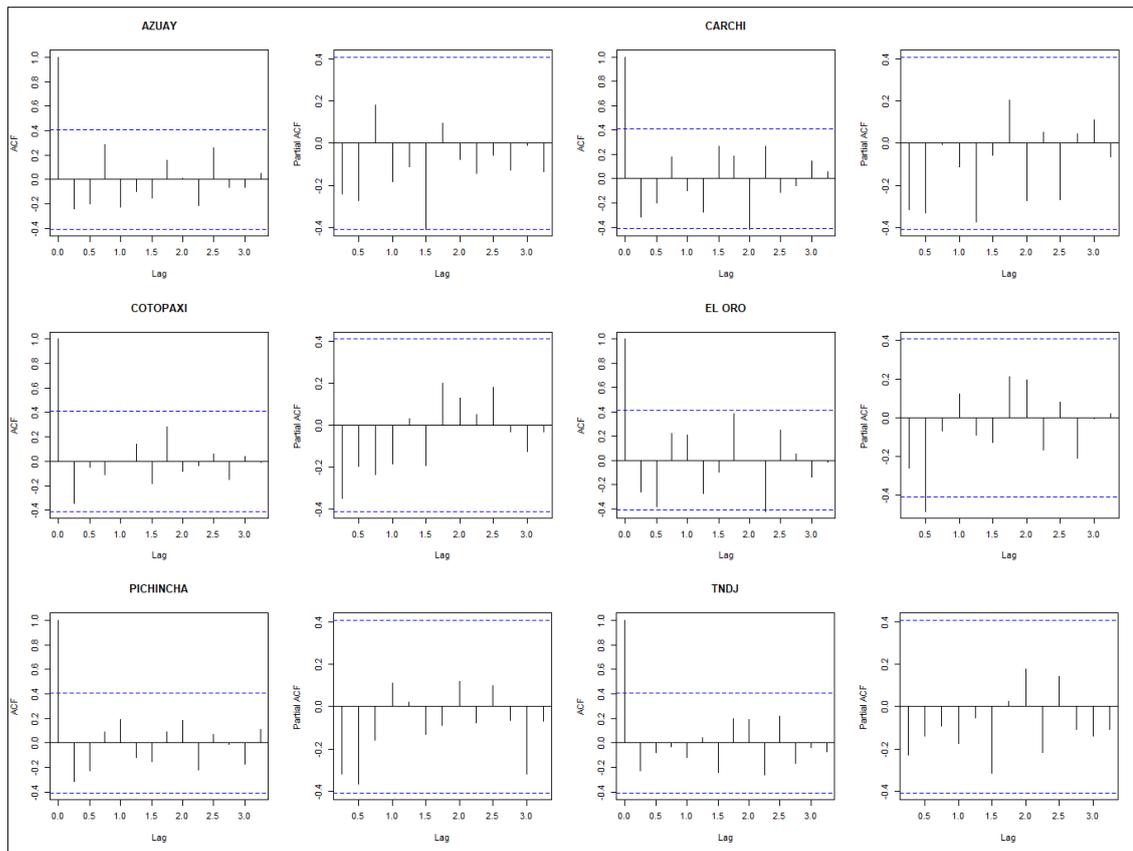


Gráfico 14 -3: Función de autocorrelación simple y parcial diferenciada para la tasa de desempleo juvenil provincial y nacional del Ecuador 2013-2018

Realizado por: Tierra, Tania. 2020

Tabla 14-3: Test Dickey-Fuller para las series diferenciadas

Provincia	Estadístico DF_r	p-valor	Nivel De Significancia	Región de rechazo	Decisión
Azuay	-6.38	0.01	0.05	Se rechaza H_0	Serie estacionaria
Carchi	-6.03	0.01	0.05	Se rechaza H_0	Serie estacionaria
Cotopaxi	-6.28	0.01	0.05	Se rechaza H_0	Serie estacionaria
El Oro	-5.95	0.01	0.05	Se rechaza H_0	Serie estacionaria
Pichincha	-6.38	0.01	0.05	Se rechaza H_0	Serie estacionaria
Tasa Nacional	-5.17	0.01	0.05	Se rechaza H_0	Serie estacionaria

Realizado por: Tierra, Tania. 2020

3.3.2 Modelo ARIMA del desempleo juvenil en el periodo 2013 - 2018

En este apartado se analiza las fases de estimación y verificación del modelo en la cual se utiliza el software R Studio para determinar los valores de los parámetros $(p, q)(P, Q)$ del modelo ARIMA mediante la función *auto.arima*, para la verificación del modelo se realiza el análisis de los residuos mediante la prueba de Box-Ljung para verificar la independencia, el test de Shapiro-Wilk para la normalidad y para la homocedasticidad el test de Dickey-Fuller a un nivel de significancia del 5%, además para la selección del mejor modelo ARIMA se lo realizó mediante el criterio de selección AIC, los resultados de estos análisis se presentan en la siguiente tabla.

Tabla 15-3: Modelos *ARIMA* obtenidos para las 24 provincias del Ecuador

Provincia	SARIMA						SUPUESTOS			AIC
	<i>p</i>	<i>d</i>	<i>q</i>	<i>P</i>	<i>D</i>	<i>Q</i>	Normalidad	Independencia	Homocedasticidad	
Azuay	1	0	0	2	0	0	Normales	Independientes	Estacionarios	96.14
Bolívar	0	0	0	0	0	1	Normales	Independientes	Estacionarios	130.79
Cañar	0	0	0	0	0	1	No son normales	Independientes	Estacionarios	135.84
Carchi	0	0	0	0	0	1	Normales	Independientes	Estacionarios	138.78
Cotopaxi	0	0	0				No son normales	Independientes	Estacionarios	103.47
Chimborazo	0	0	0				No son normales	Independientes	Estacionarios	121.52
El Oro	3	0	0	2	0	0	Normales	Independientes	Estacionarios	100.8
Esmeraldas	0	0	0				No son normales	Independientes	Estacionarios	142.98
Guayas	3	1	0	2	0	0	Normales	Independientes	Estacionarios	106.56
Imbabura	0	0	0	2	1	0	Normales	Independientes	Estacionarios	94.08
Loja	0	0	0				Normales	Independientes	Estacionarios	125.27
Los Ríos	0	0	0				Normales	Independientes	Estacionarios	113.3
Manabí	1	1	0				Normales	Independientes	Estacionarios	107.79
Morona Santiago	0	1	1				Normales	Independientes	Estacionarios	108.52
Napo	0	0	0				No son normales	Independientes	Estacionarios	104.48
Pastaza	0	0	0				No son normales	Independientes	Estacionarios	138.13
Pichincha	3	1	0	2	0	0	Normales	Independientes	Estacionarios	108.34
Tungurahua	3	0	0	2	0	0	Normales	Independientes	Estacionarios	106.44
Zamora Chinchipe	0	0	0				No son normales	Independientes	Estacionarios	141.81
Sucumbíos	0	0	0	1	0	0	Normales	Independientes	Estacionarios	140.83
Orellana	0	0	0	1	0	1	Normales	Independientes	Estacionarios	127.49
Santo Domingo	0	0	0				Normales	Independientes	Estacionarios	120.46
Santa Elena	2	1	0	1	0	1	Normales	Independientes	Estacionarios	109.16
Zonas no delimitadas	0	0	0	1	0	0	No son normales	Independientes	Estacionarios	145.63
Tasa Nacional	1	0	0	1	0	1	No son normales	Independientes	Estacionarios	70.67

Realizado por: Tierra, Tania. 2020

En la tabla anterior se puede ver que para las provincias de Azuay, Bolívar, Cañar, Carchi, El Oro, Guayas, Imbabura, Pichincha, Tungurahua, Sucumbíos, Orellana, Santa Elena, las zonas no delimitadas y la tasa nacional de desempleo juvenil se ajustaron un modelo $ARIMA(p, d, q)(P, D, Q)_4$, mientras que para el resto de las provincias se ajustó un modelo $ARIMA(p, d, q)$; los residuos de cada modelo cumplen con los supuestos de normalidad, independencia y homocedasticidad, a diferencia de los residuos de las series pertenecientes a las provincias de Cañar, Cotopaxi, Chimborazo, Esmeraldas, Napo, Pastaza, Zamora Chinchipe y las Zonas no delimitadas que no se distribuyen normalmente, por lo que para realizar la fase de pronósticos se debe aplicar el método Bootstrap para estas series dado que no se distribuyen normalmente.

Para la elección del mejor modelo se lo hace mediante el principio de parsimonia y analíticamente se toma en cuenta los resultados obtenidos mediante el cálculo del AIC, con los valores más bajos obteniendo así los siguientes modelos donde Y_t es la tasa de desempleo:

Modelo SARIMA para la provincia de Azuay

Series: ((az))				
ARIMA(1,0,0)(2,0,0)[4] with non-zero mean				
Coefficients:				
	ar1	sar1	sar2	mean
	0.5384	-0.3053	-0.1104	5.4219
s.e.	0.1781	0.2640	0.2801	0.4651

$$Y_t = C + 0.54Y_{t-1} - 0.30\omega_1\varepsilon_{t-1} - 0.11\omega_2\varepsilon_{t-2}$$

$$C = 5.42 \times 1 - (0.54 - 0.30 - 0.11) = 4.71$$

$$Y_t = 4.71 + 0.54Y_{t-1} - 0.30\omega_1\varepsilon_{t-4} - 0.11\omega_2\varepsilon_{t-4}$$

$$AIC = 96.14$$

Modelo ARIMA para la tasa nacional de desempleo juvenil

Series: ((tnac))				
ARIMA(1,0,0)(1,0,1)[4] with non-zero mean				
Coefficients:				
	ar1	sar1	smal	mean
	0.3672	-0.1272	-0.0095	9.4045
s.e.	0.2110	1.1358	1.1717	0.2428

$$Y_t = 7.23 + 0.37Y_{t-1} - 0.13\omega_1\varepsilon_{t-1} - 0.009\varepsilon_{t-1}$$

$$AIC = 70.67$$

Para la fase de pronóstico también se tomará en cuenta las demás provincias pertenecientes a las ciudades más grandes del Ecuador como es El Oro, Tungurahua, Guayas y Pichincha.

Modelo ARIMA para la provincia de El Oro

Series: ((or))						
ARIMA(3,0,0)(2,0,0)[4] with non-zero mean						
Coefficients:						
	ar1	ar2	ar3	sar1	sar2	mean
	0.3974	-0.2215	0.3586	-0.0042	0.0923	8.0722
s.e.	0.2350	0.2797	0.3026	0.3351	0.2971	0.7227

$$Y_t = 3.06 + 0.40Y_{t-1} - 0.22Y_{t-2} + 0.36Y_{t-3} - 0.004\omega_1\varepsilon_{t-1} + 0.09\omega_2\varepsilon_{t-2}$$

$$AIC = 100$$

Modelo ARIMA para la provincia de Tungurahua

```

ARIMA(3,0,0)(2,0,0)[4] with non-zero mean
Coefficients:
      ar1      ar2      ar3      sar1      sar2      mean
      0.4839 -0.0063 -0.3540  0.0170  0.4258  6.3791
s.e.  0.2401  0.2452  0.2022  0.2019  0.2481  0.5554
    
```

$$Y_t = 2.78 + 0.48Y_{t-1} - 0.006Y_{t-2} - 0.35Y_{t-3} + 0.02\omega_1\varepsilon_{t-1} + 0.42\omega_2\varepsilon_{t-2}$$

$$AIC = 106.44$$

Modelo ARIMA para la provincia de Guayas

```

Series: ((gu))
ARIMA(3,1,0)(2,0,0)[4] with drift
Coefficients:
      ar1      ar2      ar3      sar1      sar2      drift
      -0.6119 -0.7563 -0.4286 -0.6212 -0.3613 -0.1457
s.e.  0.2004  0.2243  0.2364  0.2751  0.2782  0.0916
    
```

$$Y_t = -0.53 - 0.62Y_{t-1} - 0.75Y_{t-2} - 0.43Y_{t-3} - 0.62\omega_1\varepsilon_{t-1} - 0.36\omega_2\varepsilon_{t-2}$$

$$AIC = 106.56$$

Modelo SARIMA para la provincia de Pichincha

```

Series: ((pi))
ARIMA(3,1,0)(2,0,0)[4] with drift
Coefficients:
      ar1      ar2      ar3      sar1      sar2      drift
      -0.9059 -0.9372 -0.7610 -0.5957 -0.3828  0.4155
s.e.  0.2570  0.2279  0.2833  0.4335  0.4245  0.0927
    
```

$$Y_t = 1.87 - 0.90Y_{t-1} - 0.94Y_{t-2} - 0.76Y_{t-3} - 0.60\omega_1\varepsilon_{t-1} - 0.38\omega_2\varepsilon_{t-2}$$

$$AIC = 108.34$$

3.3.3 Pronósticos

Para esta última fase de pronósticos se realizará en dos partes la primera se aplicará la función *forecast* para los modelos que cumplen normalidad, pertenecientes a las provincias de Azuay, El Oro, Tungurahua, Guayas y Pichincha, y la función *bld.mbb.bootstrap* para el modelo de la tasa nacional de desempleo que no lo cumplen el supuesto de normalidad.

Tabla 16-3: Pronósticos de las provincias más grandes del Ecuador

Año	Trimestre	Azuay	El Oro	Tungurahua	Guayas	Pichincha
2019	1 ^{ro}	5.20 (2.12; 8.28)	7.82 (4.52; 11.12)	5.20 (1.64; 8.76)	9.86 (5.68; 14.04)	16.52 (12.48; 20.56)
	2 ^{do}	4.96 (1.46; 8.46)	8.47 (4.92; 12.02)	5.61 (1.65; 9.57)	8.62 (4.14; 13.10)	16.81 (12.76; 20.87)
	3 ^{ro}	4.85 (1.24; 8.46)	8.91 (5.36; 12.47)	6.68 (2.64; 10.72)	9.07 (4.58; 13.55)	17.24 (13.19; 21.30)
	4 ^{to}	5.19 (1.55; 8.84)	8.22 (4.58; 11.87)	7.84 (3.71; 11.98)	9.44 (4.81; 14.06)	17.32 (13.20; 21.43)
2020	1 ^{ro}	5.27 (1.56; 8.98)	8.04 (4.31; 11.78)	8.10 (3.85; 12.35)	8.55 (3.92; 13.18)	18.58 (14.39; 22.78)
	2 ^{do}	5.33 (1.60; 9.05)	8.26 (4.52; 11.99)	6.12 (1.80; 10.43)	8.50 (3.83; 13.17)	18.49 (14.29; 22.70)
	3 ^{ro}	5.32 (1.59; 9.05)	8.40 (4.66; 12.14)	6.21 (1.89; 10.53)	7.97 (3.27; 12.68)	18.68 (14.47; 22.88)
	4 ^{to}	5.48 (1.75; 9.21)	8.21 (4.46; 11.97)	5.19 (0.86; 9.52)	8.16 (3.40; 12.92)	19.36 (15.10; 23.62)

Realizado por: Tierra, Tania. 2020

Con los mejores modelos estimados se ha calculado las tasas de desempleo juvenil para cada trimestre del año 2019 y 2020, para la provincia de Azuay para el 4^{to} trimestre del 2019 la tasa de desempleo estimada es de 5.19% y oscila entre [1.56%, 8.84%] con una confianza del 95%. Según los datos del INEC 2019 la tasa de desempleo juvenil para el mismo trimestre fue de 6.13% el mismo que se encuentran dentro de los intervalos de confianza y además no se aleja de la realidad, en el gráfico 15-3 se puede observar la evolución de las tasas de desempleo juvenil pronosticadas por ejemplo para la provincia de Pichincha se observa que la tasa de desempleo tendrá una tendencia creciente terminando así el 2020 con una tasa de aproximadamente el 19.36%.

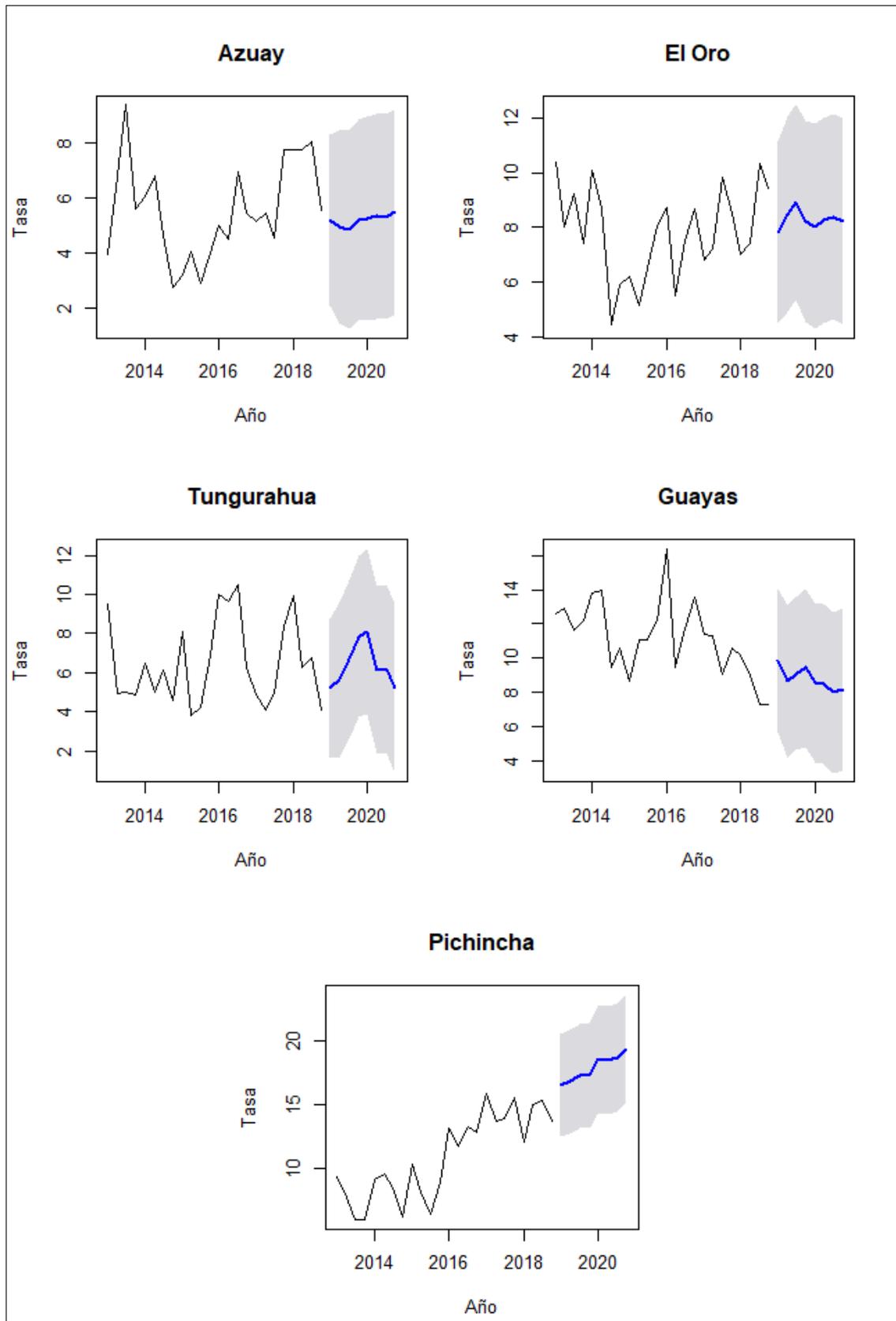


Gráfico 15 -3: Pronósticos de la tasa de desempleo juvenil provincial 2019-2020
 Realizado por: Tierra, Tania. 2020

Para estimar la tasa nacional de desempleo se lo realiza mediante el método Bootstrap dado que los residuos del modelo no cumplen con el supuesto de normalidad para lo cual se obtiene los siguientes pronósticos:

Tabla 17-3: Pronósticos de la tasa nacional de desempleo juvenil

Año	Trimestre	Tasa de desempleo nacional
2019	1 ^{ro}	8.67 (7.55; 10.26)
	2 ^{do}	8.42 (7.54; 9.54)
	3 ^{ro}	8.41 (7.40; 9.16)
	4 ^{to}	8.36 (7.41; 8.96)
2020	1 ^{ro}	8.59 (7.56; 10.12)
	2 ^{do}	8.46 (7.56; 9.56)
	3 ^{ro}	8.41 (7.42; 9.18)
	4 ^{to}	8.37 (7.41; 8.98)

Realizado por: Tierra, Tania. 2020

Como mejor modelo que explique la evolución del desempleo juvenil en el Ecuador se estimó un modelo $SARIMA(1,0,0)(1,0,1)_4$ cuyos pronósticos calculados a un corto plazo tienen un comportamiento constante (gráfico 16-3), es decir se estima que la tasa nacional de desempleo juvenil en Ecuador no siga creciendo ni tenga cambios bruscos en su comportamiento, sin embargo, la tasa nacional de desempleo juvenil seguirá manteniéndose en porcentajes altos, mismo que para el 4^{to} trimestre del 2019 aproximadamente será de 8.36%. Según datos del INEC 2019 para el mismo período la tasa de desempleo juvenil fue de 8.45% el mismo que se encuentra dentro de los intervalos de variación del modelo estimado, lo cual muestra que las predicciones dadas por el modelo resultan ser bastante exitosas, esto se constata en la diferencia entre la tasa estimada y la tasa real para el período del 4^{to} trimestre del 2019.

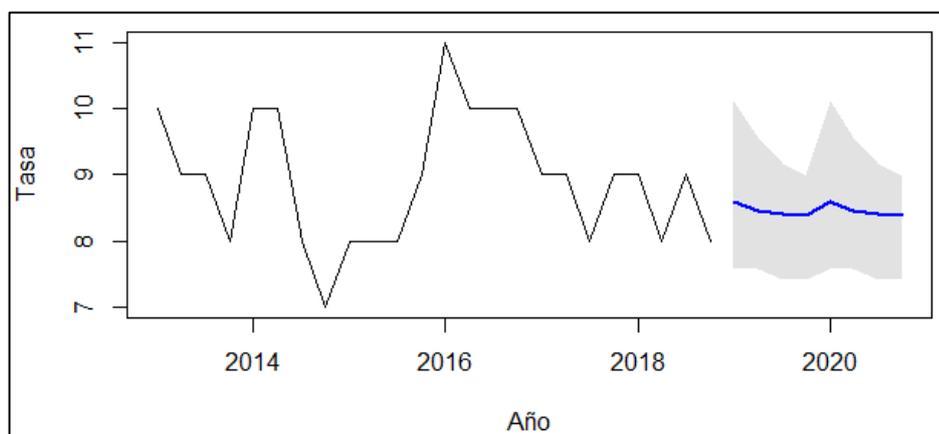


Gráfico 16 -3: Pronósticos de la tasa de desempleo juvenil nacional 2019-2020
Realizado por: Tierra, Tania . 2020

3.4 Análisis de regresión logit

El objetivo del siguiente modelo logit es estimar la probabilidad de que un joven de 18 a 29 años ecuatoriano se encuentre desempleado o no y los factores que determinan este fenómeno.

3.4.1 Modelo para el desempleo juvenil

Para este análisis se consideró como:

Variable dependiente o explicativa:

$$Y = \text{desempleo} \text{ (1: Desempleado, 0: No desempleado)}$$

Variables independientes:

Tabla 18-3: Codificación de las variables independientes

Variable	Codificación: 0	Codificación: 1	Codificación: 3	Codificación: 4
X ₁ : Sexo	Mujer	Hombre		
X ₂ : Edad	Esta variable es cuantitativa y puede recibir valores de 18 a 29 años			
X ₃ : Parentesco	Otros parientes	Jefe		
X ₄ : Estado Civil	Sin pareja estable	Con pareja		
X ₅ : Asiste a clases	No	Si		
X ₆ : Nivel de instrucción	Ninguna	Primaria	Secundaria	Superior y más
X ₇ : Ingreso laboral	Ingreso menor al SBU (356.50)	Ingreso mayor al SBU (356.50)		
X ₈ : Rama de la actividad	No especifica	Sector primario	Sector secundario	Sector terciario
X ₉ : Grupo de ocupación	No especifica	Ocupación no manual	Ocupación manual	
X ₁₀ : Recibe Seguro social	No	Si		

Realizado por: Tierra, Tania. 2020

Para la estimación del modelo logit se realizó un modelo de probabilidad para cada año utilizando el entorno de computación R Studio con la función *glm* y se obtiene los siguientes resultados:

Modelo 1: Año 2013

$$\Pr(Y = \text{Desempleo}) = \frac{1}{1 + \exp^{-(-0.69 - 0.56X_1 + 0.02X_2 - 0.28X_3 - 0.51X_4 - 0.58X_5 + 1.21X_6 - 2.12X_7 - 0.65X_8 - 2.78X_9 - 0.71X_{10})}}$$

Odd ratio= 2.02

Modelo 2: Año 2014

$$\Pr(Y = \text{Desempleo}) = \frac{1}{1 + \exp^{-(23.49 - 0.38X_1 + 0.005X_2 - 0.07X_3 - 0.21X_4 - 0.14X_5 + 0.77X_6 + 0.33X_7 - 0.65X_8 - 49.55X_9 + 0.19X_{10})}}$$

Odd ratio= 9.50E-14

Modelo 3: Año 2015

$$\Pr(Y = Desempleo) = \frac{1}{1 + \exp^{-(22.65 - 0.36X_1 + 0.03X_2 - 0.09X_3 - 0.26X_4 - 0.17X_5 + 0.80X_6 + 0.42X_7 - 0.86X_8 - 48.85X_9 + 0.15X_{10})}}$$

Odd ratio= 1.14E-13

Modelo 4: Año 2016

$$\Pr(Y = Desempleo) = \frac{1}{1 + \exp^{-(22.92 - 0.32X_1 + 0.03X_2 - 0.06X_3 - 0.262 - 0.17X_5 + 0.78X_6 + 0.39X_7 - 0.58X_8 - 49.91X_9 + 0.18X_{10})}}$$

Odd ratio= 3.26E-13

Modelo 5: Año 2017

$$\Pr(Y = Desempleo) = \frac{1}{1 + \exp^{-(22.84 - 0.33X_1 + 0.04X_2 - 0.05X_3 - 0.28 - 0.11X_5 + 0.71X_6 + 0.37X_7 - 0.58X_8 - 49.86X_9 + 0.23X_{10})}}$$

Odd ratio= 2.19E-13

Modelo 6: Año 2018

$$\Pr(Y = Desempleo) = \frac{1}{1 + \exp^{-(22.32 - 0.39X_1 + 0.03X_2 - 0.03X_3 - 0.26 - 0.18X_5 + 0.821X_6 + 0.39X_7 - 0.95X_8 - 48.57X_9 + 0.07X_{10})}}$$

Odd ratio= 1.11E-13

Se ha estimado 6 modelos de los cuales el modelo más significativo es el primero dado que explica que un joven ecuatoriano tiene 2.02 veces más riesgo de encontrarse desempleado, a diferencia de los demás modelos que explican que existe un riesgo casi nulo en encontrarse desempleado, por lo cual se toma el primer modelo como el más óptimo para explicar la influencia de los factores sociodemográficos y sociolaborales en el desempleo juvenil, el cual indica que los ingresos, el nivel de instrucción y la edad son factores sociodemográficos más significativos en el modelo.

Aplicación del modelo de regresión logit

Para validar el modelo seleccionado (Modelo 1) en el contexto real se ha tomado los datos de dos jóvenes que se encuentran en dos escenarios diferentes en donde han señalado las siguientes características:

Tabla 19-3: Resultados de la aplicación del modelo de regresión logit

Persona	Variable	Codificación	P(Y=1)
Persona 1	X ₁ : Sexo	0 (Mujer)	94.7%
	X ₂ : Edad	26 (Edad)	
	X ₃ :Parentesco	0 (Otros parientes)	
	X ₄ : Estado Civil	0 (Sin pareja estable)	
	X ₅ : Asiste a clases	1 (Si)	
	X ₆ : Nivel de instrucción	3 (Superior y más)	
	X ₇ : Ingreso laboral	0 (Menor a un SBU)	
	X ₈ :Rama de la actividad	0 (No especifica)	
	X ₉ :Grupo de ocupación	0 (No especifica)	
	X ₁₀ :Recibe Seguro social	0 (No)	
Persona 2	X ₁ : Sexo	1 (Hombre)	0.8%
	X ₂ : Edad	28 (Edad)	
	X ₃ :Parentesco	1 (Jefe de hogar)	
	X ₄ : Estado Civil	1 (Con pareja)	
	X ₅ : Asiste a clases	0 (No)	
	X ₆ : Nivel de instrucción	3 (Superior y más)	
	X ₇ : Ingreso laboral	1 (Ingreso mayor a un SBU)	
	X ₈ :Rama de la actividad	2 (Sector secundario)	
	X ₉ :Grupo de ocupación	1 (Ocupación no manual)	
	X ₁₀ :Recibe Seguro social	1 (Si)	

Realizado por: Tierra, Tania. 2020

El modelo determina una probabilidad del 94.7% que un joven se encuentre desempleado en base a que sea mujer, tenga 26 años, no sea jefe de hogar, este sin pareja, asista a clases, tenga un nivel académico superior, sus ingresos sean menores al SBU de \$357.50 y que además no especifique la rama y grupo de ocupación laboral y no reciba un seguro social.

A diferencia que existe la probabilidad del 0.8% de que un joven se encuentre desempleado, si es hombre, se encuentra en la edad de 28 años y es jefe de hogar, además tiene pareja estable, no asiste a clases, posee un nivel académico superior, recibe un ingreso mayor al SBU, labora dentro del económico sector secundario, la actividad que realiza es no manual y recibe un seguro social.

Los valores estimados arrojados por el modelo de regresión logit discrimina bastante bien a los jóvenes que se encuentran desempleados en base a la codificación de las variables influyentes.

CONCLUSIONES

Para cumplir con el primer objetivo de la investigación se realizó un análisis exploratorio de datos para describir la evolución de la tasa de desempleo juvenil en Ecuador tabla 1-1, el Anexo B se presenta las tasas de desempleo para las 24 provincias. La tasa de desempleo juvenil nacional promedio durante el período de análisis es de 8.7% la cual representa al tercer trimestre del 2015, mientras que a nivel de provincia el más significativo es de 29.80% perteneciente a la provincia de Esmeraldas y corresponde al primer trimestre del 2013, además se evidenció que a lo largo del período en estudio la tasa de desempleo ha tenido un comportamiento estable sin embargo, los valores han sido notablemente altos, mismos que dan a notar que el segmento de las personas de entre 18 y 29 años son los más sensibles a las crisis económicas y a los cambios de políticas administrativas.

Se ha aplicado la metodología Box-Jenkins para determinar un modelo $SARIMA(0,1,1)(1,0,1)_4$ óptimo que mejor estima la tasa nacional de desempleo juvenil, mientras que para las provincias pertenecientes a las ciudades más grandes del Ecuador se ajustaron los modelos $SARIMA(1,0,0)(2,0,0)_4$ para Azuay, para El Oro y Tungurahua $SARIMA(3,0,0)(2,0,0)_4$, para Pichincha y Guayas se estimó $SARIMA(3,1,0)(2,0,0)_4$. También se aplicó la técnica Bootstrap aplicada a la serie de la tasa de desempleo juvenil nacional por no satisfacer el supuesto de normalidad, además, las series originales se ajustaron correctamente a dicho modelo, sin presentar cambios bruscos en su comportamiento, aunque el valor pronosticado para las series de las tasas de desempleo juvenil nacional y provincial aún se mantienen en cantidades significativamente grandes.

Se seleccionó como mejor modelo probabilístico al modelo logit del año 2013, el cual muestra que las variables seleccionadas para el estudio fueron significativas, además indica que los factores determinantes del desempleo juvenil de los jóvenes ecuatorianos se relacionan con las variables sociodemográficas. Los factores que más peso tienen en el modelo son: el nivel de instrucción académica y la edad.

Se determinó que si un joven ecuatoriano, cumple con las características de: ser mujer, estar en una edad de 26 años, no ser jefe de hogar, se encuentre sin pareja, asista a clases, tenga un nivel académico superior, sus ingresos sean menores al SBU de \$357.50 y que además no especifique el sector económico y grupo de ocupación laboral y no cuente con un seguro social, tendrá una probabilidad del 94.7% de encontrarse desempleado.

RECOMENDACIONES

Investigar modelos probabilísticos: logit y/o probit para el análisis de los factores sociodemográficos y/o sociolaborales que inciden en el desempleo juvenil en el Ecuador.

Socializar con entes gubernamentales la evolución y los cambios a futuro de la tasa desempleo juvenil estudiada, para la toma de decisiones en medidas de contingencia que eviten el aumento de la misma.

Promover y motivar investigaciones en la utilización de modelos probabilísticos aplicando metodología Box-Jenkins u otras como Bootstrap en la carrera de Ingeniería en Estadística Informática de la ESPOC

BIBLIOGRAFÍA

ABELLANA, R.; & FARRAN, A. *Identificación, impacto y tratamiento de datos perdidos y atípicos en epidemiología nutricional*. REVISTA ESPAÑOLA DE NUTRICION COMUNITARIA [En línea], 2015, n.º. 2, pp. 188-194. [Consulta: 14 enero 2019]. ISSN 1135-3074. DOI 10.14642/RENC.2015.21.sup1.5065. Disponible en: <http://doi.org/10.14642/RENC.2015.21.sup1.5065>.

AHN, N.; & UGIDOS-OLAZABAL, A. *Duration of unemployment in spain: relative effects of unemployment benefit and family characteristics: unemployment in spain*. Oxford Bulletin of Economics and Statistics [En línea], 2009, vol. 57, n.º. 2, pp. 249-264. [Consulta: 21 junio 2019]. ISSN 03059049. DOI 10.1111/j.1468-0084.1995.mp57002006.x. Disponible en: <http://doi.wiley.com/10.1111/j.1468-0084.1995.mp57002006.x>.

AMÓN, I. *Guía metodológica para la selección de técnicas de depuración de datos* [En línea]. (Tesis). (Maestría). Universidad Nacional de Colombia. Medellín. 2010. pp. 1-120. [Consulta: 5 diciembre 2019]. Disponible en: <http://bdigital.unal.edu.co/2033/1/71644758.20101.pdf>.

ARAYA, C. *Modelos de clases latentes en tablas poco ocupadas: una contribución basada en Bootstrap* [En línea]. (Tesis). (Doctoral). UNIVERSIDAD DE SALAMANCA. Salamanca-España. 2010. Disponible en: <http://hdl.handle.net/10366/83209>.

BALLESTEROS, E. *La estadística descriptiva como herramienta de análisis en la investigación social. Ejemplo de «actividad práctica» adaptada a la enseñanza del Espacio Europeo de Educación Superior*. Números [En línea], 2013, vol. 84, pp. 85-98. ISSN 1887-1984. Disponible en: https://eprints.ucm.es/23413/1/Revista_N%C3%BAmeros_2013.pdf.

BATANERO, E.; & GODINO, J.D. *Análisis exploratorio de datos: sus posibilidades en la enseñanza secundaria*. [En línea], 1991, vol. 9, pp. 25-31. Disponible en: <https://www.ugr.es/~batanero/pages/ARTICULOS/anaexplora.pdf>.

BONILLA, M. *Esmeraldas tiene el mayor desempleo en la frontera norte*. Diario El Comercio [En línea]. GRUPO EL COMERCIO C.A. 2018. 11 mayo 2018. [Consulta: 28 enero 2019]. Disponible en: <http://www.elcomercio.com/actualidad/esmeraldas-desempleo-fronteranorte-ecuador-sucumbios.html>.

BOWERMAN, B.L.; et al. *Pronósticos, Series de Tiempo Y Regresión: Un Enfoque Aplicado* [En línea]. S.l.: International Thomson Editores, S.A. de C.V. 2007. ISBN 978-970-686-606-6. Disponible en: <https://books.google.com.ec/books?id=Wf25wAEACAAJ>.

CACUANGO, D.; & LUCERO, P. *Desempleo juvenil: estimación de la probabilidad de que un joven azuayo entre 18 y 24 años trabaje, año 2012* [En línea]. (Tesis). (Economista). Universidad de Cuenca. Cuenca. 2014. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/19868>.

CALAFATI, R. *Estrategias para el tratamiento de datos faltantes («missing data») en estudios con datos longitudinales* [En línea]. (Tesis). (Máster Universitario en Bioinformática y Bioestadística). Universitat Oberta de Catalunya (UOC). España. 2017. [Consulta: 28 enero 2020]. Disponible en: <https://www.semanticscholar.org/paper/Estrategias-para-el-tratamiento-de-datos-faltantes-Calafati/fe8ee907620690fa17424fef2053d766f9c7d0b0>.

CÁRDENAS, N. *Tasa de desempleo de largo plazo en el Ecuador entre 2007 Y 2010* [En línea]. (Tesis). (Ingeniero Matemático). Escuela Politécnica Nacional. Quito. 2011. pp. 1-89. Disponible en: <http://bibdigital.epn.edu.ec/handle/15000/4284>.

CAYUELA, L. *Modelos lineales generalizados (GLM)*. Universidad de Granada [En línea], 2010, pp. 30. Disponible en: https://www.academia.edu/6847226/Modelos_lineales_generalizados_GLM.

CHASCO, C. *Métodos gráficos del análisis exploratorio de datos espaciales*. Universidad Autónoma de Madrid [En línea], 2003, pp. 3-6. Disponible en: <http://www.asepelt.org/ficheros/File/Anales/2003%20-%20Almeria/asepeltPDF/93.PDF>.

CHEN, C.Y.; et al. *Economía laboral contemporánea: teorías y políticas* [En línea]. S.I.: Academia Nacional de Ciencias Económicas. 1990. ISBN 978-980-6149-65-6. Disponible en: <https://books.google.com.ec/books?id=3bpEAAAAYAAJ>.

CUADRADO, J.; et al. *Política Económica. Elaboración, objetivos e instrumentos*. Cuarta. Madrid [España]: McGRAW-HILL / INTERAMERICANA DE ESPAÑA. S. L. 2010. ISBN 978-84-481-7476-7.

Encuesta Nacional de Empleo Manual del Encuestador, 2018. S.I. s.n.

ESPINOZA, M. *Inserción laboral de los jóvenes con énfasis en la formación dual a nivel nacional. Periodo 2010-2015* [En línea]. (Tesis). (Economista). Pontificia Universidad Católica del Ecuador. Quito. 2016. Disponible en: <http://repositorio.puce.edu.ec/handle/22000/12616>.

FAWCETT, C., *Los jóvenes latinoamericanos en transición: Un análisis sobre el desempleo juvenil en América Latina y el Caribe*. [En línea]. 2002. [Consulta: 3 diciembre 2018]. Disponible en: <https://publications.iadb.org/es/publicacion/14948/los-jovenes-latinoamericanos-en-transicion-un-analisis-sobre-el-desempleo-juvenil>.

FIUZA PÉREZ, M.D.; & RODRÍGUEZ PÉREZ, J.C. *La regresión logística: una herramienta versátil*. Nefrología [En línea], 2000, vol. 20, n.º. 6, pp. 495-500. ISSN 02116995. Disponible en: <http://www.revistanefrologia.com/es-la-regresion-logistica-una-herramienta-articulo-X0211699500035664>.

GOICOECHEA, A. *Imputación basada en árboles de clasificación*. EUSTAT [En línea], 2002, pp. 1-76. Disponible en: https://www.eustat.eus/documentos/datos/ct_04_c.pdf.

GOMEZJURADO JARRIN, M.C. *Sectores de la economía ecuatoriana desde una perspectiva empresarial: aplicación de la Matriz Boston Consulting Group (BCG)*. Revista Publicando [En línea], 2016, vol. 3, n.º. 8, pp. 1-29. [Consulta: 1 junio 2020]. ISSN 1390-9304. Disponible en: https://revistapublicando.org/revista/index.php/crv/article/view/262/pdf_176.

GONZÁLEZ, J.; & FABIO, N. *Distribución de la estadística de Jarque y Bera para la prueba de normalidad en una serie temporal estacionaria con datos faltantes*. Entre Ciencia e Ingeniería [En línea], 2008, n.º. 4, pp. 99-114. ISSN 1909-8367. Disponible en: <http://revistas.ucp.edu.co/index.php/entrecienciaeingenieria/article/view/792/784>.

GOVERNMENT STATISTICAL SERVICE. *Report of the Task Force on Imputation*. Reino Unido: Office for National Statistics. 3, 1996.

GUERRERO, V.M. *Análisis Estadístico de Series de Tiempo Económicas*. S.I.: International Thomson Editores, S. A. de C. V. 2003. ISBN 978-970-686-326-3.

GUJARATI, D.; & PORTER, D. *Econometría* [En línea]. 5ta. S.I.: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V. 2010. ISBN 978-607-15-0294-0. [Consulta: 28 enero 2019]. Disponible en: https://scalleruizunp.files.wordpress.com/2015/04/econometria_-_damodar_n-_gujarati.pdf.

HANKE, J.E.; & WICHERN, D.W. *Pronósticos en los negocios*. Novena. México, D.F.: Pearson Educación. 2010. ISBN 978-607-442-700-4.

HÉRNANDEZ, R.; et al. *Metodología de la Investigación* [En línea]. Quinta. México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V. 2010. ISBN 978-607-15-0291-9. Disponible en: https://www.esup.edu.pe/descargas/dep_investigacion/Methodologia%20de%20la%20investigaci%C3%B3n%20ta%20Edici%C3%B3n.pdf.

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INEC), *Clasificación Nacional de Ocupaciones (CIUO 08). REV. 4.0.* [En línea]. Unidad de Análisis de Síntesis: Instituto Nacional de Estadística y Censos. 2012. Disponible en: <https://aplicaciones2.ecuadorencifras.gob.ec/SIN/metodologias/CIUO%2008.pdf>.

KOTZ, S.; & JOHNSON, N.L. *Breakthroughs in Statistics* [En línea]. Nueva York: Springer-Verlag. Perspectivas en la estadística, Springer Ser. Statistics Persp. Kotz, S. (Eds): avances, 1997. ISBN 978-0-387-94989-5. [Consulta: 28 enero 2020]. Disponible en: <https://www.springer.com/gp/book/9780387949895>.

LARROULET, C.; & MOCHÓN, F. *Economía*. 1a. Madrid [España]: McGraw-Hill Interamericana de España S.A U. 1995. ISBN 978-84-481-1718-4. ECO-02

Ley de la juventud, 2001 [En línea]. S.l. s.n. [Consulta: 28 enero 2020]. Disponible en: <https://www.acnur.org/fileadmin/Documentos/BDL/2008/6609.pdf>.

LÓPEZ, P.; & FACHELLI, S., *Análisis de regresión logística*. Metodología de la investigación social cuantitativa [En línea]. 1 era. Universitat Autònoma de Barcelona. 2016. Bellaterra (Cerdanyola del Vallès), pp. 56. Disponible en: <http://ddd.uab.cat/record/163570>.

MANZANO, D. *Mercado de trabajo en España: análisis, evolución y perspectivas* [En línea]. (Tesis). (Administración y Dirección de Empresa). Universidad Miguel Hernández. S.l. 2015. Disponible en: <http://dspace.umh.es/bitstream/11000/2064/1/TFG%20Manzano%20Rocamora,%20Daniel.pdf>.

MARTÍNEZ PORRAS, A.M. *Los determinantes del desempleo juvenil durante el periodo del 2015 en Colombia* [En línea]. Trabajo de Grado. Universidad Católica de Colombia. Facultad de Ciencias Económicas y Administrativas. Programa de Economía. Bogotá, Colombia. 2018. pp. 1-19. [Consulta: 23 marzo 2019]. Disponible en: <https://repository.ucatolica.edu.co/handle/10983/15978>.

MCCULLAGH, P.; & NELDER, J.A. *Generalized linear models*. 2nd ed. Boca Raton: Chapman & Hall/CRC. Monographs on statistics and applied probability, 37, 1989. ISBN 978-0-203-75373-6. QA276 .M38 1998

Metodología para la medición del empleo en Ecuador, 2016. S.l. s.n.

MITRAKOS, T.; et al. *Determining factors of youth unemployment in Greece with emphasis on tertiary education graduates*. Economic Bulletin [En línea], 2010, n°. 33, pp. 21-62. [Consulta: 28 diciembre 2018]. Disponible en: <https://ideas.repec.org/a/bog/econbl/y2010i33p21-62.html>.

MULLO, H.; & MARCATOMA, J. *Desempleo juvenil Ecuador*. Caribeña de Ciencias Sociales [En línea], 2018, n°. marzo. [Consulta: 28 enero 2019]. ISSN 2254-7630. Disponible en: <https://www.eumed.net/rev/caribe/2018/03/desempleo-juvenil-ecuador.html>.

OCAÑA, F., *Tratamiento estadístico de outliers y datos faltantes*. studylib.es [En línea]. 2019. [Consulta: 2 diciembre 2019]. Disponible en: <https://studylib.es/doc/4834194/1.-datos-atipicos-o-outliers>.

PLATEK, R., 1986. *Metodología y Tratamiento de la no respuesta. Seminario Internacional de Estadística en EUSKADI. Cuaderno 10* [En línea]. S.I. [Consulta: 24 enero 2019]. Disponible en: https://www.eustat.eus/prodsvol/vol10_c.html.

Provincias del Ecuador y sus Capitales. Provincias del Ecuador [En línea], 2020. [Consulta: 11 agosto 2019]. Disponible en: <https://provinciasecuador.com/>.

RAMÍREZ, J. *El desempleo juvenil, un problema estructural y global: el papel de las Organizaciones de la Sociedad Civil*. [En línea], 2002, n.º. 2, pp. 1-18. Disponible en: https://www.researchgate.net/publication/241760258_El_desempleo_juvenil_un_problema_estructural_y_global_El_papel_de_las_organizaciones_de_la_sociedad_civil.

RODRÍGUEZ, R. *OIT: solo uno de cada cinco jóvenes trabaja*. [En línea]. Quito-Ecuador.2018. 20 mayo 2018. [Consulta: 28 diciembre 2018]. Disponible en: <https://www.pressreader.com/article/281612421058444>.

ROSALES, R.; et al. *EJC 22: Metodología box – jenkins* [En línea]. Universidad de Los Andes. S.I. 2008. pp. 1-5. Disponible en: https://economia.uniandes.edu.co/files/profesores/ramon_rosales_alvarez/docs/econometria2/Salidas%20y%20Ejercicios/EJC202220Metodologia20Box20-20Jenkins.pdf.

ROSATI, G. *Construcción de un modelo de imputación para variables de ingreso con valores perdidos a partir de ensamble learning. Aplicación en la encuesta permanente en hogares (EPH)*. Saberes [En línea], 2017, vol. 9, n.º. 1, pp. 1-93. ISSN 91-111. Disponible en: <http://www.saberes.fcecon.unr.edu.ar/index.php/revista/article/view/132/389>.

SANTANA, J.; & FARFÁN, E. *El arte de programar en R: un lenguaje para la estadística* [En línea]. 1 era. México: Instituto Mexicano de Tecnología del Agua. 2014. ISBN 978- 607-9368-15-9. Disponible en: https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf.

SOLANO, O.; et al. *Análisis de Diagnóstico en el Modelo de Regresión Logística: Una aplicación*. Pesquimat [En línea], 2014, vol. 10, n.º. 1, pp. 55-70. [Consulta: 13 enero 2020]. ISSN 1609-8439, 1560-912X. DOI 10.15381/pes.v10i1.9431. Disponible en: <http://revistasinvestigacion.unmsm.edu.pe/index.php/matema/article/view/9431>.

USECHE, L.; & MESA, D. *Una introducción a la Imputación de valores perdidos*. Terra Nueva Etapa [En línea], 2006, vol. XXII, n.º. 31, pp. 127-151. ISSN 1012-7089. Disponible en: <https://www.redalyc.org/articulo.oa?id=72103106>.

VICENTE, I. *Determinantes del desempleo juvenil en Ecuador: un enfoque regional usando modelos logit multinivel, periodo 2012-2016* [En línea]. (Tesis). (Economista). s.n. Loja-Ecuador. 2018. pp. 1-125. [Consulta: 28 enero 2019]. Disponible en: <https://dspace.unl.edu.ec/handle/123456789/21551>.

ZEGARRA, J. *Un análisis del desempleo juvenil en Perú* [En línea]. (Tesis). (Economista). Universidad de La Laguna. Perú. 2015. pp. 1-36. [Consulta: 25 enero 2020]. Disponible en: <https://riull.ull.es/xmlui/handle/915/1291>.

ANEXOS

ANEXO A: AVAL DE LA INVESTIGACIÓN



ESPOCH
ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

INEC
INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS

Riobamba 18 de febrero de 2019

Señorita
Tania Tierra Vilema

De mi consideración:

Por la presente notifico a usted, que el Instituto Nacional de Estadísticas y Censos (INEC), a través del Convenio Marco de Cooperación Interinstitucional con la Escuela Superior Politécnica de Chimborazo (ESPOCH), entrega el **AVAL** y consiente el uso de la información proveniente de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) para la realización del trabajo de titulación denominado **"ANÁLISIS DE SERIES DE TIEMPO Y REGRESIÓN LOGIT DEL DESEMPLEO JUVENIL EN EL ECUADOR, PERIODO 2013 - 2018"**.

Además la institución patrocinadora declara conocer y aceptar los términos y condiciones previstas para la ejecución del Proyecto, estando conformes con todas aquellas actividades que se prevean realizar.

Sin otro particular, saludo a usted atentamente,

Alexaucha Marcátoma Tixi
Jessica Alexandra Marcátoma Tixi
COORDINADORA INSTITUCIONAL CONVENIO INEC – ESSPOCH

ESCUELA SUPERIOR
POLITECNICA DE CHIMBORAZO
DIRECCION DE VINCULACION
18 FEB 2019
CONVENIOS INTERINSTITUCIONALES

ANEXO B: TASAS DEL DESEMPLEO JUVENIL EN ECUADOR NACIONAL Y PROVINCIAL 2013-2018

Tabla 20-3: Tasas del desempleo juvenil nacional y provincial de Ecuador 2013-2018

AÑO	TRI MES TRE	PROVINCIAS																								T. Nacional
		EC-A	EC-B	EC-F	EC-C	EC-X	EC-H	EC-O	EC-E	EC-G	EC-I	EC-L	EC-R	EC-M	EC-S	EC-N	EC-Y	EC-P	EC-T	EC-Z	EC-U	EC-D	EC-SD	EC-SE	EC-ZND	
2013	1 ^{ro}	3.96	14.28	19.34	6.29	2.76	12.01	10.37	29.80	12.59	5.32	6.75	7.69	11.58	7.40	4.06	19.40	9.32	9.48	15.25	6.25	0.00			8.39	10.57
	2 ^{do}	6.81	5.52	7.81	5.39	2.89	4.87	8.03	11.72	12.93	10.84	8.78	11.15	7.69	5.54	5.27	1.35	7.77	4.92	8.02	10.45	5.34			13.13	9.11
	3 ^{ro}	9.39	4.64	6.01	0.00	2.38	8.97	9.24	12.12	11.61	16.54	9.65	7.66	13.95	11.95	6.37	4.53	6.05	5.03	18.27	0.00	0.00	0.00	9.04	18.37	9.24
	4 ^{to}	5.59	2.42	6.45	6.09	3.17	5.58	7.43	15.44	12.18	9.93	10.08	8.78	10.55	5.07	8.95	1.07	5.99	4.86	1.07	8.82	8.12	7.75	13.42	0.00	8.64
2014	1 ^{ro}	6.04	13.58	12.63	11.10	2.99	2.74	10.07	12.04	13.80	12.63	10.06	12.56	11.72	7.52	19.04	7.66	9.21	6.45	10.57	8.70	11.09	13.54	13.82	7.03	10.71
	2 ^{do}	6.82	5.81	8.16	9.64	3.74	2.00	8.71	13.15	13.95	10.45	9.90	11.94	9.17	4.15	9.50	6.06	9.58	4.99	6.45	7.24	7.07	6.42	18.07	5.88	10.08
	3 ^{ro}	4.63	9.77	6.55	16.90	3.33	5.11	4.45	8.06	9.43	13.16	6.90	12.01	8.90	4.09	11.02	6.51	8.51	6.16	6.54	11.55	7.65	7.43	13.17	8.25	8.46
	4 ^{to}	2.74	5.72	4.26	9.66	5.32	3.51	5.93	12.69	10.54	9.38	10.71	9.08	8.66	5.32	6.11	7.34	6.18	4.58	6.88	13.60	7.74	8.63	9.43	1.84	7.88
2015	1 ^{ro}	3.22	11.07	2.98	9.84	4.82	2.01	6.22	17.52	8.64	4.49	7.10	9.27	9.77	4.44	7.97	5.91	10.31	8.12	4.80	10.90	7.87	11.68	11.53	0.00	8.50
	2 ^{do}	4.04	5.70	8.31	10.24	3.98	1.57	5.17	19.20	11.04	8.75	7.47	8.93	9.78	2.80	10.30	3.88	8.20	3.82	4.64	12.37	9.99	10.03	14.70	1.72	8.70
	3 ^{ro}	2.88	5.60	10.44	13.97	4.91	7.51	6.53	17.99	11.05	12.35	13.12	7.58	9.11	3.73	11.65	5.25	6.43	4.22	6.39	10.66	6.00	6.51	12.48	4.78	8.70
	4 ^{to}	4.00	8.98	7.80	9.49	3.80	4.68	8.09	18.30	12.18	9.66	13.11	7.70	12.31	4.92	9.33	7.44	8.91	6.92	7.79	9.57	10.00	10.30	7.60	0.00	9.72
2016	1 ^{ro}	5.01	3.41	4.39	12.60	5.26	7.99	8.73	12.96	16.39	10.09	13.55	16.68	9.00	3.44	6.88	7.01	13.19	10.03	4.37	10.26	10.54	17.30	9.53	10.10	11.73
	2 ^{do}	4.50	4.29	3.65	11.15	10.20	5.03	5.51	18.71	9.41	11.70	13.59	10.63	12.43	5.97	7.20	9.12	11.77	9.64	7.33	8.94	4.07	10.40	10.75	2.82	10.01
	3 ^{ro}	6.97	5.62	10.01	4.54	7.97	9.09	7.52	15.65	11.72	13.47	14.52	11.42	8.52	4.13	8.06	13.72	13.31	10.47	2.88	20.58	7.38	10.23	9.56	1.01	10.75
	4 ^{to}	5.45	7.00	11.54	14.70	4.26	2.55	8.67	15.76	13.60	10.43	11.29	9.46	7.68	3.42	6.89	6.25	12.79	6.27	8.91	13.02	10.37	7.47	8.11	6.36	10.43
2017	1 ^{ro}	5.13	3.36	5.02	13.18	2.81	1.92	6.84	9.93	11.40	6.73	11.77	12.55	6.41	0.60	4.08	9.31	15.83	4.83	9.70	10.36	5.31	6.28	5.94	2.57	9.12
	2 ^{do}	5.45	3.54	5.28	9.74	6.02	4.30	7.21	14.50	11.28	9.83	3.21	12.48	10.25	3.14	4.56	3.32	13.68	4.08	6.70	13.43	3.52	5.95	9.18	8.95	9.26
	3 ^{ro}	4.55	3.94	6.66	7.74	5.34	7.89	9.83	13.17	9.04	10.51	15.36	6.66	3.08	6.83	4.99	8.03	13.94	4.94	2.33	17.05	11.29	5.75	9.70	3.38	8.47
	4 ^{to}	7.75	6.48	8.57	10.99	3.73	2.02	8.51	15.16	10.58	11.60	6.96	9.23	5.49	2.29	5.28	3.35	15.59	8.40	6.55	16.77	6.41	5.30	8.04	0.00	9.61
2018	1 ^{ro}	7.75	14.38	15.02	1.43	7.51	1.31	7.04	14.79	10.20	7.15	6.44	11.17	7.31	3.88	2.80	0.00	12.12	9.91	6.54	7.57	2.30	8.21	12.32	7.08	9.31
	2 ^{do}	7.74	0.60	7.63	3.43	3.73	4.01	7.41	19.02	8.98	10.24	9.00	8.75	4.18	5.77	19.99	4.59	14.97	6.29	2.73	4.65	3.21	6.89	12.84	8.07	8.88
	3 ^{ro}	8.04	3.62	13.21	11.03	5.93	2.41	10.31	22.34	7.24	13.30	7.02	6.94	3.71	1.02	4.38	5.59	15.28	6.75	6.24	13.30	5.99	5.01	6.28	0.51	9.22
	4 ^{to}	5.56	5.19	11.10	13.56	1.76	5.59	9.42	13.65	7.30	5.19	7.22	7.15	5.00	0.77	8.38	7.64	13.70	4.10	17.51	13.79	8.04	10.76	6.06	1.19	8.25

Fuente: Encuesta ENEMDU 2013-2018

Realizado por: Tierra, Tania. 2020

ANEXO C: CÓDIGO R

```
#####LECTURA DATOS#####
library(RColorBrewer)
library(foreign)
library(car)
setwd("F:/TESIS/ANALISIS_DATOS/Bases depuradas/2014")
# LECTURA DE BASE DE DATOS DE SPSS MARZO----
Mar14 <- read.spss("MARZO_14.sav", to.data.frame = TRUE)
#RECODIFICACION DE VARIABLES---
desem <- as.numeric(Mar14$desem)
Mar14$desem <- as.factor(iffelse(desem==1,c("Desempleo"),c("No desempleo")))
table(Mar14$desem)
sexo <- as.numeric(Mar14$Sexo)
Mar14$Sexo <- as.factor(iffelse(sexo==1,c("Hombre"),c("Mujer")))
table(Mar14$Sexo)
Mar14$Edad <- factor(recode(Mar14$Edad, "18:21=1; 22:25=2; 26:29=3"),
  labels = c("18-21", "22-25", "26-29"))
table(Mar14$Edad)
paren <- as.numeric(Mar14$Parentesco)
Mar14$Parentesco <- as.factor(iffelse(paren==1,c("Jefe"),c("Otros")))
table(Mar14$Parentesco)
Estcivil <- as.numeric(Mar14$Estcivil)
Mar14$Estcivil <- as.factor(iffelse(Estcivil==1,c("Con pareja"),c("Sin pareja es
table")))
table(Mar14$Estcivil)
Asclases <- as.numeric(Mar14$Asistclases)
Mar14$Asistclases <- factor(iffelse(Asclases==1,c("Si"),c("No")))
table(Mar14$Asistclases)
Nivelinst <- as.numeric(Mar14$Nivelinst)
table(Nivelinst)
Nivel2 <- vector()
Nivel2[Nivelinst == 1] <- "Ninguno"
Nivel2[Nivelinst ==2] <- "Primaria"
Nivel2[Nivelinst ==4] <- "Primaria"
Nivel2[Nivelinst ==5] <- "Primaria"
Nivel2[Nivelinst ==6] <- "Secundaria"
Nivel2[Nivelinst ==7] <- "Secundaria"
Nivel2[Nivelinst >7] <- "Superior y más"
Nivelinst <- as.factor(Nivel2)
Mar14$Nivelinst <- Nivelinst
table(Mar14$Nivelinst)
ingrl<- as.numeric(Mar14$ingrl)
table(ingrl)
ingre2 <- vector()
ingre2[ingrl >393] <- "Mayor o igual al SBU"
ingre2[ingrl <394] <- "Menor al SBU"
ingrl <- as.factor(ingre2)
Mar14$ingrl <- ingrl
table(Mar14$ingrl)
rama<- as.numeric(Mar14$rama)
table(rama)
rama2 <- vector()
rama2[rama ==1] <- "Primario"
rama2[rama ==2] <- "Secundario"
rama2[rama ==3] <- "Secundario"
rama2[rama ==6] <- "Secundario"
rama2[rama ==4] <- "Terciario"
rama2[rama ==5] <- "Terciario"
rama2[rama >6] <- "Terciario"
rama2[rama ==22] <- "No especifica"
rama <- as.factor(rama2)
Mar14$rama <- rama
table(Mar14$rama)
grupo<- as.numeric(Mar14$grupo)
table(grupo)
```

```

grupo2 <- vector()
grupo2[grupo <5] <- "Ocupaciones no Manuales"
grupo2[grupo >4] <- "Ocupaciones Manuales"
grupo2[grupo ==11] <- "No especifica"
grupo <- as.factor(grupo2)
Mar14$grupo <- grupo
table(Mar14$grupo)
Seguro <- as.numeric(Mar14$Seguro)
Mar14$Seguro <- factor(iffelse(Seguro==1,c("Si"),c("No")))
table(Mar14$Seguro)

# LECTURA DE BASE DE DATOS DE SPSS JUNIO----
Jun14 <- read.spss("JUNIO_14.sav", to.data.frame = TRUE)
#RECODIFICACION DE VARIABLES---
desem <- as.numeric(Jun14$desem)
Jun14$desem <- as.factor(iffelse(desem==1,c("Desempleo"),c("No desempleo")))
sexo <- as.numeric(Jun14$Sexo)
Jun14$Sexo <- as.factor(iffelse(sexo==1,c("Hombre"),c("Mujer")))
Jun14$Edad <- factor(recode(Jun14$Edad, "18:21=1; 22:25=2; 26:29=3"),
                    labels = c("18-21", "22-25", "26-29"))
paren <- as.numeric(Jun14$Parentesco)
Jun14$Parentesco <- as.factor(iffelse(paren==1,c("Jefe"),c("Otros")))
Estcivil <- as.numeric(Jun14$Estcivil)
Jun14$Estcivil <- as.factor(iffelse(Estcivil==1,c("Con pareja"),c("Sin pareja es
table")))
Asclases <- as.numeric(Jun14$Asistclases)
Jun14$Asistclases <- factor(iffelse(Asclases==1,c("Si"),c("No")))
Nivelinst <- as.numeric(Jun14$Nivelinst)
Nivel2 <- vector()
Nivel2[Nivelinst == 1] <- "Ninguno"
Nivel2[Nivelinst ==2] <- "Primaria"
Nivel2[Nivelinst ==4] <- "Primaria"
Nivel2[Nivelinst ==5] <- "Primaria"
Nivel2[Nivelinst ==6] <- "Secundaria"
Nivel2[Nivelinst ==7] <- "Secundaria"
Nivel2[Nivelinst >7] <- "Superior y mas"
Nivelinst <- as.factor(Nivel2)
Jun14$Nivelinst <- Nivelinst
ingr1<- as.numeric(Jun14$ingr1)
ingre2 <- vector()
ingre2[ingr1 >393] <- "Mayor o igual al SBU"
ingre2[ingr1 <394] <- "Menor al SBU"
ingr1 <- as.factor(ingre2)
Jun14$ingr1 <- ingr1
rama<- as.numeric(Jun14$rama)
rama2 <- vector()
rama2[rama ==1] <- "Primario"
rama2[rama ==2] <- "Secundario"
rama2[rama ==3] <- "Secundario"
rama2[rama ==6] <- "Secundario"
rama2[rama ==4] <- "Terciario"
rama2[rama ==5] <- "Terciario"
rama2[rama >6] <- "Terciario"
rama2[rama ==22] <- "No especifica"
rama <- as.factor(rama2)
Jun14$rama <- rama
grupo<- as.numeric(Jun14$grupo)
grupo2 <- vector()
grupo2[grupo <5] <- "Ocupaciones no Manuales"
grupo2[grupo >4] <- "Ocupaciones Manuales"
grupo2[grupo ==11] <- "No especifica"
grupo <- as.factor(grupo2)
Jun14$grupo <- grupo
Seguro <- as.numeric(Jun14$Seguro)
Jun14$Seguro <- factor(iffelse(Seguro==1,c("Si"),c("No")))
# LECTURA DE BASE DE DATOS DE SPSS SEPTIEMBRE----
Sep14 <- read.spss("SEPTIEMBRE_14.sav", to.data.frame = TRUE)
#RECODIFICACION DE VARIABLES---

```

```

desem <- as.numeric(Sep14$desem)
Sep14$desem <- as.factor(iffelse(desem==1,c("Desempleo"),c("No desempleo")))
sexo <- as.numeric(Sep14$Sexo)
Sep14$Sexo <- as.factor(iffelse(sexo==1,c("Hombre"),c("Mujer")))
Sep14$Edad <- factor(recode(Sep14$Edad, "18:21=1; 22:25=2; 26:29=3"),
  labels = c("18-21", "22-25", "26-29"))
paren <- as.numeric(Sep14$Parentesco)
Sep14$Parentesco <- as.factor(iffelse(paren==1,c("Jefe"),c("Otros")))
Estcivil <- as.numeric(Sep14$Estcivil)
Sep14$Estcivil <- as.factor(iffelse(Estcivil==1,c("Con pareja"),c("Sin pareja es
table")))
Asclases <- as.numeric(Sep14$Asistclases)
Sep14$Asistclases <- factor(iffelse(Asclases==1,c("Si"),c("No")))
Nivelinst <- as.numeric(Sep14$Nivelinst)
Nivel2 <- vector()
Nivel2[Nivelinst == 1] <- "Ninguno"
Nivel2[Nivelinst ==2] <- "Primaria"
Nivel2[Nivelinst ==4] <- "Primaria"
Nivel2[Nivelinst ==5] <- "Primaria"
Nivel2[Nivelinst ==6] <- "Secundaria"
Nivel2[Nivelinst ==7] <- "Secundaria"
Nivel2[Nivelinst >7] <- "Superior y mas"
Nivelinst <- as.factor(Nivel2)
Sep14$Nivelinst <- Nivelinst
ingr1<- as.numeric(Sep14$ingr1)
ingre2 <- vector()
ingre2[ingr1 >393] <- "Mayor o igual al SBU"
ingre2[ingr1 <394] <- "Menor al SBU"
ingr1 <- as.factor(ingre2)
Sep14$ingr1 <- ingr1
rama<- as.numeric(Sep14$rama)
rama2 <- vector()
rama2[rama ==1] <- "Primario"
rama2[rama ==2] <- "Secundario"
rama2[rama ==3] <- "Secundario"
rama2[rama ==6] <- "Secundario"
rama2[rama ==4] <- "Terciario"
rama2[rama ==5] <- "Terciario"
rama2[rama >6] <- "Terciario"
rama2[rama ==22] <- "No especifica"
rama <- as.factor(rama2)
Sep14$rama <- rama
grupo<- as.numeric(Sep14$grupo)
grupo2 <- vector()
grupo2[grupo <5] <- "Ocupaciones no Manuales"
grupo2[grupo >4] <- "Ocupaciones Manuales"
grupo2[grupo ==11] <- "No especifica"
grupo <- as.factor(grupo2)
Sep14$grupo <- grupo
Seguro <- as.numeric(Sep14$Seguro)
Sep14$Seguro <- factor(iffelse(Seguro==1,c("Si"),c("No")))
# LECTURA DE BASE DE DATOS DE SPSS DICIEMBRE----
Dic14 <- read.spss("DICIEMBRE_14.sav", to.data.frame = TRUE)
#RECODIFICACION DE VARIABLES---
desem <- as.numeric(Dic14$desem)
Dic14$desem <- as.factor(iffelse(desem==1,c("Desempleo"),c("No desempleo")))
sexo <- as.numeric(Dic14$Sexo)
Dic14$Sexo <- as.factor(iffelse(sexo==1,c("Hombre"),c("Mujer")))
Dic14$Edad <- factor(recode(Dic14$Edad, "18:21=1; 22:25=2; 26:29=3"),
  labels = c("18-21", "22-25", "26-29"))
paren <- as.numeric(Dic14$Parentesco)
Dic14$Parentesco <- as.factor(iffelse(paren==1,c("Jefe"),c("Otros")))
Estcivil <- as.numeric(Dic14$Estcivil)
Dic14$Estcivil <- as.factor(iffelse(Estcivil==1,c("Con pareja"),c("Sin pareja es
table")))
Asclases <- as.numeric(Dic14$Asistclases)
Dic14$Asistclases <- factor(iffelse(Asclases==1,c("Si"),c("No")))
Nivelinst <- as.numeric(Dic14$Nivelinst)

```

```

Nivel2 <- vector()
Nivel2[Nivelinst == 1] <- "Ninguno"
Nivel2[Nivelinst ==2] <- "Primaria"
Nivel2[Nivelinst ==4] <- "Primaria"
Nivel2[Nivelinst ==5] <- "Primaria"
Nivel2[Nivelinst ==6] <- "Secundaria"
Nivel2[Nivelinst ==7] <- "Secundaria"
Nivel2[Nivelinst >7] <- "Superior y mas"
Nivelinst <- as.factor(Nivel2)
Dic14$Nivelinst <- Nivelinst
ingr1<- as.numeric(Dic14$ingr1)
ingre2 <- vector()
ingre2[ingr1 >394] <- "Mayor o igual al SBU"
ingre2[ingr1 ==394] <- "Mayor o igual al SBU"
ingre2[ingr1 <394] <- "Menor al SBU"
ingr1 <- as.factor(ingre2)
Dic14$ingr1 <- ingr1
rama<- as.numeric(Dic14$rama)
rama2 <- vector()
rama2[rama ==1] <- "Primario"
rama2[rama ==2] <- "Secundario"
rama2[rama ==3] <- "Secundario"
rama2[rama ==6] <- "Secundario"
rama2[rama ==4] <- "Terciario"
rama2[rama ==5] <- "Terciario"
rama2[rama >6] <- "Terciario"
rama2[rama ==22] <- "No especifica"
rama <- as.factor(rama2)
Dic14$rama <- rama
grupo<- as.numeric(Dic14$grupo)
grupo2 <- vector()
grupo2[grupo <5] <- "Ocupaciones no Manuales"
grupo2[grupo >4] <- "Ocupaciones Manuales"
grupo2[grupo ==11] <- "No especifica"
grupo <- as.factor(grupo2)
Dic14$grupo <- grupo
Seguro <- as.numeric(Dic14$Seguro)
Dic14$Seguro <- factor(ifelse(Seguro==1,c("Si"),c("No")))
#***** ANALISIS DESCRIPTIVO VAR. CUALITATIVAS ----
####TABLAS DE FRECUENCIA----
# DESEMPLEO----
#MARZO
FDMA14<- as.data.frame(table(Mar14$desem)) #tabla de frecuencia
FRMAR14<- round(prop.table(FDMA14$Freq)*100,2) #La frecuencia relativa en porce
ntaje
#JUNIO
FDJU14<- as.data.frame(table(Jun14$desem)) #tabla de frecuencia
FRJUN14<- round(prop.table(FDJU14$Freq)*100,2) #La frecuencia relativa en porce
ntaje
#SEPTIEMBRE
FDSE14<- as.data.frame(table(Sep14$desem)) #tabla de frecuencia
FRSEPT14<- round(prop.table(FDSE14$Freq)*100,2) #La frecuencia relativa en porc
entaje
#DICIEMBRE
FDDI14<- as.data.frame(table(Dic14$desem)) #tabla de frecuencia
FRDIC14<- round(prop.table(FDDI14$Freq)*100,2) #La frecuencia relativa en porce
ntaje
TAB14<-cbind(FDMA14,FRMAR14,FDJU14$Freq,FRJUN14,FDSE14$Freq,FRSEPT14,FDDI14$Fr
eq,FRDIC14 ); TAB14
# SEXO----
#MARZO
FDMA14S<- as.data.frame(table(Mar14$Sexo)) #tabla de frecuencia
FRMAR14S<- round(prop.table(FDMA14S$Freq)*100,2) #La frecuencia relativa en por
centaj
#JUNIO
FDJU14S<- as.data.frame(table(Jun14$Sexo)) #tabla de frecuencia
FRJUN14S<- round(prop.table(FDJU14S$Freq)*100,2) #La frecuencia relativa en por
centa

```

```

#SEPTIEMBRE
FDSE14S<- as.data.frame(table(Sep14$Sexo)) #tabla de frecuencia
FRSEPT14S<- round(prop.table(FDSE14S$Freq)*100,2) #La frecuencia relativa en porcenta
centaj
#DICIEMBRE
FDDI14S<- as.data.frame(table(Dic14$Sexo)) #tabla de frecuencia
FRDIC14S<- round(prop.table(FDDI14S$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
TAB14S<-cbind(FDMA14S,FRMAR14S,FDJU14S$Freq,FRJUN14S,FDSE14S$Freq,FRSEPT14S,FD
DI14S$Freq,FRDIC14S); TAB14S
# EDAD----
#MARZO
FDMA14E<- as.data.frame(table(Mar14$Edad)) #tabla de frecuencia
FRMAR14E<- round(prop.table(FDMA14E$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
#JUNIO
FDJU14E<- as.data.frame(table(Jun14$Edad)) #tabla de frecuencia
FRJUN14E<- round(prop.table(FDJU14E$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
#SEPTIEMBRE
FDSE14E<- as.data.frame(table(Sep14$Edad)) #tabla de frecuencia
FRSEPT14E<- round(prop.table(FDSE14E$Freq)*100,2) #La frecuencia relativa en porcenta
centaj
#DICIEMBRE
FDDI14E<- as.data.frame(table(Dic14$Edad)) #tabla de frecuencia
FRDIC14E<- round(prop.table(FDDI14E$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
TAB14E<-cbind(FDMA14E,FRMAR14E,FDJU14E$Freq,FRJUN14E,FDSE14E$Freq,FRSEPT14E,FD
DI14E$Freq,FRDIC14E); TAB14E
# PARENTESCO----
#MARZO
FDMA14P<- as.data.frame(table(Mar14$Parentesco)) #tabla de frecuencia
FRMAR14P<- round(prop.table(FDMA14P$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
#JUNIO
FDJU14P<- as.data.frame(table(Jun14$Parentesco)) #tabla de frecuencia
FRJUN14P<- round(prop.table(FDJU14P$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
#SEPTIEMBRE
FDSE14P<- as.data.frame(table(Sep14$Parentesco)) #tabla de frecuencia
FRSEPT14P<- round(prop.table(FDSE14P$Freq)*100,2) #La frecuencia relativa en porcenta
centaj
#DICIEMBRE
FDDI14P<- as.data.frame(table(Dic14$Parentesco)) #tabla de frecuencia
FRDIC14P<- round(prop.table(FDDI14P$Freq)*100,2) #La frecuencia relativa en porcentaj
centaj
TAB14P<-cbind(FDMA14P,FRMAR14P,FDJU14P$Freq,FRJUN14P,FDSE14P$Freq,FRSEPT14P,FD
DI14P$Freq,FRDIC14P); TAB14P
# ESTADO CIVIL----
#MARZO
FDMA14EC<- as.data.frame(table(Mar14$Estcivil)) #tabla de frecuencia
FRMAR14EC<- round(prop.table(FDMA14EC$Freq)*100,2) #La frecuencia relativa en porcent
centaj
#JUNIO
FDJU14EC<- as.data.frame(table(Jun14$Estcivil)) #tabla de frecuencia
FRJUN14EC<- round(prop.table(FDJU14EC$Freq)*100,2) #La frecuencia relativa en porcent
centaj
#SEPTIEMBRE
FDSE14EC<- as.data.frame(table(Sep14$Estcivil)) #tabla de frecuencia
FRSEPT14EC<- round(prop.table(FDSE14EC$Freq)*100,2) #La frecuencia relativa en porcent
centaj
#DICIEMBRE
FDDI14EC<- as.data.frame(table(Dic14$Estcivil)) #tabla de frecuencia
FRDIC14EC<- round(prop.table(FDDI14EC$Freq)*100,2) #La frecuencia relativa en porcent
centaj
TAB14EC<-cbind(FDMA14EC,FRMAR14EC,FDJU14EC$Freq,FRJUN14EC,FDSE14EC$Freq,FRSEPT
14EC,FDDI14EC$Freq,FRDIC14EC); TAB14EC
# ASISTE A CLASES----

```

```

#MARZO
FDMA14AC<- as.data.frame(table(Mar14$Asistclases)) #tabla de frecuencia
FRMAR14AC<- round(prop.table(FDMA14AC$Freq)*100,2) #La frecuencia relativa en p
orcent
#JUNIO
FDJU14AC<- as.data.frame(table(Jun14$Asistclases)) #tabla de frecuencia
FRJUN14AC<- round(prop.table(FDJU14AC$Freq)*100,2) #La frecuencia relativa en p
orcent
#SEPTIEMBRE
FDSE14AC<- as.data.frame(table(Sep14$Asistclases)) #tabla de frecuencia
FRSEPT14AC<- round(prop.table(FDSE14AC$Freq)*100,2) #La frecuencia relativa en
porcen
#DICIEMBRE
FDDI14AC<- as.data.frame(table(Dic14$Asistclases)) #tabla de frecuencia
FRDIC14AC<- round(prop.table(FDDI14AC$Freq)*100,2) #La frecuencia relativa en p
orcent
TAB14AC<-cbind(FDMA14AC, FRMAR14AC, FDJU14AC$Freq, FRJUN14AC, FDSE14AC$Freq, FRSEPT
14AC, FDDI14AC$Freq, FRDIC14AC); TAB14AC
# NIVEL DE INSTRUCCION----
#MARZO
FDMA14NI<- as.data.frame(table(Mar14$Nivelinst)) #tabla de frecuencia
FRMAR14NI<- round(prop.table(FDMA14NI$Freq)*100,2) #La frecuencia relativa en p
orcen
#JUNIO
FDJU14NI<- as.data.frame(table(Jun14$Nivelinst)) #tabla de frecuencia
FRJUN14NI<- round(prop.table(FDJU14NI$Freq)*100,2) #La frecuencia relativa en p
orcent
#SEPTIEMBRE
FDSE14NI<- as.data.frame(table(Sep14$Nivelinst)) #tabla de frecuencia
FRSEPT14NI<- round(prop.table(FDSE14NI$Freq)*100,2) #La frecuencia relativa en
porce
#DICIEMBRE
FDDI14NI<- as.data.frame(table(Dic14$Nivelinst)) #tabla de frecuencia
FRDIC14NI<- round(prop.table(FDDI14NI$Freq)*100,2) #La frecuencia relativa en p
ortaje
TAB14NI<-cbind(FDMA14NI, FRMAR14NI, FDJU14NI$Freq, FRJUN14NI, FDSE14NI$Freq, FRSEPT
14NI, FDDI14NI$Freq, FRDIC14NI); TAB14NI
# INGRESO LABORAL----
#MARZO
FDMA14IL<- as.data.frame(table(Mar14$ingrl)) #tabla de frecuencia
FRMAR14IL<- round(prop.table(FDMA14IL$Freq)*100,2) #La frecuencia relativa en p
orcent
#JUNIO
FDJU14IL<- as.data.frame(table(Jun14$ingrl)) #tabla de frecuencia
FRJUN14IL<- round(prop.table(FDJU14IL$Freq)*100,2) #La frecuencia relativa en p
orcent
#SEPTIEMBRE
FDSE14IL<- as.data.frame(table(Sep14$ingrl)) #tabla de frecuencia
FRSEPT14IL<- round(prop.table(FDSE14IL$Freq)*100,2) #La frecuencia relativa en
porcen
#DICIEMBRE
FDDI14IL<- as.data.frame(table(Dic14$ingrl)) #tabla de frecuencia
FRDIC14IL<- round(prop.table(FDDI14IL$Freq)*100,2) #La frecuencia relativa en p
orcent
TAB14IL<-cbind(FDMA14IL, FRMAR14IL, FDJU14IL$Freq, FRJUN14IL, FDSE14IL$Freq, FRSEPT
14IL, FDDI14IL$Freq, FRDIC14IL); TAB14IL
# RAMA DE ACTIVIDAD----
#MARZO
FDMA14RA<- as.data.frame(table(Mar14$rama)) #tabla de frecuencia
FRMAR14RA<- round(prop.table(FDMA14RA$Freq)*100,2) #La frecuencia relativa en p
orcentaje
#JUNIO
FDJU14RA<- as.data.frame(table(Jun14$rama)) #tabla de frecuencia
FRJUN14RA<- round(prop.table(FDJU14RA$Freq)*100,2) #La frecuencia relativa en p
orcentaje
#SEPTIEMBRE
FDSE14RA<- as.data.frame(table(Sep14$rama)) #tabla de frecuencia

```

```

FRSEPT14RA<- round(prop.table(FDSE14RA$Freq)*100,2) #La frecuencia relativa en
porcentaje
#DICIEMBRE
FDDI14RA<- as.data.frame(table(Dic14$rama)) #tabla de frecuencia
FRDIC14RA<- round(prop.table(FDDI14RA$Freq)*100,2) #La frecuencia relativa en p
orcentaje
TAB14RA<-cbind(FDMA14RA, FRMAR14RA, FDJU14RA$Freq, FRJUN14RA, FDSE14RA$Freq, FRSEPT
14RA, FDDI14RA$Freq, FRDIC14RA); TAB14RA
# GRUPO DE OCUPACION----
#MARZO
FDMA14GO<- as.data.frame(table(Mar14$grupo)) #tabla de frecuencia
FRMAR14GO<- round(prop.table(FDMA14GO$Freq)*100,2) #La frecuencia relativa en p
orcent
#JUNIO
FDJU14GO<- as.data.frame(table(Jun14$grupo)) #tabla de frecuencia
FRJUN14GO<- round(prop.table(FDJU14GO$Freq)*100,2) #La frecuencia relativa en p
orcent
#SEPTIEMBRE
FDSE14GO<- as.data.frame(table(Sep14$grupo)) #tabla de frecuencia
FRSEPT14GO<- round(prop.table(FDSE14GO$Freq)*100,2) #La frecuencia relativa en
porcen
#DICIEMBRE
FDDI14GO<- as.data.frame(table(Dic14$grupo)) #tabla de frecuencia
FRDIC14GO<- round(prop.table(FDDI14GO$Freq)*100,2) #La frecuencia relativa en p
orcent
TAB14GO<-cbind(FDMA14GO, FRMAR14GO, FDJU14GO$Freq, FRJUN14GO, FDSE14GO$Freq, FRSEPT
14GO, FDDI14GO$Freq, FRDIC14GO); TAB14GO
# SEGURO----
#MARZO
FDMA14SS<- as.data.frame(table(Mar14$Seguro)) #tabla de frecuencia
FRMAR14SS<- round(prop.table(FDMA14SS$Freq)*100,2) #La frecuencia relativa en p
orcent
#JUNIO
FDJU14SS<- as.data.frame(table(Jun14$Seguro)) #tabla de frecuencia
FRJUN14SS<- round(prop.table(FDJU14SS$Freq)*100,2) #La frecuencia relativa en p
orcent
#SEPTIEMBRE
FDSE14SS<- as.data.frame(table(Sep14$Seguro)) #tabla de frecuencia
FRSEPT14SS<- round(prop.table(FDSE14SS$Freq)*100,2) #La frecuencia relativa en
porcen
#DICIEMBRE
FDDI14SS<- as.data.frame(table(Dic14$Seguro)) #tabla de frecuencia
FRDIC14SS<- round(prop.table(FDDI14SS$Freq)*100,2) #La frecuencia relativa en p
orcent
TAB14SS<-cbind(FDMA14SS, FRMAR14SS, FDJU14SS$Freq, FRJUN14SS, FDSE14SS$Freq, FRSEPT
14SS, FDDI14SS$Freq, FRDIC14SS); TAB14SS

#*****IMPUTACION DE DATOS FALTANTES----
#funcion para obtener la moda
moda=getmode <- function(v){
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
#MARZO
#Llamar a la base de datos
datosMAR14 <- Mar14$ingrl
#Cuento cuantos datos faltantes tengo
sum(is.na(datosMAR14))
### limpiar la base de NAS
datlimMAR14 <- na.omit(datosMAR14)
##Calculo la moda
modMAR14 <- moda(datlimMAR14);modMAR14
### Lleno los espacios vacíos con el valor de la moda
Mar14$ingrl[is.na(Mar14$ingrl)] <- modMAR14
#JUNIO
#Llamar a la base de datos
datosJUN14 <- Jun14$ingrl
#Cuento cuantos datos faltantes tengo

```

```

sum(is.na(datosJUN14))
### limpiar la base de NAS
datlimJUN14 <- na.omit(datosJUN14)
##Calculo la moda
modJUN14 <- moda(datlimJUN14);modJUN14
### Lleno los espacios vacíos con el valor de la moda
Jun14$ingrl[is.na(Jun14$ingrl)] <- modJUN14
#SEPTIEMBRE
#Llamar a la base de datos
datosSEP14 <- Sep1414$ingrl
#Cuento cuantos datos faltantes tengo
sum(is.na(datosSEP14))
### limpiar la base de NAS
datlimSEP14 <- na.omit(datosSEP14)
##Calculo la moda
modSEP14 <- moda(datlimSEP14);modSEP14
### Lleno los espacios vacíos con el valor de la moda
Sep14$ingrl[is.na(Sep14$ingrl)] <- modSEP14
#DICIEMBRE
#Llamar a la base de datos
datosDIC14 <- Dic1414$ingrl
#Cuento cuantos datos faltantes tengo
sum(is.na(datosDIC14))
### limpiar la base de NAS
datlimDIC14 <- na.omit(datosDIC14)
##Calculo la moda
modDIC14 <- moda(datlimDIC14);modDIC14
### Lleno los espacios vacios con el valor de la moda
Dic14$ingrl[is.na(Dic14$ingrl)] <- modDIC14
#Vuelvo a calcular las frecuencias
#MARZO
FDMA14IL<- as.data.frame(table(Mar14$ingrl)) #tabla de frecuencia
FRMAR14IL<- round(prop.table(FDMA14IL$Freq)*100,2) #La frecuencia relativa en p
orcent
#JUNIO
FDJU14IL<- as.data.frame(table(Jun14$ingrl)) #tabla de frecuencia
FRJUN14IL<- round(prop.table(FDJU14IL$Freq)*100,2) #La frecuencia relativa en p
orcent
#SEPTIEMBRE
FDSE14IL<- as.data.frame(table(Sep14$ingrl)) #tabla de frecuencia
FRSEPT14IL<- round(prop.table(FDSE14IL$Freq)*100,2) #La frecuencia relativa en
porcen
#DICIEMBRE
FDDI14IL<- as.data.frame(table(Dic14$ingrl)) #tabla de frecuencia
FRDIC14IL<- round(prop.table(FDDI14IL$Freq)*100,2) #La frecuencia relativa en p
orcent
TAB14IL<-cbind(FDMA14IL,FRMAR14IL,FDJU14IL$Freq,FRJUN14IL,FDSE14IL$Freq,FRSEPT
14IL,FDDI14IL$Freq,FRDIC14IL); TAB14IL
#####
###
#
# ----- MODELOS ARIMA -----
#
#####
###
library("ggplot2")
library("fpp2")
library("fpp")
library("urca")
library("cowplot")
library("readxl")
library("forecast")
library("nortest")

# DATOS trimestrales de desempleo juvenil 2013-2018

#
# ----- SERIE DE TIEMPO DE LAS 24 PROVINCIAS-----

```

```

#
datos1 <- read_excel("SERIES%.xlsx")
name <- read_excel("SERIES.xlsx", sheet = "NOMBRES")
# GRAFICA DE SERIE TEMPORAL DE LAS 24 PROVINCIAS
dats <- ts(datos1[,3:27], start=c(2013, 1), end = c(2016,12), frequency=4)
plot(dats, plot.type="single", lty=1, xlab="Año", ylab="Tasa", xlim=c(2013,2018), cex=2,
      ylim=c(0,45), col=c("tan", "yellow", "blue", "green", "royalblue1", "violet", "burlywood4",
      "chartreuse3", "slategray", "turquoise", "tomato", "slateblue", "thistle", "cornflowerblue",
      "sienna1", "plum2", "seagreen", "palevioletred1", "deepskyblue", "purple", "powderblue", "red",
      "tan4", "peru", "black"))
legend(x=c(2013,2018), y=c(30,45), c("EC-A", "EC-B", "EC-F", "EC-C", "EC-X", "EC-H", "EC-O", "EC-E",
      "EC-G", "EC-I", "EC-L", "EC-R", "EC-M", "EC-S", "EC-N", "EC-Y", "EC-P", "EC-T", "EC-Z", "EC-U",
      "EC-D", "EC-SD", "EC-SE", "EC-ZND", "TASA NACIONAL"), lty= 1, cex=0.8, col=c("tan", "yellow",
      "blue", "green", "royalblue1", "violet", "burlywood4", "chartreuse3", "slategray", "turquoise",
      "tomato", "slateblue", "thistle", "cornflowerblue", "sienna1", "plum2", "seagreen",
      "palevioletred1", "deepskyblue", "purple", "powderblue", "red", "tan4", "peru", "black"),
      ncol=5, x.intersp = 0.03, y.intersp = 0.5, box.col = "white")
#Formando series de tiempo-----
az <- ts(datos1[,3], frequency = 4, start = (2013))
bo <- ts(datos1[,4], frequency = 4, start = (2013))
ca <- ts(datos1[,5], frequency = 4, start = (2013))
car <- ts(datos1[,6], frequency = 4, start = (2013))
co <- ts(datos1[,7], frequency = 4, start = (2013))
ch <- ts(datos1[,8], frequency = 4, start = (2013))
or <- ts(datos1[,9], frequency = 4, start = (2013))
es <- ts(datos1[,10], frequency = 4, start = (2013))
gu <- ts(datos1[,11], frequency = 4, start = (2013))
im <- ts(datos1[,12], frequency = 4, start = (2013))
lo <- ts(datos1[,13], frequency = 4, start = (2013))
ri <- ts(datos1[,14], frequency = 4, start = (2013))
ma <- ts(datos1[,15], frequency = 4, start = (2013))
mo <- ts(datos1[,16], frequency = 4, start = (2013))
na <- ts(datos1[,17], frequency = 4, start = (2013))
pa <- ts(datos1[,18], frequency = 4, start = (2013))
pi <- ts(datos1[,19], frequency = 4, start = (2013))
tu <- ts(datos1[,20], frequency = 4, start = (2013))
zam <- ts(datos1[,21], frequency = 4, start = (2013))
suc <- ts(datos1[,22], frequency = 4, start = (2013))
ore <- ts(datos1[,23], frequency = 4, start = (2013))
sdo <- ts(datos1[3:24,24], frequency = 4, start = (2013))
sel <- ts(datos1[3:24,25], frequency = 4, start = (2013))
zn <- ts(datos1[,26], frequency = 4, start = (2013))
tnac <- ts(datos1[,27], frequency = 4, start = (2013))
#FUNCION DE AUTOCORRELACION SIMPLE Y PARCIAL de las provincias mas grandes del
ecuador -----
win.graph()
par(mfrow=c(2,4))
acf(diff(az), main="AZUAY"); pacf(diff(az), main=""); acf(diff(car), main="CARCHI")
; pacf(diff(car), main="")
acf(diff(co), main="COTOPAXI"); pacf(diff(co), main=""); acf(diff(or), main="EL ORO")
; pacf(diff(or), main="")
acf((pi), main="PICHINCHA"); pacf((pi), main=""); acf((tnac), main="TNDJ"); pacf((tnac), main="")

#TEST DE ESTACIONARIEDAD Dickey-Fuller-----
AZUAY <- adf.test(az, alternative = "stationary", k = 0); AZUAY
BO <- adf.test(bo, alternative = "stationary", k = 0); BO
CA <- adf.test(ca, alternative = "stationary", k = 0); CA

```

```

CAR <- adf.test(car, alternative = "stationary",k = 0);CAR
COT <- adf.test(co, alternative = "stationary",k = 0);COT
CHI <- adf.test(ch, alternative = "stationary",k = 0);CHI
OR <- adf.test(or, alternative = "stationary",k = 0);OR
ES <- adf.test(es, alternative = "stationary",k = 0);ES
GU <- adf.test(gu, alternative = "stationary",k = 0);GU
IM <- adf.test(im, alternative = "stationary",k = 0);IM
LO <- adf.test(lo, alternative = "stationary",k = 0);LO
RI <- adf.test(ri, alternative = "stationary",k = 0);RI
MA <- adf.test(ma, alternative = "stationary",k = 0);MA
MO <- adf.test(mo, alternative = "stationary",k = 0);MO
NAP <- adf.test(na, alternative = "stationary",k = 0);NAP
PA <- adf.test(pa, alternative = "stationary",k = 0);PA
PI <- adf.test(pi, alternative = "stationary",k = 0);PI
TU <- adf.test(tu, alternative = "stationary",k = 0);TU
ZAM <- adf.test(zam, alternative = "stationary",k = 0);ZAM
SUC <- adf.test(suc, alternative = "stationary",k = 0);SUC
ORE <- adf.test(ore, alternative = "stationary",k = 0);ORE
SDO <- adf.test(sdo, alternative = "stationary",k = 0);SDO
SEL <- adf.test(sel, alternative = "stationary",k = 0);SEL
ZN <- adf.test(zn, alternative = "stationary",k = 0);ZN
TNAC <- adf.test(tnac, alternative = "stationary",k = 0);TNAC

#TEST DE ESTACIONARIEDAD Dickey-Fuller SERIE DIFERENCIADA-----
AZDIFF <- adf.test(diff(pi), alternative = "stationary",k = 0);AZDIFF
CARDIFF<- adf.test(diff(car), alternative = "stationary",k = 0);CARDIFF
COTDIFF<- adf.test(diff(co), alternative = "stationary",k = 0);COTDIFF
ORDIFF <- adf.test(diff(or), alternative = "stationary",k = 0);ORDIFF
GUDIFF <- adf.test(diff(gu), alternative = "stationary",k = 0);GUDIFF
PIDIFF <- adf.test(diff(pi), alternative = "stationary",k = 0);PIDIFF
TNACDIFF <- adf.test(diff(tnac), alternative = "stationary",k = 0);TNACDIFF
#FUNCION DE AUTOCORRELACION SIMPLE Y PARCIAL DIFERENCIADAS
par(mfrow=c(4,4))
acf(diff(az));pacf(diff(az),main="");acf(diff(car));pacf(diff(car),main="")
acf(diff(co));pacf(diff(co),main="");acf(diff(or));pacf(diff(or),main="")
acf(diff(gu));pacf(diff(gu),main="");acf(diff(pi));pacf(diff(pi),main="")
acf(diff(tnac),main="Tasa Nacional");pacf(diff(tnac),main="")
#
# SELECCIÓN MODELOS ARIMA AUTOMÁTICOSS----
#
ajuste_a1<- auto.arima((az), stepwise=T, approximation=T); ajuste_a1
ajuste_a2<- auto.arima((bo), stepwise=F, approximation=F); ajuste_a2
ajuste_a3<- auto.arima((ca), stepwise=FALSE, approximation=FALSE); ajuste_a3
ajuste_a4<- auto.arima((car), stepwise=FALSE, approximation=FALSE); ajuste_a4
ajuste_a5<- auto.arima((co), stepwise=F, approximation=T); ajuste_a5
ajuste_a6<- auto.arima((ch), stepwise=FALSE, approximation=FALSE); ajuste_a6
ajuste_a7<- auto.arima((or), stepwise=F, approximation=T); ajuste_a7
ajuste_a8<- auto.arima((es), stepwise=FALSE, approximation=FALSE); ajuste_a8
ajuste_a9<- auto.arima((gu), stepwise=F, approximation=T); ajuste_a9
ajuste_a10<- auto.arima((im), stepwise=FALSE, approximation=FALSE); ajuste_a10
ajuste_a11<- auto.arima((lo), stepwise=FALSE, approximation=FALSE); ajuste_a11
ajuste_a12<- auto.arima((ri), stepwise=FALSE, approximation=FALSE); ajuste_a12
ajuste_a13<- auto.arima((ma), stepwise=FALSE, approximation=FALSE); ajuste_a13
ajuste_a14<- auto.arima((mo), stepwise=FALSE, approximation=FALSE); ajuste_a14
ajuste_a15<- auto.arima((na), stepwise=FALSE, approximation=FALSE); ajuste_a15
ajuste_a16<- auto.arima((pa), stepwise=FALSE, approximation=FALSE); ajuste_a16
ajuste_a17<- auto.arima((pi), stepwise=F, approximation=T); ajuste_a17
ajuste_a18<- auto.arima((tu), stepwise=F, approximation=T); ajuste_a18
ajuste_a19<- auto.arima((zam), stepwise=FALSE, approximation=FALSE); ajuste_a19
ajuste_a20<- auto.arima((suc), stepwise=T, approximation=T); ajuste_a20
ajuste_a21<- auto.arima((ore), stepwise=T, approximation=T); ajuste_a21
ajuste_a22<- auto.arima((sdo), stepwise=FALSE, approximation=FALSE); ajuste_a22
ajuste_a23<- auto.arima((sel), stepwise=T, approximation=T); ajuste_a23
ajuste_a24<- auto.arima((zn), stepwise=T, approximation=T); ajuste_a24
ajuste_a25<- auto.arima((tnac), stepwise=FALSE, approximation=T); ajuste_a25
checkresiduals(ajuste_a8)
#Análisis de los residuos del modelo----
shapiro.test(ajuste_a1$residuals);shapiro.test(ajuste_a2$residuals)

```

```

shapiro.test(ajuste_a3$residuals);shapiro.test(ajuste_a4$residuals)
shapiro.test(ajuste_a5$residuals);shapiro.test(ajuste_a6$residuals)
shapiro.test(ajuste_a7$residuals);shapiro.test(ajuste_a8$residuals)
shapiro.test(ajuste_a9$residuals);shapiro.test(ajuste_a10$residuals)
shapiro.test(ajuste_a11$residuals);shapiro.test(ajuste_a12$residuals)
shapiro.test(ajuste_a13$residuals);shapiro.test(ajuste_a14$residuals)
shapiro.test(ajuste_a15$residuals);shapiro.test(ajuste_a16$residuals)
shapiro.test(ajuste_a17$residuals);shapiro.test(ajuste_a18$residuals)
shapiro.test(ajuste_a19$residuals);shapiro.test(ajuste_a20$residuals)
shapiro.test(ajuste_a21$residuals);shapiro.test(ajuste_a22$residuals)
shapiro.test(ajuste_a23$residuals);shapiro.test(ajuste_a24$residuals)
shapiro.test(ajuste_a25$residuals)
# INDEPENDENCIA (ALEATORIEDAD)
Box.test(ajuste_a1$residuals,type = "Ljung-Box");Box.test(ajuste_a2$residuals,t
ype = "Ljung-Box")
Box.test(ajuste_a3$residuals,type = "Ljung-Box");Box.test(ajuste_a4$residuals,t
ype = "Ljung-Box")
Box.test(ajuste_a5$residuals,type = "Ljung-Box");Box.test(ajuste_a6$residuals,t
ype = "Ljung-Box")
Box.test(ajuste_a7$residuals,type = "Ljung-Box");Box.test(ajuste_a8$residuals,t
ype = "Ljung-Box")
Box.test(ajuste_a9$residuals,type = "Ljung-Box");Box.test(ajuste_a10$residuals,
type = "Ljung-Box")
Box.test(ajuste_a11$residuals,type = "Ljung-Box");Box.test(ajuste_a12$residuals
,type = "Ljung-Box")
Box.test(ajuste_a13$residuals,type = "Ljung-Box");Box.test(ajuste_a14$residuals
,type = "Ljung-Box")
Box.test(ajuste_a15$residuals,type = "Ljung-Box");Box.test(ajuste_a16$residuals
,type = "Ljung-Box")
Box.test(ajuste_a17$residuals,type = "Ljung-Box");Box.test(ajuste_a18$residuals
,type = "Ljung-Box")
Box.test(ajuste_a19$residuals,type = "Ljung-Box");Box.test(ajuste_a20$residuals
,type = "Ljung-Box")
Box.test(ajuste_a21$residuals,type = "Ljung-Box");Box.test(ajuste_a22$residuals
,type = "Ljung-Box")
Box.test(ajuste_a23$residuals,type = "Ljung-Box");Box.test(ajuste_a24$residuals
,type = "Ljung-Box")
Box.test(ajuste_a25$residuals,type = "Ljung-Box")

#Prueba de estacionariedad de Dickey-Fuller para los residuales del modelo sele
ccionado
adf.test(ajuste_a1$residuals,alternative = "stationary",k = 0);adf.test(ajuste_
a2$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a3$residuals,alternative = "stationary",k = 0);adf.test(ajuste_
a4$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a5$residuals,alternative = "stationary",k = 0);adf.test(ajuste_
a6$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a7$residuals,alternative = "stationary",k = 0);adf.test(ajuste_
a8$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a9$residuals,alternative = "stationary",k = 0);adf.test(ajuste_
a10$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a11$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a12$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a13$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a14$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a15$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a16$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a17$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a18$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a19$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a20$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a21$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a22$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a23$residuals,alternative = "stationary",k = 0);adf.test(ajuste
_a24$residuals,alternative = "stationary",k = 0)
adf.test(ajuste_a25$residuals,alternative = "stationary",k = 0)

```

```

#=====PRONOSTICOS=====
=
#*****SERIES QUE CUMPLEN NORMALIDAD*****
#PRONOSTICOS PARA LAS PROVINCIAS MAS GRANDEL DEL ECUADOR
datosp <- read_excel("SERIES%.xlsx")
#AZUAY
az <- ts(datosp[,3], frequency = 4,start = (2013))
ajuste_ap<- auto.arima((az)), stepwise=T, approximation=T); ajuste_ap
pronosticoaz<- (forecast(ajuste_ap,8,level=95))
matproaz <-data.frame(pronosticoaz$mean,pronosticoaz$lower,pronosticoaz$upper)
#El Oro
orp <- ts(datosp[,9], frequency = 4,start = (2013))
ajuste_op<- auto.arima((or)), stepwise=F, approximation=T); ajuste_op
pronosticop<- (forecast(ajuste_op,8,level=95))
matproor <-data.frame(pronosticop$mean,pronosticop$lower,pronosticop$upper)
#GUAYAS
gup <- ts(datosp[,11], frequency = 4,start = (2013))
ajuste_gup<- auto.arima((gu)), stepwise=F, approximation=T); ajuste_gup
pronosticogu<- (forecast(ajuste_gup,8,level=95))
matprogu <-data.frame(pronosticogu$mean,pronosticogu$lower,pronosticogu$upper)
#PICHINCHA
pip <- ts(datosp[,19], frequency = 4,start = (2013))
ajuste_pip<- auto.arima((pi)), stepwise=F, approximation=T); ajuste_pip
pronosticopi<- (forecast(ajuste_pip,8,level=95))
matpropi <-data.frame(pronosticopi$mean,pronosticopi$lower,pronosticopi$upper)
#TUNGURAHUA
tup <- ts(datosp[,20], frequency = 4,start = (2013))
ajuste_tup<- auto.arima((tu)), stepwise=F, approximation=T); ajuste_tup
pronosticotu<- (forecast(ajuste_tup,8,level=95))
matprotu <-data.frame(pronosticotu$mean,pronosticotu$lower,pronosticotu$upper)
#GRAFICO DE LOS PRONOSTICOS
win.graph()
par(mfrow=c(2,2))
plot(pronosticoaz, main="Azuary",ylab="Tasa", xlab="Año")
plot(pronosticop, main="El Oro",ylab="Tasa", xlab="Año")
plot(pronosticotu, main="Tungurahua",ylab="Tasa", xlab="Año")
plot(pronosticogu, main="Guayas",ylab="Tasa", xlab="Año")
plot(pronosticopi, main="Pichincha",ylab="Tasa", xlab="Año")

#*****SERIES QUE NO CUMPLEN NORMALIDAD*****
bootstrapped_series <- bld.mbb.bootstrap(ts((datos1$NACIONAL*100),frequency = 4
,start = (2013)), 1000)
fit <- baggedModel(ts((datos1$NACIONAL*100),frequency = 4,start = (2013),end =
2020),bootstrapped_series)
fcast <- forecast(fit,h=8)
plot(fcast,xlab="Año",ylab="Tasa",main="Pronósticos de la tasa de desempleo juv
enil nacional")
checkresiduals(fcast, lag=28)
auto.arima(fcast$fitted)
shapiro.test(fcast$residuals)
Box.test(fcast$residuals,type = "Ljung-Box")
adf.test(fcast$residuals,alternative = "stationary",k = 0)

#####
#
#          -----          REGRESION LOGIT-----
#
#####

#Cargar la base de datos
data<-read_excel("2013.xlsx")
attach(data)
#*****AÑO 2013*****
#1. Definir las variables dependiente e independientes
y<-cbind(desem) # DESEMPLEO
summary(y)
x <-cbind(Sexo, Edad, Parentesco,Estcivil, Asistclases,Nivelinst, ingrl, rama,g
rup, Seguro )

```

```

summary(x_)

#-----
#2. Estimación de los parámetros para diferentes especificaciones

#Estimacion del modelo LOGIT
logit<-glm(y~x_, family=binomial (link="logit"))
summary(logit)

#-----
#3. ODD ratio
#a. Medio
p_promedio<-mean(predict(logit, type="link"))
odd_medio_logit<-exp(p_promedio)
#odd= 0.020249
#b. Para cada variable
odd_logit<-exp(logit$coefficients)

#-----
#4. Efectos marginales promedio

fdp_logit<-mean(dlogis(predict(logit, type="link")))
em_logit<-fdp_logit*coef(logit)

#-----
#5. Cálculo de las probabilidades pronosticadas

p_gorro_logit<-predict(logit, type="response")

```



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

**DIRECCIÓN DE BIBLIOTECAS Y RECURSOS PARA EL
APRENDIZAJE Y LA INVESTIGACIÓN**



UNIDAD DE PROCESOS TÉCNICOS
REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 31 / 08 /2020

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: Tania Fernanda Tierra Vilema
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Ingeniería en Estadística Informática
Título a optar: Ingeniera en Estadística Informática
f. Analista de Biblioteca responsable: Lic. Luis Caminos Vargas Mgs.

**LUIS
ALBERTO
CAMINOS
VARGAS**

Firmado digitalmente por
LUIS ALBERTO CAMINOS
VARGAS
Nombre de
reconocimiento (DN):
c=EC, l=RIOBAMBA,
serialNumber=060276697
4, cn=LUIS ALBERTO
CAMINOS VARGAS
Fecha: 2020.08.31
15:50:20 -05'00'



0273-DBRAI-UPT-2020