

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
Facultad de Ciencias Naturales y Matemáticas



TRABAJO FINAL DE LA MATERIA INTEGRADORA

**“Determinación de Perfiles para los Ciudadanos de acuerdo a su
Condición Laboral”**

Previa a la obtención del Título de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

Presentado por:

MARÍA GABRIELA LIGUA GUILLÉN

GUAYAQUIL – ECUADOR

AÑO: 2017

AGRADECIMIENTO

Agradezco a Rey de Reyes y Señor de Señores (Dios) porque a lo largo de este camino, me ha dado inteligencia, creatividad y sabiduría para culminar mi carrera universitaria y materia integradora con éxito.

A los pilares fundamentales de mi vida que son: mis padres, mis hermanos y Luisa.

A mi confidente fiel, amigo y novio Gilmar Pivaque, que durante mi etapa universitaria estuvo ayudándome y siempre dando lo mejor de él para que ambos terminemos juntos esta etapa.

A mis profesores M. Sc. Heydi Roa y M. Sc. Sonia Reyes por compartir sus conocimientos durante este proyecto.

A mis compañeros y al Centro Estadístico.

DEDICATORIA

Dedico este proyecto primeramente a Dios porque sin él no lo hubiese logrado, por mostrarme su fidelidad y su misericordia cada día, por haberse fortalecido con su palabra cuando he estado a punto de caer, por haberme amado y sobre todo por ser el padre que siempre soñé.

A mi padre Nelson Ligua, que siempre me ha guiado y día a día se esmera por darme lo mejor y ser mi ejemplo a seguir y a mi madre Priscila Guillén que a pesar de nuestra distancia física siempre ha mostrado su preocupación, amor y guía.

A Luisa por enseñarme valores y cuidarme en todo este tiempo.

A mis primas Diana y Karla.

Mi hermana Samantha Ligua que me inspira ser un ejemplo en su vida y su personalidad alegre nuestras vidas.

A mi novio que me ha motivado para salir adelante, por ser esa mano amiga y día a día demostrarme su amor y paciencia.

Finalmente, a mis amigas de colegio.

DECLARACIÓN EXPRESA

“La responsabilidad del contenido desarrollado en la presente propuesta de la materia integradora corresponde exclusivamente a:

María Gabriela Ligua Guillén.

Y el patrimonio intelectual del mismo a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”.

.....
María Gabriela Ligua Guillén

RESUMEN

El desempleo es un problema social que no afecta solo a ciudadanos sino a una nación entera, por ello es importante realizar constantes estudios para proponer soluciones que puedan atacar a este suceso. Por otro lado, en estos últimos dos años se ha incrementado el subempleo, por lo que si no se plantea una solución a tiempo podría causar grandes estragos a nivel nacional.

Basándose en la Encuesta de Empleo, Desempleo y Subempleo (ENEMDU) en este estudio se desea conocer los factores socioeconómicos y demográficos afectan a la condición laboral (empleo adecuado, subempleo, desempleo, inactivo), por ello los datos del estudio abarca el segundo trimestre junio (2016) considerando las ciudades de Guayaquil, Quito, Cuenca y Ambato.

Para ello se abarca las variables demográficas y socioeconómicas que se considerará en el estudio, y por último los métodos estadísticos que nos ayudará a explicar la condición laboral, entre ellos: árbol de clasificación CHAID, análisis de clases latentes y análisis de correspondencia múltiple.

Respecto a los resultados se obtuvo que las variables demográficas que afectan a la condición laboral son: Género, estado civil, ciudad, edad, etnia, sexo, parentesco, instrucción, área mientras que las variables socio-económicas son: Ocupación, tipo de contrato en el caso que se encuentra laborando, los ingresos del trabajo y per-cápita.

Uno de los principales hallazgos que se encontró en este estudio, es que el nivel de instrucción de una persona influye en la condición laboral es decir que a mayor nivel de instrucción, está propenso a tener un buen empleo entre otros factores que a lo largo de este estudio se detallan.

Palabras Claves: Subempleo, desempleo, empleo adecuado, inactivos, análisis de correspondencia múltiple, análisis de clases latentes, CHAID

ABSTRACT

Unemployment is a social problem that affects not only the citizens but a whole nation, so it is important to carry out constant studies to propose solutions to attack this event. On the other hand, in the last two years, underemployment has increased, so if a solution is not proposed in time, it could cause great damage at the national level.

Based on the Survey of Employment, Unemployment and Underemployment (ENEMDU) in this study it is desired to know the socioeconomic and demographic factors affect the labor condition (adequate employment, underemployment, unemployment, inactive), therefore study data covers the second quarter June (2016) considering the cities of Guayaquil, Quito, Cuenca and Ambato.

To do so, it covers the demographic and socioeconomic variables to be considered in the study, and finally the statistical methods that will help explain the labor condition, including: CHAID classification tree, latent class analysis and multiple correspondence analysis.

Regarding the results it was obtained that the demographic variables that affect the labor condition are: Gender, marital status, city, age, ethnicity, sex, kinship, education, area while the socio-economic variables are: Occupation, type of contract In the case of work, income from work and per capita.

One of the main findings found in this study is that the level of education of a person influences the work condition is to say that at a higher level of education, is likely to have a good job among other factors that throughout This study are detailed.

Keywords: *Underemployment, unemployment, adequate employment, inactive, multiple correspondence analysis, latent class analysis, CHAID*

ÍNDICE GENERAL

RESUMEN	i
<i>ABSTRACT</i>	ii
ÍNDICE GENERAL	iii
ÍNDICE DE TABLAS	vii
ÍNDICE DE GRÁFICOS	ix
ABREVIATURA	xi
CAPÍTULO 1.....	1
1. Introducción	1
1.1 Definición del Problema	1
1.2 Justificación	1
1.3 Antecedentes	1
1.4 Alcance	2
1.5 Objetivos	3
1.5.1 Objetivos Generales.....	3
1.5.2 Objetivos Específicos.....	3
CAPÍTULO 2.....	4
2. Metodología y Materiales	4
2.1 Encuesta de Empleo, Subempleo y Desempleo (ENEMDU)	4
2.1.1 Rotación de la muestra a través del tiempo	5
2.1.2 Definiciones	6
2.2 Variables a Usar	7
2.3 Tablas de Contingencia	8
2.4 Gráficos	9

2.4.1 Diagrama de Cajas.....	9
2.4.2 Histograma	9
2.5 Probabilidad Condicional.....	9
2.6 Árboles de decisión.....	9
2.6.1 Algoritmo CHAID	10
2.6.2 Validación Cruzada K intentos.	11
2.7 Análisis de correspondencia múltiple.	11
2.7.1 Objetivos	12
2.7.2 Método.....	12
2.7.3 Coseno Cuadrado	13
2.7.4 Contribución	13
2.7.5 Matriz Burt.....	14
2.7.6 Inercia.....	14
2.8 Análisis de Clases Latentes	15
2.8.1 Estimación de los parámetros.	17
2.8.2 Elección del mejor modelo	18
CAPÍTULO 3.....	19
3. Resultados	19
3.1 Análisis univariado.	19
3.1.1 Variables Categóricas.....	19
3.1.1.1 Condición Laboral	19
3.1.1.2 Género	19
3.1.1.3 Género	19
3.1.1.4 Ciudad	20

3.1.1.5 Relación Parentesco	20
3.1.1.6 Estado Civil	21
3.1.1.7 Etnia	21
3.1.1.8 Trabajó la semana pasada	21
3.1.1.9 Buscó trabajo el mes anterior	22
3.1.1.10 Ocupación	22
3.1.2 Variables Continuas	23
3.1.2.1 Edad	23
3.1.2.2 Horas trabajadas la semana pasada	24
3.1.2.3 Semanas que busca trabajo	26
3.2 Análisis Bivariado.	28
3.2.1 Área	28
3.2.2 Ciudad	29
3.2.3 Sexo	31
3.2.4 Parentesco	32
3.2.5 Instrucción.....	33
3.2.6 Edad.....	35
3.2.7 Estado Civil.....	36
3.2.8 Etnia.....	37
3.3 Análisis de Empleo Adecuado-Subempleo.	38
3.4 Desempleados	44
3.5 Análisis de Inactivos.....	48
CAPÍTULO 4.....	51
4. Discusión y conclusiones	51

4.1	Conclusiones.....	52
4.2	Recomendaciones	54
	BIBLIOGRAFÍA	55
	APÉNDICE	57
	APÉNDICE A: Código de R de análisis de Correspondencia	57
	APÉNDICE B: Código de R de análisis de clases latentes	58
	APÉNDICE C: Árbol CHAID – Estado Laboral.....	59

ÍNDICE DE TABLAS

Tabla 2.1: Dominios de los sectores mensuales	4
Tabla 2.2: Dominios de los sectores trimestrales	5
Tabla 3.1: Frecuencia de Buscar Trabajo.....	22
Tabla 3.2: Frecuencia de Ocupación	22
Tabla 3.3: Estadísticas Descriptivas de Edad	23
Tabla 3.4: Frecuencia de Edad	23
Tabla 3.5: Estadísticas Descriptivas de Horas trabajadas la semana pasada	24
Tabla 3.6: Frecuencia de Horas Trabajadas.....	25
Tabla 3.7: Estadística Descriptiva de Semanas que busca trabajo	26
Tabla 3.8: Frecuencia de Semana Busca Trabajo.....	26
Tabla 3.9: Análisis Bi-variado Área vs Condición Laboral.....	28
Tabla 3.10: Análisis Bi-variado entre Ciudad y Condición Laboral	29
Tabla 3.11: Análisis Bi-variado entre Sexo y Condición Laboral.....	31
Tabla 3.12: Análisis Bi-variado entre Parentesco y Condición Laboral.....	32
Tabla 3.13: Análisis Bi-variado entre Instrucción y Condición Laboral.....	33
Tabla 3.14: Análisis Bi-variado entre Edad y Condición Laboral.....	35
Tabla 3.15: Análisis Bi-variado Estado Civil y Condición Laboral	36
Tabla 3.16: Análisis Bi-variado entre Etnia y Condición Laboral.....	37
Tabla 3.17: Observados vs Pronosticado.....	38
Tabla 3.18: Dimensiones del ACM – Empleo Adecuado y Subempleo.....	41
Tabla 3.19: Elección de número de clases latentes – BIC	44
Tabla 3.20: Clase-Condicional Ciudad	45
Tabla 3.21: Clase-Condicional Sexo	45
Tabla 3.22: Clase-Condicional Estado Civil	45
Tabla 3.23: Clase-Condicional Nivel de Instrucción	45
Tabla 3.24: Clase-Condicional Ocupación	46

Tabla 3.25: Clase-Condional Semanas que Busca Trabajo	46
Tabla 3.26: Clase-Condional Parentesco	46
Tabla 3.27: Clase-Condional Categoría Edad	46
Tabla 3.28: Clase-Condional Experiencia laboral	46
Tabla 3.29: Estimación de la proporción que comparte cada clase.....	47
Tabla 3.30: Observados vs Pronósticos – Inactivos	48

ÍNDICE DE GRÁFICOS

Gráfico 3.1: Gráfico de Barras de Condición Laboral.....	19
Gráfico 3.2: Gráfico de Barras de Género	19
Gráfico 3.3: Gráfico de Barras de Ciudad	20
Gráfico 3.4: Gráfico de Barras de Relación de Parentesco	20
Gráfico 3.5: Gráfico de Barras de Estado Civil.....	21
Gráfico 3.6: Gráfico de Barras de Etnia	21
Gráfico 3.7: Gráfico de Barras de Trabajó la semana pasada	22
Gráfico 3.8: Gráfico de Barras de Buscar Trabajo	22
Gráfico 3.9: Gráfico de Barras de Ocupación.....	23
Gráfico 3.10: Histograma de Edad	23
Gráfico 3.11: Diagrama de Cajas de Edad.....	24
Gráfico 3.12: Diagrama de Cajas de Horas Trabajadas	25
Gráfico 3.13: Histograma de Horas Trabajadas	25
Gráfico 3.14: Diagrama de Cajas de Busca Trabajo	27
Gráfico 3.15: Histograma de Busca Trabajo	27
Gráfico 3.16: Gráfico de Barras de las Distribución Condicional entre área y condición laboral	28
Gráfico 3.17: Gráfico de Barras de la Distribución condicional entre Ciudad y Condición Laboral.....	29
Gráfico 3.18: Gráfico de Barras de la Distribución Condicional entre Sexo y Condición Laboral.....	31
Gráfico 3.19: Gráfico de Barras de la Distribución Conjunta entre Parentesco y Condición Laboral.....	32
Gráfico 3.20: Gráfico de Barras de la Distribución Condicional entre Instrucción y Condición Laboral	34
Gráfico 3.21: Gráfico de Barras de la Distribución Condicional entre Edad y Condición Laboral.....	35

Gráfico 3.22: Gráfico de Barras de la Distribución Condicional entre Estado Civil – Condición Laboral.....	36
Gráfico 3.23: Gráfico de Barras de la Distribución Condicional entre Etnia – Condición Laboral.....	37
Gráfico 3.24: Árbol de Clasificación (CHAID) – Empleo adecuado y Subempleo.....	40
Gráfico 3.25: Análisis de Correspondencia Múltiple Variables Demográficas y Condición Laboral.....	43
Gráfico 3.26: Árbol de Clasificación (CHAID) de inactivos	50

ABREVIATURA

ACM	Análisis de Correspondencia Múltiple
ENEMDU	Encuesta Nacional de Empleo, Desempleo y Subempleo
INEC	Instituto Nacional de Estadística y Censos
OIT	Organización Internacional del Trabajo
PET	Población en Edad de Trabajar

CAPÍTULO 1

1. INTRODUCCIÓN

1.1 Definición del Problema

Es relevante conocer la situación de empleo, desempleo y subempleo que afronta nuestro país para ello, el Instituto Nacional de Estadísticas y Censos (INEC) por medio de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) permite tener información trimestral.

Tanto en nuestro país como a nivel mundial existe una alta tasa de desempleo, por consiguiente, puede afectar en la salud causando depresión, sentimiento de culpabilidad y desesperación además en nuestra sociedad puede ocasionar discriminación, comportamientos violentos e incluso actos delictivos.

Por otro lado, durante el periodo de septiembre 2015 a junio 2016 se ha incrementado la tasa de subempleados, y disminuido la tasa de personas con empleo por lo que esto causa preocupación a nivel nacional.

1.2 Justificación

La finalidad del presente trabajo es determinar los factores socio-económicos y demográficos que afecta a la condición laboral. A través de métodos estadísticos, se pretende difundir la información que pueda ser usada para intentar contrarrestar el índice de desempleo, subempleo e inactivos.

1.3 Antecedentes

Según la publicación que realizó INEC en los trimestres de marzo y junio (2016), la tasa de empleo adecuado ha disminuido representando en marzo 2016 el 40% y junio 2016 el 41%, en comparación al año 2015 cuya tasa eran de 43 y 46% respectivamente. Por otro lado, el

subempleo ha incrementado, en marzo está en un 16% y junio el 17% mientras, en el 2015 llegaba a 13.4%. Desde marzo 2016, el desempleo incrementó 5,7% en comparación al 2015 que alcanzaba 3,8%.

Basándose en las ciudades de Guayaquil, Quito, Cuenca y Ambato, la tasa de empleo adecuado ha disminuido en Guayaquil, Quito y Ambato mientras que en Cuenca hubo un incremento en el trimestre de marzo 2016 así mismo la tasa de subempleo ha incrementado en las ciudades de Guayaquil, Quito y Ambato donde Guayaquil tiene la mayor tasa de subempleo mientras en Cuenca hubo una disminución en el periodo de marzo 2016, toda esta información se la comparó con los trimestres de marzo y junio 2015. El desempleo en estas ciudades ha incrementado en relación a los dos primeros trimestres 2015.

Por otro lado, de acuerdo al género, la tasa de empleo adecuado y subempleo incrementó en los hombres; y desempleo aumentó en mujeres. Las personas con subempleo en marzo y junio de 2016, ha incrementó en personas que trabajan por cuenta propia.

1.4 Alcance

Los datos de este estudio fueron obtenidos por la Encuesta Nacional de Empleo, Desempleo y Subempleo propuesta por el INEC donde el número de observaciones a investigar son 10319 considerando la Población en Edad de Trabajar (PET) donde este grupo comprenden a las personas de 15 años en adelante y a su vez abarca las ciudades de Guayaquil, Quito, Cuenca y Ambato durante el segundo trimestre del 2016.

1.5 Objetivos

1.5.1 Objetivos Generales

Determinar las variables socio-económicas y demográficas que influyen en la condición laboral.

1.5.2 Objetivos Específicos

1. Determinar las variables que inciden en la condición laboral.
2. Identificar las principales características socio-económicas y demográficas de las personas con condición laboral de inactividad.
3. Encontrar las variables socio-económicas y demográficas que repercuten en las siguientes condiciones laborales: empleo adecuado y subempleo.
4. Identificar comportamientos de las personas sin empleo.

CAPÍTULO 2

2. METODOLOGÍA Y MATERIALES

2.1 Encuesta de Empleo, Subempleo y Desempleo (ENEMDU)

La Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) tiene como fin proporcionar información actualizada de la población económicamente activa además de sus perfiles demográficos, sociales y económicos y al mismo tiempo difundir esta información relacionada a la condición de desempleo, empleo y subempleo.

El diseño muestral que usa ENEMDU es probabilístico-bietápico donde la unidad primaria de selección es el sector censal (UPM) mientras la secundaria son las viviendas seleccionadas dentro de cada UPM.

El INEC utiliza una sub-muestra del marco maestro y el esquema de rotación de viviendas fijado para las encuestas trimestrales de ENEMDU que se puso en práctica desde septiembre 2003. Adicionalmente las mejoras se plantean en el momento de recaudar información por lo que creó dominios tanto trimestrales, de forma continua y permanente cada mes del año.

La población objetivo de este estudio son los hogares particulares dentro de Ecuador incluidas las zonas “No delimitadas”. Esta encuesta excluye a las viviendas colectivas, flotantes y sectores con población indígena. Desde marzo 2014, la muestra para ENEMDU mensual es de 471 sectores que son:

Tabla 2.1: Dominios de los sectores mensuales

Dominios	Sectores
Quito	113
Guayaquil	125
Cuenca	75
Machala	66
Ambato	92
Total	471

Fuente: Instituto Nacional de Estadísticas y Censo - INEC

Mientras que la encuesta trimestral es de 1379 sectores muestrales:

Tabla 2.2: Dominios de los sectores trimestrales

Dominios	Sectores
Quito	113
Guayaquil	125
Cuenca	75
Machala	66
Ambato	92
Resto Sierra Urbana	141
Resto Costa Urbana	148
Amazonía Urbana	148
Sierra Rural	227
Costa Rural	177
Amazonía Rural	74
Total	1379

Fuente: Instituto Nacional de Estadísticas y Censo - INEC

Con respecto a esto, se puede analizar el comportamiento del mercado laboral a través del tiempo en dominios similares con encuestas trimestrales y mensuales.

2.1.1 Rotación de la muestra a través del tiempo

ENEMDU tiene periodicidad trimestral y mensual por eso, es importante determinar como la muestra puede ser sustituida de un mes a otro. Un objetivo de la encuesta trimestral es medir cambios de empleos, desempleos entre trimestres y meses. Por ello, no es deseable mantener las mismas viviendas en las muestras por varios años, porque la tasa de respuesta disminuye y aquellas respuestas pueden estar sesgadas.

Para ello, se recomienda la rotación recomendada por la Organización Internacional del Trabajo (OIT) para encuestas trimestrales, llamadas 2-2-2; bajo esta rotación un panel de viviendas seleccionadas en una muestra, el 25% se mantiene en la muestra por dos trimestres seguidos después ya no se las considera durante 2 trimestres y nuevamente se las retoma en los 2 últimos trimestres.

ENEMDU trimestral seguirá usando este tipo de rotación, se considera la rotación cada 2 años. La última rotación se efectuó en septiembre 2013.

2.1.2 Definiciones

Población Económicamente Inactiva (PEI)

Este grupo es conformado por aquellas personas con edades mayores de 15 años que no están empleadas, no desean tener trabajo y no están disponibles para trabajar. Este grupo lo conforman los jubilados, estudiantes, ama de casa, incapacitados entre otros.

Población Económicamente Activa (PEA)

Son aquellas personas con edades superior a 15 años, que han trabajado al menos una hora. En este grupo se encuentra el empleo adecuado, subempleo y desempleo.

Empleo adecuado

Personas con empleo que durante la semana generan ingresos mayores o iguales al salario básico y que el número de horas trabajadas a la semana son mayor o igual a 40 horas independiente del deseo y disponibilidad de trabajar horas extras.

Subempleo

Es aquel grupo que tienen empleo, pero sus ingresos como mucho llega al sueldo básico y/o trabajaron menos de la jornada propuesta por el gobierno y tienen deseo y disponibilidad de trabajar horas extras.

Desempleo

Se considera a las personas que tienen edades superiores a 15 años que no estuvieron empleadas. Se incluye en este grupo a las personas que han realizado gestiones hace un mes para encontrar trabajo o emprender algún negocio.

2.2 Variables a Usar

Sección 1: Información General

- Actualmente asiste a clases: Si y No.
- Nivel de Instrucción: Ninguna, Primaria, Secundaria, Superior y Postgrado.
- Área: Urbana, Rural.
- Sexo. Femenino, Masculino.
- Etnia.
- Estado Civil: Soltero, Casado o unido, Separado o Divorciado y Viudo.
- Relación con el Jefe de hogar: Jefe y No Jefe donde esta categoría abarca conyugue, hijo o hija, yerno o nuera, empleadas doméstica y otros parientes.

Sección 2: Características Ocupacionales

Las siguientes preguntas que se presenta a continuación sirven de criterio para clasificar a una persona de acuerdo a su condición ocupado, desocupado o inactivo según el INEC.

- ¿Usted trabajó la semana pasada? Si, No.
- Número de horas que trabajó la semana pasada.
- En el caso que no haya trabajo la semana pasada ¿Usted realizó una gestión para buscar trabajo? Si, No.
- ¿Aunque no haya trabajado la semana pasada tiene algún trabajo o negocio al cual seguro va a volver?
- ¿Cuál es la razón por la cual no busca trabajo?, en el caso de que en esta pregunta una persona contestó que no tiene deseos o necesidades de trabajar, no tiene tiempo, su conyugue o familia no le permite, está enfermo o discapacitado y no está en edad de trabajar se le procede a clasificar según el criterio que tiene INEC para las personas inactivas.

Para los desempleados se considera las siguientes preguntas:

- ¿Ha trabajado anteriormente? Si, No
- Número de semanas que busca trabajo.

Sección 2: Ocupación Principal

- Categoría de Ocupación: Empleado en relación dependencia, cuenta propia, jornalero o peón, patrono, empleada doméstica.
- Contrato de Trabajo: Por horas, por jornal, por obra, contrato temporal, contrato permanente, nombramiento.

Sección 3: Ingresos

- Ingresos del trabajo.
- Ingresos per-cápita.

2.3 Tablas de Contingencia

Esta metodología permite analizar la relación de dependencia e independencia entre dos variables cualitativas

Para aplicar esta prueba se debe construir la siguiente tabla:

n: Número total de observaciones

f: Número de filas

c: Número de columnas

f_i: Total de resultados en la fila *i*

c_j: Total de resultados en la columna *j*

n_{ij}: Total de resultados observados en la fila *i* y columna *j*

Donde se plantea la siguiente hipótesis:

H₀: Entre *V₁* y *V₂* son independientes

Vs

H_a: Entre *V₁* y *V₂* no son independientes

Donde el estadístico de prueba es:

$$\chi^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (2.1)$$

Tiene una distribución Chi-Cuadrado con $\nu = (f - 1)(c - 1)$

Para obtener e_{ij} considerando las marginales de cada fila y columna.

$$e_{ij} = \frac{f_i * c_j}{n} \quad (2.2)$$

Con un nivel de significancia α se define la región de rechazo de H_0 :

$$\chi^2 > \chi_{\alpha}^2$$

Para poder aplicar esta prueba debe considerarse que la frecuencia esperada e_{ij} debe ser mayor o igual a 5.

2.4 Gráficos

2.4.1 Diagrama de Cajas

Se describe gráficamente el rango de los datos es decir, el máximo y el mínimo, rango intercuartílico representado por Q3 y Q1, datos atípicos y ubicación de los cuartiles.

2.4.2 Histograma

Representa gráficamente la distribución de frecuencia de los datos donde se dibujan rectángulos cuya base corresponde a cada intervalo y la altura a la frecuencia donde esta altura puede ser la frecuencia o la frecuencia relativa.

2.5 Probabilidad Condicional

Sea A y B eventos del espacio muestral, la probabilidad condicional del evento A dado B se lo expresa de la siguiente manera:

$$P(A|B) = \frac{P(A \cup B)}{P(B)} \quad P(B) \neq 0 \quad (2.3)$$

2.6 Árboles de decisión

Comenzando por la época de los setenta, los investigadores desarrollaron procedimientos para generar árboles de decisión como: AID (Sonquist-1971); MAID (Gillo-1972); THAID (Morgan and Messenger-1973) y CHAID (Kass-1980). Actualmente esta es una técnica de Minería de Datos.

2.6.1 Algoritmo CHAID

La metodología CHAID “Chi-Squared Automatic Interaction Detection” este algoritmo es útil para describir un conjunto de datos, predecir comportamientos, etc. Debe considerarse que las variables deben ser categóricas y en el caso de que en el estudio haya variables continuas, se debe categorizar.

CHAID consiste en realizar dos o más divisiones a partir de una variable independiente que mejor explique a la variable dependiente, para ello, utiliza el criterio del valor p de la distribución Chi-Cuadrado de modo que cada nodo secundario está conformado por un grupo de valores homogéneos de la variable seleccionada. Se debe considerar que no se realiza ninguna división si el valor p ajustado de la variable independiente que mejor explique a la variable objetivo o dependiente no es menor que cierto criterio de división.

Valor p ajustado más conocido como el ajuste de Bonferroni consiste en ajustar el nivel de significancia en relación al número de pruebas estadísticas realizadas al mismo tiempo en los datos.

Este algoritmo se detiene cuando se cumple las siguientes condiciones:

- Se alcanza la profundidad máxima del árbol.
- No satisface con el número mínimo de caso en un nodo para ser padre.
- No cumple con el número de casos para ser un nodo secundario o hijo.

La ejecución de este algoritmo abarca tres pasos:

Fusión: Considera la variable independiente y une las categorías no significativas. Para cada categoría de la última variable independiente resultará en un nodo hijo si esa variable es utilizada para dividir el nodo. Este paso calcula el p ajustado.

División: La mejor división para cada predictor es usada en el paso anterior, donde este paso selecciona el mejor predictor para dividir el nodo.

Detención: Comprueba si el árbol puede seguir dividiéndose o no.

2.6.2 Validación Cruzada K intentos.

Para conocer si el árbol que estamos usando es bueno o no, usaremos una de las técnicas de validación como la validación cruzada que consiste en dividir el conjunto de datos en K conjuntos disjuntos de igual tamaño es decir, sub-muestras donde una parte de los datos es conformada por los datos de pruebas mientras el restante de entrenamientos y este se van repitiendo en k intentos. Como ventaja, cada conjunto de datos de prueba que se realice, es independiente del otro.

Finalmente se realiza el promedio de los riesgos de los resultados de todos los árboles para obtener un único riesgo. El número de intentos que se propone es de 10 intentos.

Esta prueba es precisa dado por los k intentos, pero las desventajas que presenta son la complejidad computacional y la otra es que los datos de entrenamiento pueden sobreponerse perjudicando a la estimación de la cantidad de variación que tendría si cada conjunto de entrenamiento fuese completamente independiente del otro.

2.7 Análisis de correspondencia múltiple.

Es una técnica multivariada de reducción de dimensiones entre categorías de datos nominales mediante la medida de asociación Chi-Cuadrado.

2.7.1 Objetivos

El análisis de correspondencia tiene como objetivos:

- Asociaciones entre categorías de columnas o filas.
- Asociación entre categorías de filas y columnas es decir que buscar la relación entre ambas categorías.

2.7.2 Método

El método que se usa en este proyecto es el de análisis de correspondencia múltiple que consiste en K variables nominales y cada variable nominal tiene J_k niveles donde la suma de J_k es igual a J . La matriz X tiene dimensión $I \times J$. AL momento de realizar el análisis de correspondencia en la matriz de indicadores, proporciona dos conjuntos de puntuaciones de factores: uno para filas y otro para columnas.

Donde el gran total de la tabla es denotado por N y el cálculo de la matriz de probabilidades es realiza a través de:

$$Z = N^{-1}X \quad (2.4)$$

Se denota \mathbf{r} , el vector del total de filas de Z y \mathbf{c} , el vector del total de columnas y $D_c = \text{diag}\{c\}$ y $D_r = \text{diag}\{r\}$

Las puntuaciones de los factores son obtenidas a partir de la siguiente descomposición de valor singular.

$$D_r^{-\frac{1}{2}}(Z - \mathbf{r}\mathbf{c}^T)D_c^{-\frac{1}{2}} = P\Delta Q^T \quad (2.5)$$

Δ : Es la matriz diagonal de los valores singulares¹ y $\Lambda = \Delta^2$ es la matriz de los valores propios.

Las puntuaciones de los factores de las filas son obtenidas por:

$$F = D_r^{-\frac{1}{2}}P\Delta \quad (2.6)$$

¹ Valor singular es la $\sqrt{\lambda_i}$ donde λ_i es el valor propio.

Las puntuaciones de los factores de las columnas son obtenidas por:

$$G = D_c^{-\frac{1}{2}} Q \Delta \quad (2.7)$$

La distancia Chi-cuadrado para las filas y columnas de sus respectivos baricentros² son obtenidas por las siguientes expresiones:

$$d_r = \text{diag}\{FF^T\} \quad (2.8)$$

$$d_c = \text{diag}\{GG^T\} \quad (2.9)$$

2.7.3 Coseno Cuadrado

El coseno al cuadrado de la fila i y el factor l es obtenido por:

$$O_{i,l} = \frac{f_{i,l}^2}{d_{r,l}^2} \quad (2.10)$$

El coseno al cuadrado de la columna j y el factor l es obtenido por:

$$O_{j,l} = \frac{g_{j,l}^2}{d_{j,l}^2} \quad (2.11)$$

Para las denotaciones de $d_{j,l}^2$ y $d_{i,l}^2$ son respectivamente el j -th elemento de d_c y el i -th elemento de d_r .

El coseno cuadrático ayuda a localizar los factores importantes para cada observación o variable dada.

2.7.4 Contribución

La contribución de la fila i y el factor l es obtenido por:

$$t_{i,l} = \frac{f_{i,l}^2}{\lambda_l} \quad (2.12)$$

La contribución de la columna j y el factor l es obtenido por:

$$t_{j,l} = \frac{g_{j,l}^2}{\lambda_l} \quad (2.13)$$

² Llamado también centroide es el punto de intersección de las medianas de dicha figura geométrica.

Donde dicha contribución ayuda a diagnosticar las observaciones y variables más importantes para dicho factor.

2.7.5 Matriz Burt

Se denomina matriz Burt a la matriz de dimensión $J \times J$ expresada por la siguiente ecuación:

$$B = X^T X \quad (2.14)$$

Tanto la matriz de Burt y el análisis de X dan los mismos factores aunque computacionalmente el análisis de X es más fácil.

Dicha matriz toma un papel teórico importante porque los valores propios obtenidos en este análisis dan una mejor aproximación de la inercia explicada de los factores en comparación a los valores propios de X .

2.7.6 Inercia

Si se considera hallar la inercia con la expresión matemática de análisis de correspondencia simple, tiene como consecuencia que la inercia está artificialmente inflada y por lo tanto, el porcentaje de inercia explicado por la primera dimensión está subestimado.

Para ello se plantea una expresión de los valores propios corregidos donde se obtiene de la siguiente manera:

$$\lambda_c^l = \begin{cases} \left(\frac{K}{K-1}\right) \left(\lambda_l - \frac{1}{K}\right)^2 & \lambda_l > \frac{1}{K} \\ 0 & \lambda_l \leq \frac{1}{K} \end{cases} \quad (2.15)$$

Para una mejor estimación de la inercia, se usará la expresión (2.16) propuesta por Greenacre (1993) donde este sugiere evaluar el porcentaje de inercia relativa al promedio de la inercia sin considerar

la diagonal de los bloques de la matriz de Burt. Este promedio de la inercia es denotado por.

$$\bar{\xi} = \frac{K}{K-1} \times \left(\sum_l \lambda_l^2 - \frac{J-K}{K^2} \right) \quad (2.16)$$

Acorde a esta aproximación, el porcentaje de la inercia se expresa como:

$$\tau_c = \frac{\lambda_c}{\bar{\xi}} \quad (2.17)$$

2.8 Análisis de Clases Latentes

Basándose en la publicación Linzer y Lewis (2011), definieron que el análisis de clases latentes es una técnica estadística del análisis multivariado de datos categóricos. Cuando los datos observados toman una serie de respuestas categóricas como por ejemplo, una encuesta de opinión pública, consumo de un producto, datos de votaciones a nivel individual entre otros enfoques de estudio. A menudo es de interés investigar fuentes de confusión entre las variables observadas, identificar y caracterizar grupo de clases similares y aproximar la distribución de observaciones a través de muchas variables de interés. Dicha metodología es útil para lograr estos objetivos.

El modelo básico de análisis de clases latentes es un modelo de mezcla finita, que supone que las distribuciones de componentes son tablas de clasificación cruzada de múltiples vías con todas las variables mutuamente independientes. Este modelo fue propuesto por Lazarsfeld (1950) bajo el nombre de análisis de estructuras latentes mientras que Agresti (2002) encuentra la conexión entre modelo de clases latentes y modelos de mezcla finita.

Considerando J variables categóricas (*Las variables manifiestas*), cada una de las cuales contiene K_j posibles resultados, para individuos $i = 1, \dots, N$. Las *variables manifiestas* podrían tener diferentes números de respuesta denotado por j .

Denotado como Y_{ijk} los valores observados de las J variables manifiestas tal que $Y_{ijk} = 1$ si el individuo i da la k th respuesta a la j th variable, y $Y_{ijk} = 0$ si no, donde $j = 1, \dots, J$ y $k = 1, \dots, K_j$.

Sea π_{jrk} denota la probabilidad clase-condicional donde un individuo en la clase $r = 1, \dots, R$ (donde R se fija en el modelo), producido por k th resultado de la j th variable. Dentro de cada clase para cada variable manifiesta, se cumple $\sum_{k=1}^{K_j} \pi_{jrk} = 1$.

Por otra parte, denota p_r las proporciones de mezcla de R que consiste en proporcionar los pesos en la suma ponderada de las tablas de los componentes con $\sum_r p_r = 1$. Donde los valores p_r son tomados como referencias de la probabilidad "Priori" de pertenencia a una clase latente y esta representa la probabilidad incondicional, que un individuo podría pertenecer a cada clase antes de tomar en cuenta las respuestas, Y_{ijk} provienen de las variables manifiestas.

La función de densidad de probabilidad en todas las clases es:

$$P(Y_i|\pi, p) = \sum_{r=1}^R p_r \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \quad (2.18)$$

Donde

$$f(Y_i, \widehat{\pi}_r) = \prod_{j=1}^J \prod_{k=1}^{K_j} (\widehat{\pi}_{jrk})^{Y_{ijk}} \quad (2.19)$$

La estimación de los parámetros por el modelo de clase latente es p_r y π_{jrk} .

Dada las estimaciones de \widehat{p}_r y $\widehat{\pi}_{jrk}$ de p_r y π_{jrk} respectivamente. La probabilidad posterior de cada individuo que pertenece a cada clase

condicionada a los valores observados de las variables manifiestas donde se la puede obtener por la fórmula de Bayes.

$$\hat{P}(r_i|Y_i) = \frac{\hat{p}_r f(Y_i; \hat{\pi}_r)}{\sum_{q=1}^R \hat{p}_q f(Y_i; \hat{\pi}_q)} \quad (2.20)$$

Observación

Es importante tener en cuenta el número de estimaciones de los parámetros independientes ya que esta metodología incrementa rápidamente con R , J y K_j .

Dado estos valores, el número de parámetros es $R \sum_j (K_j - 1) + (R - 1)$. Si este número excede el número total de observaciones o una menos que el número total de celdas en la tabla de clasificación cruzada de las *variables manifiestas*, entonces el modelo de clase latente está sin identificar.

2.8.1 Estimación de los parámetros.

Para los estimadores de los modelos de clases latentes, se considerará la función de máxima verosimilitud.

$$\ln L = \sum_{i=1}^N \ln \sum_{r=1}^R p_r \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}} \quad (2.21)$$

Con respecto a p_r y π_{jrk} , se usa el algoritmo de expectativa-maximización (EM) Dempster, Laird y Rubin (1977).

El algoritmo EM es un procedimiento interactivo. Comienza con unos valores iniciales arbitrarios de \hat{p}_r y $\hat{\pi}_{jrk}$ y los etiqueta como \hat{p}_r^{old} y $\hat{\pi}_{jrk}^{old}$.

En el paso de maximización, actualiza los parámetros estimados por la función de máxima verosimilitud dando estos posteriores $\hat{P}(r_i|Y_i)$ con:

$$\hat{p}_r^{new} = \frac{1}{N} \sum_{i=1}^N \hat{P}(r_i|Y_i) \quad (2.22)$$

Como las nuevas probabilidades priori y $\hat{\pi}_{jr}^{new}$ como las nuevas probabilidades condicionales de resultado de clase:

$$\hat{\pi}_{jr}^{new} = \frac{\sum_{i=1}^N Y_{ij} \hat{P}(r_i|Y_i)}{\sum_{i=1}^N \hat{P}(r_i|Y_i)} \quad (2.23)$$

El algoritmo repite estos pasos, actualizando los valores hasta que la log-verosimilitud general alcanza un máximo y deja de incrementar más allá de un valor arbitrariamente pequeño.

2.8.2 Elección del mejor modelo

El análisis de clases latente es beneficiosa para el análisis de cluster, para ello es importante seleccionar el mejor modelo y determinar el número apropiado de clases latentes para un conjunto de datos dado. Se puede comenzar por aplicar un modelo completo de independencia con $R=1$ y después ir incrementando hasta alcanzar el nivel adecuado.

Para ello, hay dos criterios de elección de la mejor modelo que son: AIC (Akaike Information Criterion) propuesto por Akaike (1973) y BIC (Bayesian Information Criterion) planteado por Schwartz (1978). Se considera el modelo adecuado aquel que tenga el menor valor en ambos criterios o que tenga el menor valor en el AIC o BIC.

AIC

$$AIC = -2\Lambda + 2\Phi \quad (2.24)$$

BIC

$$BIC = -2\Lambda + \Phi \ln N \quad (2.25)$$

Λ : El logaritmo de máxima-verosimilitud del modelo.

Φ : El número total de parámetros a estimar.

CAPÍTULO 3

3. RESULTADOS

3.1 Análisis univariado.

3.1.1 Variables Categóricas

3.1.1.1 Condición Laboral

A modo de interpretación, respecto a las personas idóneas para trabajar, el 45,3% tiene un empleo adecuado, inactivos con un 41,2%, subempleo con un 9,2% y por último, los desempleados con un 4,3%.

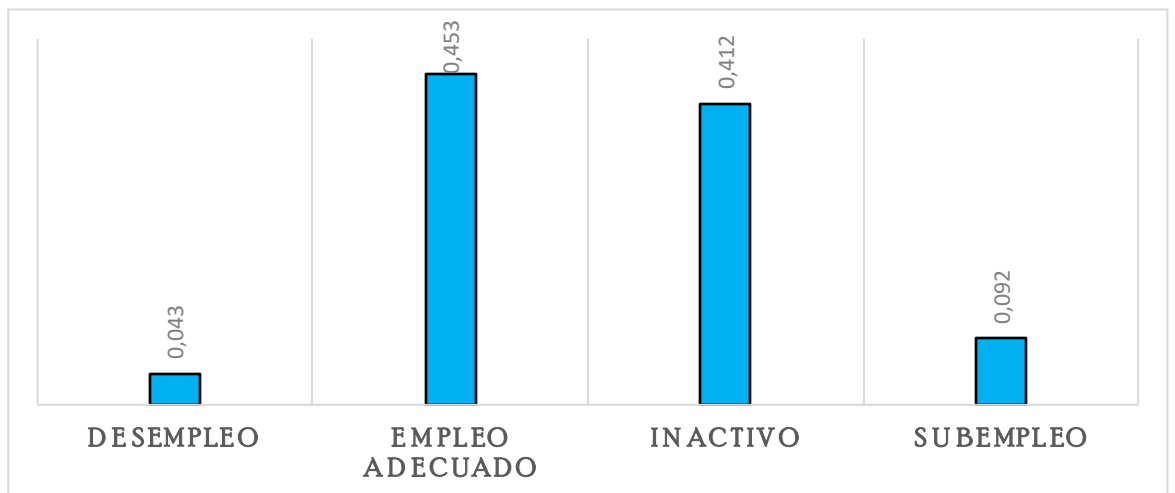


Gráfico 3.1: Gráfico de Barras de Condición Laboral.

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.3 Género

En la Gráfica 3.2, se observa que en la población apta para trabajar el 52% son mujeres y por otro lado, el 48% son hombres.

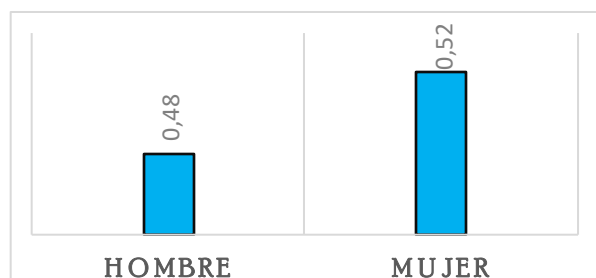


Gráfico 3.2: Gráfico de Barras de Género

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.4 Ciudad

En el caso de la Gráfico 3.3, La ciudad de Guayaquil tiene mayor número de personas disponibles para laborar representando el 32% por otro lado, el 27% pertenecen a la ciudad de Quito.

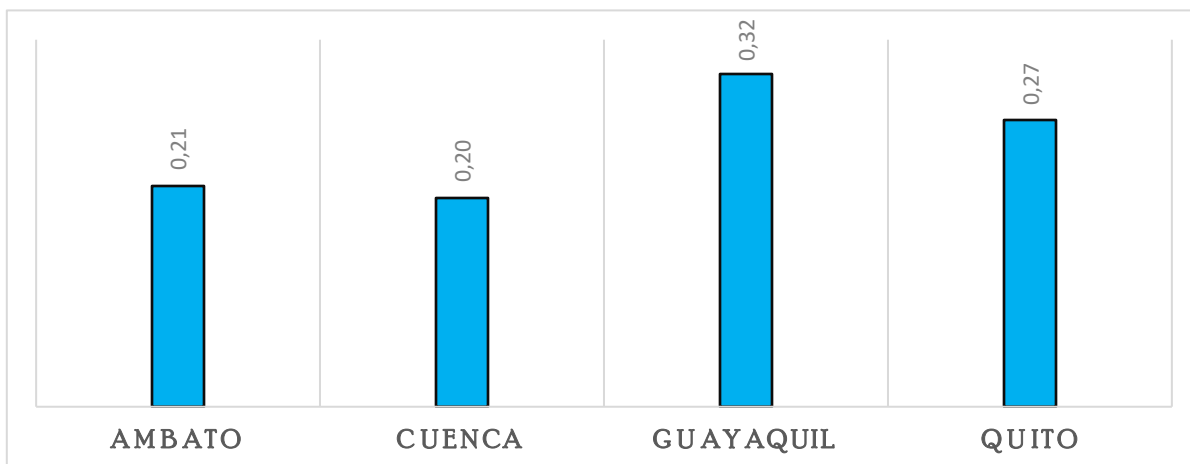


Gráfico 3.3: Gráfico de Barras de Ciudad

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.5 Relación Parentesco

Relación de parentesco con el jefe de hogar como resultado, el 39% son jefes de hogar en cambio los que no son jefes está conformados por conyugue, hijo o hija, yerno o nuera, nieto y padres con un 61%. (Ver Gráfico 3.4)

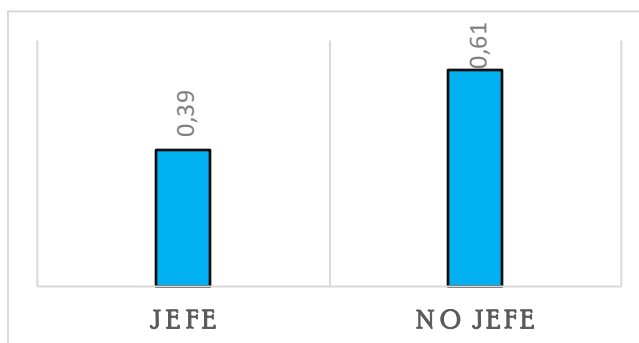


Gráfico 3.4: Gráfico de Barras de Relación de Parentesco

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.6 Estado Civil

Enfocándonos en el estado civil, el 53% de las personas están casados o unidos seguido de las personas solteras con un 34%.

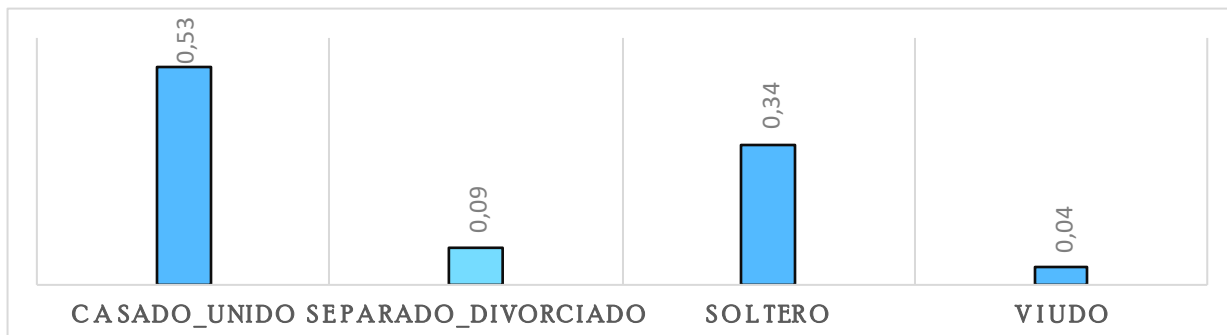


Gráfico 3.5: Gráfico de Barras de Estado Civil

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.7 Etnia

Con respecto a la etnia, el 93% de las personas se considera mestizo.

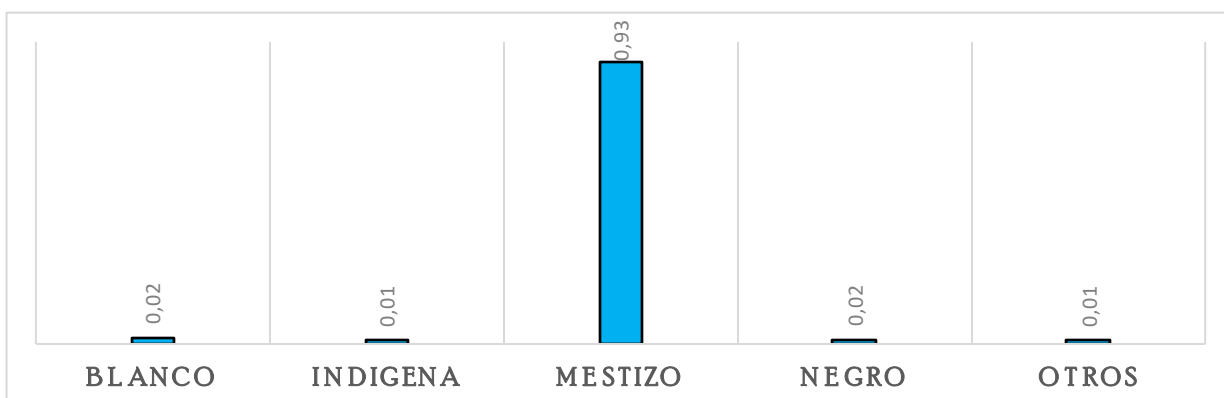


Gráfico 3.6: Gráfico de Barras de Etnia

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.8 Trabajó la semana pasada

En la Gráfico 3.7, se concluye que el 53% sí, laboró la semana pasada frente a un 47% que no trabajó.

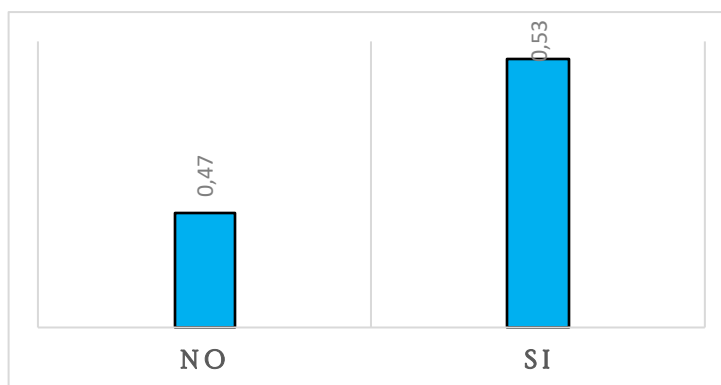


Gráfico 3.7: Gráfico de Barras de Trabajo la semana pasada

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.9 Buscó trabajo el mes anterior

Considerando el caso de las personas que respondieron no en la pregunta si laboraron la semana pasada, como resultado el 91% no buscó trabajo sin embargo, el 9% realizaron gestiones para obtener un empleo.

Tabla 3.1: Frecuencia de Buscar Trabajo

Buscó Trabajo	Frecuencia	Frecuencia Relativa
Si	401	0,09
No	4296	0,91
Total	4697	1

Fuente: Encuesta Nacional de Empleo, Desempleo y

Subempleo, 2016.

Elaboración Propia.

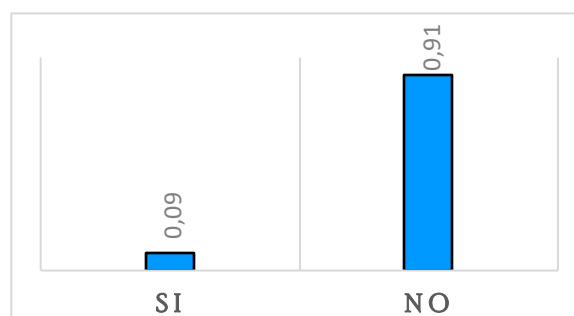


Gráfico 3.8: Gráfico de Barras de Buscar Trabajo

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.1.1.10 Ocupación

Tabla 3.2: Frecuencia de Ocupación

Ocupación	Frecuencia	Frecuencia Relativa
Cuenta Propia	1408	0,24
Empleada domestica	193	0,03
Empleado R.D	3872	0,65
Jornalero o Peon	159	0,03
Patrono	303	0,05
Total	5935	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

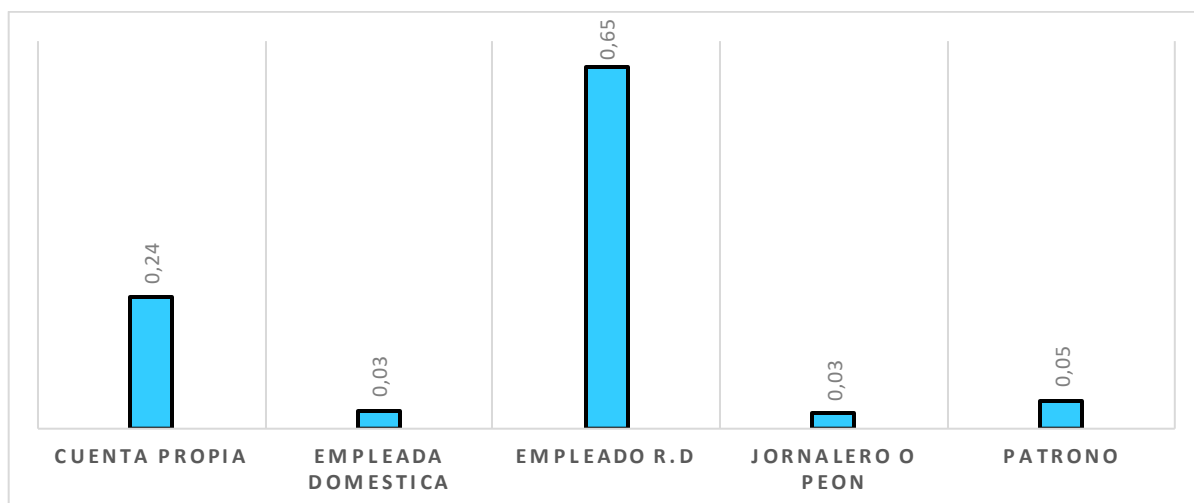


Gráfico 3.9: Gráfico de Barras de Ocupación

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Entre las ocupaciones de las personas activas se puede observar que el 64% de ellos son empleados de relación dependiente por otro lado, el 23% trabajan por cuenta propia. Ver Gráfico 3.9 y Tabla 3.2.

3.1.2 Variables Continuas

3.1.2.1 Edad

Tabla 3.3: Estadísticas Descriptivas de Edad

Media	Mediana	Moda	Desviación estándar	Mínimo	Máximo
40,12	37	15	18,29	15	98

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.4: Frecuencia de Edad

Edad	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
[15-25)	2510	0,24	0,24
[25-35)	2103	0,20	0,45
[35-45)	1873	0,18	0,63
[45-55)	1561	0,15	0,78
[55-65)	1044	0,10	0,88
[65-75)	642	0,06	0,94
[75-85)	427	0,04	0,98
[85-99)	159	0,02	1
Total	10319		

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

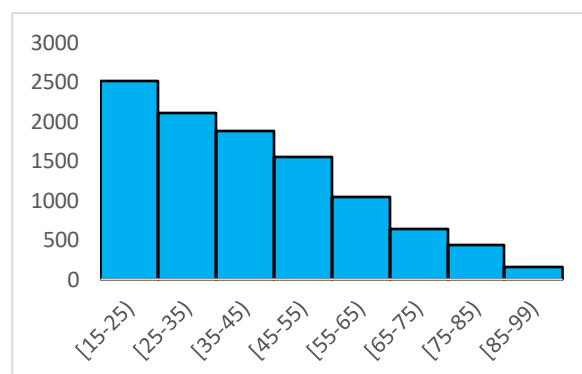


Gráfico 3.10: Histograma de Edad

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

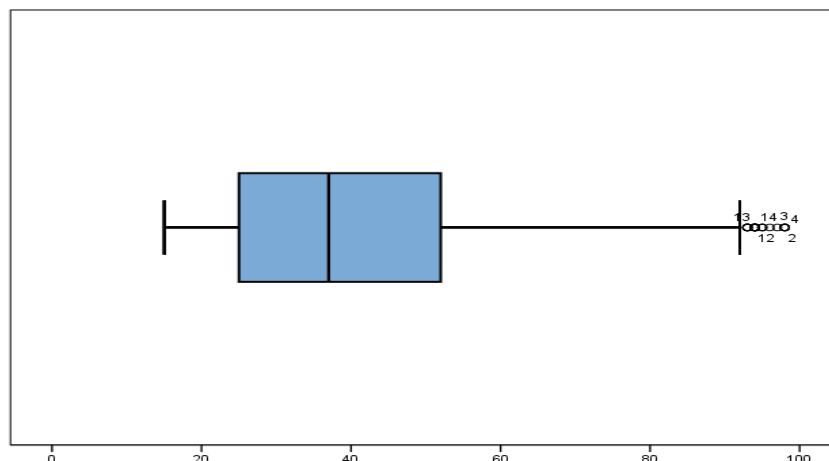


Gráfico 3.11: Diagrama de Cajas de Edad

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Con respecto a la Tabla 3.3, la edad promedio de las personas que están aptas para laborar está entre $40,12 \pm 18,29$ años. En este estudio, la edad mínima es de 15 mientras que la máxima es de 98 años. Podemos decir que 50% de las personas tienen a lo mucho 37 años. La edad más frecuente es de 15 años.

Tanto en la Tabla 3.4 como en la Gráfico 3.10, la edad con mayor frecuencia en este grupo está entre 15 a 24 años con un 24%.

Por último, en la Gráfico 3.11 se presencian datos aberrantes hay que mencionar además que el 75% de las personas tienen como mucho 52 años.

3.1.2.2 Horas trabajadas la semana pasada

Tabla 3.5: Estadísticas Descriptivas de Horas trabajadas la semana pasada

Media	Mediana	Moda	Desviación estándar	Mínimo	Máximo
40,48	40	40	12,67	2	90

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.6: Frecuencia de Horas Trabajadas

Horas Trabajadas		Frecuencia	Frecuencia Relativa	Frecuencia R. Acumulada
[2-13)		224	0,04	0,04
[13-24)		377	0,07	0,11
[24-35)		373	0,07	0,18
[35-46)		3274	0,58	0,76
[46-57)		825	0,15	0,90
[57-68)		386	0,07	0,97
[68-79)	145	0,03	0,997	
[79-91)	18	0,003	1	
Total	5622	1		

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

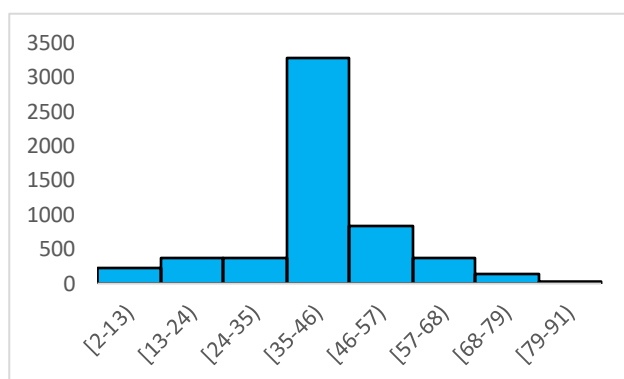


Gráfico 3.13: Histograma de Horas Trabajadas

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

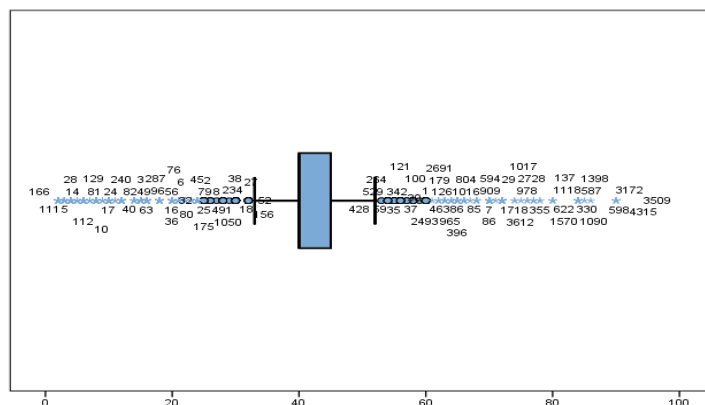


Gráfico 3.12: Diagrama de Cajas de Horas Trabajadas

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Conforme a la pregunta que se planteó si laboró la semana pasada, se desea estudiar el número de horas la semana es por esto, que la Tabla 3.5 representa las horas trabajadas cuyo promedio está entre $40,48 \pm 12,67$ no obstante, hubo personas que laboraron solo dos horas mientras el máximo de horas que trabajaron fueron 90. La mitad de las personas trabajaron como máximo 40 horas laborales. La mayoría de las personas laboran las 40 horas establecidas por el gobierno.

A continuación, en la Tabla 3.6 como en la Gráfico 3.13, el 58% de las personas que trabajaron la semana pasada laboraron entre 35 a 45 horas.

Finalmente, en la Gráfico 3.12, el 75% de las personas como máximo labora 45 horas semanales sin embargo, en este gráfico se observan numerosos datos aberrantes.

3.1.2.3 Semanas que busca trabajo

Tabla 3.7: Estadística Descriptiva de Semanas que busca trabajo

Media	Mediana	Moda	Desviación estándar	Mínimo	Máximo
18,09	12	52	16,96	1	52

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

. Elaboración Propia.

Tabla 3.8: Frecuencia de Semana Busca Trabajo

Semanas	Frecuencia	Frecuencia Relativa	Frecuencia R. Acumulada
[1-7)	127	0,32	0,32
[7-13)	99	0,25	0,56
[13-19)	19	0,05	0,61
[19-25)	68	0,17	0,78
[25-31)	8	0,02	0,80
[31-37)	15	0,04	0,8379
[37-43)	1	0,002	0,8404
[43-53)	64	0,16	1
Total	401	1	

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

. Elaboración Propia.

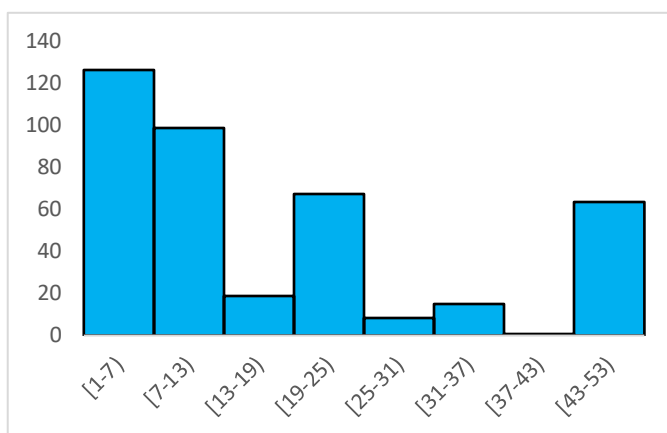


Gráfico 3.15: Histograma de Busca Trabajo

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

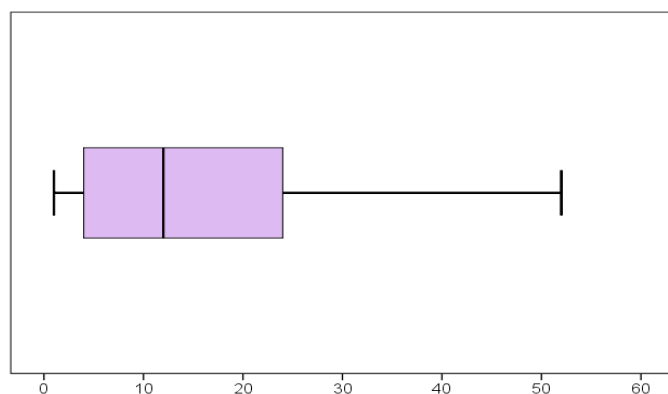


Gráfico 3.14: Diagrama de Cajas de Busca Trabajo

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Evaluando a las personas que buscaron trabajo el mes pasado, se puede afirmar que en promedio llevan $18,09 \pm 16,96$ semanas. También hubo personas que llevan 1 semana buscando trabajo mientras otras llevan 52 semanas. Podemos decir que el 50% como mucho ha buscado ofertas laborales durante 12 semanas. Las semanas con mayor frecuencia son 52. Ver Tabla 3.7

En el Gráfico 3.15 y la Tabla 3.8, el 32% de las personas que se encuentran sin trabajo llevan entre 1 a 6 semanas seguido de un 25% de 7 a 12 semanas.

Para terminar en la Gráfico 3.14, vemos que no hay datos aberrantes y también hay que acotar que el 75% de las personas lleva como mucho 24 semanas buscando trabajo.

3.2 Análisis Bivariado.

En esta sección, se consideran los casos de las variables que entre ellas son influyentes con la condición laboral es decir, cuyo valor p es inferior a 0.05.

3.2.1 Área

Tabla 3.9: Análisis Bi-variado Área vs Condición Laboral

Área	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Rural	6	40	36	26	108
Urbana	441	4634	4214	922	10211
Total	447	4674	4250	948	10319

Área	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Rural	0,0006	0,0039	0,0035	0,0025	0,0105
Urbana	0,0427	0,4491	0,4084	0,0893	0,9895
Total	0,0433	0,4530	0,4119	0,0919	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.
Elaboración Propia.

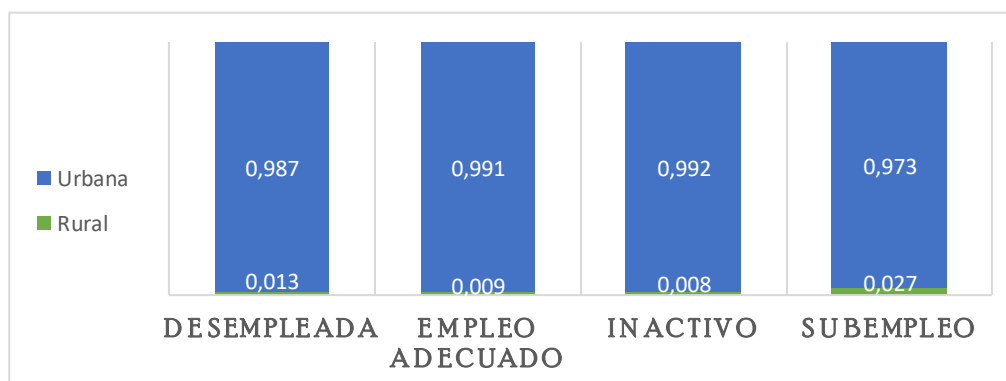


Gráfico 3.16: Gráfico de Barras de las Distribución Condicional entre área y condición laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

En la tabla 3.9, la probabilidad de que una persona tenga empleo adecuado y viva en un área urbana es 0.45 seguido por las personas

que se encuentran en condición inactiva y viven en la misma zona cuya probabilidad es de 0.41.

En la Gráfico 3.16, indicamos que es mayor la probabilidad de desempleo, inactividad, empleo adecuado, subempleo en personas que residen en zonas urbanas.

3.2.2 Ciudad

Tabla 3.10: Análisis Bi-variado entre Ciudad y Condición Laboral

Ciudad	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Ambato	97	951	877	247	2172
Cuenca	54	969	887	150	2060
Guayaquil	142	1391	1346	391	3270
Quito	154	1363	1140	160	2817
Total	447	4674	4250	948	10319

Ciudad	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Ambato	0,009	0,092	0,085	0,024	0,210
Cuenca	0,005	0,094	0,086	0,015	0,200
Guayaquil	0,014	0,135	0,130	0,038	0,317
Quito	0,015	0,132	0,110	0,016	0,273
Total	0,043	0,453	0,412	0,092	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.
Elaboración Propia.

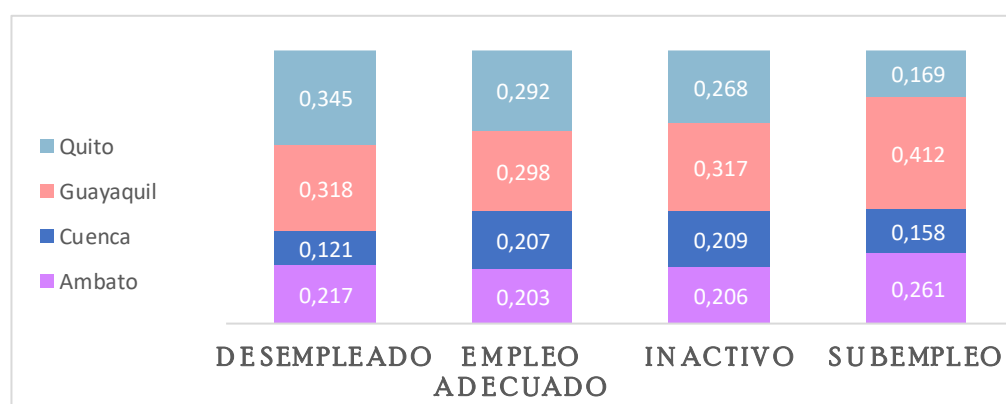


Gráfico 3.17: Gráfico de Barras de la Distribución condicional entre Ciudad y Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

En la tabla 3.10, el 14% de las personas que residen en Guayaquil tienen un empleo adecuado hay que mencionar además, que la probabilidad de que resida en Quito y tenga un empleo adecuado es de 0.13.

La probabilidad de que una persona se encuentre desempleada es aproximadamente 0.2 mayor si esta se encuentra en Quito o Guayaquil en comparación a los que viven en Cuenca sin embargo, es de 0.1 mayor si esta vive en Ambato.

En cuanto a los ciudadanos que tienen un empleo adecuado, la probabilidad aumenta 0.1 cuando habitan en Guayaquil o Quito en relación a los que habitan en Cuenca o Ambato.

Enfocándose en las personas inactivas, la probabilidad aumenta 0.09 para las personas que viven en Guayaquil con respecto a los que viven en Cuenca o Ambato, en cambio se incrementa en un 0.05 cuando viven en Quito.

Finalmente, la probabilidad de que un ciudadano que habite en Guayaquil es de 0.412 en el caso de que sean personas con condición laboral subempleadas. (Ver Gráfico 3.17)

3.2.3 Sexo

Tabla 3.11: Análisis Bi-variado entre Sexo y Condición Laboral

Sexo	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Hombre	196	2839	1422	494	4951
Mujer	251	1835	2828	454	5368
Total	447	4674	4250	948	10319

Sexo	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Hombre	0,019	0,275	0,138	0,048	0,480
Mujer	0,024	0,178	0,274	0,044	0,520
Total	0,043	0,453	0,412	0,092	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.
Elaboración Propia.

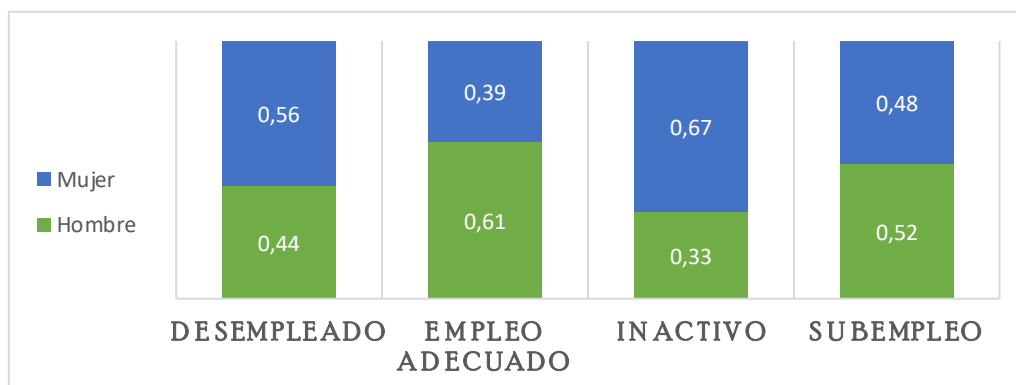


Gráfico 3.18: Gráfico de Barras de la Distribución Condicional entre Sexo y Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Comparando entre el género y condición laboral, El 21.5% de las personas tienen un empleo adecuado y son hombres en cambio, el 27.4% de las personas está conformado por aquellas que no se encuentran laborando y son mujeres. (Ver Tabla 3.11)

En la Gráfico 3.18, dada una persona desempleada, la probabilidad es 0,56 cuando son mujeres mientras en el caso de las personas que tienen un empleo adecuado la probabilidad es 0.61 cuando son hombres.

En el caso de la condición laboral inactiva, su probabilidad es 0.67 cuando son mujeres, por último, siempre y cuando sean subempleados, su probabilidad es 0.52 para los de género masculino.

3.2.4 Parentesco

Tabla 3.12: Análisis Bi-variado entre Parentesco y Condición Laboral

Parentesco	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Jefe	106	2480	917	481	3984
No jefe	341	2194	3333	467	6335
Total	447	4674	4250	948	10319

Parentesco	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Jefe	0,010	0,240	0,089	0,047	0,386
No jefe	0,033	0,213	0,323	0,045	0,614
Total	0,043	0,453	0,412	0,092	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

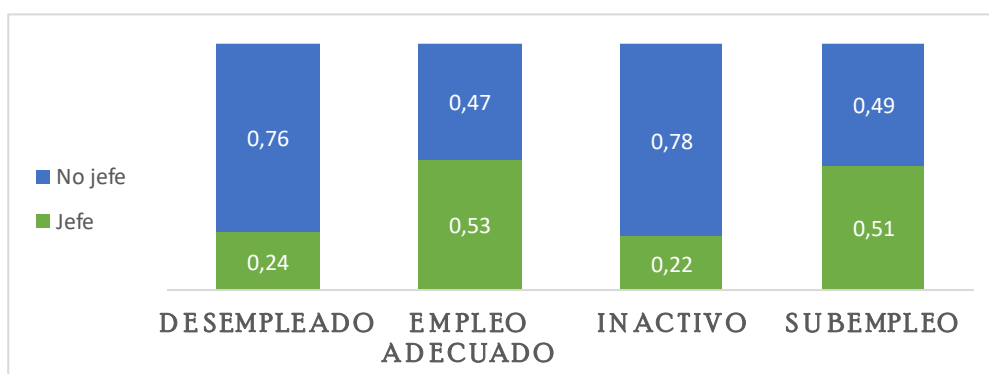


Gráfico 3.19: Gráfico de Barras de la Distribución Conjunta entre Parentesco y Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Es necesario resaltar los resultados obtenidos en la tabla 3.12, teniendo en cuenta que el 24% está conformado por las personas que son jefe de hogar y tiene un empleo adecuado por otro lado, el 32.3% de las personas tienen condición inactiva y no son jefe de hogares.

A modo de interpretación en la Gráfico 3.19, dado que una persona este desempleada, el 76% no son jefes de hogares mientras en el caso de las personas inactivas, la probabilidad es 0.78 cuando no son jefes de hogares.

Por último, considerando el caso de las personas con empleo adecuado y subempleo ambas probabilidades están entre 0.51 a 0.53 cuando son jefe de hogar.

3.2.5 Instrucción

Tabla 3.13: Análisis Bi-variado entre Instrucción y Condición Laboral

Instrucción	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Ninguno	2	13	151	9	175
Postgrado	11	256	38	9	314
Primaria	56	676	862	250	1844
Secundaria	203	1782	2286	469	4740
Superior	175	1947	913	211	3246
Total	447	4674	4250	948	10319

Instrucción	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Ninguno	0,0002	0,0013	0,0146	0,0009	0,017
Postgrado	0,0011	0,0248	0,0037	0,0009	0,030
Primaria	0,0054	0,0655	0,0835	0,0242	0,179
Secundaria	0,0197	0,1727	0,2215	0,0455	0,459
Superior	0,0170	0,1887	0,0885	0,0204	0,315
Total	0,0433	0,4530	0,4119	0,0919	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.



Gráfico 3.20: Gráfico de Barras de la Distribución Condicional entre Instrucción y Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

En la Tabla 3.13 se muestran los resultados con mayor relevancia, donde el 18.9% de las personas tiene un empleo adecuado y tienen título universitario, en cambio el 22.2% son personas inactivas y han tenido estudios secundarios.

En la Gráfico 3.20, considerando las personas desempleadas, la probabilidad es 0.454 para los que obtuvieron estudios en el colegio mientras en el caso de los que tienen empleo adecuado dicha probabilidad es de 0.417 cuando tienen estudios universitarios. Finalmente considerando a una persona inactiva y las que tienen subempleo, la probabilidad se encuentra entre 0.538 y 0.495 para las personas con estudios de secundaria.

3.2.6 Edad

Tabla 3.14: Análisis Bi-variado entre Edad y Condición Laboral

Edad	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
[15-18)	3	3	828	3	837
[18-30)	231	1226	1290	249	2996
[30-45)	139	1930	423	356	2848
[45-65)	68	1411	738	305	2522
Más de 65	6	104	971	35	1116
Total	447	4674	4250	948	10319

Edad	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
[15-18)	0,0003	0,0003	0,0802	0,0003	0,081
[18-30)	0,0224	0,1188	0,1250	0,0241	0,290
[30-45)	0,0135	0,1870	0,0410	0,0345	0,276
[45-65)	0,0066	0,1367	0,0715	0,0296	0,244
Más de 65	0,0006	0,0101	0,0941	0,0034	0,108
Total	0,0433	0,4530	0,4119	0,0919	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

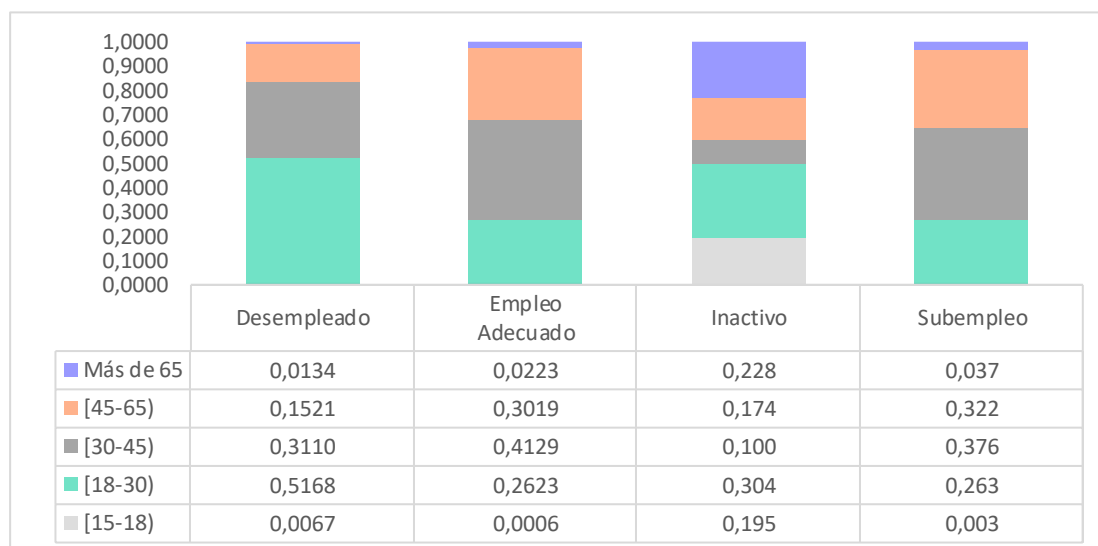


Gráfico 3.21: Gráfico de Barras de la Distribución Condicional entre Edad y Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

En la tabla 3.14, el 19% de las personas tienen un empleo adecuado y tienen entre 30 a 44 años.

Considerando a los desempleados, la probabilidad es alta representando el 0.52 para las personas que tiene entre 18 a 29 años por otro lado, la probabilidad que tienen las personas con edades entre 30 a 44 años es de 0.4129 siempre y cuando tenga empleo adecuado. (Ver Gráfico 3.21)

3.2.7 Estado Civil

Tabla 3.15: Análisis Bi-variado Estado Civil y Condición Laboral

Estado_Civil	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Casado_Unido	193	3061	1673	519	837
Separado_Divorciado	50	455	269	156	2996
Soltero	203	1096	1968	250	2848
Viudo	1	62	340	23	2522
Total	447	4674	4250	948	10319

Estado_Civil	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Casado_Unido	0,019	0,297	0,162	0,050	0,528
Separado_Divorciado	0,005	0,044	0,026	0,015	0,090
Soltero	0,020	0,106	0,191	0,024	0,341
Viudo	0,000	0,006	0,033	0,002	0,041
Total	0,0433	0,4530	0,4119	0,0919	1,000

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

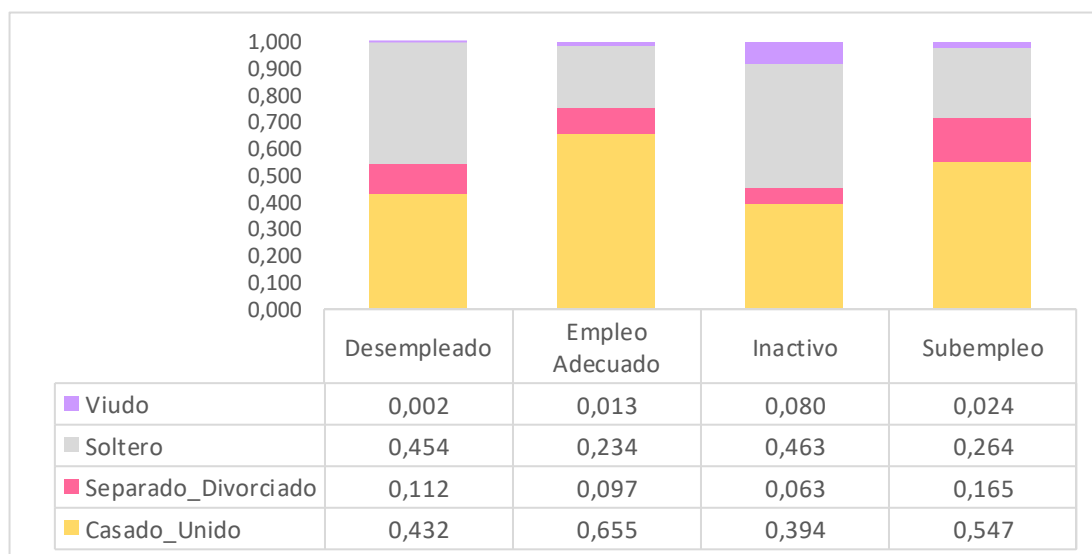


Gráfico 3.22: Gráfico de Barras de la Distribución Condicional entre Estado Civil – Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

En la tabla 3.15, el 29.7% de las personas tienen un empleo adecuado y están casadas o unidas. Considerando a los desempleados o inactivos, la probabilidad está entre 0.45 a 0.46 para las personas son solteras por otro lado, dado que una persona tenga un empleo adecuado la probabilidad de que esté casada o unida es de 0.655 mientras las personas con empleo inadecuado, la probabilidad es de 0.547 cuando están unidas o casadas. (Ver Gráfico 3.22)

3.2.8 Etnia

Tabla 3.16: Análisis Bi-variado entre Etnia y Condición Laboral

Etnia	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Blanco	24	90	112	10	175
Indígena	6	69	59	16	314
Mestizo	401	4381	3980	877	1844
Negro	11	70	63	17	4740
Otros	5	64	36	28	3246
Total	447	4674	4250	948	10319

Etnia	Condición Laboral				Total
	Desempleado	Empleo Adecuado	Inactivo	Subempleo	
Blanco	0,002	0,009	0,011	0,001	175
Indígena	0,001	0,007	0,006	0,002	314
Mestizo	0,039	0,425	0,386	0,085	1844
Negro	0,001	0,007	0,006	0,002	4740
Otros	0,000	0,006	0,003	0,003	3246
Total	447	4674	4250	948	10319

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia

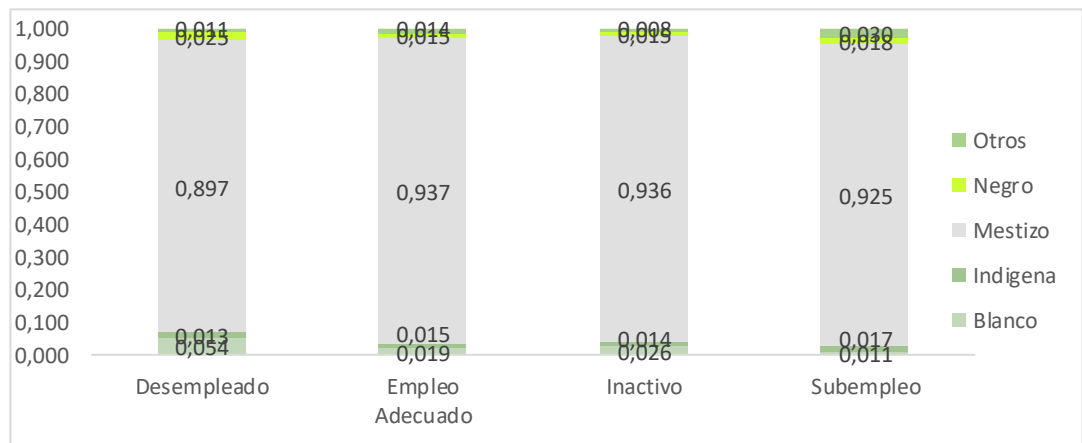


Gráfico 3.23: Gráfico de Barras de la Distribución Condicional entre Etnia – Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

En la tabla 3.16, el 42,5% de las personas tienen un empleo adecuado y son mestizas. En la Gráfico 3.23, dado que se encuentran desempleadas, subempleadas, con empleo adecuado o inactiva es muy probable que ocurra en las personas que se consideran mestizas.

3.3 Análisis de Empleo Adecuado-Subempleo.

Para conocer la influencia de las variables socio-económicas y demográficas en las personas que tienen un empleo adecuado y subempleo, se consideró dos metodologías que permiten conocer dicho comportamiento y solo las personas con edades superiores a 18 años.

Variables Socio-Económico

Por medio de la metodología CHAID con un riesgo o error de 3.8%, se permite diagnosticar perfiles socio-económicos.

Tabla 3.17: Observados vs Pronosticado

Observado	Pronosticado		
	EMPLEO ADECUADO	SUBEMPLO	Porcentaje correcto
EMPLEO ADECUADO	4640	33	99,30%
SUBEMPLO	176	703	80,00%
Porcentaje global	86,70%	13,30%	96,20%

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Como observamos en la tabla 3.17, la metodología da una buena clasificación a las personas con empleo adecuado con un 99,30% y subempleadas con 80%.

Interpretación

La variable que mejor explica la condición laboral es los ingresos que se obtienen en el trabajo. Con respecto a una persona que tiene un **empleo adecuado** es explicado por los nodos 1 y 3. Basándonos en el valor Chi cuadrado como el nodo 1 es superior, se tomará dicho nodo para describir a este grupo con un 95,8%. Por ello, este nodo representa los

ingresos comprendidos entre \$367 y \$585 y a su vez este nodo se vincula con la variable de ocupación, destacándose el nodo 6 por su alta pureza hacia la explicación de las personas con empleo adecuado, donde el 98,4% de las personas con un buen empleo lo conforman aquellas que laboran como empleados en relación dependencia, patronos y empleadas domésticas, luego este nodo se ramifica por la variable contrato de trabajo, donde el nodo 8 es seleccionado dado que ilustra el mayor porcentaje de las personas que tienen un empleo adecuado, concluyendo que el 99,1% de las personas con un buen empleo tienen contrato permanente, trabajan por jornal, por horas, por obras y tienen nombramiento.

Finalmente, la información acerca de las personas que tienen **empleo inadecuado (Subempleo)** es explicada por dos grupos. El grupo 1 es conformado por el nodo 2 que representa a las personas que generan como máximo \$150, como resultado el 100% son subempleadas mientras el nodo 4 se refiere a las personas que producen ingresos entre \$151 y \$366 donde el 78,1% son subempleadas y a su vez este nodo se asocia con la variable de ocupación, donde el nodo 9 es seleccionado dado que explica el mayor porcentaje de subempleados. Del 78.1% de subempleados, el 98,6% de ellos trabajan por cuenta propia, patrono, empleada doméstica, jornalero o peón. (Ver Gráfico 3.24)

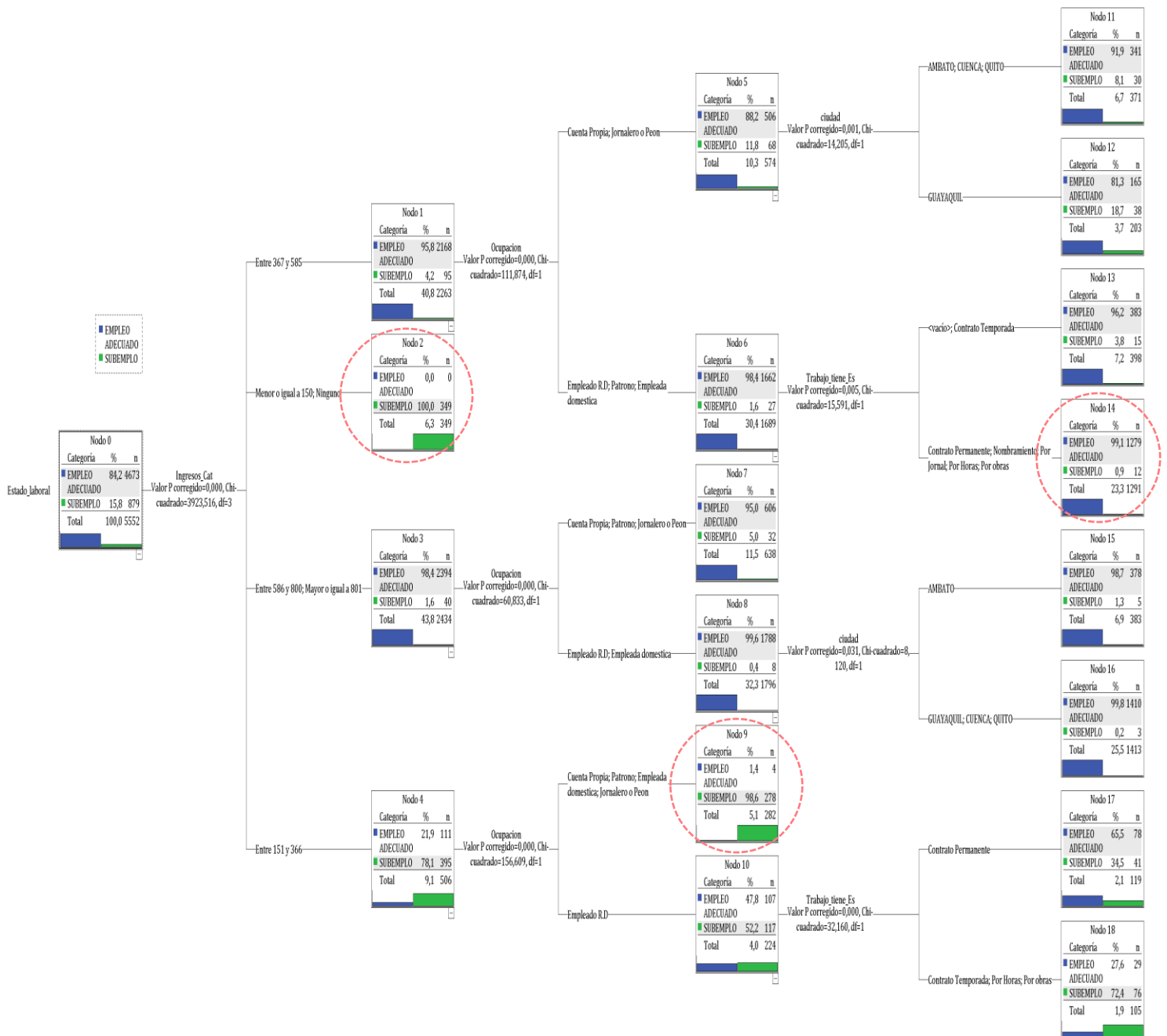


Gráfico 3.24: Árbol de Clasificación (CHAID) – Empleo adecuado y Subempleo

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Variables Demográficas

En esta sección se analizarán las variables que influyen en la condición laboral para ello, primero se realizó una prueba de independencia entre las variables y aquellas variables que sean significativas a la condición de empleo adecuado y subempleo es decir, con un valor p inferior a 0.05 se las tomará para realizar el respectivo análisis de correspondencia múltiple excluyendo el área.

- Sexo
- Estado Civil
- Nivel de Instrucción
- Edad

Tabla 3.18: Dimensiones del ACM – Empleo Adecuado y Subempleo

Dimensiones	Valor Propio	% Explicación
1	0.29853	12.4389
2	0.26148	10.8954
Total		23.3343

*Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.
Elaboración Propia.*

Interpretación

La variabilidad de los datos está representada por el 23.33% según la metodología ACM, donde se concluye que las personas que tienen un empleo adecuado son hombres, tienen instrucción secundaria, superior y postgrado con edades entre 30 y 44 años y con estado civil casado o unido.

Comparando con la metodología CHAID con un riesgo de 15.8%, el nivel de instrucción influye en que una persona tenga un empleo adecuado es decir, que a medida que una persona adquiera más estudios esta propensa a tener un empleo adecuado.

Basándose en el criterio Chi-cuadrado se explicará el nodo 1, dado que es el nodo con el mayor valor por lo tanto, el 80.5% de las personas que tienen empleo tienen un nivel de instrucción secundaria (nodo1) y a su vez este nodo se ramifica por la variable género donde el nodo 5 es el nodo que mayor explica al empleo adecuado con un 84.4% cuando es hombre y este nodo se asocia a la variable estado civil, donde el nodo 12 es el nodo con mayor explicación a las personas con un buen empleo con un 86.8% cuando son casadas o unidad y viuda y por último este nodo se asocia a la edad, donde el nodo 17 es representando por 91.8% de las personas que tienen empleo adecuado con edades entre 30 y 44 años. (Ver Apéndice C)

Por otro lado, las personas que tienen un empleo inadecuado (Subempleadas) suelen estar separadas o divorciadas. (Ver Gráfico 3.25).

Análisis de Correspondencia

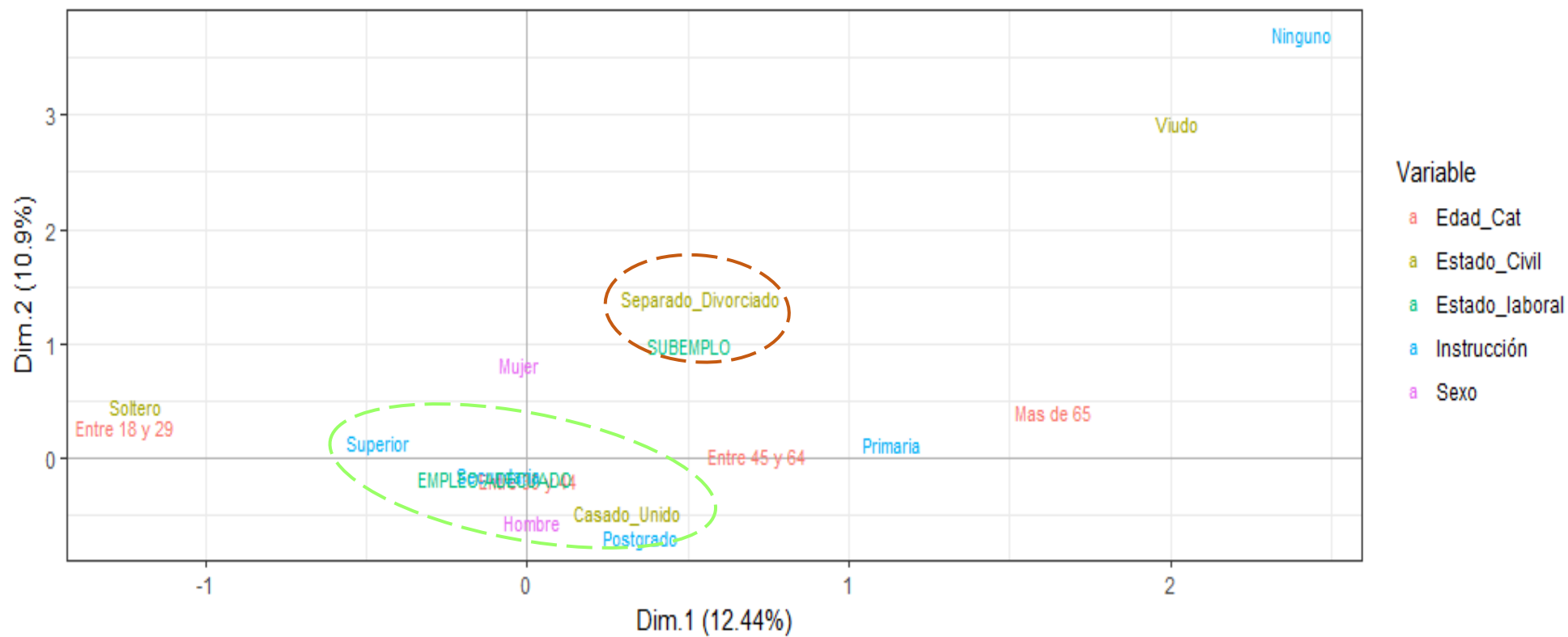


Gráfico 3.25: Análisis de Correspondencia Múltiple Variables Demográficas y Condición Laboral

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

3.4 Desempleados

Para conocer el comportamiento de los desempleados, se toma en cuenta a las personas mayores de 18 años considerando las siguientes variables:

- Ciudad.
- Estado Civil.
- Nivel de Instrucción.
- Sexo
- Ocupación
- Semanas busca trabajo
- Parentesco
- Edad
- Experiencia Laboral

Para la elección del número de clases latentes que explica el comportamiento de los desempleados, se usará el criterio BIC porque tiende a elegir modelos menos complejos es decir, con menos números de parámetros. En cambio, el criterio AIC elige modelos complejos que además son sensibles cuando se tiene un número amplio de parámetros a estimar.

Tabla 3.19: Elección de número de clases latentes – BIC

Nº de Clases	BIC
1	7795.047
2	7522.25
3	7367.62
4	7406.567

*Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.
Elaboración Propia.*

Considerando los resultados de la tabla 3.19, según el criterio del BIC, es recomendable seleccionar 3 clases latentes debido a que esta clase tiene el BIC bajo en comparación a los otros números de clases.

A continuación, examinamos la estimación de la probabilidad de respuesta clase-condicional.

Tabla 3.20: Clase-Condional Ciudad

	Ambato	Cuenca	Guayaquil	Quito
Clase 1	0.1813	0.1216	0.3888	0.3082
Clase 2	0.2844	0.0901	0.1705	0.4550
Clase 3	0.2258	0.1774	0.2742	0.3226

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.21: Clase-Condional Sexo

	Hombre	Mujer
Clase 1	0.3470	0.6530
Clase 2	0.7318	0.2682
Clase 3	0.3548	0.6452

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.22: Clase-Condional Estado Civil

	Casado_Unido	Separado_Divorciado	Soltero	Viudo
Clase 1	0.3795	0.0892	0.5277	0.0036
Clase 2	0.7325	0.2120	0.0556	0.0000
Clase 3	0.1290	0.0484	0.8226	0.0000

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.23: Clase-Condional Nivel de Instrucción

	Ninguno	Primaria	Secundaria	Superior	Postgrado
Clase 1	0.0000	0.0730	0.5242	0.3755	0.0273
Clase 2	0.0189	0.2917	0.2815	0.3749	0.0330
Clase 3	0.0000	0.0645	0.4194	0.5161	0.0000

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.24: Clase-Condición Ocupación

	Cuenta Propia	Empleada Doméstica	Empleado R. D	Jornalero o Peón	Ninguna	Otros	Patrono
Clase 1	0.0668	0.04	0.8591	0.0274	0	0.0036	0.0031
Clase 2	0.2608	0.00	0.6392	0.0799	0	0.0000	0.0201
Clase 3	0.0000	0.00	0.0000	0.0000	1	0.0000	0.0000

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.25: Clase-Condición Semanas que Busca Trabajo

	Ninguno	Entre 1 a 4	Entre 5 a 8	Entre 9 a 12	Más de 13
Clase 1	0.0683	0.2911	0.1814	0.0898	0.3694
Clase 2	0.2002	0.1599	0.1236	0.0879	0.4285
Clase 3	0.0806	0.2419	0.1290	0.1452	0.4032

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.26: Clase-Condición Parentesco

	Jefe	No Jefe
Clase 1	0.0913	0.9087
Clase 2	0.7348	0.2652
Clase 3	0.0484	0.9516

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.27: Clase-Condición Categoría Edad

	Entre 18 a 29	Entre 30 a 44	Entre 45 a 64	Más de 65
Clase 1	0.6073	0.3479	0.0415	0.0034
Clase 2	0.0000	0.3521	0.5906	0.0573
Clase 3	0.8387	0.0968	0.0645	0.0000

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.28: Clase-Condición Experiencia laboral

	Si	No
Clase 1	1	0
Clase 2	1	0
Clase 3	0	1

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Tabla 3.29: Estimación de la proporción que comparte cada clase

Clase 1	Clase 2	Clase 3
0.6275	0.2393	0.14

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

Interpretación

El primer grupo está representado por el 62.75% de las personas desempleadas cuyas características son las siguientes: El 38.88% vive en Guayaquil, 65.30% son mujeres, 52.77% son solteros, 52.42% tienen nivel de instrucción secundaria, 85.91% han trabajado como empleados en relación de dependencia, 36,93% lleva más de 13 semanas buscando trabajo, 90.87% no son jefe del hogar, 60.73% tienen edades entre 18 a 29 años y por último, el 100% de ellos han trabajado anteriormente.

Por otro lado, el segundo explica el 23,93% de los desempleados que tiene en común lo siguiente: El 45.50% es de Quito, 73.18% son hombres, 73.25% están casados o unidos, 37.49% han alcanzado un nivel de instrucción superior, 63.9% han laborado como empleados en relación de dependencia, 42,85% lleva más de 13 semanas buscando trabajo, 73.48% son jefe de hogar, 59.06% tienen edades entre 45 a 64 años y finalmente el 100% han tenido experiencia laboral.

Por último, la tercera clase latente que explica el 14% de las personas sin empleo, el 32,23% es de Quito, 64.52% son mujeres, 82.26% son solteros, 51.61% tiene nivel de instrucción superior, el 40.42% lleva más de 15 semanas buscando trabajo, 95.16% no

son jefe del hogar, 83.87% tienen edades entre 18 a 29 años y el 100% de ellas no han tenido experiencia laboral.

3.5 Análisis de Inactivos.

Esta sección se enfocará en el estudio de la clasificación de las personas inactivas: estudiantes, amas de casa, incapacitados, jubilados y otros.

A través de la metodología CHAID con un error o riesgo de 22.4%, se destacarán ciertas características de las personas inactivas.

Tabla 3.30: Observados vs Pronósticos – Inactivos

Observado	Pronosticado					Porcentaje correcto
	Ama_casa	Estudiante	Incapacitado	Jubilado	Otro	
Ama_casa	1083	65	63	65	0	84,9%
Estudiante	37	1659	0	1	0	97,8%
Incapacitado	113	35	81	100	0	24,6%
Jubilado	150	8	28	446	0	70,6%
Otro	79	67	16	107	0	0,0%
Porcentaje global	34,8%	43,6%	4,5%	17,1%	0,0%	77,8%

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

Elaboración Propia.

A modo de interpretación, en la tabla 3.30 se detalla la comparación entre los valores observados y pronosticados o clasificados por dicha metodología, donde se otorga una buena clasificación para amas de casa con un 84.9%, estudiantes 97.8% y jubilados 70.6%.

Interpretación

La variable que mejor explica la clasificación de personas inactivas es la variable edad. Como observamos la Gráfico 3.26, en el nodo

1, están representadas las personas inactivas mayores de 65 años, donde el 48.1% son **jubilados** cuyo nodo se asocia con la variable ingresos per-cápita. En el nodo 8, se explica la categoría **jubilados** con mayor pureza, donde el 68.8% de las personas jubiladas tiene ingresos per-cápita superior a \$551.

En segunda instancia, en los nodos 2 y 3 que estudian a las **amas de casa**, se tomará el nodo 2, debido a que el valor Chi-Cuadrado es mayor. Este nodo contiene el 64,5% de las personas inactivas que son **amas de casa** que tienen edades entre 45 y 64 años cuyo nodo se ramifica con la variable género. Donde el nodo más puro es el nodo 11 en el que el 76,8% equivale al género femenino. El nodo 11 se ramifica con la variable instrucción, donde el nodo 22 es el nodo que mejor explica la instrucción de las **amas de casa** destacándose que el 80,3% como máximo han alcanzado un nivel de educación secundaria.

Finalmente, la característica de los **estudiantes** está explicada por los nodos 4 y 5, donde el nodo que tiene un mayor valor Chi-Cuadrado es el nodo 4 por ello, se considerará este nodo para describir a los **estudiantes** explicando el 71.5% de esta categoría. El nodo 4 es representado por las personas que tienen edades entre 18 a 29 años y dicho nodo se ramifica por la variable asistencia a clases, donde el 98.9% de los estudiantes asisten a clases.

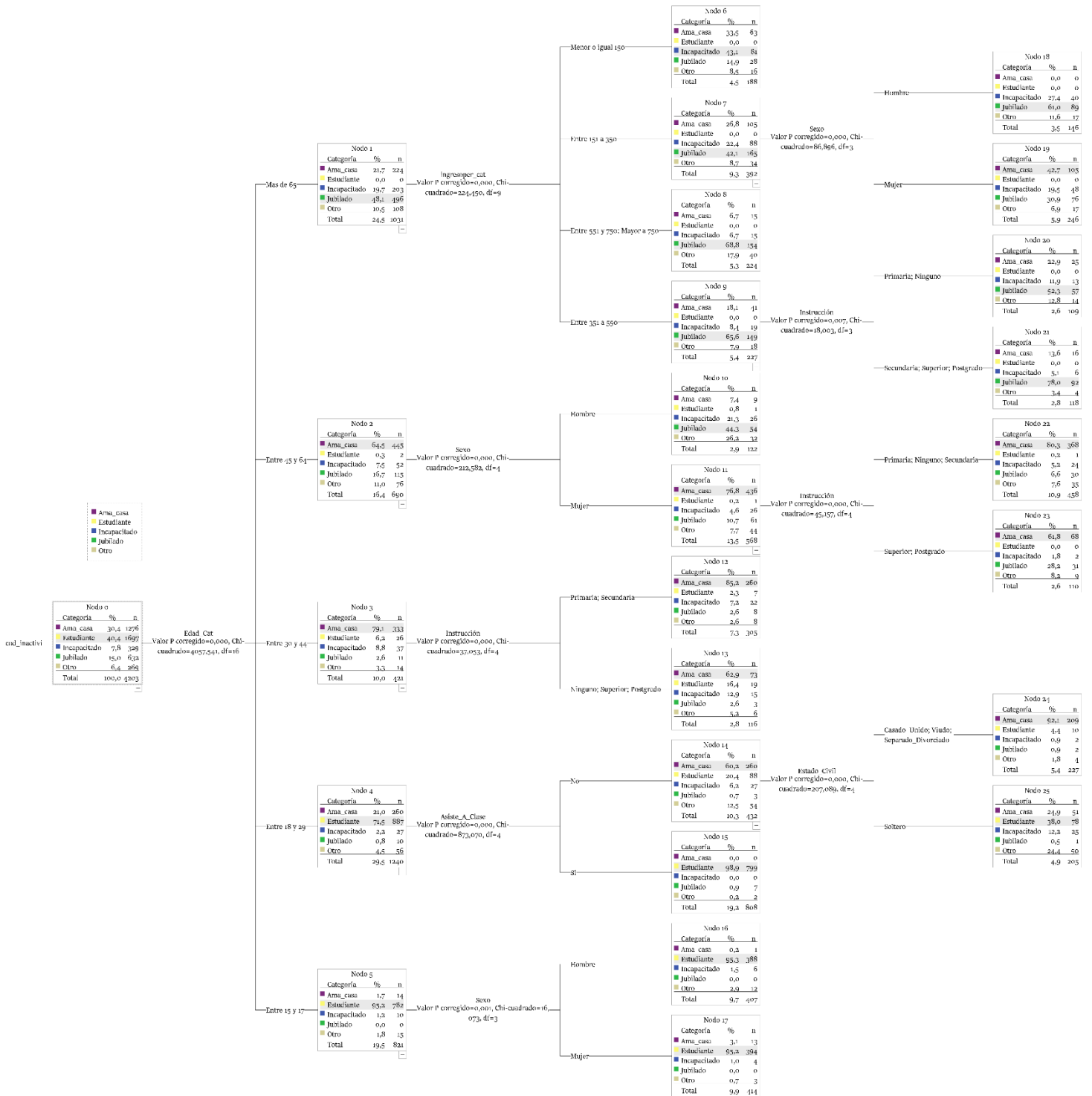


Gráfico 3.26: Árbol de Clasificación (CHAD) de inactivos

Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, 2016.

CAPÍTULO 4

4. DISCUSIÓN Y CONCLUSIONES

(Cabeza & Duque, 2007) sostiene que el nivel de instrucción influye donde dice que una persona que tenga un nivel de instrucción superior, la probabilidad de que quede desempleada es baja comparando con el presente estudio, de los tres grupos que se formaron con el análisis de clases latentes, el último grupo representando el 14% de los desempleados, la mitad de este grupo de desempleados tienen dicha instrucción por lo que aún esta conclusión se mantiene no obstante, las mujeres siguen propensas a quedar desempleadas (INEC, 2016).

(Sarmiento, 2015) Aunque este autor se basó en la provincia de Pichincha

destaca que los jóvenes están propensos a quedar desempleados y a medida que tengan una mayor edad van adquiriendo empleo en nuestro estudio, coincide dicho resultado donde los jóvenes entre 18 a 29 años y mujeres se encuentran desempleados es el grupo más propenso mientras los que tienen edades superiores a 30 años tienen empleo.

Con respecto a la inactividad (Sarmiento, 2015) concluye que las personas inactivas por lo general son mujeres, la mayoría que no tiene ningún nivel de instrucción, viudos y padres o suegros mientras en el estudio en este grupo de inactivos, la mayoría son mujeres, no son jefes de hogar por otro lado, se descubre que las amas de casa a lo mucho han culminado la secundaria.

(Cevallos, 2011) aporta al análisis de subempleados en el año 2010, que las personas que no tienen instrucción y las que han alcanzado la

primaria se encuentran en este estado, son montubios y negros, la ciudad más afectada es Machala seguida de Ambato mientras en nuestro estudio, aunque no abarca Machala pero la ciudad con mayor empleo inadecuado es Guayaquil con un 41,2%, la mitad de los subempleados tienen instrucción secundaria y están casados o unidos.

4.1 Conclusiones

De acuerdo con las personas que tienen condición para laborar se tiene que el 45.3% tienen empleo adecuado y el 41.2% son personas inactivas.

Las variables que influye en la condición laboral son: área, ciudad, sexo, parentesco con el jefe de hogar, nivel de instrucción, estado civil, etnia y edad.

Existe mayor probabilidad que las mujeres se encuentren desempleadas o inactivas, por otro lado, es más probable un hombre tenga un empleo adecuado.

Considerando al parentesco con el jefe de hogar, dado que una persona se encuentre desempleada o inactiva, la probabilidad es alta para las personas que no son jefe de hogar.

La probabilidad que una persona este subempleada o inactiva, si es de instrucción secundaria es 0.5 y 0.54 respectivamente.

Dado una persona que tenga empleo adecuado y esté subempleadas, la probabilidad es alta para las personas que se encuentran casadas o unidas.

En cuanto a la edad, las personas cuyas edades oscilan entre 18 a 30 años la probabilidad que se encuentren desempleadas es 0.5.

Comportamiento de la condición laboral Empleo Adecuado – Subempleo se halló lo siguiente:

- Las personas que tienen un empleo adecuado tiene ingresos entre \$367 y \$587, actualmente laboran como empleados en relación de dependencia, patronos y empleadas domésticas; y aquellas cuyo contrato de trabajo son por nombramiento, contratos permanentes, trabajan por horas, por jornal y por obras. Por otro lado, el nivel de instrucción influye en que una persona tenga empleo adecuado, a medida que tenga un mayor nivel de instrucción. Las personas con un buen empleo que tengan instrucción secundaria por lo general son hombres, con estado civil casado o unidos y con edades entre 30 y 44 años.
- Las personas que tienen un empleo inadecuado generan ingresos entre \$151 y \$366 y cuya ocupación son los patronos, empleadas domésticas, jornaleros o peón y los que trabaja por cuenta propia y por lo general son separados o divorciados.

Con respecto a las personas desempleadas:

- El grupo más representativo de desempleados con un 62,75% tiene las siguientes características: El 38.88% vive en Guayaquil, 36.93% lleva más de 13 semanas buscando trabajo, la mayoría son mujeres, han trabajado como

empleados en relación de dependencia, no son jefe de hogar, tienen edades entre 18 a 29 años mientras que la mitad de desempleados son solteros y tienen nivel de instrucción secundaria y por último todos han tenido experiencia laboral.

Por último, las personas inactivas:

- Las personas jubiladas (con edad superior a 65 años) tienen ingresos per-cápita superior a \$551.
- La mayoría de las amas de casa tienen edades entre 45 a 64 años, son mujeres y como máximo han alcanzado un nivel de instrucción secundaria.
- La gran parte de estudiantes tienen edades 18 a 29 años y se encuentran asistiendo a clases.

4.2 Recomendaciones

- Realizar un estudio a profundidad para determinar que otro tipo de factores afecten al subempleo.
- Desarrollar un estudio exhaustivo para aquellas personas que laboran por cuenta propia, empleadas domésticas, patronos, jornalero o peón que permita determinar los factores que impida tener ingresos superiores o iguales al salario básico.
- Crear programas para ampliar plazas de empleo para las personas jóvenes y en especial en mujeres.
- Diagnosticar las razones por la que los empleados en relación de dependencia están propensos a quedar desempleados.
- Considerar en un futuro estudio si la actividad de laboral afecte al empleo adecuado.

BIBLIOGRAFÍA

- Abdi Hervé, D. V. (2007). Multiple Correspondence Analysis. *The University of Texas at Dallas*.
- Cabeza, M. V., & Duque, E. A. (2007). La dinámica laboral y las condiciones socioeconómicas en el Ecuador período 2003-2006. *Tesis de Grado*. Guayaquil, Ecuador .
- Castillo, R., & Rosero, J. (2015). Empleo y Condición de Actividad en Ecuador. *Revista de Estadísticas y Metodologías*.
- Cevallos, A. A. (Septiembre de 2011). Variables causantes del fenómeno subempleo en el Ecuador, durante el período 2007-2010. *Tesis de Grado*. Quito, Ecuador.
- Garcés, C., Albán, A., & Troya, P. (2014). *Ecuador en Cifras*. Obtenido de http://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/archivos_ENEMDU/Metodologia_Disenio_Muestral-ENEMDU.pdf
- INEC. (2016). *Encuesta Nacional de Empleo Desempleo y Subempleo*. Ecuador: INEC.
- INEC. (s.f.). *Ecuador en cifras*. Obtenido de http://www.inec.gob.ec/estadisticas/?option=com_content&view=article&id=130&Itemid=108
- Kasprzyk, I. (2010). Latent class models in the R software. *Universitatis Lodziensis Folia Oeconomía*.
- Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*.

- María, P. L. (2014). Comparación interactiva de modelos de minería de datos utilizando técnicas de visualización. *Tesis de Maestría*. Argentina : Facultad de Ciencias Exactas y Naturales.
- María, R., & Ruth, B. (2013). Como aplicar árboles de clasificación en SPSS. *Revista d'Innovació i Recerca en Educació*.
- Ojeda, L. R. (2007). *Probabilidad y Estadística básica para Ingenieros*. Guayaquil: Espol.
- Rokach, L., & Maimon, O. (2008). *Data Mining with Decision Trees*. World Scientific Publishing.
- Sarmiento, J. d. (2015). Aplicación de análisis multivariado en la Encuesta Nacional de Empleo, Subempleo y Desempleo (ENEMDU) del INEC, provincia de Pichincha para los años 2012 y 2013. *Tesis de Grado*. Guayaquil, Ecuador.

APÉNDICE

APÉNDICE A: Código de R de análisis de Correspondencia

```
#Análisis de correspondencia en ggplot2
library("FactoMineR")
library("ggplot2")
MCA_alt=function(dats){
  fit=MCA(dats,graph = F)
  dims=paste0("Dim.",c(1,2),"          (" ,round(fit$eig$percentage          of
variance'[1:2],2),"%" ,")")
  cats=apply(dats,function(x) length(levels(x)))
  fit_vars_df = data.frame(fit$var$coord, Variable = rep(names(cats), cats))
  fit_obs_df = data.frame(fit$ind$coord)
  fit_vars_df$nam=rownames(fit_vars_df)
  plot.mca=
  ggplot(data =fit_vars_df, aes(x = Dim.1, y = Dim.2)) +
  geom_hline(yintercept = 0, colour = "gray70") +
  geom_vline(xintercept = 0, colour = "gray70") +
  xlab(dims[1])+ylab(dims[2])+
  geom_density2d(colour = "gray80") +
  geom_text(data = fit_vars_df, size=3,
            aes(x = Dim.1, y = Dim.2,
                label = nam, colour = Variable)) +
  ggtitle("Análisis de Correspondencia") +
  scale_colour_discrete(name = "Variable")+theme_bw()
  return(list(fit=fit,plt=plot.mca))
}
#Ejemplo
dat=read.delim("clipboard",T)
fit=MCA_alt(dat)
```

```
plot(fit$plt)
```

APÉNDICE B: Código de R de análisis de clases latentes

```
datosdesempleados=read.delim("clipboard",T)
```

```
attach(datosdesempleados)
```

```
library(scatterplot3d)
```

```
library(MASS)
```

```
library(poLCA)
```

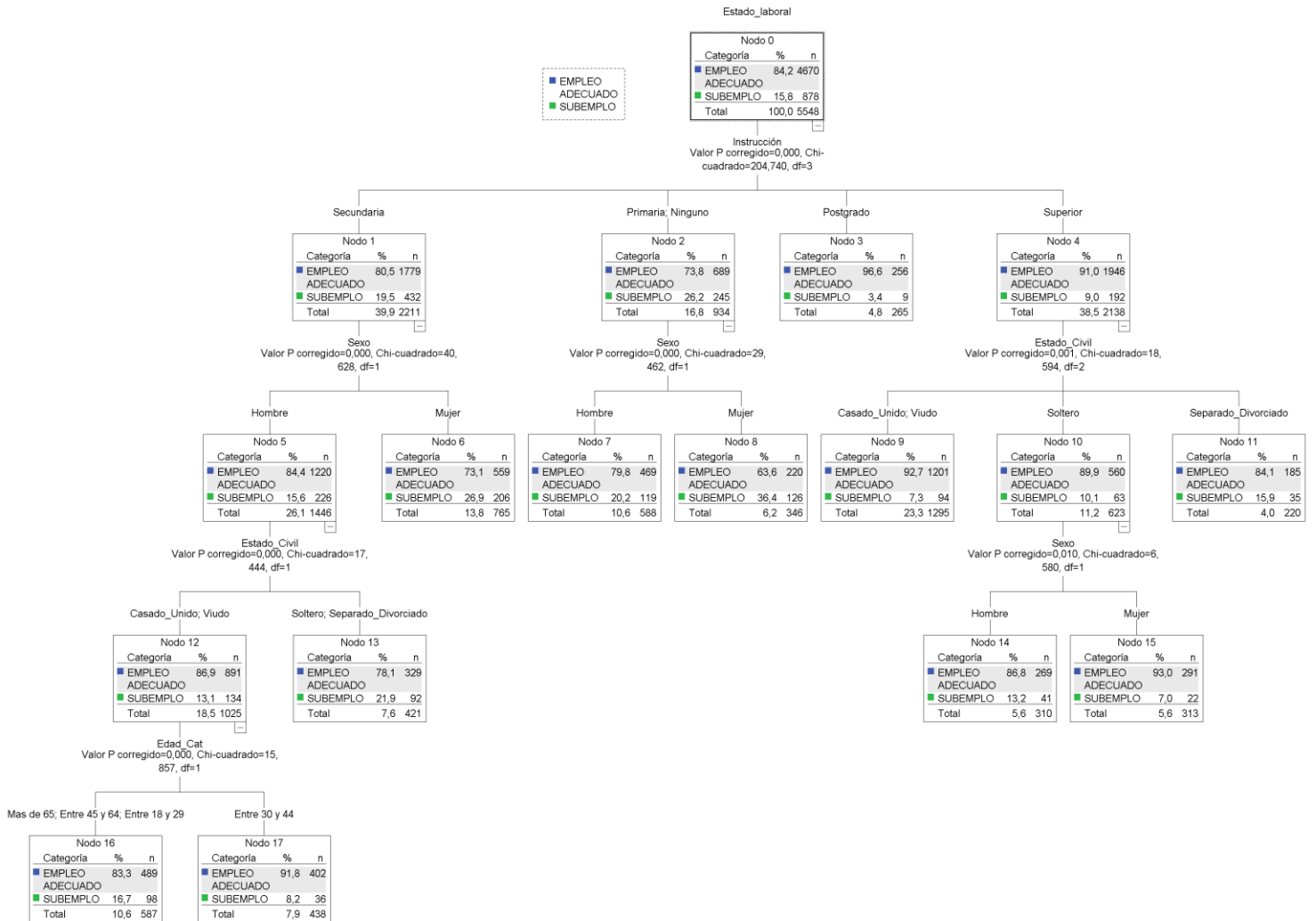
```
## análisis de clases latentes desempleados
```

```
f1<-
```

```
cbind(ciudad,Sexo,Estado_Civil,Instrucción,Ocupacion,SemanasCatBt,Parentezco,Edad_Cat,Traba_Anterior)~1
```

```
gss.lc2 <- poLCA(f1,datosdesempleados,nclass=3,graphs = T)
```

APÉNDICE C: Árbol CHAID – Estado Laboral



Riesgo

Método	Estimación	Típ. Error
Resustitución	,158	,005
Validación cruzada	,158	,005

Métodos de crecimiento: CHAID

Variable dependiente: Estado_laboral

Clasificación

Observador	Pronosticado		
	EMPLEO ADECUADO	SUBEMPLO	Porcentaje correcto
EMPLEO ADECUADO	4670	0	100,0%
SUBEMPLO	878	0	0,0%
Porcentaje global	100,0%	0,0%	84,2%

Métodos de crecimiento: CHAID

Variable dependiente: Estado_laboral