ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

CARACTERIZACIÓN DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA EN EL ECUADOR

PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DE TÍTULO DE INGENIERA MATEMÁTICA

PAOLA GEOVANNA TROYA BASTIDAS

paotroyab@gmail.com

DIRECTOR: MAT. LUIS GUASGUA

lguasguaa@gmail.com

Quito, agosto 2011

DECLARACIÓN

Yo, Paola Geovanna Troya Bastidas, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

PAOLA GEOVANNA TROYA BASTIDAS

AGRADECIMIENTOS

Agradezco a Dios por ser mi guía y compañero en el transcurso de mi vida.

A mi mami Piedad, por su entrega total de amor y cariño, por ser más que mi madre, una gran amiga y compañera y además por ser un motivo especial para sentir que compartir todo lo hermoso de la vida a su lado es una gran bendición.

A mi director Luis, un agradecimiento sincero y un sentimiento de respeto y consideración por ser un gran ser humano; por su paciencia y colaboración, puesto que sin su ayuda habría sido difícil la culminación de este proyecto, a mis lectores por su tiempo y valiosa contribución.

Paola

DEDICATORIA

Dedico este trabajo a Dios puesto que entiendo claramente que el mundo gira en torno a su voluntad y solo él es dueño de nuestro destino además es el único que sabe lo que habrá de suceder para alcanzar el bienestar de nosotros sus hijos.

También una dedicatoria a la virgen María, la madre de todos nosotros, a quien con toda la fe admiro, respeto y en quien confío pues nunca me ha abandonado , para ti madre mía con toda la humildad te dedico este trabajo.

En especial a mi mami querida, porque gracias a su entrega, a su esfuerzo diario he conseguido muchas cosas, ha sido un modelo de perseverancia y paciencia para alcanzar un objetivo, por su entereza por salir adelante, por sus consejos y por su gran espíritu de heroísmo y lealtad para conmigo.

A mi papi Carlos, desde lo más profundo de mi corazón, aunque ya no se encuentre a mi lado, se que desde el cielo me seguirá guiando y protegiendo como siempre, le dedico mi esfuerzo en honor a las grandes alegrías y tristezas que compartimos juntos y por su gran ejemplo de amor y dedicación.

Paola

INDICE

LISTA DE GRÁFICOS……i

LISTA DE TABLAS……ii

LISTA DE ANEXOS……iv

RESUMEN……v

ABSTRACT……vi

INTRODUCCIÓN……1

CAPÍTULO 1……2

LA POBLACIÓN ECONOMICAMENTE ACTIVA EN EL ECUADOR (PEA)……2

1.1 Clasificación actual de la población…… 2

1.2 Mercado laboral ……3

1.3 Evolución de la Población Económicamente Activa…… 3

1.4 Estados de Ocupación de la PEA ……5

1.5 CLASIFICACIÓN DE LA POBLACIÓN……10

1.5.1 Población en Edad de Trabajar (PET)……10

1.5.2 Población Económicamente Activa y Población Económicamente Inactiva……11

1.6 DISTRIBUCIÓN DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA (PEA)……12

1.7 INDICADORES LABORALES DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA……13

1.7.1 TASAS DE PARTICIPACIÓN……13

1.7.2 TASAS DE OCUPACIÓN……14

1.7.3 TASAS DE SUBEMPLEO……14

1.7.4 TASAS DE DESEMPLEO……15

1.7.5 TASAS DE SECTORIZACIÓN DE LA PEA……15

1.7.6 TASAS DE SUBUTILIZACIÓN……16

CAPÍTULO 2……17

TÉCNICAS DE ESTADÍSTICA MULTIVARIANTE……17

2.1 MODELOS DE REGRESIÓN DE RESPUESTA CUALITATIVA……19

2.1.1 Modelo Lineal de Probabilidad……19

2.1.2 Regresión Logística……21

2.1.2.1 Regresión Logística Binomial 22
2.1.2.1.1 Estimación del Modelo Logit.....23
2.1.2.1.2 Estadístico de Wald.....27
2.1.2.1.3 Interpretación del modelo ogístico.....28
2.1.2.1.4 Medidas de confiabilidad del modelo.....30
2.1.2.1.5 Estadísticos influenciales para regresión logística.....31
2.1.2.2 Regresión Logística Multinomial..... 31
2.1.2.2.1 Estimación de parámetros..... 34
2.1.2.2.2 Significatividad global del modelo..... 35
2.1.2.2.3 Significatividad del efecto de cada variable regresora..... 35
2.1.2.2.4 Significatividad de cada parámetro..... 35
2.1.2.2.5 Intervalo de confianza para los parámetros..... 35
2.1.2.2.6 Calidad del ajuste..... 36
2.2 ANÁLISIS DISCRIMINANTE..... 38
2.2.1 Modelo matemático..... 39
2.2.2 Extracción de las funciones discriminantes..... 40
2.2.3 Procedimiento Matricial..... 41
2.2.4 Clasificación entre dos poblaciones..... 43
2.2.4.1Poblaciones Normales..... 45
2.2.4.2 Cálculo de Probabilidades de Error..... 45
2.2.4.3 Probabilidades a posteriori..... 46
2.2.5 Poblaciones Desconocidas. Caso General..... 47
2.2.5.1 Cálculo de Probabilidades de error..... 48
CAPÍTULO 3.....50
CARACTERIZACIÓN DE LA PEA UTILIZANDO MÉTODOS DE ESTADÍSTICA MULTIVARIANTE..... 50
3.1 Definición de Variables..... 50
3.2 Análisis descriptivo de las variables..... 50
3.2.1 Ocupación Total y por Sexo..... 50
3.2.2 Subocupación Total y por Sexo..... 51
3.2.3 Desocupación total y por sexo..... 52

3.2.4 Tasa de Participación Global Urbana (TPG)…… 53
3.2.5 Tasa de Participación Global Urbana según sexo (TPG)…… 54
3.2.6 Principales Ramas de Actividad en el Área Urbana…… 54
3.2.7 Evolución del Desempleo en las áreas Urbanas de las Regiones Naturales…….. 55
3.2.8 Evolución del Subempleo Global Urbano, según Regiones Naturales…… 56
3.2.9 PEA SEGUN GRUPOS ETARIOS…… 57
3.3 Variables Utilizadas…… 59
3.4 Interpretación y Resultados…….. 61
3.4.1 Modelos de Regresión Logística Multinomial…….... 61
3.4.1.1 Pruebas de significancia de las variables independientes……. 63
3.4.1.2 Calidad de predicción…… 64
3.4.1.3 Interpretación de las tasas de cambio por categoría……69
3.4.2 Análisis Discriminante…….. 73
3.4.2.1 Interpretación de las funciones discriminantes……. 76
3.4.2.2 Validación de los Resultados…… 81
CAPÍTULO 4…… 82
Conclusiones y Recomendaciones……. 82
REFERENCIAS…… 92

LISTA DE GRÁFICOS

Gráfico 1: Población en Edad de Trabajar Urbana y Rural.

Gráfico 2: PEA y PEI Urbana.

Gráfico 3: PEA y PEI Rural.

Gráfico 4: PEA Urbana.

Gráfico 5: PEA Rural.

Gráfico 6: Ocupación Total y por Sexo.

Gráfico 7: Subocupación Total y por Sexo.

Gráfico 8: Desocupación Total y por Sexo.

Gráfico 9: Tasa de Participación Global Urbana.

Gráfico 10: Tasa de Participación Global Urbana según Sexo.

Gráfico 11: Principales Ramas de Actividad en el Área Urbana.

Gráfico 12: Evolución del Desempleo en las áreas urbanas de las Regiones Naturales.

Gráfico 13: Evolución del Subempleo Global Urbano por Regiones Naturales.

Gráfico 14: Mapa Territorial.

Gráfico 15: Ocupados Plenos.

Gráfico 16: Subempleados.

Gráfico 17: Desempleados.

Gráfico 18: Estados de Ocupación.

LISTA DE TABLAS

Tabla 1: Tasa de Ocupación Plena según Grupos Etarios

Tabla 2: Tasa de Subempleo según Grupos Etarios

Tabla 3: Tasa de Desempleo según Grupos Etarios

Tabla 4: Tabla de Clasificación del Modelo Logístico Multinomial 2007

Tabla 5: Tabla de Clasificación del Modelo Logístico Multinomial 2008

Tabla 6: Tabla de Clasificación del Modelo Logístico Multinomial 2009

Tabla 7: Tabla de Clasificación del Modelo Logístico Multinomial 2010

Tabla 8: Tabla de Clasificación del Modelo Logístico Multinomial Global

Tabla 9: Coeficientes del Modelo

Tabla 10: Tasas de cambio de subempleo y desempleo con respecto a la ocupación plena

Tabla 11: Coeficientes del Modelo por Categoría

Tabla 12: Tamaños Muestrales de cada grupo

Tabla 13: Variables en el análisis

Tabla 14: Lambda de Wilks

Tabla 15: Comparaciones de grupos por pares

Tabla 16: Valores propios

Tabla 17: Lambda de Wilks. Contraste de las funciones

Tabla 18: Funciones en los centroides de los grupos

Tabla 19: Coeficientes estandarizados de las funciones discriminantes canónicas.

Tabla 20: Matriz de estructura.

Tabla 21: Probabilidades previas para los grupos.

Tabla 22: Resultados de la clasificación.

LISTA DE ANEXOS

Anexo 1: Modelo Logístico Multinomial para la Caracterización de la PEA .

Anexo 2: Coeficientes Odds Ratio.

Anexo 3: Resultados de Análisis Discriminante.

Anexo 4: Metodología de la Encuesta de Empleo y Desempleo del INEC.

Anexo 5: Organización Internacional del Trabajo OIT.

Anexo 6: Clasificación de la Población.

Anexo 7: Evolución del Mercado Laboral Nacional.

RESUMEN

Este trabajo tiene como objetivo encontrar las características más relevantes de la Población Económicamente Activa, mediante el uso de técnicas de estadística multivariante.
Debido a la naturaleza de las variables se va a utilizar un modelo de regresión logística multinomial y análisis discriminante.
Se basa en la teoría de la Población Económicamente Activa y en las Encuestas de Empleo y Desempleo que realiza el Instituto Nacional de Estadísticas y Censos.
Consta de cuatro capítulos, en el primer capítulo se tiene una descripción teórica de la Población Económicamente Activa, en el segundo presentamos una explicación de la teoría de estadística multivariante, a continuación en el tercer capítulo, tenemos la implementación de los modelos y su interpretación correspondiente y para finalizar se presenta las características que permitirán diferenciar el estado de ocupación en el que se puedan encontrar las personas.

ABSTRACT

This thesis has an objective finding the most important characteristics of Economic Active Population, through multivariate statistics techniques.
Because of nature of variables it will use a regression multinomial logistic model and discriminator analysis.
It´s base on Economic Active Population and Employ and Unemployment Investigation which realize Instituto Nacional de Estadísticas y Censos.
This contains four chapters, the first has a theorical description about Economic Active Population, in the second, it presents an explanation of statistic multivariate theory, in the next chapter we have implementation of models and it´s correspondence interpretation and finally characteristics which will permit to differ occupation state on being people.

INTRODUCCIÓN

La Constitución Política de la República establece, entre los objetivos de la economía, la eliminación de la indigencia, la superación de la pobreza, la reducción del desempleo y subempleo, el mejoramiento de la calidad de vida de los habitantes y la distribución equitativa de la riqueza. Según la carta política, le corresponde al Estado, en el marco de una economía social de mercado, incentivar el pleno empleo, el mejoramiento de los salarios reales y el aumento de la productividad, así como otorgar subsidios específicos a la población que los necesite.

Por lo tanto, garantizar el empleo y además lograr que éste sea adecuado y bien remunerado, es deseable social e individualmente, entonces el disminuir el desempleo, el subempleo y mejorar los salarios reales; son objetivos de política muy importantes.

De tal manera, la fuerza de trabajo, o el potencial de una persona de generar ingresos laborales, se convierte en el mayor activo que un hogar o una persona posee para vivir una vida digna y escapar de la pobreza.

La población total de nuestro país se divide en dos grupos: Población Económicamente Activa y Población Económicamente Inactiva.

Para el presente trabajo las personas de mayor interés son quienes se encuentran dentro de la actividad, es decir aquellos que conforman el mercado laboral o forman parte de los tres estados de ocupación que conforman la PEA (Población Económica Activa).

Al final se presentara una análisis de las variables más significativas que determinaran las razones del porque las personas pueden encontrarse en cada uno de dichos estados y la variación de los indicadores de empleo que presenta el Instituto Nacional de Estadísticas y Censos, trimestral y anualmente.

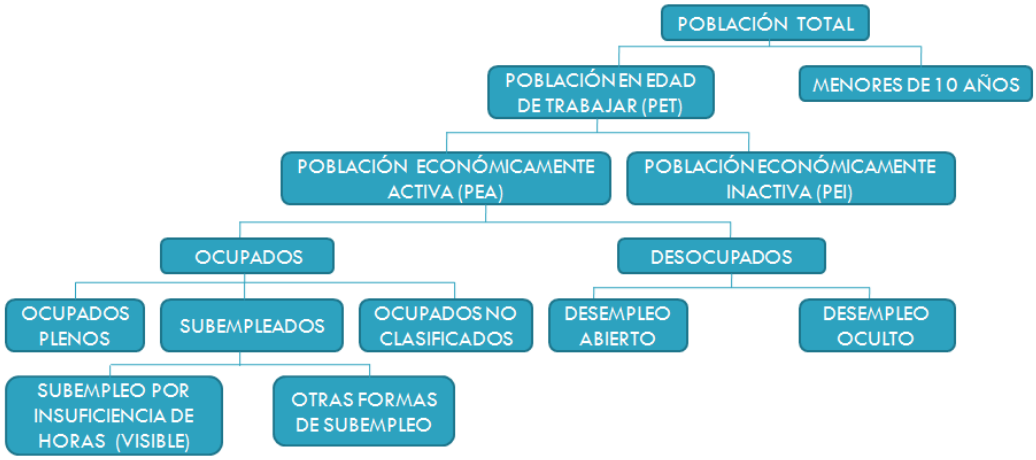
CAPÍTULO 1

LA POBLACIÓN ECONÓMICAMENTE ACTIVA EN EL ECUADOR (PEA)

La población, es un conjunto de todas las unidades o elementos que hacen parte de un todo, en este caso; las viviendas, los hogares y las personas que conforman los hogares y que habitan en las viviendas ubicadas en las áreas urbanas y rurales.

La población total, se divide en dos grupos: población en edad de trabajar y población menor a 10 años.
La población en edad de trabajar se subdivide en: Población Económicamente Activa y Población Económicamente Inactiva.

1.1 Clasificación actual de la población



Fuente: Instituto Nacional de Estadísticas y Censos, Publicación ENEMDU Junio 2010, Ronda 28.

Se considera población económicamente activa (PEA) al grupo de personas que, dentro de las edades productivas, se encuentran ocupadas o desocupadas.
Entre las personas activas tenemos a quienes se encuentran empleadas, subempleadas y desempleadas.

Entre las inactivas se incluye a quienes realizan actividades de estudio o quehaceres domésticos, a los jubilados o pensionistas, y a las personas que están impedidas para trabajar.

El concepto de Población Económicamente Activa (PEA), comprende una parte de la población dedicada a la producción de bienes y servicios de una sociedad, tipos de actividades que se pueden realizar a partir de una cierta edad[1], sin embargo la realidad es diferente ya que existen personas que desde muy corta edad trabajan debido a la escases de recursos en cada hogar.

1.2 Mercado laboral

El mercado laboral o mercado de trabajo; es una relación en la cual las personas intercambian servicios de trabajo; quienes compran servicios de trabajo son las empresas, es decir conforman la demanda de trabajo. Los que venden servicios de trabajo son los trabajadores y ellos conforman la oferta de trabajo.

Entonces se define al mercado laboral como un mercado donde interactúan la demanda y la oferta de trabajo que además se caracteriza por:

- Capacidad que tengan los empleadores para contratar a uno o más trabajadores.
- Leyes de Contratación Laboral.
- Salario o remuneración.
- Pago de beneficios a los trabajadores.
- Jornada Laboral o carga horaria de trabajo.
- Participación de los trabajadores en las utilidades.
- Indemnización al término de un contrato laboral.

Las características mencionadas son las cláusulas que forman parte de un contrato de trabajo las cuales se compromete a cumplir tanto el empleado como el empleador.

---

[1] OIT, "El límite de edad mínima usado para definir la población en edad de trabajar debe ser fijado por cada país, en función de sus circunstancias nacionales, como la edad de escolaridad obligatoria, la edad mínima de admisión en el empleo y las proporciones de trabajo infantil. Algunos países también utilizan un límite de edad máxima para definir este concepto, aunque las normas internacionales no lo recomiendan".De acuerdo a éstas características en el Ecuador, el INEC, estableció la edad mínima de 10 años y con respecto a la edad máxima no se considera ningún límite.

1.3 Evolución de la Población Económicamente Activa

En sociedades como la ecuatoriana no todas las personas que están en edad de trabajar laboran normalmente (están plenamente ocupadas) o están abiertamente desempleadas. Además de las dos situaciones normales se presenta también el caso de personas que aparentemente están trabajando pero no lo hacen de manera adecuada, ya sea porque laboran menos tiempo que lo que exige una jornada semanal de trabajo o porque, pese a hacerlo durante ese tiempo o inclusive un mayor número de horas, no perciben una apropiada remuneración por tal labor.

Estas situaciones generalmente son catalogadas como subempleo por tiempo o subempleo por ingresos, a lo cual se podría añadir el subempleo por calificación, cuando ciertas personas realizan una labor que les exige un menor nivel de capacitación que el que poseen los individuos que a tal trabajo se dedican.[2]

Pese a la aparente claridad de estas modalidades del denominado subempleo, es útil y conveniente evaluar adecuadamente lo que puede estar aconteciendo detrás de estas situaciones ya que, por ejemplo, bajo la modalidad de subempleo por tiempo se presume que existe desperdicio de la capacidad productiva de la mano de obra y la consecuente solución parecería ser incrementar el tiempo de trabajo, más lo que puede acontecer en la realidad es que quien labora pocas horas o menor número de días lo hace por propia decisión en vista que su tiempo de dedicación le genera suficientes ingresos y el tiempo restante lo dedica a otro quehacer que no se considera económicamente productivo, por ejemplo efectuar estudios o atender las tareas del hogar.
Por el contrario, las personas que laboran mucho más tiempo que la jornada catalogada como normal podrían estar deseando trabajar más horas diarias o semanales y, en efecto, pueden hacerlo en jornadas de diez o doce horas diarias durante seis o siete días a la semana, lo cual no implica mejor condición de

---

[2]Consejo Nacional de Población y Fondo de Población de las Naciones Unidas, Caracterización de la Fuerza de Trabajo Ecuatoriana y su Evolución, CONADE-FNUAP, Proyecto ECU – 80 –P04, Quito, s/f, pp. 105-106.

ocupación que las anteriores, sino que encubren la mala retribución a su tiempo de trabajo y ocultamiento de la ineficaz distribución de los ingresos.[3]

Dicho de otra manera, el denominado subempleo por ingresos en lugar de denotar desperdicio de la capacidad productiva, puede estar ocultando improductividad del sistema económico o explotación a la fuerza de trabajo. Y el subempleo por tiempo, que evidencia pocas horas de dedicación al trabajo y por tanto desperdicio de capacidad productiva, podría evidenciar afanes de trabajar más pero no a causa del real ánimo de hacerlo sino porque la explotación laboral repercute en que estos trabajadores perciban ingresos tan bajos que les impiden satisfacer sus necesidades primordiales.

Además suele añadirse la condición de subempleo por calificación con el cual, en lugar de destacar la alta calificación que tienen los que poseen mayor nivel educativo al compararlos con los elementos de un mismo grupo de trabajo, se destaca el desperdicio de capacidad productiva de unos trabajadores, comparativamente con quienes desempeñan una labor semejante aunque carecen de las necesarias calificaciones. Es decir, la carencia de calificación de unos determina que los otros aparenten ineficacia o desperdicio de sus capacidades.

Por cierto, en estas situaciones sí suelen haber casos en que la capacidad productiva se desperdicia, ya que, pese a tener niveles de instrucción universitaria para ejercer como médicos o abogados, por ejemplo, laboran de vendedores de almacén o transportistas de vehículos, lo cual ejemplifica el desperdicio de su nivel de conocimientos y su capacidad productiva, pero también denota irracionalidad en los esquemas educativos a más de deficiencias en el aparato productivo o empresarial y la obvia carencia de una planificación socioeconómica general de la sociedad que es objeto de análisis.

---

[3] Véase más ampliamente en el libro de Pachano Simón, **Población, Migración y Empleo en el Ecuador**, Serie de Antología de las Ciencias Sociales, el artículo de Gordillo Montalvo José Estudio crítico del denominado Subempleo en el Ecuador, ILDIS, Quito, 1988., pp. 267-286.

De acuerdo a una normativa internacional para las investigaciones sobre mercado laboral y empleo, las recomendaciones se las emite desde la Oficina Internacional de Trabajo OIT en Ginebra, la forma como las elabora son las reuniones de expertos estadísticos internacionales llamadas Conferencias Internacionales de Estadísticos de Trabajo CIET, realizadas en Ginebra con una periodicidad de cada dos años.

Adicionalmente, siguiendo las recomendaciones de OIT-CIET, las reuniones de expertos gubernamentales de los países de la Comunidad Andina de Naciones CAN organizadas por la división de estadística de la Secretaría General, emite también recomendaciones de armonización para la producción estadística en materia de empleo, las que deben ser acogidas por sus países miembros.

Según éstas condiciones y a la realidad de nuestro país, el Instituto Nacional de Estadísticas y Censos, en la metodología de la Encuesta de Empleo utiliza los conceptos siguientes para los estados de ocupación de la Población Económicamente Activa.

1.4 Estados de Ocupación de la PEA

1.4.1 Ocupados[4]

Son aquellas personas de 10 años y más que trabajaron al menos una hora en la semana de referencia o pese a que no trabajaron, tienen trabajo del cual estuvieron ausentes por motivos tales como: vacaciones, enfermedad, licencia por estudios, etc. Se considera ocupadas también a aquellas personas que realizan actividades dentro del hogar por un ingreso, aunque las actividades desarrolladas no guarden las formas típicas de trabajo asalariado o independiente.

---

[4] Definiciones tomadas de la Publicación de la Encuesta ENEMDU, Junio 2010.Instituto Nacional de Estadísticas y Censos.

1.4.2 Ocupados Plenos[4]

La población con ocupación plena está constituida por personas ocupadas de 10 años y más, que trabajan como mínimo la jornada legal de trabajo y tienen ingresos iguales o superiores al salario unificado legal y no desean trabajar más horas (no realizaron gestiones).

1.4.3 Desempleados[4]

Son las personas de 10 años y más que no tienen trabajo, pero que tomaron medidas concretas para buscar un trabajo asalariado o independiente en las últimas cuatro semanas y además, están disponibles para trabajar.

El desempleo total constituye la suma del desempleo abierto más el oculto, como están definidos a continuación.

Desempleo abierto[4].-Personas de 10 años y más, que en el período de referencia presentan simultáneamente las siguientes características:

- Sin empleo, no ocupado en la semana pasada.
- Buscaron trabajo, hicieron gestiones concretas para conseguir empleo o para establecer algún negocio en las cuatro semanas anteriores.

Desempleo oculto[4].- Personas de 10 años y más que en la semana de referencia presentan simultáneamente las siguientes características:

- Sin empleo, no ocupado en la semana pasada.
- No buscaron trabajo, (no hicieron gestiones concretas para conseguir empleo o para establecer algún negocio en las cuatro semanas anteriores), por alguna de las siguientes razones:
    - Tiene un trabajo esporádico u ocasional.
    - Tiene un trabajo para empezar inmediatamente.

- Espera respuesta por una gestión en una empresa o negocio propio.
- Espera respuesta de un empleador o de otras gestiones efectuadas para conseguir empleo.
- Espera cosecha o temporada de trabajo.
- Piensa que no le darán trabajo o se cansó de buscar.
- No cree poder encontrar.
- Disponible para trabajar.

1.4.3.1 Características del Desempleo[8]

Existen dos clases de desempleo: voluntario e involuntario. En el punto de equilibrio, todos los trabajadores que lo deseen encuentran un empleo, pero habrá una cierta cantidad de personas que no estarán dispuestas a trabajar por no encontrar los salarios que maximicen su bienestar, esto es el desempleo voluntario. Si algún factor externo como, recesiones en el ciclo económico, sindicatos o gobierno impide la generación de lugares de trabajo, aparecerá el desempleo involuntario.

La tasa de desempleo, en un punto del tiempo, es la fracción de personas que quieren trabajar pero no pueden encontrar empleo. Esta tasa, está en principio compuesta de diferentes tipos de desempleo:

- **Desempleo friccional**

Dado el hecho de que los empleos tienden a destruirse o desaparecer, y que toma tiempo para que los desempleados temporales encuentren empleadores nuevos que sean apropiados, siempre habrá una tasa positiva friccional de desempleo[5].

Se trata de un desempleo voluntario, en la medida en que los desempleados deciden tomarse su tiempo al buscar trabajo, tratando de maximizar su bienestar, al comparar los costos del desempleo con el valor presente de los futuros ingresos generados vía salario.

---

[5]Mankiw, Gregory; Principles of Macroeconomics; South-Western College Pub; 4 edition 2006.

Para aliviar este tipo de desempleo se requiere mejorar los sistemas de información, por ejemplo, mediante la creación de un sistema de intermediación laboral que suministre información sobre las vacantes existentes por tipo de calificación, de esta forma, se lograría disminuir el tiempo requerido para encontrar empleo.[6]

Cuando el trabajador no acepta ciertas colocaciones porque tiene la expectativa de conseguir otra mejor si espera un tiempo prudencial, se habla de desempleo de precaución o especulativo. En todo caso el desempleo friccional, en las sociedades modernas, que usan ampliamente diversos medios de comunicación, resulta bastante reducido en términos porcentuales.

- **Desempleo cíclico[7]**

En todas las economías existen períodos de contracción y otros de expansión de la economía. Se producen shocks tanto a la demanda como a la oferta, y los cambios adversos de este tipo causan reducciones en la demanda de trabajo que a su vez hacen que muchos trabajadores pierdan sus empleos, este desempleo se conoce como cíclico (Uc). En este sentido es temporal –se espera que dure no más de dos años[8].

Cuando el desempleo cíclico es igual a cero, se dice que la economía se encuentra en una situación de pleno empleo, en el sentido en que la demanda de fuerza de trabajo es igual a la oferta. En esta situación no habría presiones de costos laborales que lleven a elevar o a disminuir la tasa de inflación. Este nivel se conoce como la "tasa desempleo de inflación estable".

Además de estos tipos de desempleo se puede hablar de desempleo estacional, para referirse al que se produce por la demanda fluctuante que existe en ciertas actividades, como la agricultura, por ejemplo. En este caso existen períodos del año económico que requieren de mucha mano de obra -como la temporada de cosecha, por ejemplo- y otros en que la demanda se reduce notablemente. El desempleo estacional se hace menor cuando las personas tienen posibilidades de ocuparse en otras ramas de actividad durante el período en que desciende la

---

[6] Aguinaga C, Criollo C, Mercado de trabajo: diseño y metodología de investigación, pag.21.
[7] Víctor Aguiar (2007): El Mercado Laboral Ecuatoriano: Propuesta de una reforma Edición electrónica, pag.10-11.
[8] Larraín, F; Sachs, Jeffrey; Macroeconomía en la economía global; Segunda Edición; Pearson-Prentice Hall; Chile 2004

demanda de trabajo. Suele hablarse también de desempleo tecnológico cuando éste es producido por cambios en los procesos productivos que hacen que las habilidades que poseen los trabajadores no resulten ya útiles, en tanto que puede no haber suficientes personas entrenadas en las nuevas técnicas como para satisfacer la demanda.

El desempleo, además, es producido también por factores de tipo legal o institucional: cuando las autoridades de un país fijan un salario mínimo impiden que se ofrezcan colocaciones a un nivel inferior a éste, con lo que no es posible el descenso de los salarios reales y se dificulta el crecimiento de la demanda de trabajo; del mismo modo opera la fijación de salarios diferenciales para diversas categorías de trabajadores y cualquier limitación al desplazamiento de la mano de obra, ya sea geográfica o según ramas de actividad. Las limitaciones que pueden imponer los sindicatos y grupos de presión a la contratación de mano de obra -según edad, sexo o nacionalidad- o los requisitos formales de aprendizaje que se establezcan operan en el mismo sentido. Por último, se ha comprobado que el propio seguro de desempleo, cuando ofrece compensaciones tan altas o tan largas que desestimulan la búsqueda de trabajo, aumenta el desempleo global de un país, y lo mismo sucede cuando se establecen subsidios directos a sectores pobres de la población que estarían, de otro modo, dispuestos a conseguir una colocación. En estos casos las personas reciben un ingreso que, aunque menor que el salario, se ofrece sin la contraprestación de su trabajo; la relación coste / beneficio se altera así de modo sustancial, por lo que se reduce notablemente la oferta de trabajo. La combinación de estas compensaciones de pobreza o desempleo con la fijación de un salario mínimo alientan notablemente el desempleo de una sociedad y estimulan en muchos casos, paralelamente, la emergencia de la economía informal.

- **Desempleo estructural**[9]

Se origina en los desajustes entre la calificación de la oferta laboral y los requerimientos de los empleadores, dentro de un área geográfica común; o en los desequilibrios regionales entre la oferta y la demanda de fuerza de trabajo. En presencia de perfecta movilidad de mano de obra y de capacitación y educación adecuadas, esta variante no debería existir, sin embargo, los costos de capacitación y de transporte de la mano de obra (de una región a otra) son elevados. En consecuencia, el desempleo estructural puede mantenerse y crecer en el tiempo.

Dos factores pueden aclarar la presencia de desempleo estructural:

• Cambios en la estructura de producción, que desajustan la calificación de la mano de obra y los requerimientos de ciertos empleos. Esta situación impide a las personas que buscan empleo llenar las vacantes disponibles. Los procesos de ajuste económico realizados en América Latina han propiciado este tipo de desempleo, al obligar a los empresarios a reconstituir o tecnificar la producción, generando una nueva demanda de mano de obra, más calificada. Esto puede causar, simultáneamente, excesos de oferta de mano de obra entre la población activa no calificada.

• Los desbalances regionales, como cambios en la estructura de la producción, pueden propiciar excesos de demanda de trabajo en unas regiones, con desempleo en otras. Si la movilidad de la mano de obra fuese perfecta, se produciría una rápida migración y desaparecería el desempleo. Sin embargo, los flujos de información interregional, suelen ser deficientes. Además, la movilización representa costos elevados (a los que habría que añadir los costos de cambio de vivienda). Por último, los temores psicológicos y las barreras culturales contribuyen a inhibir la movilidad regional.

El desempleo estructural se explica, entonces, por los altos costos requeridos para calificar mano de obra o para movilizarla. La forma más eficiente de contrarrestar este problema sería recurrir a programas de recalificación de los

---

[9] Víctor Aguiar (2007): El Mercado Laboral Ecuatoriano: Propuesta de una reforma Edición electrónica, pag.11.

cesantes, así como a la readecuación de los programas educativos para los futuros aspirantes. Sistemas de información ágiles y, programas de apoyo a la migración contribuirían a reducir este tipo de desempleo.[10]

1.4.4 Subempleados[4]

Se considera personas subempleadas a quienes se encuentran dentro de los siguientes grupos:

Subempleados por insuficiencia de horas[4].- Personas que trabajan menos de 40 horas a la semana, y están buscando trabajar más horas ya sea en el trabajo actual o en otro.

Otras formas de subempleo[4].- Personas que cumplen cualquiera de las siguientes características:

- Trabajan 40 horas o más, tienen ingresos superiores al salario unificado legal [11] y están disponibles para trabajar más horas.

- Personas ocupadas que trabajan 40 horas o más, tienen ingresos menores al salario unificado legal.
- Personas ocupadas que trabajan menos de 40 horas, tienen ingresos menores al salario unificado legal y no están dispuestos o disponibles a trabajar más.

1.5 CLASIFICACIÓN DE LA POBLACIÓN

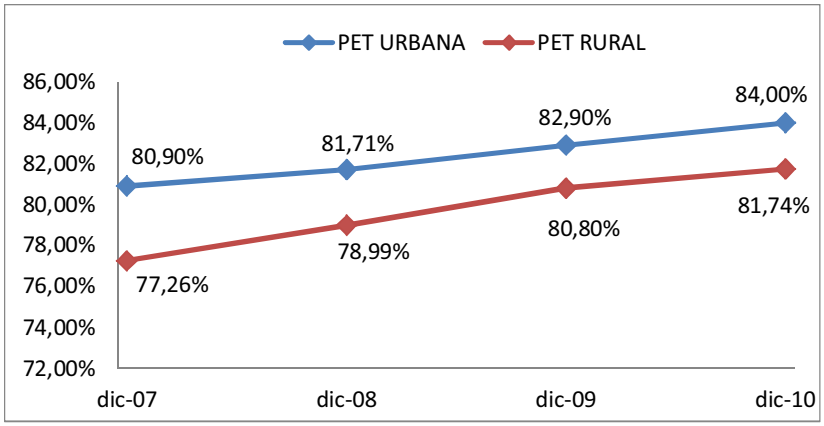
1.5.1 Población en Edad de Trabajar (PET)

A diciembre del 2010[12], la PET del área urbana se ubicó en 84,0%[13], y en el área rural (81,74%)[13]; las dos series mantienen una tendencia creciente.

---

[10] (Mercado de Trabajo: Diseño y Metodología de la Investigación, págs. 20-21)
[11] El salario unificado legal tiene las siguientes categorías: Trabajadores en general = $264, Empleadas Domésticas =$264, valores que se actualizan cada año por el Consejo nacional de salarios en función de lo dispuesto del Código de Trabajo. Para los Empleados Públicos= $500, de acuerdo a las disposiciones de la SENRES.
[12] Se toma en cuenta el análisis a diciembre de cada año, porque es en este mes en el cual la encuesta de empleo que realiza el INEC, tiene cobertura a nivel nacional.

La población menor a 10 años que sumada a la PET representa el total de la población del país, fue de 16,0% para el área urbana y 18,3% para el área rural del país, respectivamente.



Fuente: Bases de la ENEMDU Dic 2007-Dic 2010
Elaboración: Paola Troya

Gráfico 1: PET URBANA Y RURAL

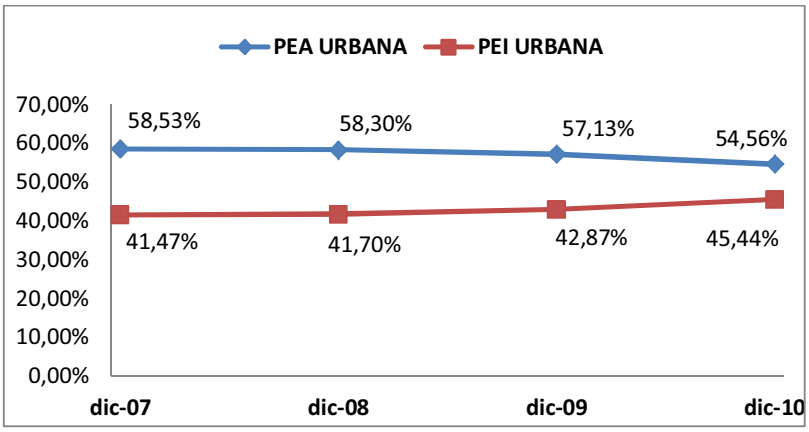
1.5.2 Población Económicamente Activa (PEA) y Población Económicamente Inactiva (PEI)

Desde diciembre 2007[12] la PEA en el área urbana (58,53%) ha mantenido una tendencia constante hasta diciembre 2009 (57,13%) y a diciembre 2010(54,56%) tiende a disminuir; y la PEI (41,47% a diciembre 2007), de la misma manera con tendencia constante hasta diciembre 2009 (42,87%) y al alza a diciembre 2010 (45,44%). (Gráfico 2)[14].

En el área rural desde diciembre 2007 la PEA (63,23%) ha mantenido una tendencia a la baja hasta diciembre 2010(55,83%) y la PEI (36,77% a diciembre 2007), de la misma manera con tendencia a subir a diciembre 2010 (44,17%). (Gráfico 3)[14].

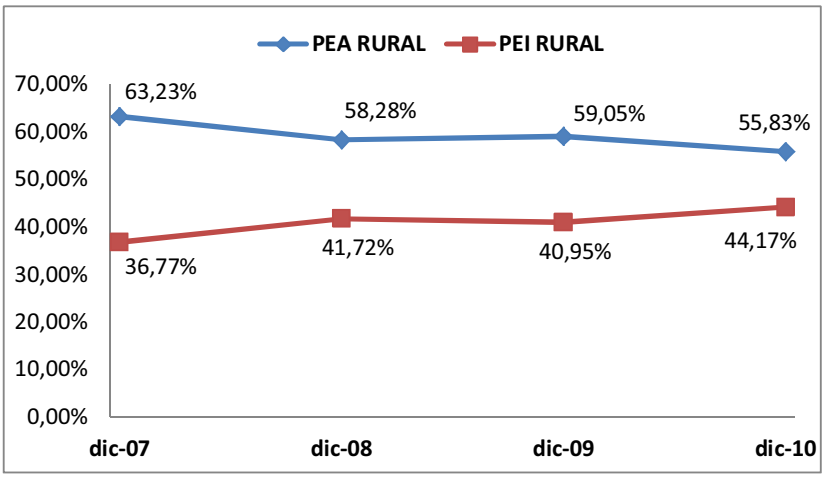
---

[13] Porcentaje con respecto a la población total.
[14] Los porcentajes del gráfico 2 y gráfico 3 son calculados con respecto a la Población en Edad de Trabajar.



Fuente: Bases de la ENEMDU Dic 2007-Dic 2010.INEC
Elaboración: Paola Troya

Gráfico 2: PEA Y PEI URBANA



Fuente: Bases de la ENEMDU Dic 2007-Dic 2010.INEC
Elaboración: Paola Troya

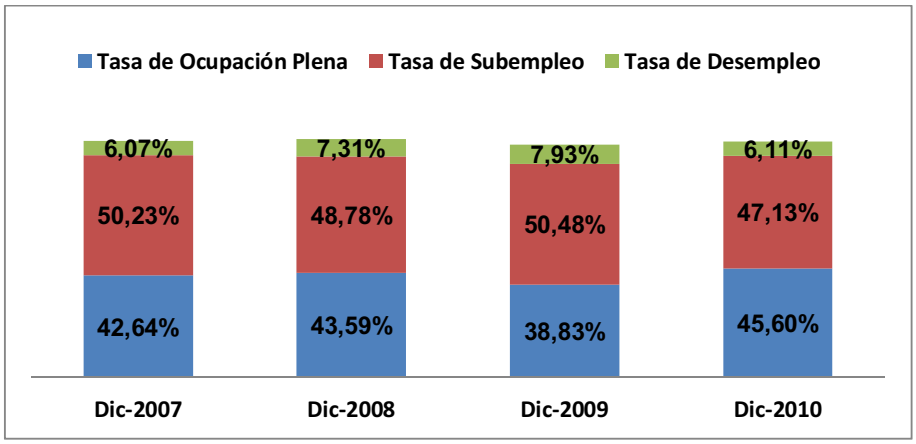
Gráfico 3: PEA Y PEI RURAL

1.6 DISTRIBUCIÓN DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA (PEA)

Al mes de diciembre 2010[12], la población subocupada representa en el área urbana y rural del país el 47,13% y 76,1% respectivamente, seguida por los ocupados plenos que registraron en los sectores urbano y rural el 45,6% y 21,0% respectivamente. Finalmente, los desocupados urbanos fueron del 6,11% y los rurales del 2,8%.[15]

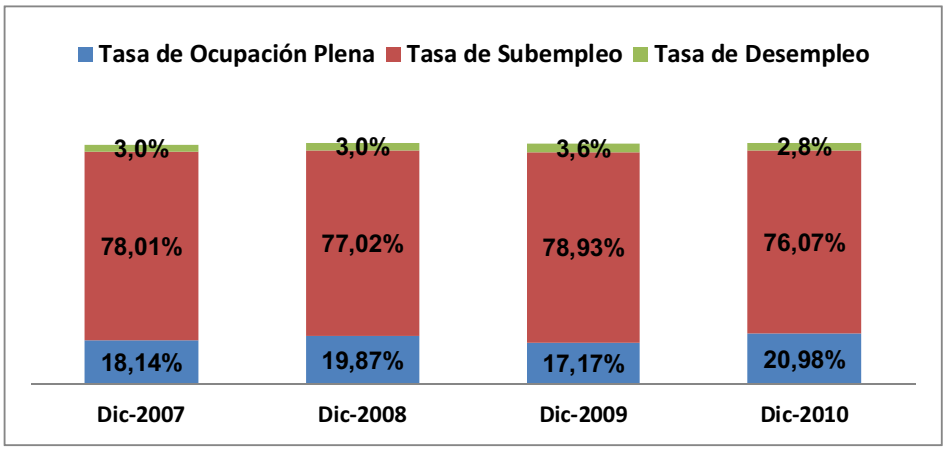
---

[15] Los porcentajes de los gráficos 4 y 5 son calculados con respecto a la Población Económicamente Activa.



Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración: Paola Troya

Gráfico 4: PEA URBANA



Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración: Paola Troya

Gráfico 5: PEA RURAL

1.7 INDICADORES LABORALES DE LA POBLACIÓN ECONÓMICAMENTE ACTIVA[16]

Con los datos de población y sus características en el mercado laboral se puede obtener varios indicadores de acuerdo al análisis que se requiera.

A continuación presentamos los indicadores más importantes para situación laboral:

[16] Las definiciones y cálculos de los indicadores son tomadas de la Metodología de la Encuesta de Empleo, Desempleo y Subempleo, ronda Diciembre 2008,pag 39, INEC.

1.7.1 TASAS DE PARTICIPACIÓN

a. Tasa de Participación Bruta.-porcentaje de la Población Económicamente Activa (PEA) respecto de la Población Total (PT).

$$TPB = \frac{PEA}{PT} \times 100$$

b. Tasa de Participación Global.- porcentaje de la Población Económicamente Activa (PEA) respecto de la Población en Edad de Trabajar (PET). Esta tasa resulta más adecuada para medir la participación ya que aísla fenómenos de tipo demográfico.

$$TPG = \frac{PEA}{PET} \times 100$$

c. Tasa de Participación Específica.- porcentaje de la Población Económicamente Activa de un determinado grupo (jóvenes, mujeres, población con educación primaria, etc.) respecto de la población de ese grupo.

$$TPE_i = \frac{PEA_i}{PTi} \times 100$$

1.7.2 TASAS DE OCUPACIÓN

a. Tasa de Ocupación Bruta.- porcentaje del total de ocupados (O) respecto de la Población en Edad de Trabajar (PET).

$$TOB = \frac{O}{PET} \times 100$$

b. Tasa de Ocupación Global.- porcentaje del total de ocupados (O) respecto a la Población Económicamente Activa (PEA).

$$TOG = \frac{O}{PEA} \times 100$$

c. Tasa de Ocupados Plenos.- porcentaje del total de ocupados plenos (OP) respecto a la Población Económicamente Activa (PEA).

$$TOP = \frac{OP}{PEA} \times 100$$

1.7.3 TASAS DE SUBEMPLEO

a. Tasa de Subempleo por insuficiencia de horas (Visible).- porcentaje de los subempleados visibles (SV) respecto de la Población Económicamente Activa (PEA).

$$TSV = \frac{SV}{PEA} \times 100$$

b. Tasa de Otras Formas de Subempleo.- porcentaje de la cantidad de ocupados que está clasificado en otras formas de subempleo (OFS) respecto de la Población Económicamente Activa (PEA).

$$TOFS = \frac{OFS}{PEA} \times 100$$

c. Tasa de Subempleo Bruta.- Corresponde a la suma de la tasa de Subempleo Visibles (TSV) y de Otras Formas de Subempleo (TOFS).

$$TSB = TSV + TOFS$$

d. Tasa de Subempleo Global.- porcentaje de la suma de los Subempleados Visibles, más los Subempleados de otras formas, respecto del total de Ocupados.

$$TSG = \frac{SV + OFS}{OCUPADOS} \times 100$$

1.7.4 TASAS DE DESEMPLEO

a. Tasa de Desempleo.- porcentaje del número de desocupados (D) respecto de la población económicamente activa (PEA).

$$TD = \frac{D}{PEA} \times 100$$

a.1. Tasa de Desempleo Abierto.- porcentaje del número de desempleados abiertos (DA) respecto de la población económicamente activa (PEA).

$$TDA = \frac{DA}{PEA} \times 100$$

a.2. Tasa de Desempleo Oculto.- porcentaje del número de desempleados ocultos (DO) respecto de la población económicamente activa (PEA).

$$TDO = \frac{DO}{PEA} \times 100$$

b. Tasa de Cesantes.- porcentaje de los cesantes (C) respecto de la Población Económicamente Activa (PEA).

$$TC = \frac{C}{PEA} \times 100$$

1.7.5 TASAS DE SECTORIZACIÓN DE LA PEA

a. Tasa de Ocupados del Sector Formal.- porcentaje de número de ocupados del Sector Formal (SF) respecto de la Población Económicamente Activa (PEA).

$$TOSF = \frac{SF}{PEA} \times 100$$

b. **Tasa de Ocupados del Sector Informal.-** porcentaje del número de ocupados del Sector Informal (SI) respecto de la Población Económicamente Activa (PEA).

$$TOSI = \frac{SI}{PEA} \times 100$$

c. **Tasa de No clasificados por sector.-** porcentaje del número de ocupados no clasificados por sector (NCS) respecto de la Población Económicamente Activa (PEA).

$$TNCS = \frac{ONCS}{PEA} \times 100$$

d. **Tasa de Ocupados del Servicio Doméstico.-** porcentaje del número de ocupados del Servicio Doméstico (SD) respecto de la Población Económicamente Activa (PEA).

$$TOSD = \frac{SD}{PEA} \times 100$$

1.7.6 TASAS DE SUBUTILIZACIÓN

a. **Tasa de Subutilización Bruta.-** Es la suma de la tasa de Subempleo Bruta (TSB) y la Tasa de Desempleo (TD).

$$TS_U B = TSB + TD$$

b. **Tasa de Subutilización Neta.-** Es la suma de la tasa de desempleo equivalente al subempleo (TDE) y la tasa de Desempleo (TD).

$$TS_U N = TDE + TD$$

CAPÍTULO 2

TÉCNICAS DE ESTADÍSTICA MULTIVARIANTE

El análisis de datos multivariante comprende el estudio estadístico de dos o más variables que pueden estar relacionadas, y tiene los siguientes objetivos[17]:

- Resumir los datos mediante un pequeño conjunto de nuevas variables, construidas como transformaciones de las originales, con la mínima pérdida de información.
- Encontrar grupos en los datos, si existen.
- Clasificar nuevas observaciones en grupos definidos.
- Relacionar dos conjuntos de variables.

La descripción de una situación cualquiera, implica la utilización de muchas variables, la cual se simplifica mediante la construcción de uno o varios índices o indicadores que la resumen.

Por ejemplo, el crecimiento de los precios en una economía se resume en un índice de precios, la calidad de una universidad o de un departamento se resume en unos pocos indicadores, las dimensiones del cuerpo humano se resumen en la ropa de confección en unas pocas variables indicadoras del conjunto.

Cuando se dispone estos indicadores, se tiene varias ventajas: si son pocos pueden ser representados gráficamente y comparar distintos conjuntos de datos o instantes en el tiempo; simplifican el análisis al permitir trabajar con un número menor de variables; si las variables indicadoras pueden interpretarse, se mejoraría el conocimiento de la realidad estudiada.
El análisis multivariante de datos proporciona métodos objetivos para conocer cuántas variables indicadoras (o factores), son necesarias para describir una realidad compleja y determinar su estructura.

---

[17] Debido a que las variables se consideran en forma simultánea, las técnicas de análisis multivariante permiten realizar interpretaciones más complejas a las que surgen mediante la utilización de métodos univariados.

El segundo objetivo es identificar grupos si existen. Por ejemplo, si se tiene un conjunto de variables en empresas, se espera que los datos indiquen una división de las empresas en grupos en función de su rentabilidad, su eficacia comercial o su estructura productiva. En muchas situaciones los grupos son desconocidos a priori, entonces se requiere disponer de un procedimiento objetivo para obtener los grupos existentes y clasificar las observaciones. Un tercer objetivo relacionado con el anterior aparece cuando los grupos están bien definidos a priori y se quiere clasificar nuevas observaciones. Por ejemplo, para clasificar a clientes que solicitan créditos como fiables o no, personas como enfermas o no, o diseñar una máquina que clasifique monedas o billetes en clases prefijadas.

Un problema distinto es relacionar dos conjuntos de variables, por ejemplo, al disponer de un conjunto de variables de capacidad intelectual y otros de resultados profesionales, se necesita relacionar ambos conjuntos de variables. En particular, los dos grupos de variables pueden corresponder a las mismas variables medidas en dos momentos distintos en el tiempo o en el espacio y se pretende ver la relación entre ambos conjuntos.

Las técnicas de análisis multivariante tienen aplicaciones en todos los campos científicos y comenzaron desarrollándose para resolver problemas de clasificación en Biología, se extendieron para encontrar variables indicadoras y factores en Psicometría, Marketing y las Ciencias sociales y han alcanzado una gran aplicación en Ingeniería y Ciencias de la computación como herramientas para resumir la información y diseñar sistemas de clasificación automática y de reconocimiento de patrones.

Con bases a los anteriores objetivos, al igual que la naturaleza de cada variable incluida en el estudio (cualitativa o cuantitativa) y la relación existente entre ellas (explicativas y explicadas) se han desarrollado diferentes técnicas de análisis multivariante de datos, de las cuales las más utilizadas son:

- Métodos Factoriales: Análisis Factorial, Análisis de Componentes. Principales y Análisis de Correspondencias.
- Escalamiento Multidimensional.
- Regresión Lineal Simple y Múltiple.
- Regresión Logística.
- Análisis Discriminante.
- Análisis Conjunto.
- Análisis de Varianza y de Covarianza: Anova, Ancova, Manova y Mancova.
- Segmentación: Métodos de Agrupamiento Jerárquicos y No Jerárquicos.

(Peña, 2002, págs. 13-17)
En la actualidad, el software que se utiliza para la estadística y econometría incluyen procedimientos para aplicar estas técnicas del análisis de datos.
Para el presente trabajo se va a utilizar dos de las técnicas antes mencionadas: Regresión Logística Multinomial y Análisis Discriminante.

2.1 MODELOS DE REGRESIÓN DE RESPUESTA CUALITATIVA[18]

Se consideran a los modelos de regresión en los cuales la variable dependiente o de respuesta puede ser de naturaleza cualitativa dicotómica o de categoría múltiple.

2.1.1 Modelo Lineal de Probabilidad[18]

Para establecer el concepto se considera el siguiente modelo simple como ejemplo:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{2.a}$$

[18] Damodar N.Gujarati.4ta edición. McGrawhill. Econometría, pág: 562-564.

Donde:[19]

X=ingresos.
Y=1 si una persona tiene un empleo.
Y=0 si una persona se encuentra desempleada.

Este es un modelo lineal de probabilidad (MLP) ya que la variable regresada es binaria,[20] esto se debe a que,$E[Y_i|X_i]$ la esperanza condicional de Y_i dado X_i, puede ser interpretada como la probabilidad condicional de que el evento suceda dado X_i; es decir, $Pr(Y_i = 1|X_i)$. Así, en el caso anterior, $E[Y_i|X_i]$ da la probabilidad de que una persona se encuentre ocupada y tenga un ingreso de una cierta cantidad X_i.
La justificación del nombre MLP para modelos como (2.a) puede ser la siguiente.

Suponiendo que $E(u_i) = 0$, se obtiene:

$$E[Y_i|X_i] = \beta_0 + \beta_1 X_i$$

(2a-1)

Ahora, permitiendo que P_i=probabilidad de que $Y_i = 1$(es decir, de que el evento ocurra) y 1-P_i=probabilidad de que $Y_i = 0$ (es decir, de que el evento no ocurra), la variable Y tiene la siguiente distribución (de probabilidad):

Y_i	$Probabilidad$
0	$1 - P_i$
1	P_i
$Total$	1

Es decir,Y_i sigue la distribución de probabilidad de Bernoulli.
Por consiguiente, la definición de esperanza matemática, se obtiene:

$$E[Y_i] = 0(1 - P_i) + 1(P_i) = P_i$$

[19] Considerando las categorías de la variable Y para el presente estudio.
[20] No se tiene que restringir la variable respuesta a dos categorías. En general se puede tener una variable de respuesta politómica o de categoría múltiple.

Comparando (2a-1) con (2a-2), se puede igualar

$$E[Y_i|X_i] = \beta_0 + \beta_1 X_i = P_i$$

(2a-2)

es decir, la esperanza condicional del modelo (2a-1) puede, de hecho, ser interpretada como la probabilidad condicional de Y_i. En general, la esperanza de una variable aleatoria Bernoulli está dada por la probabilidad de que esa variable sea igual a 1[21].

Puesto que la probabilidad P_i debe encontrarse entre 0 y 1, se tiene la restricción

$$0 \leq E(Y_i|X_i) \leq 1$$

es decir, la esperanza condicional (o probabilidad condicional) debe encontrarse entre 0 y 1.

Los modelos lineales de probabilidad plantean diversos problemas:

No normalidad de las perturbaciones u_i [22]

El supuesto de normalidad para u_i ya no se mantiene en los MLP porque, al igual que Y_i, u_i solamente toma dos valores; es decir, también siguen la distribución de Bernoulli. Para ver esto, se escribe (2a-1) como

$$u_i = Y_i - \beta_1 + \beta_2 X_i$$

(2a-3)

La probabilidad de distribución de u_i es:

	u_i	$Probabilidad$
$Cuando\ Y_i = 1$	$1 - \beta_1 - \beta_2 X_i$	P_i
$Cuando\ Y_i = 0$	$1 - \beta_1 - \beta_2 X_i$	$1 - P_i$

Obviamente, no puede suponerse que u_i esté normalmente distribuida; en realidad ésta sigue la distribución de Bernoulli.

---

[21] Si existen n intentos independientes, cada uno con una probabilidad p de éxito y con una probabilidad (1-p) de fracaso, y X de tales intentos representa el número de éxitos, entonces se dice que X sigue una distribución binomial. La media de la distribución binomial es np y su varianza np(1-p)

[22] Los Mínimos Cuadrado Ordinarios no requieren que las perturbaciones (u) estén normalmente distribuidas, se supone que lo están para fines de inferencia estadística.

Varianzas heteroscedásticas de las perturbaciones

Aun si $E(u_i) = 0$ y $E(u_i, u_j) = 0$, para $i \neq j$ (es decir no hay correlación serial), ya no es posible sostener la afirmación de que las perturbaciones en el MLP son homocedásticas.[23] Para la distribución del término del error dado en (2a-5), al aplicar la definición de varianza, se tiene que,

$$var(u_i) = P_i(1 - P_i)$$

(2a-4)

Es decir, la varianza del término de error en el MLP es heteroscedástica. Puesto que $P_i = E(Y_i|X_i) = \beta_0 + \beta_1 X_i$, la varianza de u_i depende de los valores de X y por tanto no es homoscedástica.

No cumplimiento de $0 \leq E(Y_i|X) \leq 1$

Puesto que $E(Y_i|X_i)$ en los modelos lineales de probabilidad mide la probabilidad condicional de que ocurra el evento Y dado X, ésta debe encontrarse necesariamente entre 0 y 1. Aunque a priori esto es verdadero, no hay garantía de que $\hat{Y}_i$, los estimadores de $E(Y_i|X_i)$, cumplan necesariamente esta restricción y éste es el problema real con la estimación Mínimos Cuadrados Ordinarios (MCO) del MLP. Hay dos formas de establecer si el $\hat{Y}_i$ estimado se encuentra entre 0 y 1.Una es estimar el MLP mediante el método usual MCO y determinar si el $\hat{Y}_i$ estimado se encuentra entre 0 y 1.

Si algunos valores son menores que 0 (es decir, negativos), para esos casos se supone que $\hat{Y}_i$ es cero; si son mayores de 1, se supone que son 1. El segundo procedimiento es diseñar una técnica de estimación que garantice que las probabilidades condicionales estimadas $\hat{Y}_i$ se encuentran entre 0 y 1. Los modelos logit y probit garantizarán que las probabilidades estimadas se encuentren con seguridad entre los límites lógicos 0 y 1.

---

[23] La teoría estadística lo demuestra, para una distribución Bernoulli, la media y la varianza teóricas son p y p(1-p) respectivamente, donde p es la probabilidad de éxito(es decir; que algo suceda),lo cual indica que la varianza es una función de la media. Por tanto, la varianza del error es heteroscedástica.

2.1.2 Regresión Logística[24]

Los modelos de regresión logística son modelos estadísticos en los que se desea conocer la relación entre una variable dependiente cualitativa, dicotómica (regresión logística binomial) o con más de dos valores (regresión logística multinomial); y una o más variables explicativas independientes (covariables), cualitativas o cuantitativas, siendo la ecuación inicial del modelo de tipo exponencial.

Las covariables cualitativas deben ser dicotómicas (tomando valores 0 para su ausencia y 1 para su presencia). Pero si la covariable cualitativa tuviera más de dos categorías, para su inclusión en el modelo se debe realizar una transformación de la misma en varias covariables cualitativas dicotómicas ficticias, de forma que una de las categorías se tomaría como categoría de referencia, de esta manera cada categoría entraría en el modelo de forma individual. En general, si la covariable cualitativa posee n categorías, habrá que realizar n −1 covariables ficticias.

El objetivo principal de esta técnica es el de modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. También puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (politómico).

2.1.2.1 Regresión Logística Binomial:

Es un modelo donde la variable de respuesta es dicotómica, para establecer el concepto se considera el siguiente modelo:

---

[24] http://www.seqc.es/es/Varios/7/40/Modulo_3:Regresion_Logística_y_múltiple/pág2-4

Daniel Peña. McGrawhill. Análisis de Datos Multivariante.pag 433-445

$$Prob(y = 1|x) = \frac{exp(b_0 + b_1X_1 + \ldots + b_nX_n)}{1 + exp(b_0 + b_1X_1 + \ldots + b_nX_n)}$$

(2.1)

Donde:

$Prob(y = 1|x)$ es la probabilidad de que y tome el valor 1 (probabilidad de éxito, es decir presencia de la característica estudiada), para determinados valores de las covariables X;

X es un conjunto de n covariables $\{x_1, x_2, \ldots, x_n\}$ que forman parte del modelo;
b_0 es la constante del modelo o término independiente;
b_i los coeficientes de las covariables.

Si se divide la expresión (2.1) por su complemento, es decir se construye su razón de probabilidades, se obtiene la siguiente expresión:

$$\frac{Prob(y = 1|x)}{1 - Prob(y = 1|x)} = exp\left(b_0 + \sum_{i=1}^{n} b_i\, X_i\right)$$

(2.2)

Ahora realizamos una transformación aplicando logaritmo natural a la ecuación:

$$In\left(\frac{Prob(y = 1|x)}{1 - Prob(y = 1|x)}\right) = b_0 + \sum_{i=1}^{n} b_i\, X_i$$

(2.3)

Si expresamos (2.3) de otra forma, se tiene:

$$g_i = log\,\frac{p_i}{1 - p_i} = b_0 + \sum_{i=1}^{n} b_i\, X_i$$

(2.4)

Donde, g_i, es el logaritmo de la razón de probabilidades, el cual es lineal en X y (desde el punto de vista de estimación) lineal en los parámetros; g_i es llamado logit y representa la razón de probabilidad de éxito respecto a probabilidad de fracaso, y al ser una función lineal de las variables explicativas nos facilita la estimación y la interpretación del modelo.

2.1.2.1.1 Estimación del Modelo Logit

Estimación de Máxima Verosimilitud

La estimación de los coeficientes del modelo y de sus errores estándar se realiza por el método de estimaciones de máxima verosimilitud, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente; y proporcionados por los datos de la muestra.
En un proceso iterativo, se van probando distintos valores de los parámetros b hasta que se encuentran los coeficientes que maximizan tal función de verosimilitud[25] (o sea, los coeficientes que más probablemente correspondan a los valores muestrales); tales coeficientes serán los estimados para nuestro modelo.

Supongamos que se dispone de una muestra aleatoria simple de n-elementos de una variable aleatoria p-dimensional, x, con función de densidad $f(x|\theta)$, donde $\theta = (\theta_1, \dots, \theta_r)'$ es un vector de parámetros que se supone tiene $r \leq pn$. Sea $X = (x_1, \dots, x_n)$ los datos muestrales, la función de densidad conjunta de la muestra, será, por la independencia de las observaciones:

$$f(X|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{2.4.1}$$

Cuando el parámetro θ es conocido, esta función determina la probabilidad de aparición de cada muestra. En el problema de estimación se dispone de la muestra, pero θ es desconocido.

---

[25] De hecho, no se maximiza directamente la función de verosimilitud, sino su logaritmo (log likelihood), lo que simplifica la computación. Este log likelihood es el que aparece en la salida de Stata de logit.

Considerando en la expresión de la densidad conjunta a θ como una variable y particularizando esta función para los datos observados, se obtiene una función que llamaremos función de verosimilitud $\boldsymbol{l(\hat{\theta}|X)}$**, o** $\boldsymbol{l(\theta)}$**:**

$$\boldsymbol{l(\theta|X) = l(\theta) = \prod_{i=1}^{n} f(x_i|\theta)} \qquad X\ fijo;\ \theta\ variable$$

(2.4.2)

El estimador de máxima verosimilitud, o estimador MV, es el valor de $\boldsymbol{\theta}$ que hace máxima la probabilidad de aparición de los valores muestrales efectivamente observados y se obtiene calculando el valor máximo de la función $\boldsymbol{l(\theta)}$**.** . Suponiendo que esta función es diferenciable y que su máximo no ocurre en un extremo de su dominio de definición, el máximo se obtendrá resolviendo el sistema de ecuaciones:

$$\frac{\partial l(\theta)}{\partial \theta_1} = 0$$

$$\vdots$$

$$\frac{\partial l(\theta)}{\partial \theta_r} = 0$$

El vector $\hat{\theta}$ que satisface este sistema de ecuaciones corresponderá a un máximo si la matriz hessiana de segundas derivadas H, evaluada en $\hat{\theta}$, es definida negativa:

$$H(\hat{\theta}) = \left(\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}\right)_{\theta=\hat{\theta}} \text{ definida negativa.}$$

(2.4.3)

En ese caso $\hat{\theta}$ es el estimador de máxima verosimilitud o estimador MV de θ. En la práctica suele ser más cómodo obtener el máximo del logaritmo de la función de verosimilitud:

$$L(\theta) = \ln l\,(\theta)$$

(2.4.4)

También conocida como la función soporte. Como el logaritmo es una transformación monótona, ambas funciones tienen el mismo máximo, pero trabajar con el soporte tiene dos ventajas principales.

En primer lugar pasamos del producto de densidades (2.4.2) a la suma de sus logaritmos y la expresión resultante suele ser más simple que la verosimilitud, con lo que resulta más cómodo obtener el máximo. En segundo lugar, al tomar logaritmos las constantes multiplicativas de la función de densidad, que son irrelevante para el máximo, se hacen aditivas y desaparecen al derivar, con lo que la derivada del soporte tiene siempre la misma expresión y no depende de constantes arbitrarias. En tercer lugar, el doble de la función soporte cambiada de signo proporciona un método general para juzgar el ajuste de un modelo a los datos que se denomina desviación:

$$D(\boldsymbol{\theta}) = -\mathbf{2}\boldsymbol{L}(\boldsymbol{\theta})$$

(2.4.5)

y la desviación $D(\boldsymbol{\theta})$ mide la discrepancia entre el modelo y los datos. Cuanto mayor sea el soporte,$L(\theta)$ mayor es la concordancia entre el valor del parámetro y los datos y menor la desviación. La desviación aparecerá de manera natural en el contraste de hipótesis y es una medida global de ajuste de un modelo a los datos.

Ahora, se supone una muestra aleatoria de datos (x_i, y_i), i=1,…, n. La función de probabilidades para una respuesta y_i cualquiera es:

$$P(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \qquad y_i = 0{,}1$$

y para la muestra:

$$P(y_1, \dots, y_n) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i}$$

(2.5)

Tomando logaritmos de la ecuación anterior se obtiene:

$$logP(\boldsymbol{y}) = \sum_{i=1}^{n} y_i \, log\left(\frac{p_i}{1 - p_i}\right) + \sum_{i=1}^{n} log(1 - p_i)$$

(2.6)

la cuál es la función soporte (de verosimilitud en logaritmos) que puede escribirse como:

$$logP(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i \, log \, p_i + (1 - y_i) log(1 - p_i)).$$

(2.7)

donde $\boldsymbol{\beta}' = (b_0, b_1, \dots, b_p)$ es el vector de p+1 componentes (estimadores), incluyendo la constante b_0 que determina las probabilidades p_i de aparición de cada muestra. Maximizar la verosimilitud puede expresarse como minimizar una función que mide la desviación entre los datos y el modelo.

De acuerdo a la definición de la ecuación(2.4.5), la desviación para el modelo será:

$$D(\boldsymbol{\beta}) = -2 \sum_{i=1}^{n} (y_i \log p_i + (1 - y_i) \log(1 - p_i)).$$

(2.7.1)

y hablaremos indistintamente de de maximizar el soporte o minimizar la desviación del modelo. Se define la desviación de cada dato por:

$$d_i = -2\big(y_i \log p_i + (1 - y_i) log(1 - p_i)\big)$$

(2.7.2)

y miden el ajuste del modelo al dato (y_i, x_i).En efecto, ya que los $\widehat{p_i}$ son menores que uno, sus logaritmos son negativos, por lo que la desviación es siempre

positiva. Además, en el cálculo de la desviación sólo interviene uno de sus dos términos, ya que y_i solo puede valer cero o uno. Entonces:

- Si $y_i = 1$, y la observación pertenece a la segunda población, el segundo término de la desviación es nulo y $d_i = -2\, log\, p_i$. La observación tendrá una desviación grande si la probabilidad estimada de pertenecer a la segunda población, p_i es pequeña, lo que indica que esta observación está mal explicada por el modelo.
- Si $y_i = 0$, y la observación pertenece a la primera población, sólo interviene el segundo término de la desviación $d_i = -2\, log\, (1 - p_i)$. La desviación será grande si p_i es grande, lo que indica que la probabilidad de pertenecer a la verdadera población es pequeña y el modelo ajusta mal dicho dato.

Para maximizar la verosimilitud, expresando p_i en función de los parámetros $\boldsymbol{\beta}$, en (2.6) obtenemos la función soporte:

$$logP(\boldsymbol{y}) = \sum_{i=1}^{n}\left(y_i(b_0 + b_1x_1 + \ldots + b_ix_i) + \sum_{i=1}^{n} log\left(1 - \frac{e^{b_0+\sum_{i=1}^{n} b_ix_i}}{1 + e^{b_0+\sum_{i=1}^{n} b_ix_i}}\right)\right)$$

$$logP(\boldsymbol{y}) = \sum_{i=1}^{n}\left(y_i\, x_i'\, \boldsymbol{\beta} + \sum_{i=1}^{n} log\left(\frac{1}{1 + e^{b_0+\sum_{i=1}^{n} b_ix_i}}\right)\right)$$

$$logP(\boldsymbol{y}) = \sum_{i=1}^{n}\left(y_i\, x_i'\, \boldsymbol{\beta} - \sum_{i=1}^{n} log\left(1 + e^{b_0+\sum_{i=1}^{n} b_ix_i}\right)\right)$$

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \boldsymbol{x_i'}\, \beta - \sum_{i=1}^{n} log\left(1 + e^{\boldsymbol{x_i'\beta}}\right)$$

(2.8)

de la cual se calcula la derivada para obtener los estimadores de máxima verosimilitud.

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \left(\frac{e^{x_i' \boldsymbol{\beta}}}{1 + e^{x_i' \boldsymbol{\beta}}} \right)$$

(2.9)

igualando este vector a cero y llamando $\widehat{\boldsymbol{\beta}}$ a los parámetros que satisfacen el sistema de ecuaciones:

$$\sum_{i=1}^{n} y_i x_i = \sum x_i \left(\frac{1}{1 + e^{-x_i' \widehat{\boldsymbol{\beta}}}} \right) = \sum \widehat{y_i}\, x_i$$

(2.10)

Estas ecuaciones establecen que el producto de los valores observados por las variables explicativas debe ser igual al de los valores previstos. También, que los residuos del modelo,$e_i = y_i - \widehat{y_i}$ deben ser ortogonales a las variables x.Esta condición es análoga a la obtenida en el modelo de regresión estándar, pero ahora el sistema (2.10) resultante no es lineal en los parámetros $\widehat{\boldsymbol{\beta}}$. Para obtener el valor $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MV}}$ que maximiza la verosimilitud se va a utilizar el Método de Newton-Raphson. Desarrollando el vector $\partial L(\beta)/\partial \beta$ alrededor de un punto β_a, se tiene:

$$\frac{\partial L(\beta)}{\partial \beta} = \frac{\partial L(\beta_a)}{\partial \beta} + \frac{\partial^2 L(\beta_a)}{\partial \beta \partial \beta'} (\beta - \beta_a);$$

para que el punto β_a corresponda al máximo de verosimilitud su primera derivada debe anularse. Imponiendo la condición $\partial L(\beta_a)/\partial \beta = 0$, se obtiene:

$$\beta_a = \hat{\beta} + \left(-\frac{\partial^2 L(\beta_a)}{\partial \beta \partial \beta'} \right)^{-1} \left(\frac{\partial L(\beta)}{\partial \beta} \right)$$

(2.11)

que expresa cómo obtener el punto máximo β_a,a partir de un punto próximo cualquiera β.

La ecuación depende de la matriz de segundas derivadas, que, en el óptimo, es la inversa de la matriz de varianzas y covarianzas asintótica de los estimadores de máxima verosimilitud. Para obtener su expresión, derivando por segunda vez en (2.9), se obtiene:

$$\widehat{M}^{-1} = \left(-\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta'}\right) = \sum_{i=1}^{n} x_i x_i' w_i$$

(2.12)

donde los coeficientes w_i están dados por:

$$w_i = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})^2} = p_i(1-p_i)$$

(2.13)

Sustituyendo en (2.11) las expresiones (2.12) y (2.9) y evaluando las derivadas en un estimador inicial $\hat{\beta}$, se obtiene el siguiente método para obtener un nuevo valor del estimador,β_a a partir del $\hat{\beta}$

$$\beta_a = \hat{\beta} + \left(\sum_{i=1}^{n} x_i x_i' \widehat{w}_i\right)^{-1} \left(\sum x_i\,(y_i - \hat{p}_i)\right)$$

(2.14)

donde $\hat{p}_i$ y $\widehat{w}_i$ se calculan con el valor $\hat{\beta}$. El algoritmo puede escribirse como:

$$\beta_a = \hat{\beta} + \left(X'\widehat{W}X\right)^{-1} X'\left(Y - \hat{Y}\right)$$

(2.15)

donde $\widehat{W}$ es una matriz diagonal con términos $\hat{p}_i(1-\hat{p}_i\ )$ y $\hat{Y}$ el vector de valores esperados de Y. La matriz de varianzas y covarianzas de los estimadores así obtenidos es aproximadamente, según (2.15) $\left(X'\widehat{W}X\right)^{-1}$. La ecuación (2.15) indica que se debe modificar el estimador si los residuos no son ortogonales a las variables explicativas, es decir si $X'\left(Y-\hat{Y}\right) \neq 0$. La modificación del estimador depende de esta diferencia y se reparte entre los componentes de $\hat{\beta}$ en función de su matriz de varianzas y covarianzas estimada.

La forma habitual de implementar este método es el siguiente algoritmo iterativo que proporciona en convergencia el estimador MV[26] de β .

1. Fijar un valor arbitrario inicial, $\widehat{\beta}_1$ para los parámetros y obtener el vector $\widehat{Y}_1$ para dicho valor en el modelo logit.

---

[26] Estimador de Máxima Verosimilitud

2. Definir una variable auxiliar z_i de residuos estandarizados por:

$$z_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i(1-\hat{y}_i)}} = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}$$

o vectorialmente:

$$Z = \widehat{W}^{-1/2}\left(Y - \hat{Y}\right)$$

donde $\widehat{W}$ es una matriz diagonal con términos $\hat{y}_i(1-\hat{y}_i)$.

3. Estimar por mínimos cuadrados una regresión con variable dependiente Z y matriz de regresores $T = \widehat{W}^{1/2}X$. Los parámetros estimados con ésta regresión , $\hat{b}_1$, vendrán dados por:

$$\begin{aligned}\hat{b}_1 &= (T'T)^{-1}T'Z \\ &= \left(X'\widehat{W}X\right)^{-1}X'\left(Y - \hat{Y}\right)\end{aligned}$$

y comparando con (2.15), se puede observar que b_1 estima el incremento $\beta_a - \hat{\beta}_1$ de los parámetros que nos acerca al máximo.

4. Obtener un nuevo estimador de los parámetros $\hat{\beta}_2$ del modelo logístico mediante

$$\hat{\beta}_2 = \hat{\beta}_1 + \hat{b}_1$$

5. Tomar el valor estimado resultante de la etapa anterior, $\hat{\beta}_h$ y sustituirlo en la ecuación del modelo logístico para obtener el vector de estimadores $\hat{Y}\left(\hat{\beta}_h\right) = \hat{Y}_h$. Utilizando este vector $\hat{Y}_h$ construir la matriz $\widehat{W}_h$ y las nuevas variables Z_h y T_h.

$$z_h = \widehat{W}_h^{-1/2}\left(Y - \hat{Y}_h\right),$$

$$T_h = \widehat{W}_h^{1/2}X,$$

y volver a la etapa 2.El proceso se repite hasta obtener la convergencia $\left(\hat{\beta}_{h+1} \simeq \hat{\beta}_h\right)$.

$$(T'T)^{-1}T'Z$$

2.1.2.1.2 Estadístico de Wald

Para comprobar si una variable o grupo de variables incluidas dentro de la ecuación es significativo, se puede construir un contraste de la razón de verosimilitudes comparando el máximo de la función de verosimilitud para el modelo con y sin estas variables.

Supongamos que $\beta = (\beta_1, \beta_2)$, donde β_1 tiene dimensión p-s, y β_2 tiene dimensión s. Se desea contrastar si el vector de parámetros:

$$H_O: \beta_2 = 0$$

Frente a la alternativa

$$H_1: \beta_2 \neq 0$$

El contraste de razón de verosimilitudes utiliza que $\lambda = 2L(H_1) - 2L(H_0)$, donde $L(H_1)$ es el máximo del soporte cuando se estima los parámetros bajo H_1 y $L(H_0)$ es el máximo cuando estimamos los parámetros bajo H_0. Una manera equivalente de definir el contraste es llamar $D(H_0) = -2L(\hat{\beta}_1)$ a la desviación cuando el modelo se estima bajo H_0, es decir, suponiendo que $\beta_2 = 0$, y $D(H_1) = -2L(\hat{\beta}_1\hat{\beta}_2)$ a la desviación bajo H_1. La desviación será menor con el modelo con más parámetros (la verosimilitud será siempre mayor bajo H_1 y, si H_0 es cierta, la diferencia de desviaciones, que es el contraste de verosimilitudes se distribuye como una χ_s^2 con s grados de libertad, siendo

$$\chi_s^2 = D(H_0) - D(H_1) = 2L(\hat{\beta}_1\hat{\beta}_2) - 2L(\hat{\beta}_1) \tag{2.16}$$

En particular este test puede aplicarse para comprobar si un parámetro es significativo y debe dejarse en el modelo. Sin embargo, es más habitual en estos casos comparar el parámetro estimado con su desviación típica.

Los cocientes

$$w_j = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)} \tag{2.17}$$

se denominan ***estadísticos de Wald*** y en muestras grandes se distribuyen, si el verdadero valor del parámetro es cero, como una normal estándar.

Una medida global del ajuste es:

$$R^2 = 1 - \frac{D(\hat{\beta})}{D_0} = 1 - \frac{L(\hat{\beta})}{L(\beta_0)}$$

(2.18)

donde el numerador es la desviación (verosimilitud en el máximo) para el modelo con parámetros estimados $\hat{\beta}$ y el denominador la desviación (verosimilitud) para el modelo que sólo incluye la constante β_0.En éste último caso, la estimación de la probabilidad p_i es constante para todos los datos e igual a m/n siendo m el número de elementos en la muestra con la variable y=1.Entonces ,sustituyendo en (2.7) la desviación máxima que corresponde al modelo más simple posible con sólo β_0 que asigna la misma probabilidad a todos los datos, es:

$$D_0 = -2L(\beta_0) = -2m\, log\, m - 2(n-m)log(n-m) + 2n \log n$$

(2.19)

Por otro lado, si el ajuste es perfecto, es decir todas las observaciones con $y = 1$ tienen $p_i = 1$ y las de $y = 0$ tienen $p_i = 0$,entonces, según (2.5) la desviación es cero y $L(\hat{\beta}) = 0$ y $R^2 = 1$. Por el contrario, si las variables explicativas no influyen nada la desviación con las variables explicativas será igual que sin ellas,

$$L(\hat{\beta}) = L(\beta_0) \text{ y } R^2 = 0.$$

2.1.2.1.3 Interpretación del modelo logístico

Los parámetros del modelo son:

- β_0, la ordenada en el origen, y

- $\beta_i = \{ \beta_1, \beta_2, \ldots, \beta_k \}$, las pendientes.

A veces, se utilizan también como parámetros exp (β_0) y exp (β_i), que se denominan razón de probabilidades. Estos valores indican cuánto se modifican las probabilidades por unidad de cambio en las variables *x*. De la ecuación (2.4) se deduce que:

$$O_i = \frac{p_i}{1 - p_i} = exp(\beta_o) \prod_{j=1}^{k} exp(\beta_j)^{x_j}$$

(2.20)

Supóngase que dos elementos tienen valores iguales en todas las variables menos en una.

Sean dos elementos, i, k con todos los valores de las variables iguales menos la variable h y $x_{ih} = x_{jh} + 1$. Entonces, la razón de probabilidades para estas dos observaciones es:

$$\frac{O_i}{O_k} = exp(\beta_h)$$

(2.21)

e indica cuánto se modifica la razón de probabilidades cuando la variable x_h aumenta en una unidad.

Si se considera $\hat{p}_i = 0{,}5$ en el modelo logit, entonces,

$$log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{i1} + \ ... + \beta_k x_{ik} = 0$$

(2.22)

es decir,

$$x_{i1} = -\frac{\beta_0}{\beta_1} - \sum_{j=2}^{k} \frac{\beta_j x_{ij}}{\beta_1}$$

(2.23)

donde x_{i1} representa el valor de x_1 que hace igualmente probable que un elemento cuyas variables restantes son $(x_{i2}, \dots, x_{ik})$ pertenezca a la primera o la segunda población.

Intervalos de Confianza

Una vez que se han estimado los coeficientes de regresión y sus errores estándar se calcularán los correspondientes intervalos de confianza para las estimaciones, bajo la hipótesis de que dichos coeficientes se distribuyen según respectivas distribuciones normales. Para un determinado coeficiente, su intervalo de confianza al 95 % vendrá dado por:

$$IC95\%(\beta) = [(\beta - 1.96ee) \cdot (\beta + 1.96ee)]$$

(2.24)

$$IC95\%\left(OR^{27}\right) = [exp(\beta - 1.96ee) \cdot exp(\beta + 1.96ee)]$$

(2.25)

2.1.2.1.4 Medidas de confiabilidad del modelo

- **Desvianza**

Es similar a la suma de cuadrados del error de la regresión lineal y se define como:

$$D = -2\sum_{I=1}^{n}\left(y_i log\left(\frac{\hat{p}}{y_i}\right) + (1 - y_i) log\frac{1 - \hat{p}}{1 - y_i}\right)$$

(2.26)

Si D es mayor que una χ^2 con n-p grados de libertad para un nivel de significación dado entonces el modelo logístico es confiable.

- **Criterio AIC de Akaike**

Se define como: $AIC = D + 2(p + 1)$

(2.27)

---

[27] Es la razón de probabilidades también denominada "odds-ratio"

donde p es el número de variables predictoras, y D es la desvianza.
El valor de AIC es interpretable sobre todo en la comparación, más que en sí mismo, el modelo con un AIC menor es el mejor ajustado.

- **Prueba de Bondad de ajuste de Hosmer Lemeshow**

Se define como:

$$c = \sum_{i=1}^{g} \frac{(O_i - n_i' \bar{p}_i)^2}{n_i' \bar{p}_i (1 - \bar{p}_i)}$$

(2.28)

donde g es el número de grupos;
n_i' es el número de observaciones en el i-ésimo grupo;
O_i es la suma de las Y en el i-ésimo grupo; y
$\bar{p}_i$ es el promedio de las p_i en el i-ésimo grupo.

2.1.2.1.5 Estadísticos influenciales para regresión logística

Existen varios tipos de residuales que permiten contrastar si una observación es o no influenciable.

Residuales de Pearson

Definidos como:

$$r_i = \frac{y_i - m_i \hat{p}_i}{\sqrt{m_i \hat{p}_i (1 - \hat{p}_i)}}$$

(2.29)

donde, y_i representa el número de veces que *y* =1 entre las m_i repeticiones de x_i si los valores de la variable de respuesta están agrupadas,

Si el modelo es correcto, los residuales de Pearson serán variables de media cero y varianza unidad que pueden servirnos para hacer el diagnóstico de dicho modelo. El estadístico $\chi_0^2 = \sum_{i=1}^{k} e_i^2$ permite realizar un contraste global de la bondad del ajuste. Se distribuye asintóticamente como una χ^2 con (n − k −1) grados de libertad, donde k +1 es el número de parámetros en el modelo.

En lugar de los residuos de Pearson se pueden utilizar, también, las desviaciones de las observaciones o pseudo residuos definidos por:

$$d_i = -2\left(y_i \log \widehat{p_i} + (1 - y_i) log(1 - \widehat{p_i})\right)$$

(2.30)

2.1.2.2 Regresión Logística Multinomial

El modelo logit multinomial se desarrolla cuando la variable dependiente es categórica con más de dos grupos; el procedimiento consiste en elegir una categoría como referencia y se modelan varios *logits* simultáneamente, uno para cada una de las restantes categorías respecto a la de referencia. En el presente trabajo la variable dependiente es: condición de ocupación, la misma que tiene tres categorías:

– Ocupación Plena (referencia)
– Subempleo
– Desempleo
Se modelan dos *logits* simultáneamente:
-subempleo con respecto a ocupación plena
-desempleo con respecto a ocupación plena

Se tiene una variable aleatoria dependiente Y categórica nominal politómica (3 categorías en el presente trabajo), con probabilidades p_1=p (Y=1), p_2= p (Y=2), p_3=p (Y=3)=1- p_1- p_2.
Supongamos que

$$(Y_1, Y_2) = \begin{cases} (1,0) & si\, Y = 1 \\ (0,1) & si\, Y = 2 \\ (0,0) & si\, Y = 3 \end{cases}$$

(2.31)

Las variables Y_1; Y_2 tienen una distribución de Bernoulli con E $(Y_1)=p_1$ y E $(Y_2)=p_2$, al igual que la variable dependiente en la regresión logística binaria clásica. Obviamente estas dos variables no son independientes ya que cov $(Y_1, Y_2) = p_1 p_2$. Formulamos el modelo multivariante definido por las siguientes ecuaciones:

$$p_1(X_1, X_2) = p_1 = E(Y_1) = \frac{exp(Z_1)}{1 + exp(Z_1) + exp(Z_2)}$$

(2.32)

$$p_2(X_1, X_2) = p_2 = E(Y_2) = \frac{exp(Z_2)}{1 + exp(Z_1) + exp(Z_2)}$$

(2.33)

Donde $Z_1 = \beta_{01} + \beta_{11} \cdot X_1 + \beta_{21} \cdot X_2$ y $Z_2 = \beta_{02} + \beta_{12} \cdot X_1 + \beta_{22} \cdot X_2$

Siendo $\beta_{01}, \beta_{11}, \beta_{21}, \beta_{02}, \beta_{12}, \beta_{22}$ parámetros que deseamos estimar:

$$p_3(X_1, X_2) = p_3 = 1 - p_1 - p_2 = \frac{1}{1 + exp(Z_1) + exp(Z_2)}$$

(2.34)

Con el propósito de interpretar mejor los parámetros que aparecen en el modelo, podríamos reescribir de la siguiente forma:

$$\frac{p_1}{p_3} = exp(Z_1) = exp(\beta_{01}) \cdot \left(exp(\beta_{11})\right)^{X_1} \cdot \left(exp(\beta_{21})\right)^{X_2}$$

$$\frac{p_2}{p_3} = exp(Z_2) = exp(\beta_{02}) \cdot \left(exp(\beta_{12})\right)^{X_1} \cdot \left(exp(\beta_{22})\right)^{X_2}$$

(2.35)

Al cociente p_1/p_3 se le denomina razones de probabilidad de la categoría 1 respecto de la categoría 3 y se le representa por $O_1(X_1, X_2) = O_1$ de la misma manera para O_2.

De este modo se puede concluir que la razón de probabilidad en O_1 cuando X_1 se incrementa en una unidad manteniéndose constante X_2 viene dada por

$\frac{O_1(X_1+1,X_2)}{O_1(X_1,X_2)} = exp(\beta_{11})$, que recibe el nombre de razones de probabilidad de la categoría 1 respecto de la variable X_1 y se representa por $OR_1(X_1)$; de la misma manera para $OR_1(X_2), OR_2(X_1)$ y $OR_2(X_2)$.

Las proporciones de cambio en las razones de probabilidad con respecto a cada variable regresora que, por ejemplo, para O_1 con respecto a X_1, viene dada por:

$$\frac{O_1(X_1+1;X_2) - O_1(X_1,X_2)}{O_1(X_1,X_2)} = OR_1(X_1) - 1 = exp(\beta_{11}) - 1$$

(2.36)

Y que serán representados por: $OC_1(X_1)$; $OC_1(X_2)$; $OC_2(X_1)$; $OC_2(X_2)$).

Otra manera de representar el modelo es tomando los logaritmos en ambas ecuaciones del modelo:

$$ln\left(\frac{p_1}{p_3}\right) = Z_1 = \beta_{01} + \beta_{11} \cdot X_1 + \beta_{21} \cdot X_2$$

$$ln\left(\frac{p_2}{p_3}\right) = Z_2 = \beta_{02} + \beta_{12} \cdot X_1 + \beta_{22} \cdot X_2$$

(2.37)

donde las expresiones del miembro izquierdo se denominan 'logits' (al igual que en la regresión logística binaria) y los parámetros representan las tasas de cambio en los 'logits' cuando una de las variables explicativas se incrementa en una unidad manteniéndose constante la otra.

2.1.2.2.1 Estimación de parámetros

Dada una muestra de datos $(Y_{1i,}Y_{2i,}X_{1i,}X_{2i,})$ con i=1,2,…, n podemos definir, en función de los parámetros del modelo, las funciones Z_{1i},Z_{2i}, p_{1i}, p_{2i} y abordar el problema de la estimación de los mismos mediante el método de máxima verosimilitud, de la siguiente manera:

Con el modelo planteado, la función de verosimilitud de la muestra viene dada por la siguiente expresión:

$$L = \prod_{i=1}^{n} \left(p_{1i}^{Y_{1i}} \cdot p_{2i}^{Y_{2i}} \cdot p_{3i}^{1-Y_{1i}-Y_{2i}} \right) = \prod_{i=1}^{n} \left(\left(\frac{p_{1i}}{p_{3i}} \right)^{Y_{1i}} \cdot \left(\frac{p_{2i}}{p_{3i}} \right)^{Y_{2i}} \cdot p_{3i} \right)$$

(2.38)

En lugar de trabajar con ésta expresión se utiliza la función auxiliar:

$$\Lambda = -2 \cdot ln(L) = -2 \cdot \sum_{i=1}^{n} \left(Y_{1i} \cdot ln\left(\frac{p_{1i}}{p_{3i}} \right) + Y_{2i} \cdot ln\left(\frac{p_{2i}}{p_{3i}} \right) + ln(p_{3i}) \right)$$

$$= -2 \cdot \sum_{i=1}^{n} \left(In\left(1 + exp(Z_{1i}) + exp(Z_{2i})\right) - Y_{1i} \cdot Z_{1i} - Y_{2i} \cdot Z_{2i} \right)$$

(2.39)

El problema de maximizar la verosimilitud equivale a minimizar la función auxiliar Λ_Λ y puede resolverse por métodos numéricos de forma iterativa partiendo de la estimación inicial $\beta_{11} = \beta_{21} = \beta_{12} = \beta_{22} = 0$, $\beta_{01} = ln(n_1) - ln(n - n_1 - n_2)$ y $\beta_{02} = ln(n_2) - ln(n - n_1 - n_2)$ siendo n_1 y n_2 el número de observaciones en las categorías 1 y 2 respectivamente. Estos estimadores iniciales se obtienen suponiendo que no hay una influencia de las variables regresoras en el modelo planteado y para ellos el valor inicial de la función auxiliar que debemos minimizar es:

$$\Lambda_0 = -2 \cdot \left(n_1 \cdot ln\left(\frac{n_1}{n} \right) + n_2 \cdot ln\left(\frac{n_2}{n} \right) + (n - n_1 - n_2) \cdot ln\left(\frac{n - n_1 - n_2}{n} \right) \right)$$

(2.40)

Una vez alcanzada la convergencia del método iterativo, designaremos por Λ_0 al mínimo obtenido y por $\hat{\beta}_{01}, \hat{\beta}_{11}, \hat{\beta}_{21}, \hat{\beta}_{02}, \hat{\beta}_{12}, \hat{\beta}_{22}$ a los valores estimados de los parámetros del modelo.

2.1.2.2.2 Significatividad global del modelo

Podemos contrastar la hipótesis de no existencia de un efecto significativo global de las variables regresoras teniendo en cuenta que la diferencia entre el valor

inicial y el valor final de la función auxiliar Λ tiene una distribución χ^2 con 4 grados de libertad (en general, número de regresores multiplicado por número de categorías menos una). El p-valor del test para la hipótesis nula de que no existe efecto de las variables regresoras $(\beta_{11} = \beta_{21} = \beta_{12} = \beta_{22} = 0)$ vendrá dado por $p(\chi_4^2 > \Lambda_0 - \Lambda_f)$.

2.1.2.2.3 Significatividad del efecto de cada variable regresora

Si llamamos Λ_{-1} al mínimo de la función auxiliar que se obtendría eliminando del modelo la variable $X_1(\beta_{11} = \beta_{12} = 0)$ se verifica que la diferencia entre los mínimos de la función auxiliar en el modelo reducido y en el modelo completo tiene una distribución χ^2 con 2 grados de libertad (en general, número de regresores menos uno multiplicado por número de categorías menos una). Por tanto el p-valor del test para la hipótesis nula de que no existe efecto de la variable $X_1(\beta_{11} = \beta_{12} = 0)$ vendrá dado por $p(\chi_{2>}^2 \Lambda_{-1} - \Lambda_o)$.De modo similar podríamos calcular Λ_0 (mínimo de la función auxiliar eliminando β_{01} y β_{02} del modelo) y Λ_{-2}(mínimo de la función auxiliar eliminando del modelo la variable X_2) y construir tests de hipótesis para $\beta_{01} = \beta_{02} = 0$ y $\beta_{21} = \beta_{22} = 0$, respectivamente.

2.1.2.2.4 Significatividad de cada parámetro

Teniendo en cuenta que el cuadrado de cada estimador dividido por su error estándar tiene una distribución χ^2 con 1 grado de libertad podemos construir test de hipótesis para la igualdad de cada parámetro a cero y podremos saber que estimadores de los parámetros del modelo son significativamente distintos de cero.

Por ejemplo, para el test de hipótesis $\beta_{11} = 0$ el p-valor sería $p\left(\chi_1^2 > \left(\frac{\hat{\beta}_{11}}{s.e.(\hat{\beta}_{11})}\right)^2\right)$, siendo $s.e.(\hat{\beta}_{11})$ el valor correspondiente al error estándar del estimador del parámetro β_{11}.

2.1.2.2.5 Intervalo de confianza para los parámetros

De acuerdo a la normalidad asintótica de los estimadores máximo verosímiles podemos construir, utilizando la distribución normal, intervalos de confianza asintóticos para cada uno de los parámetros del modelo y, mediante las transformaciones correspondientes, intervalos de confianza para las OR y las OC. Por ejemplo, para el parámetro β_{11}, y utilizando un grado de confianza de $1-\alpha$, tendríamos:

I.C para $\boldsymbol{\beta_{11}}$:

$$\left(\hat{\beta}_{11} - z_{\alpha/2} \cdot s.e.\left(\hat{\beta}_{11}\right), \hat{\beta}_{11} + z_{\alpha/2} \cdot s.e.\left(\hat{\beta}_{11}\right)\right) \tag{2.41}$$

I.C. para $\boldsymbol{OR_1(X_1)}$:

$$\left(exp\left(\hat{\beta}_{11} - z_{\alpha/2} \cdot s.e.\left(\hat{\beta}_{11}\right)\right), exp\left(\hat{\beta}_{11} + z_{\alpha/2} \cdot s.e.\left(\hat{\beta}_{11}\right)\right)\right) \tag{2.42}$$

I.C. para $\boldsymbol{OC_1(X_1)}$:

$$\left(exp\left(\hat{\beta}_{11} - z_{\alpha/2} \cdot s.e.\left(\hat{\beta}_{11}\right)\right) - 1, exp\left(\hat{\beta}_{11} + z_{\alpha/2} \cdot s.e.\left(\hat{\beta}_{11}\right)\right) - 1\right) \tag{2.43}$$

Siendo $z_{\propto/2}$ el valor que, en una distribución normal (0,1), verifica $p\left(Z > z_{\propto/2}\right) = {}^{\propto}/_{2}$

2.1.2.2.6 Calidad del ajuste

Al igual que en la regresión logística binaria, la calidad del ajuste en la regresión logística multinomial se mide mediante coeficientes de determinación conocidos como Pseudo-R^2, de entre los cuales los más importantes y los que nos permiten observar algunos de los paquetes estadísticos son:

El primero se basa en la función auxiliar Λ utilizada en el ajuste ,se conoce como pseudo-R^2 de Mc-Fadden y viene dado por: $R^2_{MF} = 1 - \frac{\Lambda_f}{\Lambda_o}$.

Su rango teórico de valores es $0 \leq R^2_{MF} \leq 1$, pero muy raramente su valor se aproxima a 1. Suele considerarse una buena calidad del ajuste cuando 0.2≤ $R^2_{MF} \leq 0.4$ y excelente para valores superiores.

Otros coeficientes son basados directamente en la verosimilitud L, y no en la función auxiliar Λ.El más conocido es el pseudo-R^2 de Cox-Snell, definido como:

$$R^2_{CS} = 1 - \frac{(\sqrt[n]{L_0})^2}{(\sqrt[n]{L_f})^2} = 1 - \exp(\frac{\Lambda_f - \Lambda_0}{n})$$

(2.44)

Siendo:

$L_0 = \exp\left(\frac{-\Lambda_0}{2}\right)$ y $L_f = \exp\left(\frac{-\Lambda_f}{2}\right)$

(2.45)

El rango teórico de valores para este coeficiente es:

$$0 \leq R^2_{CS} \leq 1 - \left(\sqrt[n]{L_0}\right)^2$$

Lo que le hace poco interpretable al depender de L_0. Por este motivo es preferible el pseudo-R^2 de Nagelkerke, que se define como:

$$R^2_N = \frac{R^2_{CS}}{1 - \left(\sqrt[n]{L_0}\right)^2} = \frac{1 - exp\left(\frac{\Lambda_f - \Lambda_0}{n}\right)}{1 - exp\left(-\frac{\Lambda_0}{n}\right)}$$

(2.46)

y su rango de valores es $0 \leq R^2_N \leq 1$ por lo que puede interpretarse del mismo modo que el coeficiente de determinación de la regresión lineal clásica, aunque es más difícil que alcance valores próximos a 1 .

Para comparar modelos de regresión logística multinomial con diferente número de variables regresoras suelen introducirse coeficientes Pseudo-R^2 ajustados.

El más conocido es el de Mc-Fadden, definido como:

$$Adj - R^2_{MF} = 1 - \frac{0.5.\Lambda_f + k + 1}{0.5.\Lambda_0 + 1}$$

(2.47)

siendo k el número de regresores.

Calidad en la predicción

Si a partir del modelo ajustado, clasificamos cada observación en la categoría más probable, podemos construir una matriz de clasificación de valores observados-predichos y utilizar el porcentaje de clasificaciones correctas como una medida de la calidad de predicción.

Si a partir del modelo ajustado, clasificamos cada observación en la categoría más probable, podemos construir una matriz de clasificación de valores observados-predichos y utilizar el porcentaje de clasificaciones correctas como una medida de la calidad de predicción.

2.2 ANÁLISIS DISCRIMINANTE:[28]

Supongamos que un conjunto de objetos está ya clasificado en una serie de grupos, es decir, se sabe previamente a qué grupos pertenecen; el Análisis Discriminante se puede considerar como un análisis de regresión donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. Se pretende encontrar relaciones lineales entre las variables continuas que mejor discriminen en los grupos dados a los objetos.

---

[28] Daniel Peña. McGrawHill. Análisis de Datos Multivariante. Pág. 397-406. http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema6am.pdf.pag1-6.

Un segundo objetivo es construir una regla de decisión que asigne un objeto nuevo, que no sabemos clasificar previamente, a uno de los grupos prefijados con un cierto grado de riesgo.

Es necesario considerar una serie de restricciones o supuestos:

- Se tiene una variable categórica y el resto de variables cuantitativas e independientes entre sí.

- También se puede introducir variables independientes cualitativas o categóricas, después de transformarlas en tantas variables dicotómicas (ficticias).

- Es necesario que existan al menos dos grupos, y para cada grupo se necesitan dos o más casos.

- El número de variables discriminantes debe ser menor que el número de objetos menos 2: $x_1,..., x_p$, donde $p < (n - 2)$ y n es el número de objetos.

- Ninguna variable discriminante puede ser combinación lineal de otras variables discriminantes.

- El número máximo de ***funciones discriminantes***[29] es igual al mínimo entre el número de variables y el número de grupos menos 1 (con q grupos, $(q - 1)$ funciones discriminantes).

- Las matrices de covarianzas dentro de cada grupo deben ser aproximadamente iguales. Las variables continuas deben seguir una distribución normal multivariante.

---

[29] Las funciones discriminantes son vectores ortogonales que determinan los grupos que se requieren identificar en el Análisis Discriminante.

2.2.1 Modelo Matemático

A partir de q grupos donde se asignan a una serie de objetos y de p variables medidas sobre ellos $(x_1, \ldots, x_p)$, se trata de obtener para cada objeto una serie de puntuaciones que indican el grupo al que pertenecen $(y_1, \ldots, y_m)$, de modo que sean funciones lineales de $x_1, \ldots, x_p$.

$$y_m = a_{m1}x_1 + \ \ldots + a_{mp}x_p + a_{m0}$$

(2.48)

Donde m = min (q − 1, p), tales que discriminen o separen lo máximo posible a los q grupos. Estas combinaciones lineales de las p variables deben maximizar la varianza entre los grupos y minimizar la varianza dentro de los grupos.

Descomposición de la varianza

Se puede descomponer la variabilidad total de la muestra en variabilidad dentro de los grupos y entre los grupos.

Partimos de

$$Cov(x_j, x_{j'}) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x_j})(x_{ij'} - \overline{x_{j'}})$$

(2.49)

Se puede considerar la media de la variable x_j en cada uno de los grupos $I_1, \ldots, I_q$, es decir,

$$\bar{x}_{kj} = \frac{1}{n_k}\sum_{i \epsilon I_k} x_{ij}$$

(2.50)

para k = 1, . . . , q.

De este modo, la media total de la variable x_j se puede expresar como función de las medias dentro de cada grupo. Así,

$$\sum_{i \epsilon I_k} x_{ij} = n_k \bar{x}_{kj}$$

(2.51)

y

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} = \frac{1}{n}\sum_{k=1}^{q}\sum_{i\epsilon I_k} x_{ij} = \frac{1}{n}\sum_{k=1}^{q} n_k \ \overline{x}_{kj} = \sum_{k=1}^{q}\frac{n_k}{n}\overline{x}_{kj}$$

(2.52)

Sustituyendo (2.51) y (2.52) en (2.49), se tiene:

$$Cov(x_j, x_{j'}) = \frac{1}{n}\sum_{k=1}^{q}\sum_{i\epsilon k}(x_{ij} - \overline{x}_j)\left(x_{ij'} - \overline{x}_{j'}\right) = d\left(x_j, x_{j'}\right) + e(x_j, x_{j'})$$

(2.53)

Es decir, la covarianza total es igual a la covarianza dentro de grupos más la covarianza entre grupos. Si denominamos como $t(x_j, x_{j'})$ a la covarianza total entre x_j y $x_{j'}$ (sin distinguir grupos), entonces lo anterior se puede expresar como:

$$t(x_j, x_{j'}) = d(x_j, x_j') + e(x_j, x_{j'})$$

En notación matricial esto es equivalente a

$$T = E + D$$

(2.54)

donde:

T = matriz de covarianzas total
E = matriz de covarianzas inter-grupos
D = matriz de covarianzas intra- grupos.

2.2.2 Extracción de las funciones discriminantes

La idea básica del Análisis Discriminante consiste en extraer a partir de $x_1, \dots, x_p$ variables observadas en k grupos, m funciones $y_1, \dots, y_m$ de forma:

$$y_i = a_{i1}x_1 + \ ... + a_{ip}x_p + a_{io}$$

(2.55)

donde m = min(q − 1, p), tales que corr(y_i, y_j) = 0 para todo i ≠j.
Si las variables x_1,..., x_p están tipificadas, entonces las funciones:

$$y_i = a_{i1}x_1 + \ ... + a_{ip}x_p$$

(2.56)

para i = 1,..., m, se denominan funciones discriminantes canónicas.
Las funciones y_1,..., y_m se extraen de modo que:

(i) y_1 sea la combinación lineal de x_1,..., x_p que proporciona la mayor discriminación posible entre los grupos.
(ii) y_2 sea la combinación lineal de x_1,..., x_p que proporciona la mayor discriminación posible entre los grupos, después de y_1, tal que corr (y_i, y_2) = 0.

En general, y_i es la combinación lineal de x_1,..., x_p que proporciona la mayor discriminación posible entre los grupos después de y_{i-1} y tal que corr (y_i, y_j) = 0 para j = 1,..., (i − 1).

2.2.3 Procedimiento matricial

Se sigue un método parecido al análisis factorial, así se busca una función lineal de $x_1, \ldots, x_p$: $y = a'x$, de modo que

$$Var(y) = a'Ta = a'Ea + a'Da$$

(2.57)

es decir, la variabilidad intergrupos más la variabilidad intragrupos.

Maximizar la variabilidad entre los grupos para discriminarlos mejor equivale a:

$$máx\left(\frac{a'Ea}{a'Ta}\right)$$

(2.58)

es decir, maximizar la varianza entre grupos en relación al total de la varianza.

Si consideramos la función:

$$f(a) = \frac{a'Ea}{a'Ta}$$

(2.59)

Se observa que f es una función homogénea, es decir,

$$f(a) = f(\mu a)$$

(2.60)

para todo $\mu \epsilon \mathbb{R}$

Una función sea homogénea implica que calcular $máx\left(\frac{a'Ea}{a'Ta}\right)$ equivale a calcular

$$máx(a'Ea)$$

tal que $$a'Ta = 1$$

Según los multiplicadores de Lagrange, se define:

$$L = a'Ea - \lambda(a'Ta - 1)$$

(2.61)

y se calcula su derivada:

$$\frac{\partial L}{\partial a} = 0$$

$$\frac{\partial L}{\partial a} = 2Ea - 2\lambda Ta = 0 \Longrightarrow$$

$$Ea = \lambda Ta \Longrightarrow$$

$$(T^{-1}E)a = \lambda a$$

(2.62)

Por tanto, los valores propios asociados a la primera función discriminante lo son de la matriz $T^{-1}E$ (que no es simétrica en general).

Como $Ea = \lambda Ta$,

$$a'Ea = \lambda a'Ta = \lambda$$

(2.63)

Luego si tomo el máximo valor propio, se obtendrá la función que recoge el máximo poder discriminante.

El valor propio asociado a la función discriminante indica la proporción de varianza total explicada por las m funciones discriminantes que recoge la variable y_i.

Para obtener más funciones discriminantes, se siguen sacando los valores propios de la matriz $(T^{-1}E)$ elegidos en orden decreciente:

$$\begin{array}{ccc} a_2' & \Rightarrow & a_2'x = y_2 \\ \vdots & \dots & \vdots \\ a_m' & \Rightarrow & a_m'x = y_m \end{array}$$

(2.64)

donde $m = mín(q-1, p)$

La suma de todos los valores propios,$\sum_{i=1}^{m} \lambda_i$ es la proporción de varianza total que queda explicada, o se conserva, al considerar sólo los ejes o funciones discriminantes.

Como consecuencia, el porcentaje explicado por y_i del total de varianza explicada por $y_1, \dots, y_m$ es

$$\frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} . 100\%$$

(2.65)

2.2.4 Clasificación entre dos poblaciones

Sean P_1 y P_2 dos poblaciones donde tenemos definida una variable aleatoria vectorial, x. Si se supone que x es absolutamente continua y que las funciones de densidad de ambas poblaciones, f_1 y f_2, son conocidas.

En el caso de querer clasificar un elemento nuevo x_0, Si conocemos las probabilidades a priori π_1, π_2, con $\pi_1 + \pi_2 = 1$, de que el elemento venga de cada una de las dos poblaciones, su distribución de probabilidad será:

$$f(x) = \pi_1 f_1(x) + \pi_2 f_2(x)$$

(2.66)

y una vez observado x_0 se calcula las probabilidades a posteriori de que el elemento haya sido generado por cada una de las dos poblaciones, $P(i|x_o)$,con $i = 1,2$. Estas probabilidades se calculan por el Teorema de Bayes:

$$P(1|x_0) = \frac{P(x_0|1)\pi_1}{\pi_1 P(x_0|1) + \pi_2 P(x_0|2)}$$

como $P(x_0|1) = f_1(x_0)\Delta_{x_0}$, entonces:

$$P(1|x_0) = \frac{f_1(x_0)\pi_1}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

(2.67)

y para la segunda población

$$P(2|x_0) = \frac{f_2(x_0)\pi_2}{f_1(x_0)\pi_1 + f_2(x_0)\pi_2}$$

(2.68)

Se clasifica x_0 en la población más probable a posteriori. Como los denominadores son iguales, se clasifica x_0 en P_2 si:

$$\pi_2 f_2(x_0) > \pi_1 f_1(x_0)$$

Si las probabilidades a priori son iguales, la condición de clasificar en P_2 se reduce a:

$$f_2(x_0) > f_1(x_0)$$

es decir, se clasifica a x_0 en la población más probable, o donde su verosimilitud es más alta.

Consideración de las consecuencias

Si se clasifica al elemento en el grupo 2 las posibles consecuencias son:

(a) Acertar, con probabilidad $P(2|x_0)$, en cuyo caso no hay ningún costo de penalización.

(b) Equivocarse, con probabilidad $P(1|x_0)$, en cuyo caso se incurre en el costo asociado $c(2|1)$[30].

El costo promedio, o valor esperado, de la decisión "d_2: clasificar x_0 en P_2" será:

$$E(d_2) = c(2|1)P(1|x_0) + 0P(2|x_0) = c(2|1)P(1|x_0)$$

[30] $c(i|j)$ es el costo de clasificación en P_i de una unidad que pertenece a P_j.

Análogamente, el costo esperado de la decisión "d_1: clasificar en el grupo 1" es:

$$E(d_1) = 0P(1|x_0) + c(1|2)P(2|x_0) = c(1|2)P(2|x_0)$$

Se asigna el elemento al grupo 2 si su costo esperado es menor, es decir, utilizando (2.67) y (2.68), si:

$$\frac{f_2(x_0)\pi_2}{c(2|1)} > \frac{f_1(x_0)\pi_1}{c(1|2)}$$

(2.69)

Esta condición indica que, a igualdad de los otros términos, se clasifica en la población P_2, si:

(a) Su probabilidad a priori es más alta;
(b) La verosimilitud de que x_0 provenga de P_2 es más alta;
(c) El costo de equivocación al clasificarlo en P_2 es más bajo.

2.2.4.1 Poblaciones Normales

Ahora se aplica el análisis anterior al caso en que f_1 y f_2 son distribuciones normales con distintos vectores de medias pero idéntica matriz de varianzas. Para establecer la regla con carácter general se supone que se desea clasificar un elemento genérico x. [31]

La participación óptima, es, de acuerdo con clasificación anterior, clasificar en la población P_2 si:

$$\frac{f_2(x)\pi_2}{c(2|1)} > \frac{f_1(x)\pi_1}{c(1|2)}$$

Como ambos términos son siempre positivos, tomando logaritmos y sustituyendo $f_i(x)$ por su expresión la ecuación anterior se convierte en:

$$D_1^2 - log\frac{\pi_1}{c(1|2)} > D_2^2 - log\frac{\pi_2}{c(2|1)}$$

(2.70)

---

[31] $f_i(x) = \frac{1}{(2\pi)^{p/2}|V|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu_i)'V^{-1}(x-\mu_i)\right\}$

y suponiendo iguales los costos y las probabilidades a priori, $c(1/2) = c(2/1)$; $\pi_1 = \pi_2$, la regla anterior se reduce a clasificar en la población 2 si $D_1^2 > D_2^2$ [32], es decir ,clasificar la observación en la población de cuya media esté más próxima, midiendo la distancia con la medida de Mahalanobis.

2.2.4.2 Cálculo de Probabilidades de error

La utilidad de la regla de clasificación depende de los errores esperados. Sea la distribución de la variable $z = w'x$ normal, con media $m_i = w'\mu_i$ y varianza $D^2 = (m_2 - m_1)^2$, se puede calcular las probabilidades de clasificar erróneamente una observación en cada una de las dos poblaciones.

La probabilidad de una decisión errónea cuando $x \epsilon P_1$ es:

$$P(2|1) = P\left\{z \geq \frac{m_1 + m_2}{2} |z\ es\ N(m_1; D)\right\}$$

si $y = (z - m_1)/D$, una variable aleatoria N(0,1), y Φ su función de distribución, entonces

$$P(2|1) = P\left\{y \geq \frac{\frac{m_1 + m_2}{2} - m_1}{D}\right\} = 1 - \Phi\left(\frac{D}{2}\right)$$

(2.71)

Análogamente, la probabilidad de una decisión errónea cuando $x \epsilon P_2$ es:

$$P(1|2) = P\left\{z \leq \frac{m_1 + m_2}{2} |z\ es\ N(m_2; D)\right\} = P\left\{y \leq \frac{\frac{m_1 + m_2}{2} - m_2}{D}\right\} = \Phi\left(-\frac{D}{2}\right)$$

(2.72)

y ambas probabilidades de error son idénticas, por la simetría de la distribución normal.

Por lo tanto se concluye que la regla obtenida hace iguales y mínimas las probabilidades de error y que los errores de clasificación sólo dependen de las distancias de Mahalanobis entre las medias.

---

[32] $D_i^2 = (x - \mu_i)'V^{-1}(x - \mu_i)$ es la distancia de Mahalanobis entre el punto observado ,x,y la media de la población i.

2.2.4.3 Probabilidades a posteriori

El grado de confianza al clasificar una observación depende de la probabilidad de acertar. La probabilidad a posteriori de que la observación pertenezca a la primera población se calcula con:

$$P(1|x) = \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)} = \frac{\pi_1 exp\left\{-\frac{1}{2}(x-\mu_1)' V^{-1}(x-\mu_1)\right\}}{\pi_1 exp\left\{-\frac{1}{2}(x-\mu_1)' V^{-1}(x-\mu_1)\right\} + \pi_2 exp\left\{-\frac{1}{2}(x-\mu_2)' V^{-1}(x-\mu_2)\right\}}$$

y llamando D_1^2 y D_2^2 a las distancias de Mahalanobis entre el punto y cada una de las dos medias, ésta expresión puede escribirse:

$$P(1|x) = \frac{1}{1 + \frac{\pi_2}{\pi_1} exp\left\{-\frac{1}{2}(D_2^2 - D_1^2)\right\}}$$

(2.73)

y sólo depende de las probabilidades a priori y de las distancias entre el punto y las medias de ambas poblaciones. Si $\pi_2/\pi_1 = 1$, cuanto más alejado está el punto de la primera población, es decir, cuanto mayor sea D_1^2 respecto a D_2^2, mayor será el denominador y menor será la probabilidad de que pertenezca a ella,$P(1|x)$ y al contrario.

2.2.5 Poblaciones Desconocidas. Caso General

Sea un conjunto de G poblaciones posibles, la matriz general de datos X de dimensiones n*p, n individuos y p variables), puede considerarse particionada en G matrices correspondientes a subpoblaciones, los elementos de estas submatrices x_{ijg}, donde i representa el individuo, j la variable y g el grupo o submatriz.

El número de elementos en el grupo g es n y el número total de observaciones es:

$$n = \sum_{g=1}^{G} n_g$$

(2.74)

El vector de medias dentro de cada clase o subpoblación es:

$$\bar{x}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} x_{ig}$$

(2.75)

La matriz de varianzas covarianzas para los elementos de la clase g será:

$$\hat{S}_g = \frac{1}{n_g - 1}\sum_{i=1}^{n_g}(x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)'$$

(2.76)

Si suponemos que la G subpoblaciones tienen la misma matriz de varianzas y covarianzas, su estimación centrada es:

$$\hat{S}_w = \sum_{g=1}^{G}\frac{n_g - 1}{n - G}\hat{S}_g$$

(2.77)

La matriz de sumas de cuadrados dentro de las clases viene dada por:

$$W = (n - G)\hat{S}_w$$

(2.78)

Suponiendo las probabilidades a priori y los costos de clasificación, se clasifica al elemento en el grupo que conduzca a un valor mínimo de la distancia de Mahalanobis entre el punto x y la media del grupo.

Si $\widehat{w}_g = \hat{S}_w^{-1}\bar{x}_g$ un nuevo elemento x_0 se clasifica en la población g donde

$$\min_g \left(x_0 - \bar{x}_g\right)'\hat{S}_w^{-1}\left(x_0 - \bar{x}_g\right) = \min_g \widehat{w}_g'\left(\bar{x}_g - x_0\right)$$

(2.79)

lo cual equivale a construir las variables indicadoras escalares

$$z_{g,g+1} = \widehat{w}_{g,g+1}' x_0 \qquad g = 1, \dots, G$$

(2.80)

donde

$$\widehat{w}_{g,g+1} = \hat{S}_w^{-1}\left(\bar{x}_g - \bar{x}_{g+1}\right) = \widehat{w}_g - \widehat{w}_{g+1}$$

y clasificar en g frente a g+1 si

$$\left|z_{g,g+1} - \widehat{m}_g\right| < \left|z_{g,g+1} - \widehat{m}_{g+1}\right| \tag{2.81}$$

donde $\widehat{m}_g = \widehat{w}'_{g,g+1}\bar{x}_g$.

2.2.5.1 Cálculo de Probabilidades de error

El cálculo de probabilidades de error podría hacerse sustituyendo los parámetros desconocidos por los estimados y aplicando las fórmulas antes mencionadas, pero este método no es recomendable ya que va a subestimar mucho las probabilidades de error al no tener en cuenta la incertidumbre de estimación de los parámetros. Un mejor procedimiento, que además no depende de la hipótesis de normalidad, es aplicar la función discriminante a las n observaciones y clasificarlas. En el caso de 2 grupos, obtendríamos la tabla:

		Clasificado	
		$\boldsymbol{P_1}$	$\boldsymbol{P_2}$
Realidad	$\mathbf{P_1}$	n_{11}	n_{12}
	$\mathbf{P_2}$	n_{21}	n_{22}

donde n_{ij} es el número de datos que viniendo de la población i se clasifica en j. El error aparente de la regla es:

$$Error = \frac{n_{12} + n_{21}}{n_{11} + n_{22}} = \frac{Total\ mal\ clasificados}{Total\ bien\ clasificados} \tag{2.82}$$

Este método tiende a subestimar las probabilidades de error ya que los mismos datos se utilizan para estimar los parámetros y para evaluar el procedimiento resultante. Un procedimiento mejor es clasificar cada elemento con una regla que no se ha construido usándolo.

Para ello, podemos construir n funciones discriminantes con las n muestras de tamaño $n-1$ que resultan al eliminar uno a uno cada elemento de la población y clasificar después cada dato con la regla construida sin él. Este método se conoce como validación cruzada y conduce a una mejor estimación del error de clasificación. Si el número de observaciones es muy alto, el costo computacional de la validación cruzada es alto y una solución más rápida es subdividir la muestra en k grupos iguales y realizar la validación cruzada eliminado en lugar de una observación uno de estos grupos.

CAPÍTULO 3

CARACTERIZACIÓN DE LA PEA UTILIZANDO MÉTODOS DE ESTADÍSTICA MULTIVARIANTE.

En este capítulo se desarrollará modelos de regresión logística multinomial y análisis discriminante de acuerdo a las características de las variables que vamos a utilizar.

Nuestro objetivo es predecir el comportamiento de la variable dependiente estado de ocupación (eocupa) en función del menor número de variables independientes; es decir dar a conocer según el modelo logístico cuales son las variables que influyen en el estado de ocupación de una persona; y a continuación desarrollar el análisis discriminante para obtener los grupos en los cuales se clasifica dicho estado.

3.1 Definición de Variables.

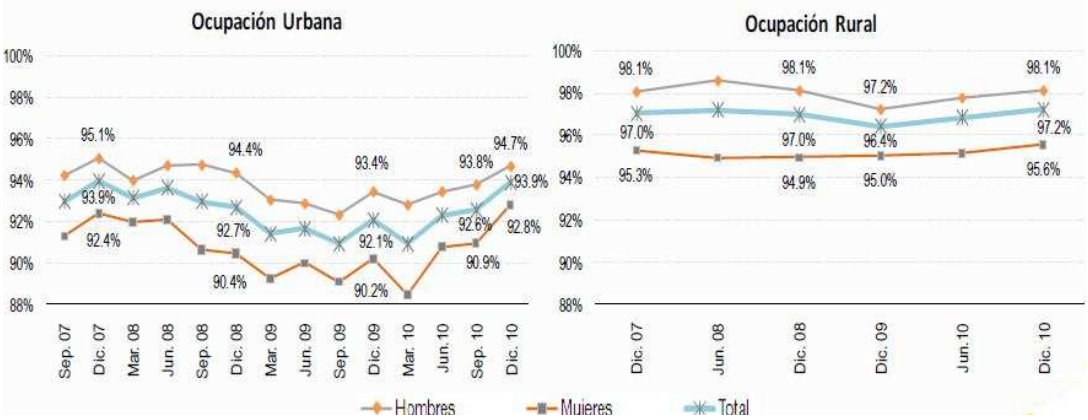
El conjunto de variables a utilizar son aquellas que sean estadísticamente significativas y que estén asociadas fuertemente con el estado de ocupación de la PEA (variable dependiente).
Las variables que tengan n categorías serán transformadas a (n-1) variables dicotómicas ficticias.

3.2 Análisis descriptivo de las variables.

3.2.1 Ocupación Total y por Sexo

De diciembre 2007 a diciembre 2008, la ocupación rural se ha mantenido en un 97%; con una tendencia constante hasta diciembre 2010 (97,2%).
En el área urbana, en diciembre 2007 la ocupación urbana es del 93,9%, y se mantiene constante hasta diciembre 2009(92.1%) a partir de este mes tiende al alza a diciembre 2010 (93.9%).

Se puede observar claramente que la ocupación rural es superior respecto al área urbana en todos los años. De igual manera, por sexo, el área rural supera a la urbana, tanto en la ocupación de hombres y mujeres. (Gráfico 8)



Fuente: Informe Trimestral del Mercado Laboral a Diciembre 2010
Elaboración: Banco Central del Ecuador

Gráfico 8: Ocupación Total y por Sexo

3.2.2 Subocupación Total y por Sexo

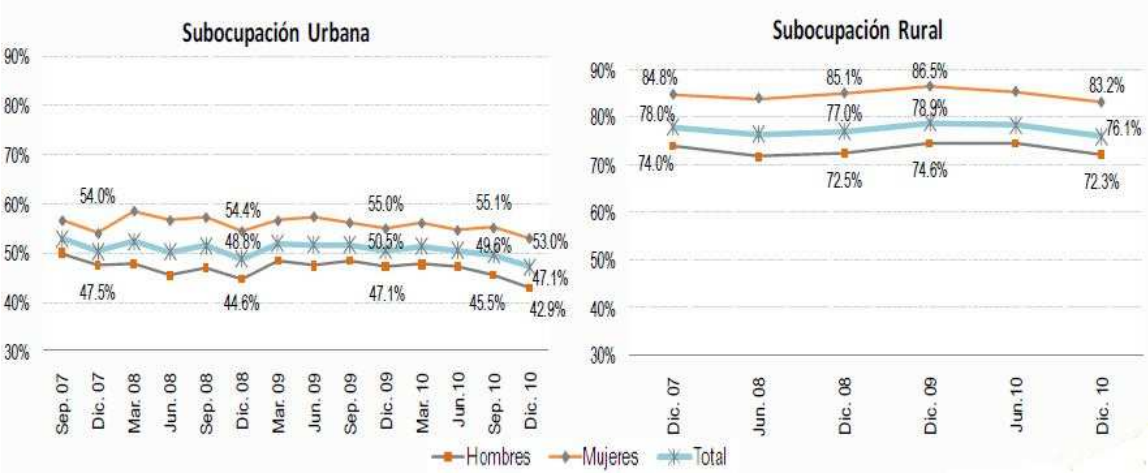
La tasa de subocupación total**,** a diciembre 2007, fue superior en el sector rural (78%) en comparación a la registrada en el área urbana (50%). Por sexo, 84,8% y 74% corresponden a la subocupación de mujeres y hombres del área rural, respectivamente; en tanto que el 54% y 47,5% corresponden a las mujeres y hombres del área urbana.

En diciembre del 2008, la tasa de subocupación total**,** fue superior en el sector rural (77%) en comparación a la registrada en el área urbana (48,8%). Por sexo, 85,1% y 72,5% corresponden a la subocupación de mujeres y hombres del área rural, respectivamente; en tanto que el 54,4% y 44,6% corresponden a las mujeres y hombres del área urbana.

En el área rural, la tasa de subocupación total, para diciembre del 2009 fue del 78,9% mayor a la del área urbana 50,5%.
Por sexo, 86,5% y 74,6% corresponden a la subocupación de mujeres y hombres del área rural, respectivamente; en tanto que el 55% y 47,1% corresponden a las mujeres y hombres del área urbana.

De la misma manera en diciembre del 2010 la tasa de subocupación fue del 76,1% para el área rural mientras que para el área urbana una tasa menor de 47,1%.

El 83,2% y el 72,3% corresponden a la subocupación de mujeres y hombres del área rural, valores superiores a las tasas del área urbana 53% para las mujeres y 42,9% para los hombres.



Fuente: Informe Trimestral del Mercado Laboral a Diciembre 2010.
Elaboración: Banco Central del Ecuador.

Gráfico 9: Subocupación Total y por Sexo.

3.2.3 Desocupación total y por sexo

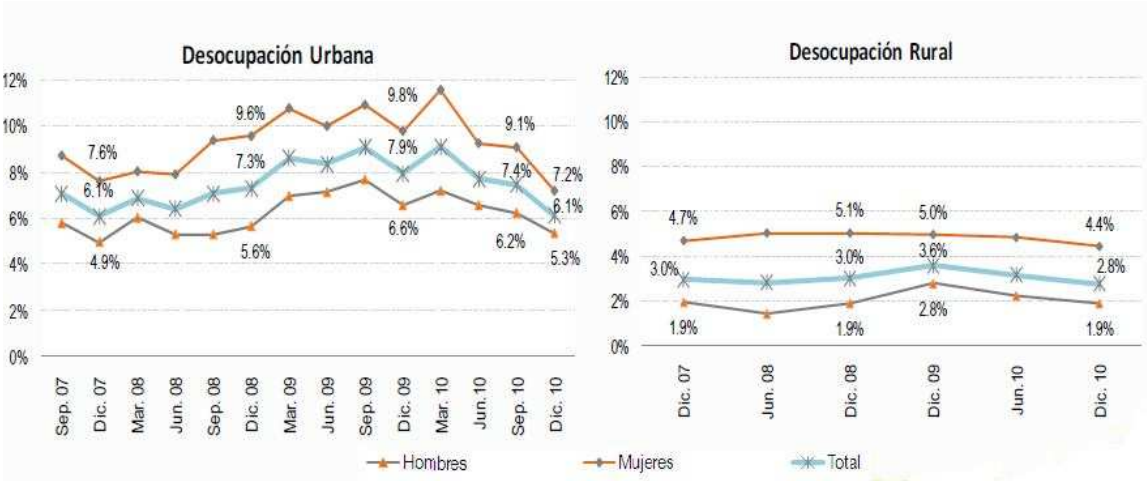
Dentro del área urbana, los desocupados en diciembre 2007 fueron del 6,1%**,** en tanto que las personas desocupadas en el área rural registraron el 3% con respecto a la PEA de cada sector. Las cifras evidencian que el desempleo en las mujeres y hombres urbanos fue superior al femenino y masculino del área rural (7,6% y 4,9% urbano; y, 4,7% y 1,9%, respectivamente).

Para diciembre 2008 en el área urbana, los desocupados fueron el 7,3%**,** en tanto que las personas desocupados rurales registraron el 3% con respecto a la PEA de cada sector. Las cifras evidencian que el desempleo en las mujeres y hombres urbanos fue superior al femenino y masculino del área rural (9,6% y 5,6% urbano; y, 5,1% y 1,9%, respectivamente).

Los desocupados en el área urbana a diciembre 2009 fueron del 7,9% y en el área rural de 3,6%. El desempleo en las mujeres y hombres urbanos fue superior

al femenino y masculino del área rural (9,8% y 6,6% urbano; y 5,0% y 2,8% respectivamente.)

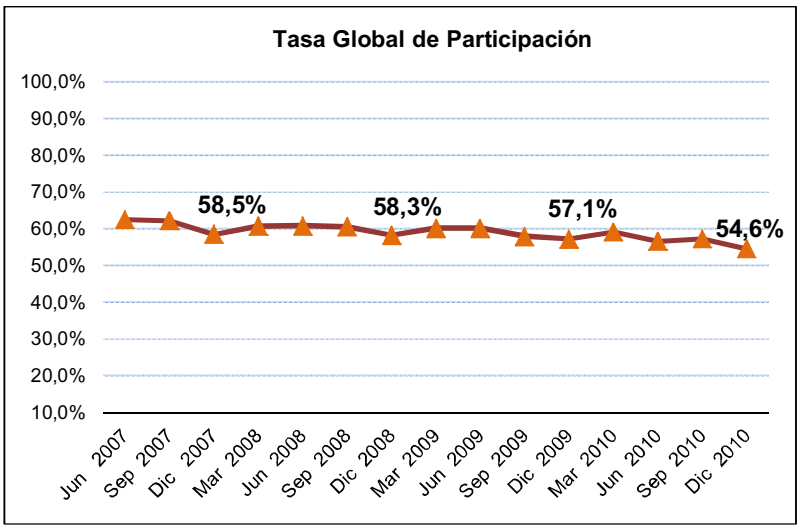
Dentro del área urbana, los desocupados fueron el 6,1%**,** en tanto que las personas desocupados rurales registraron el 2,8% con respecto a la PEA de cada sector. Las cifras evidencian que el desempleo en las mujeres y hombres urbanos fue superior al femenino y masculino del área rural (7,2% y 5,3% urbano; y, 4,4% y 1,9%, respectivamente).a diciembre 2010.



Fuente: Informe Trimestral del Mercado Laboral a Diciembre 2010.
Elaboración: Banco Central del Ecuador.

Gráfico 10: Desocupación Total y por Sexo.

3.2.4 Tasa de Participación Global Urbana (TPG)



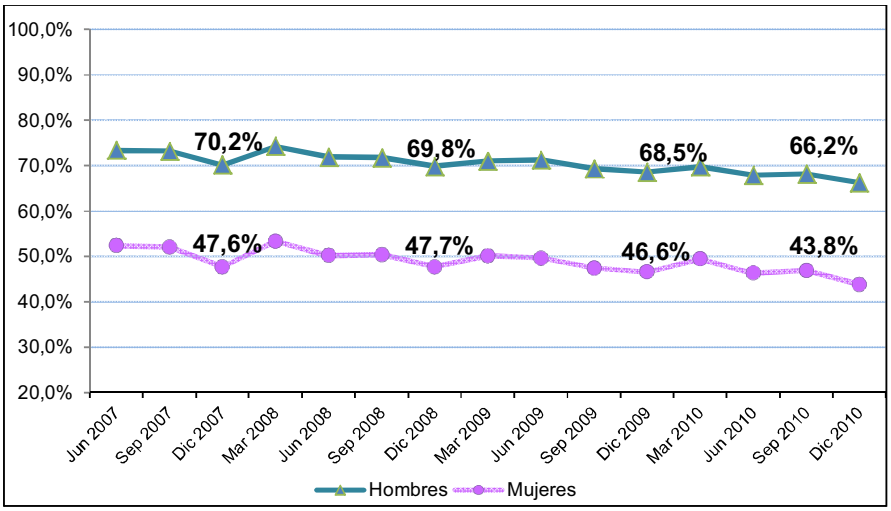
Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración: Paola Troya

Gráfico 11: Tasa de Participación Global Urbana

La Tasa de Participación Global en el año 2010 alcanza el 54,6%, es decir, por cada 100 personas que se encuentran en edad de trabajar, 55 están conformando la Población Económicamente Activa (PEA).

De diciembre del 2007 a diciembre 2009, la tasa de participación global tiene una tendencia a mantenerse constante, (58,5%, 58,3%, y 57,1% respectivamente) mientras que a diciembre del 2010 tiende a la baja, con respecto a diciembre del 2009 en 2,5 puntos porcentuales (ver Gráfica 11).

3.2.5 Tasa de Participación Global Urbana según sexo (TPG)



Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración: Paola Troya

Gráfica 12: Tasa de Participación Global Urbana según Sexo

La tasa de participación global permite cuantificar el tamaño de la fuerza de trabajo. Al observar la gráfica 12, la participación de las mujeres en la fuerza laboral es menor con respecto a la participación de los hombres.

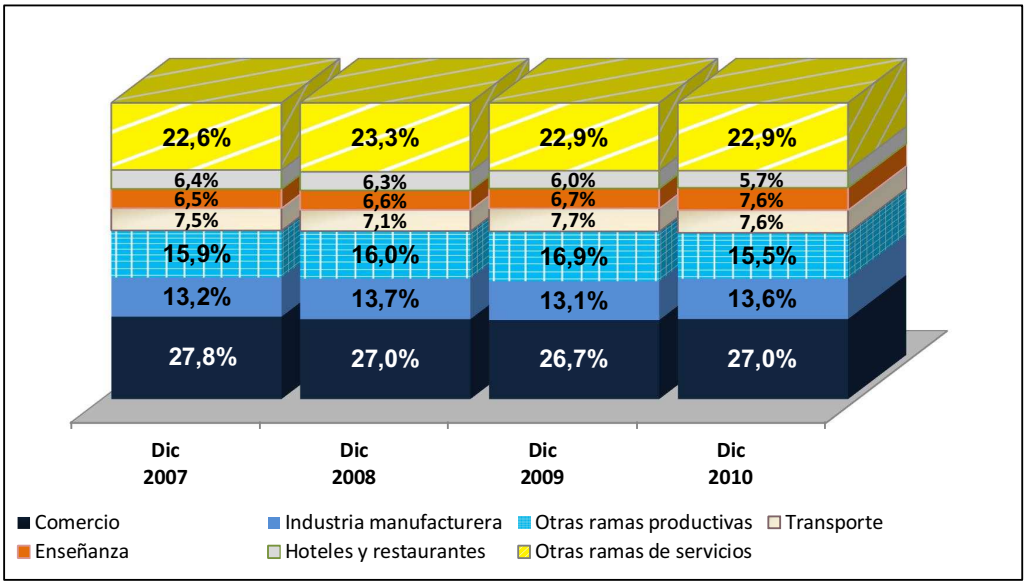
En diciembre 2010 la tasa de participación global de las mujeres alcanzó el 43,8% que al comparar con el indicador del mismo mes del año anterior es menor en 2,8 puntos porcentuales. El indicador para la población masculina alcanzó el 66,2%, es decir que por cada 100 hombres en edad de trabajar, aproximadamente 66 son económicamente activos.

Desde diciembre 2007 (47,6%), diciembre 2008(47,7%) y hasta diciembre 2009(46,6%); la tasa de participación global urbana en las mujeres ha mantenido una tendencia constante.

3.2.6 Principales Ramas de Actividad en el Área Urbana

La rama de actividad económica permite clasificar la actividad del establecimiento según la clase de bienes o servicios que produce.

De acuerdo a la gráfica 13, la mayor proporción de la población ocupada en el área urbana está en la actividad de comercio. A continuación se tienen la industria manufacturera, transporte, almacenamiento y comunicaciones; enseñanza; hoteles y restaurantes.



Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración Paola Troya

Gráfica 13:[33]Principales Ramas de Actividad en el área urbana

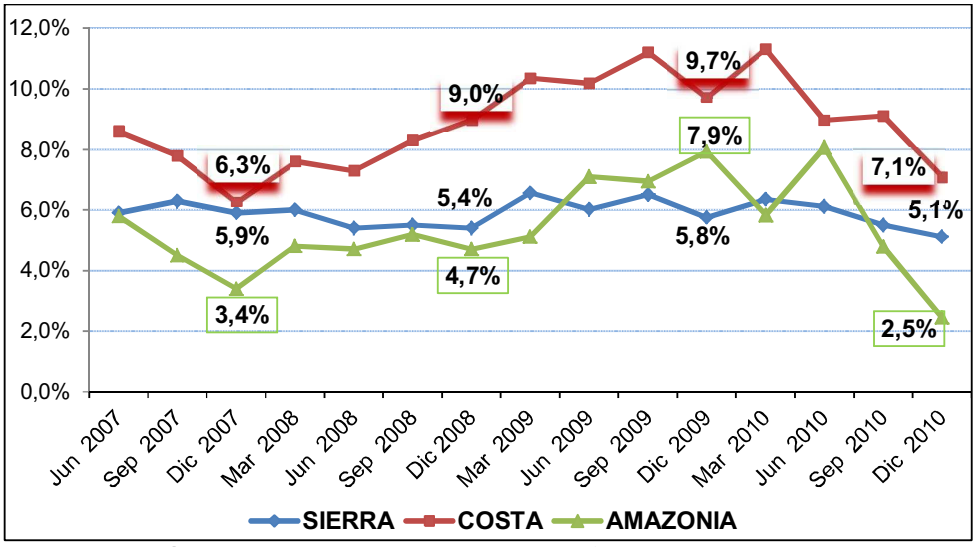
[33]Las alternativas graficadas como Otras ramas productivas (14.9%), corresponden a la agrupación de: Agricultura, Ganadería y Caza, Construcción, Pesca y Criaderos, Explotación de Minas y Canteras y Suministro de Electricidad, Gas y Agua. En la rama "Otras Ramas de Servicios" (23,0%.) está: Hogares con Servicio Doméstico, Intermediación Financiera, Actividades Inmobiliarias, Administración Pública, Servicios Sociales y de Salud, Otras Actividades Comunitarias sociales y personales. Estas ramas presentan porcentajes de distribución independientes con valores de hasta un 5 por ciento.

3.2.7 Evolución del Desempleo en las áreas Urbanas de las Regiones Naturales

En la gráfica 14 se puede observar que en la región Costa hay un mayor índice de desempleo que ha tenido una tendencia al alza a partir de diciembre 2007(6,3%) hasta diciembre 2009 (9,7%), mes a partir del cual tiende a disminuir hasta diciembre 2010(7,1%).

El segundo lugar para el desempleo es la sierra con un porcentaje de 5,9 a diciembre 2007, con una tendencia constante hasta diciembre 2010(5,1%).

En la región amazónica estaría en el tercer lugar con un índice de desempleo menor en comparación a las otras regiones en diciembre 2007(3,4%), con una tendencia a subir a diciembre 2009(7,9%) y a disminuir hasta diciembre 2010(2,5%).

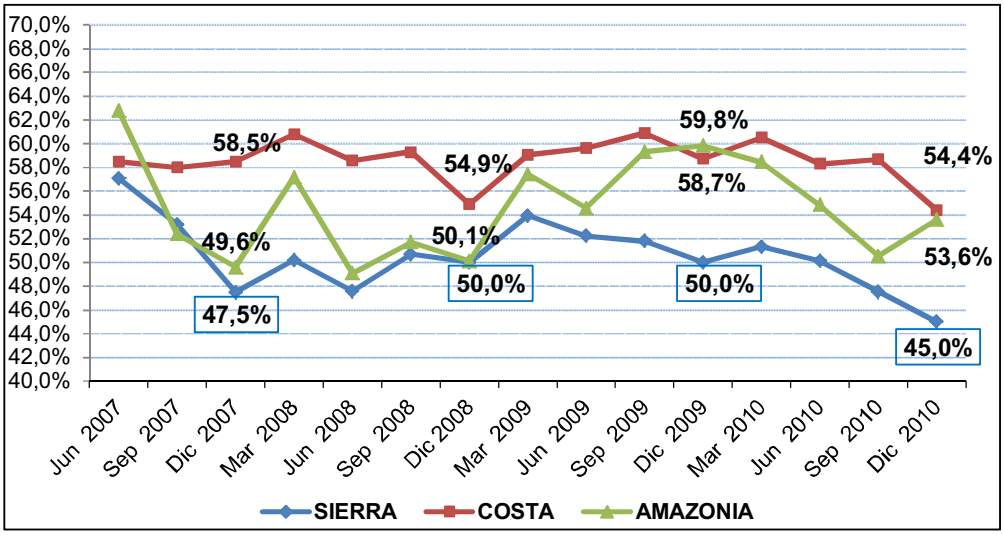


Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración Paola Troya

Gráfica 14: Evolución del Desempleo en las áreas urbanas de las Regiones Naturales

Observando las cifras desde diciembre 2007, la gráfica demuestra la altísima sensibilidad del indicador en la región costa y amazonia, no así en la región sierra que muestra un comportamiento más bien estable.

3.2.8 Evolución del Subempleo Global Urbano, según Regiones Naturales



Fuente: Evolución del Mercado Laboral a Diciembre 2010.INEC
Elaboración Paola Troya

Gráfica 15: Evolución del Subempleo Global Urbano por Regiones Naturales

El indicador de subempleo en la región Costa en el mes de diciembre 2010, alcanzó el 54,4%, valor que al comparar con el indicador de diciembre 2009 es menor en 4,3 puntos porcentuales.

El indicador de la región Amazónica como se observa en la gráfica 15; presenta una variación más pronunciada desde diciembre 2007. A diciembre del 2010 alcanzó el 53,6%, valor que es menor en 6,2 puntos porcentuales, respecto al indicador del año anterior (diciembre 2009).

3.2.9 PEA según grupos etarios*:*

Los estados de ocupación que conforman la PEA son: la ocupación plena, el subempleo y el desempleo, los cuales a continuación se describen por grupos de edad.

OCUPACION PLENA	Dic 2007	Dic 2008	Dic 2009	Dic 2010
10 a 17 años	5,16%	5,40%	3,42%	5,37%
18 a 29 años	28,56%	28,93%	25,64%	32,41%
30 a 39 años	35,86%	38,56%	34,04%	38,62%
40 a 49 años	38,33%	39,14%	34,96%	39,07%
50 a 64 años	35,12%	37,86%	33,06%	35,48%
65 años y más	21,23%	21,75%	19,32%	19,94%

Fuente: Evolución del Mercado Laboral a Diciembre 2010
Elaboración Paola Troya

Tabla 1: Tasa de Ocupación Plena

SUBEMPLEO	Dic 2007	Dic 2008	Dic 2009	Dic 2010
10 a 17 años	87,39%	86,47%	87,42%	87,74%
18 a 29 años	63,85%	61,33%	63,67%	58,27%
30 a 39 años	60,48%	58,04%	61,27%	57,44%
40 a 49 años	58,96%	58,20%	61,76%	58,30%
50 a 64 años	62,83%	60,17%	64,68%	62,72%
65 años y más	77,16%	76,87%	79,46%	79,25%

Fuente: Evolución del Mercado Laboral a Diciembre 2010
Elaboración Paola Troya

Tabla 2: Tasa de Subempleo

DESEMPLEO	Dic 2007	Dic 2008	Dic 2009	Dic 2010
10 a 17 años	7,45%	8,13%	9,16%	6,88%
18 a 29 años	7,59%	9,73%	10,69%	9,32%
30 a 39 años	3,66%	3,41%	4,69%	3,94%
40 a 49 años	2,71%	2,66%	3,29%	2,63%
50 a 64 años	2,05%	1,98%	2,25%	1,79%
65 años y más	1,60%	1,38%	1,23%	0,81%

Fuente: Evolución del Mercado Laboral a Diciembre 2010
Elaboración Paola Troya

Tabla 3: Tasa de Desempleo

Los estados de ocupación se han establecido por grupos de edad en vista de que la edad es un factor influyente para analizar dicho estado en una persona.

Tasa de Ocupación Plena

De la tabla 1, la ocupación plena en el rango de edades de 10 a 17 años tiene un porcentaje mínimo en vista de que las personas pertenecientes a ese grupo, generalmente se encuentran estudiando, o no tienen la edad suficiente como para

cumplir con los requerimientos de encontrarse ocupados plenos. En el grupo de 18 a 29 años, tenemos un porcentaje de alrededor de un 30%, puesto que es probable que éstas personas ya tengan un título profesional y se encuentren trabajando satisfactoriamente.
A partir de los 30 a 49 años, el porcentaje es de 35 a 39% en todos los años, ya que es una condición que se puede considerar que una persona ha estudiado y(o) luego de un tiempo prudencial ha conseguido trabajo.
En el grupo de 50 a 64 años el porcentaje tiende a disminuir al 35% y de 65 años en adelante al 21%, de diciembre 2007 a diciembre 2010, el porcentaje de ocupados plenos es menor en éste grupo porque a ésta edad las personas se acogen a la jubilación.

Tasa de Subempleo

De la tabla 2 se puede concluir que en el rango de edad de 10 a 17 años el porcentaje de alrededor del 86% (diciembre 2007 a diciembre 2010), es un porcentaje muy alto porque quienes se encuentran en éste grupo son más susceptibles de estar subempleados puesto que las condiciones del subempleo son poco exigentes.

En las edades de 18 a 64 años, el subempleo es alrededor del 60%, porque ante la falta de empleo, la necesidad de sobrevivir obliga a las personas a aceptar cualquier tipo de trabajo aunque no sea remunerado satisfactoriamente.

A partir de los 65 años el porcentaje es de alrededor del 78% (dic 2007-dic 2010), éste grupo es más vulnerable porque por la edad no son considerados aptos para encontrarse en un trabajo estable y bien remunerado, además a éste grupo también pertenecerían quienes se acogen a la jubilación y optan por tener una ocupación que les ubica en el subempleo.

Tasa de Desempleo

Se puede observar en la tabla 3 que, en el rango de 10 a 17 años el desempleo se ubica en alrededor de un 8%, a éste grupo pertenecerían quienes se encuentran en su etapa de estudios y quienes no han tenido la oportunidad de encontrar un trabajo, por lo que su principal limitante sería la edad que no le permitiría desenvolverse satisfactoriamente en un trabajo.

De 18 a 29 años, la tasa es de un 9% aproximadamente (dic 2007-dic 2010), es un valor considerable puesto que en éste grupo las personas ya tienen mayor probabilidad de haber terminado sus estudios y es probable que por ésta razón tengan mayor probabilidad de encontrar trabajo; además de poder ser considerados con mayores aptitudes y capacidades.

En el grupo de de 30 a 39 años el porcentaje de desempleados es de alrededor del 4% y hay una tendencia a disminuir conforme aumenta la edad de las personas.
De 40 a 49 años el desempleo es de alrededor del 3%, hasta el último grupo de 65 años y más que tiene una tasa mínima de aproximadamente el 1%.
Estas cifras nos indican que la mayoría de personas se encuentran en la ocupación plena y en el subempleo, una de las razones podría ser que a través del tiempo han aumentado las oportunidades laborales, principalmente porque el mes en el cual se realiza el análisis es temporada donde el comercio aumenta considerablemente.

3.3 Variables Utilizadas:

Variables originales en la base:

1) área :corresponde al área donde pertenecen las personas, se define el valor 0 como área rural y el valor 1 como área urbana,(variable dicotómica)
2) p02: corresponde al sexo al cual pertenece cada individuo, se define al valor 1 como hombre y el valor 0 como mujer, (variable dicotómica).

3) p03: se refiere a la edad de cada individuo. (variable numérica).
4) p07: es la pregunta referente a asistencia a clases, teniendo al valor 1 como asistencia y el valor 0 no asistencia (variable dicotómica).
5) p10a: es el nivel de instrucción, en la encuesta se considera las siguientes categorías para ésta variable:
   1ninguno
   2centro de alfabetización
   3primaria
   4educación básica
   5secundaria
   6educación media
   7superior no universitaria
   8superior universitaria
   9post-grado

Esta pregunta nos permite conocer el nivel de estudios alcanzado por la persona dentro del sistema de educación formal y reglamentada por el Ministerio de Educación o por el Consejo de Universidades y Escuelas Politécnicas, CONESUP, y que son aplicados a todos los establecimientos educacionales.[34]

6) p10b: ésta variable es numérica y corresponde al número de años aprobados por cada individuo, en el nivel de instrucción en el que se encuentre (p10a).

7) p45: es la pregunta que se refiere a la experiencia de las personas que están trabajando, cantidad de años de trabajo. (variable numérica).

8) p50: es el número de trabajos que pudiera tener una persona. (variable numérica).

---

[34] Manual del Encuestador Diciembre 2010, pag.51

9) p75: se refiere a la pregunta ¿Recibió el bono de desarrollo urbano?, siendo el valor 0 la respuesta negativa y el valor 1 la respuesta positiva.(variable dicotómica)

Variables Creadas:

10) eocupa: es la variable dependiente que contiene las 3 categorías, tomando como valor 0, la ocupación plena, 1 el subempleo y 2 el desempleo. (variable categórica).

11) escol: ésta variable es numérica y se refiere a los años de escolaridad, es decir a todos los años de estudio aprobados a lo largo de toda la vida de la persona hasta el momento de la encuesta, esta variable es construida en base a dos variables: nivel de instrucción y año aprobado.

12) Ingreso: es el salario de los individuos que trabajan, esta variable se ha tomado en cuenta como el logaritmo de los ingresos para obtener una mejor aproximación del modelo puesto que estas variables tienen valores que son muy altos y se van a comparar con variables dicotómicas que tienen valores 0 y 1.(variable numérica).

Región Natural: variable categórica, convertida en 3 variables dicotómicas que tendrían valor 1 si pertenecen a cada región y 0 caso contrario.

13) Sierra
14) Costa
15) Amazonía

Año: variable categórica, convertida en 4 variables dicotómicas que tendrían valor 1 si pertenecen al año indicado y valor 0 en caso contrario.

16) año0:2007
17) año1:2008

18) año2:2009
19) año3:2010

Nota: en la base condensada se tenía variables categóricas con valores 1 y 2, pero de manera que todas las variables se encuentren en una descripción uniforme, a todas las que eran categóricas se les definió como variables dicotómicas.

3.4 Interpretación y Resultados

3.4.1 Modelos de Regresión Logística Multinomial

Luego de analizar la teoría de los modelos de regresión logística, se procede a la estimación del modelo y la interpretación de los parámetros y sus coeficientes.
El objetivo de nuestro trabajo es estudiar las variables que influyen en los estados de ocupación (eocupa, variable dependiente categórica, 0 ocupado pleno, 1 subempleado y 2 desempleado). Como variables independientes, se emplearán distintas variables sociodemográficas tanto cuantitativas como cualitativas (área, ingresos, edad, estudios, sexo, región, experiencia laboral), lo que queremos es construir un modelo que explique la mayor cantidad posible de variabilidad de la variable dependiente con el menor número posible de variables independientes.

El modelo no puede estimar las ecuaciones para las tres alternativas, ya que una de ellas debe de servir como la base, en este caso se escogió la categoría 0 igual a ocupados plenos ya que es el mejor estado en que se puede encontrar una persona.

La estimación del modelo de regresión logística se realiza por el método de máxima verosimilitud. Este método estima los valores de los parámetros *b* de la regresión que con mayor probabilidad pueden haber generado los valores de la variable dependiente de la muestra. Se calcula una función de verosimilitud que indica cuál es la probabilidad de que para unos determinados parámetros *b* se

hayan observado los valores muestrales. En un proceso iterativo, se van probando distintos valores de los parámetros b hasta que se encuentran los coeficientes que maximizan tal función de verosimilitud[35] (o sea, los coeficientes que más verosímilmente corresponden a los valores muestrales): tales coeficientes serán los estimados para nuestro modelo.

Este proceso iterativo lo realiza el comando mlogit, cuyo formato general es el siguiente:

mlogit variable dependiente, lista variables independientes

Entonces, para nuestro modelo se tiene:

mlogit eocupa área p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa, base(0)[36]

Ya que nuestra base contiene datos de cuatro períodos (años), se realizó un modelo para cada período, para analizar el comportamiento de la PEA en cada período y un modelo global de todas las observaciones conjuntas.

En el anexo 1 se encuentra los resultados de los modelos, donde se tiene el proceso iterativo de la estimación del modelo por el método de máxima verosimilitud.

En la primera iteración todos los coeficientes valen cero, a excepción de la constante y en iteraciones sucesivas se van aproximando valores de los coeficientes que incrementan el valor de la función de verosimilitud.
Se puede observar en cada iteración que el logaritmo de la verosimilitud es mayor hasta que se aproxima más a su máximo, cada iteración añade un valor más pequeño.

---

[35] De hecho, no se maximiza directamente la función de verosimilitud, sino su logaritmo (log likelihood), lo que simplifica la computación. Este log likelihood es el que aparece en la salida de Stata de logit.
[36] En Stata para especificar una categoría de la variable dependiente como base, el momento de estimar el modelo se debe incluir la opción base(0),0 es el número al que corresponde la categoría base.

El proceso iterativo termina cuando las iteraciones no añaden nada de verosimilitud al modelo y muestra los coeficientes estimados.
Como se puede observar en las iteraciones finales la diferencia del log likelihood es mínima.
A la derecha aparece el número de observaciones y una prueba estadística de significación del modelo, basada en la distribución χ^2. Con un nivel de confianza del 95%, el modelo es significativo si el valor p es menor a 0,05.Por tanto podemos concluir que la relación entre los coeficientes del modelo y la probabilidad de encontrarse en cualquiera de los tres estados de ocupación es significativa estadísticamente.

Por último aparece en ésta columna de estadísticos el pseudo R^2, es un estadístico análogo al R^2, que indica la bondad de ajuste del modelo a los datos. Aunque no tiene la inmediatez de interpretación del R^2 de la regresión lineal, que directamente indica qué proporción de la varianza de la variable dependiente es explicado por el modelo, es una aproximación basada en una comparación de la verosimilitud del modelo sólo con la constante, con la verosimilitud del modelo con todos los parámetros estimados.

$$pseudo\ R^2 = 1 - \left(\frac{\ln L_1}{\ln L_0}\right) \tag{3.1}$$

Siendo L_1, la razón de verosimilitud del modelo del que se desea estimar la bondad y L_0, la del modelo que sólo posee la constante.

Si bien el pseudo R^2 no es tan preciso como el R^2 de la regresión lineal, es una medida útil del ajuste del modelo a los datos, y puede servir para comparar la capacidad explicativa de modelos distintos.
Bajo estos estadísticos aparecen los coeficientes, del mismo modo que en la regresión lineal. La primera columna muestra los coeficientes para cada variable de la regresión logística.
De acuerdo a los resultados (anexo 1) es elemental que a diferencia de la regresión lineal en el modelo logit no se pueden interpretar directamente los coeficientes.

En la siguiente columna aparece el error típico de cada coeficiente, su valor z y la probabilidad asociada a ese valor z (que nos indica si el coeficiente es estadísticamente significativo), y los intervalos de confianza de cada coeficiente.

3.4.1.1 Pruebas de significancia de las variables independientes

Con respecto a las variables se plantea dos pruebas de significación diferentes:
El test de z para cada coeficiente, indicando hasta qué punto cada coeficiente tiene un efecto significativo en la ecuación, de acuerdo al p valor asociado se rechaza la hipótesis nula y la variable es significativa, hay variables con un p-valor>0.05 lo que quiere decir que no serían significativas a ese nivel, en los modelos el resultado se tiene para cada categoría de la variable dependiente.

La prueba de Wald que indica la significación del modelo completo, es decir hasta qué punto la relación existente entre la variable dependiente y el conjunto de variables independientes es significativa), basada en el χ^2 .

Para realizar la prueba de Wald, utilizamos el comando mlogtest, el cual reporta estadísticos de significancia de cada una de las variables del modelo y para cada modelo (anexo 1).

La calidad del ajuste del modelo se mide mediante coeficientes de determinación pseudo-R^2, cuyo valor es alrededor de 0,6 en los cuatro períodos el cual se encuentra dentro del intervalo en el que se puede considerarse una buena calidad del ajuste.

3.4.1.2 Calidad de predicción:

Una vez que el modelo ha sido estimado, generamos la variable predictora peocupa, la cual contiene los valores ajustados para cada uno de los modelos.

La sintaxis para este procedimiento esta en el Anexo 2, se realizó para cada año y se detalla los resultados a continuación:

AÑO 2007:

Con respecto a la calidad de la predicción del modelo se presenta la tabla de clasificación, la cual nos muestra un 84,64% del total de observaciones clasificadas correctamente.

peocupa (predicción)	Eocupa			Total
	Oplenos	Subempleados	Desempleados	
0	8.011	2.380	20	10.411
1	3.127	21.422	23	24.572
2	7	69	1.574	1.650
Total	11.145	23.871	1.617	36.633

peocupa (predicción)	Eocupa		
	oplenos	Subempleados	Desempleados
0	71,88%	9,97%	1,24%
1	28,06%	89,74%	1,42%
2	0,06%	0,29%	97,34%

Tabla 4: Clasificación Modelo Logístico-año 2007

Del 100% de ocupados plenos, el modelo predice el 71,88% correctamente del grupo de ocupados plenos, del 100% de los subempleados, el modelo predice el 89,74% de las observaciones de este grupo, y del 100% de los desempleados el modelo predice el 97,34% del grupo de desempleados.

AÑO 2008:

Con respecto a la calidad de la predicción del modelo se presenta la tabla de clasificación, la cual nos muestra un 85,20% del total de observaciones clasificadas correctamente.

peocupa (predicción)	Eocupa			Total
	Oplenos	Subempleados	Desempleados	
0	8.776	2.354	18	11.148
1	2.927	20.456	7	23.390
2	11	64	1.734	1.809
Total	11.714	22.874	1.759	36.347

peocupa (predicción)	Eocupa		
	oplenos	Subempleados	desempleados
0	74,92%	10,29%	1,02%
1	24,99%	89,43%	0,40%
2	0,09%	0,28%	98,58%

Tabla 5: Clasificación Modelo Logístico-año 2008

Del 100% de ocupados plenos, el modelo predice el 74,92% correctamente del grupo de ocupados plenos, del 100% de los subempleados, el modelo predice el 89,43% de las observaciones de este grupo, y del 100% de los desempleados el modelo predice el 98,58% del grupo de desempleados.

AÑO 2009:

Con respecto a la calidad de la predicción del modelo se presenta la tabla de clasificación, la cual nos muestra un 87,1% del total de observaciones clasificadas correctamente.

peocupa (predicción)	Eocupa			Total
	Oplenos	Subempleados	Desempleados	
0	7.719	1.936	16	9.671
1	2.675	22.031	1	24.707
2	21	55	2.009	2.085
Total	10.415	24.022	2.026	36.463

peocupa (predicción)	Eocupa		
	Oplenos	**Subempleados**	**desempleados**
0	74,11%	8,06%	0,79%
1	25,68%	91,71%	0,05%
2	0,20%	0,23%	99,16%

Tabla 6: Clasificación Modelo Logístico-año 2009

Del 100% de ocupados plenos, el modelo predice el 74,11% correctamente del grupo de ocupados plenos, del 100% de los subempleados, el modelo predice el 91,71% de las observaciones de este grupo, y del 100% de los desempleados el modelo predice el 99,16% del grupo de desempleados.

AÑO 2010:

Con respecto a la calidad de la predicción del modelo se presenta la tabla de clasificación, la cual nos muestra un 87,88% del total de observaciones clasificadas correctamente.

peocupa (predicción)	Eocupa			Total
	Oplenos	**Subempleados**	**Desempleados**	
0	10.198	2.153	9	12.360
1	2.414	21.640	3	24.057
2	13	38	1.724	1.775
Total	12.625	23.831	1.736	38.192

peocupa (predicción)	Eocupa		
	oplenos	**subempleados**	**desempleados**
0	80,78%	9,03%	0,52%
1	19,12%	90,81%	0,17%
2	0,10%	0,16%	99,31%

Tabla 7: Clasificación Modelo Logístico-año 2010

Del 100% de ocupados plenos, el modelo predice el 80,78% correctamente del grupo de ocupados plenos, del 100% de los subempleados, el modelo predice el 90,81% de las observaciones de este grupo, y del 100% de los desempleados el modelo predice el 99,31% del grupo de desempleados.

La calidad de predicción del modelo global es:

Con respecto a la calidad de la predicción del modelo se presenta la tabla de clasificación, la cual nos muestra un 86,25% del total de observaciones clasificadas correctamente.

peocupa (predicción)	**Eocupa**			**Total**
	Oplenos	**Subempleados**	**desempleados**	
0	34.919	8.967	68	43.954
1	10.908	85.370	24	96.302
2	72	261	7.046	7.379
Total	45.899	94.598	7.138	147.635

peocupa (predicción)	**Eocupa**		
	oplenos	**subempleados**	**desempleados**
0	76,08%	9,48%	0,95%
1	23,77%	90,25%	0,34%
2	0,16%	0,28%	98,71%

Tabla 8: Clasificación Modelo Logístico Global

Del 100% de ocupados plenos, el modelo predice el 76,08% correctamente del grupo de ocupados plenos, del 100% de los subempleados, el modelo predice el 90,25% de las observaciones de este grupo, y del 100% de los desempleados el modelo predice el 98,71% del grupo de desempleados.

Para hacer una interpretación de las probabilidades por categoría de la variable dependiente, en primer lugar presentamos todas las observaciones por categoría:

Eocupa	Freq.	Percent	Cum.
Oplenos	45.899	31,09%	31,09
subempleados	94.598	64,08%	95,17
desempleados	7.138	4,83%	100
Total	147.635	100%	

Luego, empezamos con un modelo nulo, sin regresores (Anexo2); sus coeficientes, son las razones de probabilidad de encontrarse en el estado de ocupación 1 versus el estado 0 y el 2 ante el 0.

En este modelo se muestra lo siguiente:

$$\boldsymbol{Ln\left(\frac{\pi_1}{\pi_0}\right) = \alpha_1}$$

$$\boldsymbol{Ln\left(\frac{94598}{45899}\right) = 0.723193} \quad \textbf{(1)}$$

Para la categoría 2:

$$\boldsymbol{Ln\left(\frac{7138}{45899}\right) = -1.861} \quad \textbf{(2)}$$

Si calculamos el exp (0,723193)= 2,05, este valor indica que la probabilidad de estar subempleado ante la ocupación plena es mayor con aproximadamente el doble.

En el segundo valor tenemos: exp (-1,861)=0,15, el cual indica que la probabilidad de encontrarse en la desocupación ante la ocupación plena categoría 3 versus la categoría 0 disminuye un 85%.

Si hacemos el mismo análisis (anexo 2) para cada año tendremos el siguiente resultado:

Categorías	2007	2008	2009	2010	modelo global
subempleo	0,76	0,66	0,835	0,635	0,72
desempleo	-1,93	-1,89	-1,637	-1,984	-1,86

Tabla 9: Coeficientes del modelo

Categorías	2007	2008	2009	2010	modelo global
Subempleo	2,13827622	1,93479233	2,30481405	1,88702214	2,054433211
Desempleo	0,1451482	0,15107181	0,19456286	0,13751806	0,15567263

Tabla 10: Tasas de cambio de subempleo y desempleo con respecto a la ocupación plena

De aquí podemos demostrar que la probabilidad de encontrarse en el subempleo ante la ocupación plena aumenta un promedio del doble y la de encontrarse desempleado ante la ocupación plena, disminuye en promedio, un 85%, las dos tasas se comportan aproximadamente de manera similar en cada periodo y en el modelo global.

Ahora sí, comenzamos a añadir variables al modelo, tenemos lo siguiente:

$$\boldsymbol{Ln\left(\frac{p_{sub}}{p_{opleno}}\right) = \alpha_1 + \beta_1\ área}$$

$$\boldsymbol{ln\left(\frac{p_{des}}{p_{opleno}}\right) = \alpha_2 + \beta_2\ área}$$

Y de esta manera se presenta el modelo con todas las variables, los coeficientes β_n obtenidos en el resultado del modelo no se pueden explicar directamente porque dichos coeficientes no tienen una relación lineal con la variable dependiente ; por tal razón para poder interpretar el modelo se lo hace en base a la razón de cambio de las probabilidades, entonces se ejecuta un nuevo modelo añadiendo la opción rrr a la sintaxis de su estimación ,de la siguiente manera:

```
mlogit eocupa área p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa, base (0)rrr
```

De la misma manera para analizar el comportamiento se realizó una estimación de las tasas de variación para cada año (anexo 2), a continuación se presenta el siguiente cuadro que es el resumen de los cuatro modelos y su interpretación.

3.4.1.3 Interpretación de las tasas de cambio por categoría:

VARIABLES		2007		2008		2009		2010	
		β	RRR:exp(β)	β	RRR:exp(β)	β	RRR:exp(β)	β	RRR:exp(β)
subempleo	area	-0,1431853	0,8665935	-0,1347637	0,8739224	-0,0853701	0,9181724	-0,0926267	0,9115337
	p02	0,4041945	1,498095	0,1356566	1,145289	0,1752463	1,19154	0,1083523	1,11444
	p03	0,0272927	1,027669	0,0234515	1,023729	0,0317729	1,032283	0,0361795	1,036842
	edad2	-0,0002919	0,9997081	-0,0002726	0,9997274	-0,0003884	0,9996117	-0,0003431	0,9996569
	escol	0,0368353	1,037522	0,024137	1,024431	0,0341135	1,034702	0,0211357	1,021361
	p45	0,0047926	1,004804	0,0086467	1,008684	0,0120851	1,012158	0,0093513	1,009395
	p50	-0,5818219	0,5588792	-0,519265	0,5949577	-0,5826188	0,558434	-0,7231516	0,4852206
	p75	0,1452793	1,156362	0,2396958	1,270863	0,1709867	1,186475	0,1819901	1,199602
	ingreso	-2,832678	0,058855	-3,205485	0,0405393	-3,841141	0,0214691	-4,199127	0,0150087
	sierra	-0,1321834	0,8761803	-0,1834843	0,832365	-0,2244365	0,7989663	-0,4791954	0,6192814
	costa	-0,0212285	0,9789952	-0,1922773	0,825078	-0,2081323	0,8120996	-0,387293	0,6788921
	_cons	14,75626		17,41108		21,12393		23,40851	
desempleo	area	1,106016	3,022294	0,8423175	2,321741	0,4490433	1,566813	1,203547	3,331914
	p02	1,461646	4,313052	0,4344506	1,544115	0,6897934	1,993304	0,7924312	2,20876
	p03	0,0144645	1,01457	-0,097834	0,9067995	-0,1675298	0,8457514	-0,1184508	0,8882955
	edad2	-0,0003006	0,9996995	0,0010235	1,001024	0,0022476	1,00225	0,0013228	1,001324
	escol	0,0762979	1,079284	0,050965	1,052286	0,1199376	1,127427	0,0534209	1,054874
	p45	-19,13449	4,90E-09	-18,76308	7,10E-09	-20,11001	1,85E-09	-21,03975	7,29E-10
	p50	-29,47025	1,59E-13	-28,35381	4,85E-13	-30,88574	3,86E-14	-30,67956	4,74E-14
	p75	0,2328318	1,262169	0,2932293	1,34075	-0,6109714	0,5428233	0,2100261	1,23371
	ingreso	-3,761576	0,0232471	-3,96018	0,0190597	-4,563463	0,0104259	-4,968578	0,006953
	sierra	0,1961891	1,216757	0,0289705	1,029394	-0,1216326	0,8854736	-0,0849004	0,9186038
	costa	0,159269	1,172653	0,3465519	1,414183	0,2157201	1,240755	0,2121002	1,236272
	_cons	17,90141		23,10998		28,27648		29,99015	

Tabla 11: Coeficientes del Modelo por Categoría

Subempleo:

Para la categoría del subempleo, como ejemplo tenemos la ecuación, para el año 2007:

$$\boldsymbol{ln\frac{Pr(Y=1)}{Pr(Y=0)} = 14,75 - 0,1431área + 0,4041\ p02 + 0,0272p03 - 0,0002919\ edad2 + 0,0368\ escol + 0,0047\ p45 - 0,5818\ p50 + 0,1452\ p75 - 2,8326\ ingreso - 0,1321\ sierra - 0,0212\ costa} \quad \textbf{(3.2)}$$

De la misma manera se podrá escribir las ecuaciones para la categoría del desempleo, para los demás años; y para el modelo global.

La interpretación es, el logaritmo de la razón de encontrarse subempleado es igual a 14,75 menos 0,14 si se encuentra en el área urbana, más 0,40 si (el individuo) es hombre, más 0,027 por cada año de edad, etc.

El logaritmo de la razón es una medida de probabilidad difícil de interpretar, para lo cual es necesario transformar la ecuación logit original en una ecuación más fácilmente interpretable, que muestre la relación entre las variables dependiente e independientes del modelo de manera más comprensible.
Para nuestro modelo esta instrucción se realiza añadiendo rrr a la instrucción mlogit, luego de la tabla 11 la interpretación de la razón de probabilidades (e^{β}) se observa por cada categoría de la variable independiente, siendo así:

Los datos que muestran la columna de los cocientes de razones (RRR, e^{β}) expresa el cambio que experimenta la mencionada razón, cuando una variable independiente varía en una unidad. Si el cociente de razones asociado a una variable es superior a 1, la razón aumenta cuando aumenta el valor de la variable, por tanto, la variable tiene un efecto positivo sobre la probabilidad de ocurrencia del suceso. Si el coeficiente mostrado es inferior a 1; la razón de ocurrencia del suceso disminuye cuando aumenta en una unidad la variable independiente en cuestión.

[37] La probabilidad de estar subempleado [38] es mayor para una persona que vive en el área rural y disminuye alrededor de un 13% para quien resida en el área urbana.

Pertenecer al género femenino implica menor probabilidad de estar subempleada y aumenta en alrededor 50% en el caso de los hombres para el año 2007, pero a partir del 2008 aumenta en un 15%.
En el modelo global ser hombre incrementa la razón de estar subempleado en 1,22 veces.
Los jóvenes tienen menor riesgo de estar subempleados y este aumenta en 3% por cada año de edad en todos los períodos analizados y en el modelo global.

Tener estudios incrementa la razón de estar subempleado 1,03 veces, es decir que por cada año de escolaridad aumenta la probabilidad de estar subempleado en alrededor de un 3%.

Con respecto a la experiencia laboral el modelo nos indica que el riesgo de estar subempleado con respecto a estar ocupado pleno es aproximadamente 1, es decir que esta variable no influye para esta categoría.

Si una persona tiene más de un trabajo es más probable que este subempleado y disminuye en un 50% aproximadamente si tiene un solo trabajo.

Para las personas que reciben el bono de desarrollo humano la probabilidad de estar subempleado es 1,28 veces superior que para quienes no lo reciben, es decir que, recibir el bono de desarrollo humano incrementa la razón de estar subempleado ante la ocupación plena alrededor del 28%.

Cuando un individuo tiene ingresos, la razón de estar subempleado disminuye 0,96 veces, lo cual significa que por cada dólar adicional en el ingreso, la razón de encontrarse subempleado, disminuye alrededor del 4%.

---

[37] La interpretación es para el modelo global en todas las categorías.
[38] La razón es de la categoría en mención respecto a la categoría base que son los ocupados plenos.

Para las personas que viven en la región sierra, el riesgo de estar subempleado, es menor en un 20% que en las otras regiones.
En la región costa para el 2007 la razón de estar subempleado disminuye en un 3%, del 2008 al 2009 en alrededor del 20% y en el 2010 el 30%.

Desempleo

La razón de estar desempleado para quienes viven en el área urbana aumenta casi al triple en los años 2007 y 2010, en el 2008 aumenta al doble y en el 2009 el 56%, que para quienes residan en el área rural

Pertenecer al género femenino implica menor probabilidad de estar desempleada y aumenta cuatro veces en el año 2007 en el caso de los hombres, en el resto de períodos aumenta en aproximadamente el doble, en el modelo global aumenta 2,21 veces.

Los jóvenes tienen mayor riesgo de estar desempleados y este disminuye en 10% por cada año de edad para todos los períodos, excepto para el 2007 en el cual se refleja un aumento mínimo de la razón de estar desempleado, en menos del 1%.
En el modelo global la razón del desempleo disminuye en un 5%, por cada año de edad.

Tener un nivel de estudios incrementa la probabilidad de estar desempleado en un 5% aproximadamente por cada año adicional de escolaridad, excepto en el 2009 en el que aumenta en un 13%, en el modelo general aumenta en un 8%.

Con respecto a la experiencia laboral el modelo nos indica que la probabilidad de estar desempleado ante estar ocupado pleno para quienes tienen menos años de trabajo es alta y disminuye en casi el 100% para quienes tienen más años de trabajo, lo cual es coherente con la realidad de que cuando una persona tiene más años de experiencia es cuando tiene un trabajo más o menos estable.

De la misma manera para la variable que indica número de trabajos es mayor la probabilidad de estar desempleado para quienes tienen más de un trabajo y disminuye en alrededor del 100% para quienes tienen un trabajo.

Para las personas que no reciben el bono de desarrollo humano la probabilidad de estar desempleado es menor y aumenta para las personas que si lo reciben en un 25% para los años 2007 y 2010, en un 34% para el 2008 y disminuye en un 45% en el 2009, en el modelo global disminuye en un 3%.
Para un individuo que tiene ingresos, el riesgo de estar desempleado disminuye en un 100%.

En la sierra, el riesgo de estar desempleado, aumenta en 21% en el 2007, un 3% mientras que en el 2009 disminuye en un 12% y 8% en el 2010; en el modelo global aumenta en un 8%.

En la región costa la probabilidad de estar desempleado aumenta alrededor del 25% excepto en el 2008 que sube en un 40%; y en el modelo global aumenta un 48%.

Coeficientes no estimados

Los coeficientes y la variación de la razón de estar desempleado ante el subempleo, o de estar subempleado ante el desempleo no se puede observar en nuestro modelo porque se ha escogido como categoría base a los ocupados plenos.
Ahora, se puede calcular estos coeficientes de los resultados anteriormente obtenidos. Por ejemplo, si queremos analizar el efecto de la variable área en la razón de estar desempleado ante estar subempleado, se tiene:

$$ln\left(\frac{p_{desempleo}}{p_{subempleo}}\right) = \left(\propto_{desempleo} - \propto_{subempleo}\right) + \left(\beta_{desempleo} - \beta_{subempleo}\right)\acute{a}rea$$

Ahora, si residir en el área urbana incrementa la razón de estar desempleado ante la ocupación plena en 2,30 veces y en la razón de estar subempleado ante la ocupación plena en 0,87 veces, entonces la razón de estar desempleado ante el

subempleo tomando el subempleo como categoría base aumenta en alrededor de un 2,30-0,87=1,43.

$$\beta_{desempleado} - \beta_{subempleado} = 0{,}8370034 - (-0{,}1422916) = 0{,}979295$$

Para obtener la razón de probabilidades debemos calcular exp (0,97929)=**2,6625**, Si dividimos $\frac{2{,}309436}{0{,}8673683}$, la razón de probabilidades de estar desempleado para la razón de estar subempleado con ocupación plena como categoría base, tenemos $\mathbf{2{,}6625}$, la razón de probabilidades de estar desempleado ante estar subempleado, el mismo valor que obtuvimos aplicando la fórmula.
De la misma manera se pueden calcular los coeficientes y las razones de probabilidad tomando como categoría base cualquiera de las categorías de la variable dependiente, mediante la siguiente instrucción:

listcoef área p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa

En el anexo 1 tenemos los resultados para el modelo global, donde se puede observar que los coeficientes de las variables para la comparación de la ocupación plena al subempleo es la misma en magnitud pero con signo opuesto al coeficiente en el que se compara el subempleo con la ocupación plena; y de la misma manera para todas las comparaciones posibles de las categorías de la variable dependiente.

3.4.2 Análisis Discriminante.

El objetivo de nuestro trabajo es investigar las diferencias entre grupos de personas con la finalidad de identificar cuáles son las características determinantes de las personas de cada grupo y poder utilizarlas para estimar el grupo al que pertenecen otras personas de los que se conocen dichas características, pero no su grupo de pertenencia, es decir, si aplicamos nuestro modelo a una persona de la cual tengamos la información de sus características sociodemográficas(variables independientes) podremos saber de acuerdo a éstas, a qué grupo de estado de ocupación pertenece.

Para resolver este problema se ha determinado utilizar el análisis discriminante ya que éste permite seleccionar cuales son las variables que mejor explican la pertenencia de una persona a un grupo de ocupación, mediante la construcción de funciones discriminantes.

Existe una gran cantidad de programas de computación para resolver éste análisis, de los cuales se escogió el SPSS; ingresamos las variables utilizadas en el modelo logístico y se obtuvo los siguientes resultados:

En primer lugar se tiene la tabla 12 (anexo 3), que contiene un resumen de los casos válidos para cada variable discriminante.[39]

La tabla 13 (anexo 3) muestra las variables independientes incluidas en el modelo en cada uno de los pasos, se puede observar que ninguna variable ha quedado fuera del modelo.[40]

Con respecto a lambda de Wilks se puede ver en la tabla 14 que a medida que se van incorporando nuevas variables al modelo en cada paso, los valores de lambda de Wilks global y del estadístico F asociado a ella van disminuyendo, lo cual significa que los grupos se encuentran cada vez más distanciados.

---

[39] La tabla contiene los valores correspondientes a dos de las variables, pero el valor es igual para las demás variables de acuerdo a cada categoría.
[40] El resultado muestra la introducción de las variables una a una en 15 pasos, de los cuales se presenta el paso 15.

Paso	Introducidas	Lambda de Wilks							
						F exacta			
		Estadístico	gl1	gl2	gl3	Estadístico	gl1	gl2	Sig.
1	50. número de trabajos	,480	1	2	147632,000	79980,130	2	147632,000	,000
2	ingreso	,315	2	2	147632,000	57698,684	4	295262,000	,000
3	escol	,302	3	2	147632,000	40344,318	6	295260,000	,000
4	45. cuàntos años trabaja	,296	4	2	147632,000	30916,765	8	295258,000	,000
5	area	,292	5	2	147632,000	25127,539	10	295256,000	,000
6	costa	,290	6	2	147632,000	21061,422	12	295254,000	,000
7	2. sexo	,289	7	2	147632,000	18115,314	14	295252,000	,000
8	75. recibió el bono de desarrollo humano	,289	8	2	147632,000	15882,741	16	295250,000	,000
9	edad2	,288	9	2	147632,000	14135,764	18	295248,000	,000
10	3. edad	,288	10	2	147632,000	12731,288	20	295246,000	,000
11	sierra	,288	11	2	147632,000	11576,634	22	295244,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

a. El número máximo de pasos es 22.

b. La F parcial mínima para entrar es 3.84.

c. La F parcial máxima para salir es 2.71

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Tabla 14: Lambda de Wilks

La tabla 15 (Anexo 3) presenta estadísticos F que permiten contrastar la hipótesis de igualdad de medias entre cada dos grupos, es decir permite averiguar que grupos difieren de qué otros.

La tabla muestra todas las comparaciones posibles entre cada dos grupos y el estadístico F y su significación para esa comparación, el proceso se realizó en 11 pasos, tanto en el primero como en los demás pasos se consigue distinguir significativamente a los tres estados de ocupación entre sí.[41]

Las tablas de valores propios y de los valores lambda de Wilks, se encuentran estrechamente relacionadas y cobran mayor significado para nuestro estudio ya

[41] Se muestra el resultado del último paso.

que cuando se trabaja con más de dos grupos se obtiene más de una función discriminante, por lo que con estas tablas es posible comparar de manera global la capacidad discriminativa de cada función.
En la tabla 16, tenemos que la primera función explica el 80,4% de la variabilidad disponible en los datos y la segunda explica el 19,6%.

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1,528	80,4	80,4	,777
2	,373	19,6	100,0	,521

a. Se han empleado las 2 primeras funciones discriminantes canónicas

Tabla 16: Valores propios

El valor propio es el cociente entre la variación debida a las diferencias entre los grupos (medida mediante la suma de cuadrados inter-grupos) y la variación que se da dentro de cada grupo combinada en una única cantidad (medida mediante la suma de cuadrados intragrupos). Este estadístico permite comparar como se distribuye la dispersión inter grupos cuando existe más de una función.
Aunque un valor propio tiene un mínimo de cero, no tiene un máximo, lo cual lo hace difícilmente interpretable por sí solo. Por ésta razón se acostumbra a utilizar el estadístico de lambda de Wilks, que se encuentra estrechamente relacionado con los valores propios.

La correlación canónica es la correlación entre la combinación lineal de las variables independientes y una combinación lineal de variables indicadores (unos y ceros) que recogen la pertenencia de los sujetos a los grupos.

La correlación canónica de la primera función es alta y el valor propio de de la primera función está más cercano a uno, entonces se puede decir que las variables discriminantes de la primera función permiten diferenciar entre los grupos mejor que las de la segunda función.

El estadístico lambda de Wilks expresa la proporción de variabilidad total no debida a las diferencias entre los grupos, permite contrastar la hipótesis nula de que las medias multivariantes de los grupos (centroides) son iguales.

De acuerdo a la tabla 17 tenemos en la primera línea un valor de lambda Wilks (que corresponde a las dos funciones escogidas conjuntamente, modelo completo), tiene asociado, con 22 grados de libertad un nivel crítico asociado menor a 0,05 por lo que se rechaza la hipótesis nula y por tanto se acepta que las medias de los grupos son diferentes, por lo que se puede concluir que el modelo permite distinguir significativamente entre los grupos.

En la segunda línea se contrasta si las medias de los grupos son iguales en la segunda función discriminante, el valor lambda de wilks es cercano a uno, pero el nivel es menor a 0,05 por lo que se puede concluir que la segunda función también permite discriminar entre los grupos.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	Gl	Sig.
1 a la 2	,288	183644,676	22	,000
2	,729	46752,335	10	,000

Tabla 17: Lambda de Wilks

3.4.2.1 Interpretación de las funciones discriminantes

En la tabla 18 se muestra la ubicación de los centroides en cada una de las funciones discriminantes.

Funciones en los centroides de los grupos

Eocupa	Función 1	Función 2
Oplenos	,686	,843
Subempleados	,070	-,456
Desempleados	-5,339	,618

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

Tabla 18: Funciones en los centroides de los grupos

La primera función distingue fundamentalmente a las personas que se encuentran en el desempleo, (cuyo centroide está ubicado en la parte negativa) de las personas que se encuentran en la ocupación plena y el subempleo (cuyos centroides se encuentran en la parte positiva).

En la segunda función, el centroide de los subempleados se sitúa en la parte negativa mientras que el de los ocupados plenos se sitúa en la parte positiva, el de las personas desempleadas queda en la parte central.
Ya que la primera función explica el máximo de las diferencias existentes entre las personas que se encuentran en el desempleo y el resto, es lógico que la segunda función discrimine precisamente entre los dos grupos que han quedado más próximos en la primera.

La matriz de coeficientes estandarizados contiene dos columnas, una para cada función discriminante. Las funciones se encuentran siempre ordenadas en correspondencia con los valores propios de la tabla 16, siendo la primera función la de mayor capacidad discriminativa.
Debido a que la primera función tiene los centroides positivos para los ocupados plenos y subempleados, se puede interpretar la clasificación para estos dos grupos. Por tanto, de acuerdo a la tabla 19, se puede observar que quienes viven en el área rural, pertenecen al sexo femenino, tienen menos años de escolaridad, viven en la región amazónica, tienden a pertenecer al grupo de los subempleados.

Mientras que las personas que tengan más edad, más años de experiencia, más de un trabajo y más ingresos, se los puede clasificar como ocupados plenos.
En la segunda función tenemos dos centroides positivos para los ocupados plenos y desempleados, pero puesto que ya se tiene una clasificación para los ocupados plenos en la primera función (en la segunda función la clasificación es igual) entonces se puede explicar que quienes tengan menor edad, menos años de escolaridad tienden a estar en el grupo de los desempleados.

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función	
	1	2
Área	-,097	,191
2. sexo	-,040	,075
3. edad	,088	,212
edad2	-,138	-,158
Escol	-,009	,285
45. cuántos años trabaja	,197	-,066
50. número de trabajos	,862	-,394
75. recibió el bono de desarrollo humano	-,004	-,095
Sierra	-,045	-,013
Costa	-,127	-,068
Ingreso	,577	,671

Tabla 19: Coeficientes estandarizados de las funciones discriminantes canónicas

La matriz de estructura, tabla 20, presenta las variables ordenadas por su grado de correlación (de mayor a menor) con la función discriminante. Este orden puede ser distinto del orden en el que aparecen en otras tablas y del orden en que han sido incluidas en el análisis.

Mientras que los coeficientes estandarizados muestran la contribución neta de cada variable independiente a la función discriminante, las correlaciones muestran la relación entre cada variable y la función discriminante.

El coeficiente más alto de cada variable aparece marcado con un asterisco que indica cuál es la función con la que más correlaciona esta variable (lo que no significa que sea ésa la función en la que más discrimina la variable).

Matriz de estructura

	Función	
	1	2
50. número de trabajos	,813*	-,444
3. edad	,112*	,012
edad2	,094*	-,045
Ingreso	,471	,812*
Escol	-,008	,497*
Área	-,010	,430*
75. recibió el bono de desarrollo humano	-,019	-,275*
45. cuántos años trabaja	,147	-,175*
2. sexo	,074	,168*
Costa	-,040	,068*
Sierra	,037	-,058*

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas

Variables ordenadas por el tamaño de la correlación con la función.

*. Mayor correlación absoluta entre cada variable y cualquier función discriminante.

Tabla 20: Matriz de Estructura

En nuestro modelo, la primera función correlaciona con: número de trabajos y edad; y la segunda función correlaciona con ingreso, escolaridad, área, si obtuvo algún título por sus estudios, recibió el bono de desarrollo humano, años de experiencia, sexo, sierra, costa, año1, año2, año3.

Ahora se va a analizar la capacidad predictiva del modelo estimado:

La tabla 21 nos presenta las probabilidades previas a priori. Estas probabilidades indican que se ha dado la misma importancia relativa a todos los grupos: 0,33 (a pesar de que los subempleados constituyen más del 50% de la muestra).
Enseguida veremos qué ocurre si utilizamos probabilidades previas basadas en los tamaños de los grupos.

Eocupa	Previas	Casos utilizados en el análisis	
		No ponderados	Ponderados
Oplenos	.333	45899	45899.000
Subempleados	.333	94598	94598.000
Desempleados	.333	7138	7138.000
Total	1.000	147635	147635.000

Tabla 21: Probabilidades previas para los grupos

El gráfico 16 (anexo 3), muestra el mapa territorial .Un mapa territorial representa el territorio (espacio) que corresponde a cada uno de los grupos en el plano definido por las dos funciones discriminantes: la primera función en el eje de las abscisas y la segunda función en el eje de las ordenadas. Los centroides de cada grupo están representados por asteriscos. Para representar los centroides se utilizan las coordenadas de las tablas de centroides (tabla 18).
Observando la ubicación de los centroides en la figura 1, se puede decir que la primera función posee mayor capacidad discriminativa que la segunda, puesto que los centroides se dispersan más en la dirección horizontal.

Las secuencias de números que aparecen dividiendo el plano en territorios son los límites o fronteras impuestos por la regla de clasificación. Los números (1,2…) identifican el grupo al que corresponde cada territorio. Conviene tener en cuenta que, ya que la regla de clasificación cambia al cambiar las probabilidades previas, si se cambian esas probabilidades también cambiarán las fronteras de los territorios (el efecto concreto es que las fronteras se alejan del centroide del grupo al que se le asigna mayor probabilidad).

Para conocer el grupo pronosticado de un caso cualquiera (es decir, el grupo en el que será clasificado), basta con representar en el mapa territorial el punto definido por sus puntuaciones discriminantes en ambas funciones. El grupo pronosticado es aquel al que corresponde el territorio en el que queda ubicado el punto.

Se puede observar a los tres grupos sobre el mapa territorial del que se podría interpretar que las personas que tengan alta puntuación en la primera función

discriminante se pueden clasificar como ocupados plenos y subempleados; mientras que quienes tengan puntuaciones próximas a cero o negativas en la misma función serán clasificados como desempleados.

Con respecto a la segunda función discriminante no se distingue entre un grupo y otro puesto que los centroides se separan aproximadamente en línea recta.

El gráfico 16 muestra el diagrama de dispersión de todos los casos utilizados en el análisis sobre el plano definido por las dos funciones discriminantes. Los casos están identificados por el estado de ocupación de las personas. La mayor utilidad de éste gráfico radica en la posibilidad de identificar casos atípicos difíciles de clasificar.

Los gráficos 17al 19, muestra cada uno de los grupos por separado y el gráfico 20, es el gráfico de grupos combinado, que muestra todos los puntos separándolos según el grupo de pertenencia.

3.4.2.2 Validación de los Resultados

Resultados de la clasificación[b,c]

			Grupo de pertenencia pronosticado			
		Eocupa	oplenos	subempleados	desempleados	Total
Original	Recuento	Oplenos	39892	3581	2426	45899
		subempleados	27381	63632	3585	94598
		desempleados	0	0	7138	7138
	%	Oplenos	86,9	7,8	5,3	100,0
		subempleados	28,9	67,3	3,8	100,0
		desempleados	,0	,0	100,0	100,0
Validación cruzada[a]	Recuento	Oplenos	39889	3584	2426	45899
		subempleados	27384	63629	3585	94598
		desempleados	0	0	7138	7138
	%	oplenos	86,9	7,8	5,3	100,0
		subempleados	28,9	67,3	3,8	100,0
		desempleados	,0	,0	100,0	100,0

a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

b. Clasificados correctamente el 75,0% de los casos agrupados originales.

c. Clasificados correctamente el 75,0% de los casos agrupados validados mediante validación cruzada.

Tabla 22: Resultados de la clasificación

Una opción para verificar si los resultados son aceptables, el análisis discriminante nos permite realizar una tabla de clasificación (tabla 22), la cual indica el porcentaje de valores predichos sobre el valor real.

La tabla nos indica que en un 75,20% de las observaciones coincide el grupo inicial con el predicho, y para los ocupados plenos se ha predicho correctamente el 86,9%, para los subempleados, el 67,30% y para los desempleados, el 100%; el grupo mejor asignado a través de las funciones discriminantes.

CAPÍTULO 4

Conclusiones y Recomendaciones

En este capítulo tenemos las conclusiones y recomendaciones luego de haber realizado nuestro estudio acerca de los factores más influyentes que determinan los estados de ocupación de la PEA.

En el modelo se analizó los estados de ocupación en los que se pueden encontrar las personas, y las variables que podrían ser determinantes para ésta situación, se ha realizado el estudio alrededor de los últimos cuatro años tomando en cuenta la información de las encuestas de empleo y desempleo (INEC) a nivel Nacional, es decir las encuestas anuales que registran una muestra más representativa de la población, de manera que podamos estimar un modelo que tenga mayor precisión.

En vista de las propiedades de las variables se decidió implementar el modelo de regresión logística multinomial de manera que se distinga las características principales de la PEA, y a continuación el análisis discriminante para determinar cómo están distribuidos cada estado de ocupación.

Se escogió un gran número de variables y luego de implementar los modelos, se decidió seleccionar las más importantes, según el modelo, y de acuerdo a la teoría desarrollada, las más significativas.

Se analizó el modelo global comparando el subempleo y el desempleo con categoría base ocupación plena porque de esta manera se puede dar a conocer las variables que influyen positiva o negativamente (y en qué porcentaje) para encontrarse ocupado pleno, el mejor estado en el cual todas las personas desearían permanecer; y la importancia del análisis es saber qué factores afectan a los individuos para no tener un trabajo ,o que características se requieren para conseguirlo, entonces, los resultados se

explican para cada categoría de la variable dependiente, en base a la categoría 0(ocupados plenos) y se concluye lo siguiente:

CARACTERÍSTICAS CON RESPECTO AL SUBEMPLEO:

- Las características que indican que la probabilidad de estar subempleado aumenta son: el sexo, la edad, la escolaridad, años de experiencia y el recibir el bono de desarrollo humano.
- Con respecto al sexo, el hecho de ser hombre aumenta la posibilidad de estar subempleado con respecto a la posibilidad de estar ocupado pleno siendo mujer, es decir que dentro de la Población Económicamente Activa la probabilidad de encontrar un hombre subempleado es mayor que la de encontrar una mujer en la ocupación plena. Esto puede suceder por varias causas, por ejemplo el hecho de que dentro de los subempleados la mayor frecuencia son jefes de hogar para quienes la razón primordial de encontrar un trabajo es el sustento de su hogar y a la falta de ofertas laborales estables aceptan un trabajo que les permite entrar en el subempleo.

  Generalmente las mujeres han participado en el desarrollo del Ecuador de distintas maneras, de las cuales las más importantes han sido el trabajo doméstico y actividades dedicadas al mercado económico, pero dicha participación es visible parcialmente, si el trabajo doméstico ayudaría a la economía nacional y si este podría ser retribuido se podría hacer un análisis de que si esto sucediera mejoraría la condición general de las mujeres.

  Otra causa importante puede ser que el subempleo se enfatiza en las ramas agricultura, industrias manufactureras, construcción en las cuales predomina la participación masculina.

  Aunque las mujeres han mejorado notablemente su situación educativa, en especial quienes forman parte de la PEA, aún sufren condiciones de trabajo peores que los hombres y reciben menores ingresos.

  Estas dificultades guardan cada vez menos relación con la educación formal de las mujeres ecuatorianas, y más con la falta de capacitación adecuada y la división que existe cuando se trata de elegir una carrera

profesional, esto se relaciona con nuestra cultura y sus condiciones que aún están presentes, las cuales tienden a identificar la fuerza de trabajo femenina menos destacada que la masculina.

- El efecto de la variable edad es mínimo, por cada año de edad la razón de estar subempleado ante la ocupación plena aumenta.
  El porcentaje de subempleados que se encuentran en las edades entre 10 y 17 años es alto, desde los 18 años en adelante disminuye y mantiene una tendencia constante hasta los 65 años y a partir de esta edad aumenta nuevamente (PEA según grupos etarios, tabla 2,pag 53).
  El subempleo es mayoritario en los jóvenes en vista de que en el rango de edad de 10 a 17 años se encuentran estudiando y si trabajan las oportunidades para hacerlo no se van a presentar de manera que puedan estar ocupados plenos, por varias razones tales como: si estudia y trabaja no va a tener disponibilidad de tiempo, si solamente trabaja las condiciones de la persona no van a satisfacer completamente los requerimientos de las ofertas laborales.
  En cuanto a las personas que tienen 50 años y más el subempleo aumenta en vista de que la edad es un limitante para conseguir un empleo en nuestro país, ya que las oportunidades de tener un empleo pleno a esa edad las personas optan por aceptar cualquier trabajo que no satisfaga sus necesidades completamente pero que por lo menos les ayude a sobrevivir.

- La escolaridad tiene un efecto mínimo, aumenta la razón de estar subempleado un pequeño porcentaje, esto se debe a un grupo de personas que con un alto nivel de educación esperan encontrar la oportunidad que les permita desenvolverse en su rama, y se encuentran en el subempleo.
  Si la escolaridad es un requerimiento primordial para estar en la ocupación plena, el problema es que no todas las personas con niveles de estudio altos están siendo consideradas ni valoradas de acuerdo a esos niveles. De igual manera también se puede explicar que hay mayores posibilidades

de encontrar en el subempleo a las personas que tengan un título que personas que no lo tengan y se encuentren en la ocupación plena.

- La variable años de experiencia no tiene mayor efecto en el subempleo y el valor mínimo que indica el aumento de la posibilidad de estar en el subempleo puede verse reflejado en que tener más años de experiencia no beneficia a las personas para conseguir un trabajo ya que la experiencia es válida si esta y dicho trabajo se relacionan.

- Con respecto al bono de desarrollo humano hay más posibilidades de encontrar subempleados que reciben el bono que quienes no lo reciben y están ocupados plenos, esta situación es porque aproximadamente del total de personas que reciben el bono, el 42% son subempleadas, el 4,4% se encuentran en la ocupación plena y 44,6% pertenecen a la población económicamente inactiva.

  Si una persona está subempleada y obtiene un dinero adicional, el bono, sin hacer ninguna actividad, es probable que complete sus ingresos para poder subsistir y se conforme con lo que recibe pero también esta situación provoca que las personas se mantengan en el estado del subempleo y no hagan mayor esfuerzo para cambiar su estado laboral.
  Este bono debería convertirse en un subsidio mientras las personas puedan conseguir un trabajo o de otra manera debería ser un incentivo por actividades que permitan el desarrollo de las personas de manera que se sientan útiles a la sociedad.

- Para que el riesgo de estar subempleado disminuya influyen las variables: área, edad al cuadrado, número de trabajos, ingreso y regiones naturales.
  Si una persona que reside en el área rural y se traslada al área urbana la razón de estar subempleado disminuye en un 13% aproximadamente.

Por lo general en el área urbana hay mayores oportunidades de desarrollo social y económico y allí es donde se encuentran las empresas, industrias y demás entidades donde se pueden encontrar ofertas laborales.

- La variable edad al cuadrado presenta efectos bajos sobre el subempleo que indican que mientras la edad aumenta el subempleo disminuye, lo cual se explica de la misma manera que en la variable edad.

- Tener un trabajo disminuye la razón de estar subempleado, es decir que hay menos posibilidades de encontrar subempleados con un trabajo que quienes tienen más de un trabajo y están en la ocupación plena; ésta situación puede justificarse en vista de que si se tiene un trabajo el subempleo será menor y mucho menor si se tiene más de un trabajo, se entiende que hay mayores ingresos y más horas de trabajo para una persona que bajo éstas circunstancias es probable que lleguen a conformar la definición de ocupación plena, de acuerdo a sus aspiraciones y deseos.

- Si aumenta el nivel de ingresos la razón de estar subempleado disminuye, es decir que las personas que tienen más ingresos tienen menor riesgo de estar en el subempleo, que en la ocupación plena, pero tener un mejor salario es una necesidad que surge en las personas quienes deberían tomar conciencia de que para ganar más dinero deben cumplir con los requisitos de la oferta laboral que permitiría que esto ocurra.

- Trasladarse a la región Sierra desde cualquier otra región disminuye la razón de estar subempleado, de acuerdo a la definición de la variable sierra y a los datos, específicamente, trasladarse de la amazonia a la sierra, disminuye la posibilidad de encontrar subempleados.

- Ahora para trasladarse a la región Costa también disminuye la razón de estar subempleado pero en este caso, comparando con los datos el subempleo es menor en la Costa que en la Sierra, lo que significa que vivir

en la Costa en comparación a la Sierra implica menor posibilidad de encontrar subempleados.

CARACTERISTICAS DEL DESEMPLEO:

Las características que indican que la probabilidad de estar desempleado aumenta son: área, sexo, escolaridad, y las regiones sierra y costa.

- En el caso del desempleo por lo general hay más oportunidades de trabajo aunque temporal en el área rural, principalmente en la agricultura, por tal razón y de acuerdo a los datos el porcentaje de desempleados es mayor en el área urbana que en el área rural y así se justificaría que trasladarse a vivir del área rural al área urbana implicaría aumento de posibilidades de encontrar personas desempleadas.

- Ser hombre aumenta la probabilidad de estar subempleado ante la posibilidad en encontrar en la ocupación plena a una mujer.
  De la misma manera que se explicó en el subempleo la participación de los hombres en la fuerza laboral es mayor que en las mujeres ya que además de subsistir y desarrollarse en el campo profesional, los hombres son jefes de hogar y buscan mantener a su familia, lo cual les encamina ante la escases de ofertas laborales a entrar en el subempleo o en el peor de los casos, en el desempleo.

- El nivel de escolaridad superior aumenta el riesgo de estar desempleado, lo cual se puede explicar que un gran porcentaje de personas que han obtenido un nivel de formación superior están desempleadas, es decir que hay personas que en realidad si demuestran un esfuerzo en el aspecto del estudio pero no son tomadas en cuenta para la participación laboral.

- Trasladarse de otras regiones a la sierra aumenta el riesgo de estar desocupado, (precisamente desde la amazonia a la sierra, de acuerdo a los datos).
- En cambio para el caso de la región costa trasladarse de la sierra o amazonia a ésta región aumenta las posibilidades de estar desempleado.

  En la Costa, tradicionalmente, se registra un incremento de desempleo, debido particularmente a dos causas: la culminación del año escolar y el invierno que soporta la Región, con inundaciones y fenómenos impredecibles; estos índices aumentan ostensiblemente en las zonas urbanas de Manabí, Los Ríos, Cañar Esmeraldas, especialmente en el sector del comercio, manufactura y construcción.

- Para que el riesgo de estar desempleado disminuya influyen las variables: edad, años de experiencia, numero de trabajos., bono de desarrollo humano, ingreso.
- La edad para el desempleo no tiene mayor efecto en el año 2007, a partir de año 2008 la razón de estar desempleado disminuye un 10% conforme aumenta la edad (tabla 3.pag.53).

  De 10 a 17 años el desempleo es alto por la condiciones de las personas (de la misma manera que en el subempleo), se encuentran estudiando o con deseos de trabajar pero pocas posibilidades por las exigencias que las ofertas laborales presentan, principalmente en este caso porque la ley en nuestro país para conseguir una buena oportunidad laboral se debe tener 18 años o más.

  A partir de las personas que tienen 18 años y más el desempleo disminuye, el modelo explica que conforme aumenta la edad la posibilidad de encontrar desempleados es menor por lo general hasta los 24 años aproximadamente, las personas terminan sus estudios y tienen más oportunidades de aplicar para un trabajo de acuerdo a las exigencias de las ofertas laborales.

- En términos estadísticos el efecto de las variables referentes al número de trabajos y experiencia laboral no son significativas. No obstante, en

conjunto, todas las regresoras tienen un impacto decisivo en el estado de la desocupación; por tal razón se puede entender que quien tenga uno o más trabajos y años de experiencia tiene una probabilidad nula de encontrarse en el desempleo y como era de esperarse un ingreso bajo implica mayor riesgo de pasar al desempleo.

- Recibir el bono de desarrollo humano aumenta las posibilidades de encontrar desempleados que quienes están ocupados plenos y no lo reciben, el bono está destinado a la gente pobre y por ende a quienes por dicha condición no pueden conseguir un empleo pleno. Estas personas pueden invertir esta pequeña cantidad en un empleo informal que les permita subsistir pero encaminados a quedarse en el desempleo.
- La recuperación en el mercado laboral se debe, principalmente, a un incremento del nivel de empleo en las ciudades de Guayaquil y de Machala. Esto se debe a que estas ciudades fueron afectadas por la crisis internacional en estas ciudades durante el 2008 y el 2009.

ANALISIS DISCRIMINANTE:

- Las tablas de clasificación en los dos modelos predicen un porcentaje similar de los desempleados, y difieren en las otras dos categorías, pero la clasificación que se tomará en cuenta es la del análisis discriminante porque este era el objetivo de nuestro estudio escoger las variables influyentes con el modelo de regresión logística y agruparlas con el análisis discriminante.

- La Población Económicamente Activa se puede observar claramente (Anexo 3) en dos grupos altamente diferenciados ocupados plenos y subempleados y el segundo grupo más destacado es de los desempleados.

  Con el análisis discriminante se determinó cuáles son las variables más determinantes del subempleo (área, sexo, escolaridad, región, años de experiencia, número de trabajos) y ocupación plena (edad, sexo, área,

años de experiencia, número de trabajos e ingresos), y las que discriminan o agrupan al desempleo (área, sexo, edad, escolaridad, región).

- En general se ha podido observar que las variables analizadas determinan la situación del mercado laboral del país por la destacada variabilidad de las probabilidades de que una persona permanezca entre uno y otro estado; en el caso de las mujeres, la diferencia del desempleo entre hombres y mujeres ha disminuido, lo cual significa que la inserción laboral en el caso de las mujeres ha aumentado y ha permitido que se valore sus capacidades para mejorar el desarrollo de nuestra sociedad.

En el centro se acumulan los ocupados plenos y subempleados y se diferencian considerablemente con los desocupados, junto a los cuales los pequeños grupos de ocupados plenos y subempleados se consideran puntos atípicos, difíciles de clasificar, la mayor parte de ocupados plenos son quienes tienen puntuaciones altas en las funciones discriminantes, un alto porcentaje de subempleados se concentran alrededor del origen es decir quienes también tengan puntuaciones altas (menores que en la ocupación plena) en las variables relacionadas con el grupo y las funciones discriminantes se clasificarán como subempleados, y quienes tienen puntuaciones inferiores se clasificarán como desempleados.

La tabla de clasificación indica que el mejor grupo que se ha clasificado es el de los desempleados (100%) por lo cual en los gráficos es el más destacado, el siguiente grupo mejor clasificado es el de los ocupados plenos y en último lugar los subempleados.

Entonces se puede concluir que las variables si agrupan específicamente a los grupos, es decir caracterizan y definen a los grupos y coinciden con los resultados del modelo logístico, por tanto a partir de este modelo se puede predecir sin conocer el estado de ocupación en cuál de los estados se encuentra.

RECOMENDACIONES:

- Los niveles de estudio permiten una estabilidad laboral, puesto que hay una gran demanda de las empresas por conocimientos que ayuden su

mejor desenvolvimiento, si una persona no estudia será más probable que se encuentre en el subempleo o en el desempleo, estados en los que se puede tener trabajos temporales y con bajo sueldo. Actualmente se ha implementado la gratuidad para el estudio entonces se necesita más voluntad y esfuerzo de las personas para mantener una mejor condición de vida, no conformes con una situación que solo les permita subsistir sino aportar al desarrollo de un país.

- El bono de desarrollo humano que actualmente reciben las personas induce a estar desempleado porque quienes lo reciben no se preocupan por buscar un empleo, nuestro gobierno debería implementar otra técnica de ayuda, brindar una compensación a quienes tienen menos oportunidades de desarrollarse pero a cambio de actividades que ayuden de alguna manera a las personas a sentirse importantes para que entiendan el gran valor de sus capacidades.

- Si bien es cierto que dichas personas aportan con conocimientos actuales, a su vez carecen de experiencia, aspecto muy significativo para valorar, por tanto sería un éxito que dentro de las condiciones para entrar al mercado laboral se presenten estas dos características juntas, pero lamentablemente no es así, sería importante para quienes ofrecen un trabajo tomar en cuenta a los jóvenes como a quienes ya tengan una experiencia adquirida y estén en capacidad de trabajar.

- Tomar en cuenta esta variable permite ver que ésta es una más de las razones para que aumente la ocupación plena, pero no debería ser así; aún cuando la probabilidad de estar subempleado disminuye cuando las personas tienen más de un trabajo, se debería tomar en cuenta algunos aspectos tales como: la estabilidad de las personas y las causas que ésta situación implicaría, movilizarse de un trabajo a otro, gastos, horarios extendidos, accesibilidad para trabajar en horarios nocturnos, etc.

REFERENCIAS:

1. A.C y Trivedi P.K. (2009). *Microeconometrics using Stata.* Stata Press.
2. Acock, A. (2006). *A Gentle Introduction to Stata* (Third Edition,Stata Press ed.).
3. Adkins L.C.y Carter R. (2008). *Using Stata for Principles of Econometrics.* Wiley.
4. Aguiar Victor. (2007). El Mercado Laboral Ecuatoriano,Propuesta de una reforma. Quito, Pichincha, Ecuador: Edición Electrónica.
5. Aguinaga C. y Criollo C. (Abril de 1998). Mercado de Trabajo. *Diseño y Metodología de Investigación* . Pichincha, Ecuador: Banco Central del Ecuador, Cuaderno de Trabajo N° 117.
6. Alarco G. y Del Hierro P. (Julio,1995). *Empleo,salarios y distribución del ingreso: márgenes de política.La problemática del empleo,las remuneraciones y la distribución del ingreso en una perspectiva macroeconómica.* Lima: Fundación Friedrich.
7. Baum C.F. (2006). *An Introduction to Modern Econometrics Using Stata.* Stata Press.
8. Becker Gary S. (1964). *Human Capital:A theorical and empirical analysis,with special reference to education.* New York.
9. Blossfeld H-P,Golsch K.,Rohwer G,. (2007). *Event History Analysis with Stata.*
10. Cleves M.,Gould W. Gutierrez R., Marchenko Y,. (2002). *An Introduction to Survival Analysis using Stata.* Third Edition ,Stata Press.
11. Estela, D. (Marzo 2003). Análisis Discriminante.
12. Gujarati Damodar N. (2004). *Econometría*. Cuarta Edición.
13. Hachuel L., Boggio G.,Wojdyla D.,. (Noviembre,2005). Interpretación y Comparación de Modelos de Regresión Logística para el Estudio de la Desocupación. *Décimas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística.*
14. Hamilton L.C. (2004). Statistics with Stata for version 8.
15. Hogg and Craig,. (1994). *Introduction to Mathematical Statistics.* Prentice Hall.
16. Instituto Nacional de Estadísticas y Censos. (Junio de 2007). Metodología Nueva de la Encuesta de Empleo, Desempleo y Subempleo. Pichincha, Ecuador.
17. Instituto Nacional de Estadísticas y Censos. (Junio de 2010). Publicación de la Encuesta de Empleo, Desempleo y Subempleo. Quito, Pichincha, Ecuador.
18. Maddala, G. (1983). Limited dependent and qualitative variables in econometrics'. Cambridge University Press.
19. Peña Daniel (2002). *Análisis de Datos Multivariante.*
20. Revista Tecnológica. (2004). Incidencias y Determinantes del Desempleo en Ecuador. *17* (1).
21. http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema6am.pdf .
22. http://www.seqc.es/es/Varios/7/40/Modulo_3:Regresion_Logística_y_multiple/.

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

ESTIMACION DEL MODELO LOGISTICO MULTINOMIAL

```
clear all
set memory 512m
set matsize 800 //en stata establece el numero de variables que se puede generar
set more off
set dp comma //en stata cambia la coma en los decimales
cap log close
```

1. Establecer el directorio de trabajo

```
cd "H:\MODELO LOGISTICO"
```

2. Crear log

```
log using "logmultinomial.log", replace
use "personas2007_2010" //Abre la base de datos
```

3. MODELO LOGISTICO PARA LA CARACTERIZACION DE LA PEA

3.1 Selección del Grupo de Personas que estan dentro del mercado laboral

```
keep if pea==1 & condact!=0 & p03<99
```

3.2 Recodificación de las variables

Variable 1: Area

```
replace area=1 if area==1
replace area=0 if area==2
```

Variable 2: p02-Sexo

```
replace p02=1 if p02==1
replace p02=0 if p02==2
```

Variable 3: Construcción de la variable Edad al cuadrado

```
gen edad2=p03*p03
```

Variable 4: escolaridad

```
gen escol=0 if (p10a==1 | p10a==3)
replace escol=0 if (p10a==2 | p10b==0)
replace escol=2 if (p10a==2 | p10b==1)
replace escol=4 if (p10a==2 | p10b==2)
replace escol=5 if (p10a==2 | p10b==3)
replace escol=p10b if p10a==4
replace escol=0 if (p10a==5 & p10b==0)
replace escol= p10b-1 if (p10a==5 & p10b>=1)
replace escol=p10b+6 if p10a==6
replace escol=p10b+9 if p10a==7
replace escol=p10b+12 if (p10a==8 | p10a==9)
replace escol=p10b+17 if p10a==10
```

Variable 5: p45-¿Cuántos años trabaja?

```
replace p45=0 if p45==.
```

Variable 6: p50-Número de trabajos

```
replace p50=1 if p50==1
replace p50=0 if p50!=1
```

Variable 7: p66-Ingreso de Asalariados y o empleados domesticas

```
replace p66=0 if p66==999999 | p66==.
gen ingras=ln(p66)
replace ingras=0 if ingras==.
```

Variable 8 :p75-¿Recibió el bono de desarrollo humano?

```
replace p75=1 if p75==1
replace p75=0 if p75!=1
```

Variable 9: Construccion de logaritmo de la variable ingresos

```
replace ingrl=0 if (ingrl==-1 | ingrl==999999 | ingrl==.)
gen ingreso=ln(ingrl)
replace ingreso=0 if ingreso==.
```

Variable 10: Regiones Naturales

```
gen sierra=1 if rn==1
replace sierra = 0 if rn!=1

gen costa=1 if rn==2
replace costa = 0 if rn!=2
```

3.3 Construccion de la Variable Dependiente

```
gen eocupa=0 if condact==1
replace eocupa=1 if condact==2 | condact==3
replace eocupa=2 if condact==5 | condact==6
label define eocupa 0 oplenos 1 subempleados 2 desempleados
label values eocupa eocupa
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

3.4 ESTIMACION DEL MODELO

3.4.1 MODELO GLOBAL.-Es el modelo total de todos los datos.

```
mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0)

Iteration 0:   log likelihood = -117353,21
Iteration 1:   log likelihood = -103971,17
Iteration 2:   log likelihood = -102402,49
Iteration 3:   log likelihood = -87839,194  (backed up)
Iteration 4:   log likelihood =  -70516,86
Iteration 5:   log likelihood = -59450,677
Iteration 6:   log likelihood =  -48593,39
Iteration 7:   log likelihood =  -47292,56
Iteration 8:   log likelihood =  -46660,56
Iteration 9:   log likelihood = -46306,919
Iteration 10:  log likelihood = -46136,702
Iteration 11:  log likelihood = -46074,785
Iteration 12:  log likelihood = -46057,809
Iteration 13:  log likelihood = -46054,534
Iteration 14:  log likelihood = -46053,902
Iteration 15:  log likelihood = -46053,753
Iteration 16:  log likelihood = -46053,723
Iteration 17:  log likelihood = -46053,716
Iteration 18:  log likelihood = -46053,715
Iteration 19:  log likelihood = -46053,715
Iteration 20:  log likelihood = -46053,715
Iteration 21:  log likelihood = -46053,715

Multinomial logistic regression                   Number of obs   =     147635
                                                  LR chi2(22)     =  142598,98
                                                  Prob > chi2     =     0,0000
Log likelihood = -46053,715                       Pseudo R2       =     0,6076

------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |  -,1847266   ,0187333    -9,86   0,000    -,2214431     -,14801
         p02 |   ,1397239   ,0195362     7,15   0,000     ,1014336    ,1780142
         p03 |   ,0322142   ,0029442    10,94   0,000     ,0264437    ,0379847
       edad2 |  -,0003138    ,000034    -9,23   0,000    -,0003805   -,0002472
       escol |   ,0715554   ,0022447    31,88   0,000     ,0671559    ,0759549
         p45 |   ,0092283   ,0009011    10,24   0,000     ,0074623    ,0109944
         p50 |  -,5213281   ,0419134   -12,44   0,000    -,6034768   -,4391794
         p75 |   ,2506245    ,038795     6,46   0,000     ,1745877    ,3266613
     ingreso |  -3,202832   ,0214106  -149,59   0,000    -3,244796   -3,160868
      sierra |  -,1825864   ,0434964    -4,20   0,000    -,2678377    -,097335
       costa |  -,1426308   ,0437401    -3,26   0,001    -,2283598   -,0569017
       _cons |   16,98851   ,1322246   128,48   0,000     16,72936    17,24767
-------------+----------------------------------------------------------------
desempleados |
        area |   ,7991486   ,1460863     5,47   0,000     ,5128247    1,085473
         p02 |   ,7450607   ,1432028     5,20   0,000     ,4643884    1,025733
         p03 |  -,0508286   ,0252215    -2,02   0,044    -,1002618   -,0013955
       edad2 |    ,000603   ,0003225     1,87   0,062    -,0000292    ,0012351
       escol |   ,1118691   ,0162539     6,88   0,000     ,0800121    ,1437262
         p45 |   -20,7104   562,6738    -0,04   0,971    -1123,531     1082,11
         p50 |   -30,8985   6073,513    -0,01   0,996    -11934,77    11872,97
         p75 |  -,0221884   ,2141974    -0,10   0,917    -,4420077    ,3976308
     ingreso |  -3,974603   ,0381493  -104,19   0,000    -4,049374   -3,899831
      sierra |   ,1162799   ,3170905     0,37   0,714    -,5052062    ,7377659
       costa |   ,4092946   ,3180115     1,29   0,198    -,2139965    1,032586
       _cons |   21,78163   ,5541924    39,30   0,000     20,69544    22,86783
------------------------------------------------------------------------------
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

3.4.2 Modelo para el AÑO 2007

```
. mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0)

Iteration 0:   log likelihood = -28531,332
Iteration 1:   log likelihood =  -26355,87
Iteration 2:   log likelihood =  -21002,11
Iteration 3:   log likelihood = -20078,633
Iteration 4:   log likelihood = -15916,971
Iteration 5:   log likelihood = -12631,543
Iteration 6:   log likelihood = -12175,683
Iteration 7:   log likelihood = -12055,113
Iteration 8:   log likelihood = -11979,806
Iteration 9:   log likelihood = -11948,537
Iteration 10:  log likelihood = -11935,851
Iteration 11:  log likelihood = -11932,611
Iteration 12:  log likelihood = -11932,027
Iteration 13:  log likelihood = -11931,901
Iteration 14:  log likelihood = -11931,871
Iteration 15:  log likelihood = -11931,865
Iteration 16:  log likelihood = -11931,864
Iteration 17:  log likelihood = -11931,864
Iteration 18:  log likelihood = -11931,864

Multinomial logistic regression                   Number of obs   =      36633
                                                  LR chi2(22)     =   33198,94
                                                  Prob > chi2     =     0,0000
Log likelihood = -11931,864                       Pseudo R2       =     0,5818
------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |  -,1910076   ,0373191    -5,12   0,000    -,2641517   -,1178634
         p02 |   ,3540721   ,0386878     9,15   0,000     ,2782454    ,4298989
         p03 |     ,03152   ,0058057     5,43   0,000      ,020141     ,042899
       edad2 |  -,0003068   ,0000672    -4,57   0,000    -,0004385   -,0001751
       escol |   ,0785316   ,0043978    17,86   0,000      ,069912    ,0871512
         p45 |    ,006103   ,0017393     3,51   0,000      ,002694    ,0095119
         p50 |  -,5981874   ,0672276    -8,90   0,000    -,7299511   -,4664237
         p75 |   ,1537297   ,0781848     1,97   0,049     ,0004904     ,306969
     ingreso |  -2,745404   ,0376479   -72,92   0,000    -2,819192   -2,671615
      sierra |  -,0880984   ,0830875    -1,06   0,289    -,2509469    ,0747501
       costa |   ,0000174   ,0837741     0,00   1,000    -,1641768    ,1642116
       _cons |    13,9428   ,2279948    61,15   0,000     13,49594    14,38966
-------------+----------------------------------------------------------------
desempleados |
        area |   1,062637   ,2816191     3,77   0,000     ,5106735      1,6146
         p02 |   1,428373   ,2832935     5,04   0,000     ,8731278    1,983618
         p03 |   ,0171111   ,0434373     0,39   0,694    -,0680246    ,1022467
       edad2 |  -,0003081   ,0005621    -0,55   0,584    -,0014098    ,0007935
       escol |   ,0995202   ,0315058     3,16   0,002     ,0377699    ,1612704
         p45 |  -18,53461   451,9714    -0,04   0,967    -904,3823     867,313
         p50 |  -28,84012   4683,471    -0,01   0,995    -9208,275    9150,595
         p75 |   ,2511604   ,3923303     0,64   0,522    -,5177928    1,020114
     ingreso |  -3,676609   ,0719127   -51,13   0,000    -3,817555   -3,535662
      sierra |   ,2266142   ,6134007     0,37   0,712     -,975629    1,428857
       costa |   ,1991237   ,6170888     0,32   0,747    -1,010348    1,408596
       _cons |   17,23027   ,9728944    17,71   0,000     15,32343    19,13711
------------------------------------------------------------------------------
```

3.4.3 Modelo para el Año 2008

```
. mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0)

Iteration 0:   log likelihood =  -29184,17
Iteration 1:   log likelihood = -25514,295
Iteration 2:   log likelihood = -24218,219
Iteration 3:   log likelihood = -22998,047
Iteration 4:   log likelihood = -18068,531
Iteration 5:   log likelihood =  -17025,93
Iteration 6:   log likelihood = -13123,663
Iteration 7:   log likelihood = -11944,036
Iteration 8:   log likelihood =  -11778,49
Iteration 9:   log likelihood = -11690,525
Iteration 10:  log likelihood =   -11633,6
Iteration 11:  log likelihood = -11610,316
Iteration 12:  log likelihood = -11603,408
Iteration 13:  log likelihood = -11601,965
Iteration 14:  log likelihood = -11601,683
Iteration 15:  log likelihood = -11601,622
Iteration 16:  log likelihood = -11601,609
Iteration 17:  log likelihood = -11601,607
Iteration 18:  log likelihood = -11601,607
Iteration 19:  log likelihood = -11601,607
Iteration 20:  log likelihood = -11601,607
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      36347
                                                  LR chi2(22)     =   35165,13
                                                  Prob > chi2     =     0,0000
Log likelihood = -11601,607                       Pseudo R2       =     0,6025

------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |  -,1826595   ,0373156    -4,89   0,000    -,2557967   -,1095223
         p02 |   ,0642167   ,0390097     1,65   0,100     -,012241    ,1406743
         p03 |   ,0292016    ,005792     5,04   0,000     ,0178494    ,0405537
       edad2 |  -,0003058   ,0000667    -4,58   0,000    -,0004365    -,000175
       escol |   ,0719339    ,004477    16,07   0,000     ,0631592    ,0807085
         p45 |   ,0105102   ,0017816     5,90   0,000     ,0070183     ,014002
         p50 |  -,5327178   ,0827502    -6,44   0,000    -,6949052   -,3705305
         p75 |   ,2244856   ,0800218     2,81   0,005     ,0676458    ,3813254
     ingreso |  -3,095242   ,0418694   -73,93   0,000    -3,177304   -3,013179
      sierra |  -,1264622   ,0846617    -1,49   0,135    -,2923961    ,0394717
       costa |  -,1575153   ,0853679    -1,85   0,065    -,3248334    ,0098027
       _cons |   16,39056   ,2565051    63,90   0,000     15,88782     16,8933
-------------+----------------------------------------------------------------
desempleados |
        area |   ,8064406   ,3024224     2,67   0,008     ,2137036    1,399178
         p02 |    ,373643   ,2973584     1,26   0,209    -,2091689    ,9564548
         p03 |  -,0961188   ,0459304    -2,09   0,036    -,1861407   -,0060969
       edad2 |   ,0010193    ,000564     1,81   0,071    -,0000862    ,0021247
       escol |   ,0815488   ,0337847     2,41   0,016      ,015332    ,1477656
         p45 |  -18,59806   432,2731    -0,04   0,966    -865,8377    828,6416
         p50 |  -27,93174   2854,552    -0,01   0,992    -5622,751    5566,887
         p75 |   ,2872568   ,4516324     0,64   0,525    -,5979265     1,17244
     ingreso |  -3,854288   ,0745394   -51,71   0,000    -4,000382   -3,708193
      sierra |   ,0592975   ,5308743     0,11   0,911    -,9811971    1,099792
       costa |   ,3683123    ,541802     0,68   0,497    -,6936002    1,430225
       _cons |   22,31398   1,049617    21,26   0,000     20,25677    24,37119
------------------------------------------------------------------------------
```

3.4.4 Modelo AÑO 2009

```
 mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0)

Iteration 0:   log likelihood = -28931,194
Iteration 1:   log likelihood = -23991,979
Iteration 2:   log likelihood = -18632,831  (backed up)
Iteration 3:   log likelihood = -17924,244
Iteration 4:   log likelihood = -15113,753
Iteration 5:   log likelihood = -11193,524
Iteration 6:   log likelihood = -10559,799
Iteration 7:   log likelihood = -10397,847
Iteration 8:   log likelihood = -10298,292
Iteration 9:   log likelihood = -10247,533
Iteration 10:  log likelihood = -10231,588
Iteration 11:  log likelihood = -10226,994
Iteration 12:  log likelihood = -10225,867
Iteration 13:  log likelihood = -10225,656
Iteration 14:  log likelihood = -10225,622
Iteration 15:  log likelihood = -10225,615
Iteration 16:  log likelihood = -10225,613
Iteration 17:  log likelihood = -10225,613
Iteration 18:  log likelihood = -10225,612
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      36463
                                                  LR chi2(22)     =   37411,16
                                                  Prob > chi2     =     0,0000
Log likelihood = -10225,612                       Pseudo R2       =     0,6466

------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |  -,1254338   ,0398227    -3,15   0,002    -,2034848   -,0473828
         p02 |   ,1160393   ,0418979     2,77   0,006     ,0339209    ,1981576
         p03 |    ,036698   ,0062309     5,89   0,000     ,0244857    ,0489102
       edad2 |   -,000417   ,0000716    -5,83   0,000    -,0005572   -,0002767
       escol |   ,0740099   ,0048375    15,30   0,000     ,0645286    ,0834912
         p45 |   ,0135151    ,001956     6,91   0,000     ,0096813    ,0173488
         p50 |  -,5830292   ,0972063    -6,00   0,000    -,7735501   -,3925083
         p75 |   ,1831169   ,0800142     2,29   0,022     ,0262919    ,3399419
     ingreso |   -3,73595   ,0504847   -74,00   0,000    -3,834898   -3,637002
      sierra |  -,1849884   ,0992039    -1,86   0,062    -,3794244    ,0094476
       costa |   -,193326    ,099583    -1,94   0,052    -,3885052    ,0018531
       _cons |   20,18033   ,3117781    64,73   0,000     19,56926    20,79141
-------------+----------------------------------------------------------------
desempleados |
        area |   ,4213677   ,3158722     1,33   0,182    -,1977305    1,040466
         p02 |   ,6711727   ,3012123     2,23   0,026     ,0808075    1,261538
         p03 |  -,1663989   ,0702485    -2,37   0,018    -,3040834   -,0287144
       edad2 |   ,0022276    ,000932     2,39   0,017     ,0004008    ,0040543
       escol |   ,1254973   ,0326485     3,84   0,000     ,0615074    ,1894873
         p45 |  -18,22476   354,8702    -0,05   0,959    -713,7576    677,3081
         p50 |  -29,15942   3347,035    -0,01   0,993    -6589,227    6530,909
         p75 |  -,5607917   ,4213453    -1,33   0,183    -1,386613      ,26503
     ingreso |  -4,463764   ,0868127   -51,42   0,000    -4,633914   -4,293614
      sierra |  -,0845834   ,8074822    -0,10   0,917    -1,667219    1,498053
       costa |   ,2705828   ,8045496     0,34   0,737    -1,306305    1,847471
       _cons |   27,62203   1,448156    19,07   0,000      24,7837    30,46037
------------------------------------------------------------------------------
```

3.4.5 Modelo AÑO 2010

```
mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0)

Iteration 0:   log likelihood = -30580,881
Iteration 1:   log likelihood = -26704,268
Iteration 2:   log likelihood = -21124,464  (backed up)
Iteration 3:   log likelihood = -19010,808
Iteration 4:   log likelihood = -14842,641
Iteration 5:   log likelihood = -11911,397
Iteration 6:   log likelihood = -11558,787
Iteration 7:   log likelihood = -11424,426
Iteration 8:   log likelihood = -11346,539
Iteration 9:   log likelihood = -11300,999
Iteration 10:  log likelihood = -11287,122
Iteration 11:  log likelihood = -11281,189
Iteration 12:  log likelihood = -11280,174
Iteration 13:  log likelihood = -11280,014
Iteration 14:  log likelihood = -11279,989
Iteration 15:  log likelihood = -11279,984
Iteration 16:  log likelihood = -11279,982
Iteration 17:  log likelihood = -11279,982
Iteration 18:  log likelihood = -11279,982
Iteration 19:  log likelihood = -11279,982
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      38192
                                                  LR chi2(22)     =   38601,80
                                                  Prob > chi2     =     0,0000
Log likelihood = -11279,982                       Pseudo R2       =     0,6311

------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |  -,1470369   ,0374436    -3,93   0,000     -,220425   -,0736488
         p02 |   ,0311613   ,0391668     0,80   0,426    -,0456042    ,1079267
         p03 |   ,0422212   ,0060714     6,95   0,000     ,0303215    ,0541209
       edad2 |  -,0003736    ,000071    -5,26   0,000    -,0005127   -,0002344
       escol |   ,0738737   ,0045241    16,33   0,000     ,0650065    ,0827408
         p45 |   ,0109075   ,0018541     5,88   0,000     ,0072736    ,0145414
         p50 |  -,7470723    ,104505    -7,15   0,000    -,9518983   -,5422463
         p75 |   ,1811257   ,0774699     2,34   0,019     ,0292875    ,3329639
     ingreso |  -4,017531   ,0524408   -76,61   0,000    -4,120313   -3,914749
      sierra |  -,4210988   ,0884149    -4,76   0,000    -,5943888   -,2478089
       costa |   -,361934   ,0886324    -4,08   0,000    -,5356503   -,1882177
       _cons |   21,94428   ,3246853    67,59   0,000      21,3079    22,58065
-------------+----------------------------------------------------------------
desempleados |
        area |   1,127405   ,3491162     3,23   0,001     ,4431498     1,81166
         p02 |   ,6844282   ,3490786     1,96   0,050     ,0002466     1,36861
         p03 |  -,1127036   ,0759491    -1,48   0,138    -,2615611    ,0361538
       edad2 |   ,0013281   ,0009642     1,38   0,168    -,0005618    ,0032179
       escol |   ,1256811   ,0405165     3,10   0,002     ,0462702     ,205092
         p45 |  -21,23387   732,7756    -0,03   0,977    -1457,448     1414,98
         p50 |   -30,9261     7464,9    -0,00   0,997    -14661,86    14600,01
         p75 |   ,2352486   ,6296897     0,37   0,709    -,9989206    1,469418
     ingreso |  -4,787516   ,0974854   -49,11   0,000    -4,978584   -4,596448
      sierra |   ,0654851   ,9702933     0,07   0,946    -1,836255    1,967225
       costa |   ,2929913   ,9765085     0,30   0,764     -1,62093    2,206913
       _cons |   28,28273   1,738731    16,27   0,000     24,87487    31,69058
------------------------------------------------------------------------------
```

3.5 Prueba de Significación Estadística de las Variables

3.5.1 MODELO GLOBAL

```
mlogtest,wald

Problem determining number of categories.

**** Wald tests for independent variables (N=147635)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+-------------------------
      o.area |    140,039    2    0,000
       o.p02 |     71,078    2    0,000
       o.p03 |    129,071    2    0,000
     o.edad2 |     92,146    2    0,000
     o.escol |   1027,594    2    0,000
       o.p45 |    104,890    2    0,000
       o.p50 |    154,710    2    0,000
       o.p75 |     43,136    2    0,000
   o.ingreso |  23008,447    2    0,000
    o.sierra |     18,365    2    0,000
     o.costa |     13,459    2    0,001
     o._cons |  16607,364    2    0,000
        area |    140,039    2    0,000
         p02 |     71,078    2    0,000
         p03 |    129,071    2    0,000
       edad2 |     92,146    2    0,000
       escol |   1027,594    2    0,000
         p45 |    104,890    2    0,000
         p50 |    154,710    2    0,000
         p75 |     43,136    2    0,000
     ingreso |  23008,447    2    0,000
      sierra |     18,365    2    0,000
       costa |     13,459    2    0,001
---------------------------------------
end of do-file
```

3.5.2 MODELO AÑO 2007

```
 mlogtest,wald

Problem determining number of categories.

**** Wald tests for independent variables (N=36633)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+------------------------
      o.area |     45,408    2    0,000
       o.p02 |     99,917    2    0,000
       o.p03 |     29,533    2    0,000
     o.edad2 |     20,846    2    0,000
     o.escol |    319,968    2    0,000
       o.p45 |     12,314    2    0,002
       o.p50 |     79,173    2    0,000
       o.p75 |      3,948    2    0,139
   o.ingreso |   5550,210    2    0,000
    o.sierra |      1,379    2    0,502
     o.costa |      0,106    2    0,949
     o._cons |   3754,369    2    0,000
        area |     45,408    2    0,000
         p02 |     99,917    2    0,000
         p03 |     29,533    2    0,000
       edad2 |     20,846    2    0,000
       escol |    319,968    2    0,000
         p45 |     12,314    2    0,002
         p50 |     79,173    2    0,000
         p75 |      3,948    2    0,139
     ingreso |   5550,210    2    0,000
      sierra |      1,379    2    0,502
       costa |      0,106    2    0,949
---------------------------------------
end of do-file
```

3.5.3 MODELO AÑO 2008

```
mlogtest,wald

Problem determining number of categories.

**** Wald tests for independent variables (N=36347)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+------------------------
      o.area |     34,135    2    0,000
       o.p02 |      3,902    2    0,142
       o.p03 |     32,040    2    0,000
     o.edad2 |     25,825    2    0,000
     o.escol |    258,644    2    0,000
       o.p45 |     34,805    2    0,000
       o.p50 |     41,444    2    0,000
       o.p75 |      7,902    2    0,019
   o.ingreso |   5622,733    2    0,000
    o.sierra |      2,326    2    0,313
     o.costa |      4,259    2    0,119
     o._cons |   4126,510    2    0,000
        area |     34,135    2    0,000
         p02 |      3,902    2    0,142
         p03 |     32,040    2    0,000
       edad2 |     25,825    2    0,000
       escol |    258,644    2    0,000
         p45 |     34,805    2    0,000
         p50 |     41,444    2    0,000
         p75 |      7,902    2    0,019
     ingreso |   5622,733    2    0,000
      sierra |      2,326    2    0,313
       costa |      4,259    2    0,119
---------------------------------------

.
end of do-file
```

3.5.4 MODELO AÑO 2009

```
mlogtest,wald

Problem determining number of categories.

**** Wald tests for independent variables (N=36463)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+------------------------
      o.area |     12,684    2    0,002
       o.p02 |     11,468    2    0,003
       o.p03 |     42,399    2    0,000
     o.edad2 |     41,322    2    0,000
     o.escol |    238,481    2    0,000
       o.p45 |     47,742    2    0,000
       o.p50 |     35,974    2    0,000
       o.p75 |      8,260    2    0,016
   o.ingreso |   5597,269    2    0,000
    o.sierra |      3,487    2    0,175
     o.costa |      4,075    2    0,130
     o._cons |   4220,222    2    0,000
        area |     12,684    2    0,002
         p02 |     11,468    2    0,003
         p03 |     42,399    2    0,000
       edad2 |     41,322    2    0,000
       escol |    238,481    2    0,000
         p45 |     47,742    2    0,000
         p50 |     35,974    2    0,000
         p75 |      8,260    2    0,016
     ingreso |   5597,269    2    0,000
      sierra |      3,487    2    0,175
       costa |      4,075    2    0,130
---------------------------------------

.
end of do-file
```

3.5.5 MODELO AÑO 2010

```
. mlogtest,wald

Problem determining number of categories.

**** Wald tests for independent variables (N=38192)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+------------------------
      o.area |     28,230    2    0,000
       o.p02 |      4,238    2    0,120
       o.p03 |     52,222    2    0,000
     o.edad2 |     30,642    2    0,000
     o.escol |    269,466    2    0,000
       o.p45 |     34,611    2    0,000
       o.p50 |     51,104    2    0,000
       o.p75 |      5,478    2    0,065
   o.ingreso |   5962,700    2    0,000
    o.sierra |     22,856    2    0,000
     o.costa |     17,032    2    0,000
     o._cons |   4583,714    2    0,000
        area |     28,230    2    0,000
         p02 |      4,238    2    0,120
         p03 |     52,222    2    0,000
       edad2 |     30,642    2    0,000
       escol |    269,466    2    0,000
         p45 |     34,611    2    0,000
         p50 |     51,104    2    0,000
         p75 |      5,478    2    0,065
     ingreso |   5962,700    2    0,000
      sierra |     22,856    2    0,000
       costa |     17,032    2    0,000
---------------------------------------

.
end of do-file
```

3.6 Comparación de las variables independientes del Modelo con todas las categorías

3.6.1 MODELO GLOBAL

listcoef area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa

mlogit (N=147635): Factor Change in the Odds of eocupa

Variable: area (sd=,49789163)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	-0,98388	-6,771	0,000	0,3739	0,6127
subemple-oplenos	-0,18473	-9,861	0,000	0,8313	0,9121
desemple-subemple	0,98388	6,771	0,000	2,6748	1,6321
desemple-oplenos	0,79915	5,470	0,000	2,2236	1,4887
oplenos -subemple	0,18473	9,861	0,000	1,2029	1,0963
oplenos -desemple	-0,79915	-5,470	0,000	0,4497	0,6717

Variable: p02 (sd=,48883659)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	-0,60534	-4,250	0,000	0,5459	0,7439
subemple-oplenos	0,13972	7,152	0,000	1,1500	1,0707
desemple-subemple	0,60534	4,250	0,000	1,8319	1,3444
desemple-oplenos	0,74506	5,203	0,000	2,1066	1,4394
oplenos -subemple	-0,13972	-7,152	0,000	0,8696	0,9340
oplenos -desemple	-0,74506	-5,203	0,000	0,4747	0,6947

Variable: p03 (sd=16,436455)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	0,08304	3,306	0,001	1,0866	3,9154
subemple-oplenos	0,03221	10,942	0,000	1,0327	1,6981
desemple-subemple	-0,08304	-3,306	0,001	0,9203	0,2554
desemple-oplenos	-0,05083	-2,015	0,044	0,9504	0,4337
oplenos -subemple	-0,03221	-10,942	0,000	0,9683	0,5889
oplenos -desemple	0,05083	2,015	0,044	1,0521	2,3058

Variable: edad2 (sd=1437,7354)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	-0,00092	-2,852	0,004	0,9991	0,2676
subemple-oplenos	-0,00031	-9,226	0,000	0,9997	0,6369
desemple-subemple	0,00092	2,852	0,004	1,0009	3,7365
desemple-oplenos	0,00060	1,870	0,062	1,0006	2,3796
oplenos -subemple	0,00031	9,226	0,000	1,0003	1,5702
oplenos -desemple	-0,00060	-1,870	0,062	0,9994	0,4202

Variable: escol (sd=4,7789349)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	-0,04031	-2,495	0,013	0,9605	0,8248
subemple-oplenos	0,07156	31,878	0,000	1,0742	1,4077
desemple-subemple	0,04031	2,495	0,013	1,0411	1,2125
desemple-oplenos	0,11187	6,883	0,000	1,1184	1,7068
oplenos -subemple	-0,07156	-31,878	0,000	0,9309	0,7104
oplenos -desemple	-0,11187	-6,883	0,000	0,8942	0,5859

Variable: p45 (sd=14,27049)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	20,71963	0,037	0,971	9,96e+08	2,6e+128
subemple-oplenos	0,00923	10,242	0,000	1,0093	1,1408
desemple-subemple	-20,71963	-0,037	0,971	0,0000	0,0000
desemple-oplenos	-20,71040	-0,037	0,971	0,0000	0,0000
oplenos -subemple	-0,00923	-10,242	0,000	0,9908	0,8766
oplenos -desemple	20,71040	0,037	0,971	9,87e+08	2,3e+128

Variable: p50 (sd=,2848765)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	30,37717	0,005	0,996	1,56e+135	731,5498
subemple-oplenos	-0,52133	-12,438	0,000	0,5937	0,8620
desemple-subemple	-30,37717	-0,005	0,996	0,0000	0,0002
desemple-oplenos	-30,89850	-0,005	0,996	0,0000	0,0002
oplenos -subemple	0,52133	12,438	0,000	1,6843	1,1601
oplenos -desemple	30,89850	0,005	0,996	2,62e+136	649,2236

Variable: p75 (sd=,30713245)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	0,27281	1,291	0,197	1,3137	1,0874
subemple-oplenos	0,25062	6,460	0,000	1,2848	1,0800
desemple-subemple	-0,27281	-1,291	0,197	0,7612	0,9196
desemple-oplenos	-0,02219	-0,104	0,917	0,9781	0,9932
oplenos -subemple	-0,25062	-6,460	0,000	0,7783	0,9259
oplenos -desemple	0,02219	0,104	0,917	1,0224	1,0068

Variable: ingreso (sd=2,2940999)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	0,77177	24,359	0,000	2,1636	5,8739
subemple-oplenos	-3,20283	-149,591	0,000	0,0406	0,0006
desemple-subemple	-0,77177	-24,359	0,000	0,4622	0,1702
desemple-oplenos	-3,97460	-104,186	0,000	0,0188	0,0001
oplenos -subemple	3,20283	149,591	0,000	24,60211	552,4905
oplenos -desemple	3,97460	104,186	0,000	53,22909	119,1832

Variable: sierra (sd=,4972273)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	-0,29887	-0,949	0,343	0,7417	0,8619
subemple-oplenos	-0,18259	-4,198	0,000	0,8331	0,9132
desemple-subemple	0,29887	0,949	0,343	1,3483	1,1602
desemple-oplenos	0,11628	0,367	0,714	1,1233	1,0595
oplenos -subemple	0,18259	4,198	0,000	1,2003	1,0950
oplenos -desemple	-0,11628	-0,367	0,714	0,8902	0,9438

Variable: costa (sd=,48991481)

Odds comparing Alternative 1 to Alternative 2	b	z	P>|z|	e^b	e^bStdX
subemple-desemple	-0,55193	-1,747	0,081	0,5758	0,7631
subemple-oplenos	-0,14263	-3,261	0,001	0,8671	0,9325
desemple-subemple	0,55193	1,747	0,081	1,7366	1,3105
desemple-oplenos	0,40929	1,287	0,198	1,5058	1,2220
oplenos -subemple	0,14263	3,261	0,001	1,1533	1,0724
oplenos -desemple	-0,40929	-1,287	0,198	0,6641	0,8183

end of do-file

ANEXO 2

1.Coeficientes que indican la razón de cambio de probabilidades

1.1 MODELO NULO,SIN REGRESORES

```
tab eocupa

      eocupa |      Freq.     Percent        Cum.
-------------+-----------------------------------
     oplenos |     45.899       31,09       31,09
subempleados |     94.598       64,08       95,17
desempleados |      7.138        4,83      100,00
-------------+-----------------------------------
       Total |    147.635      100,00


. mlogit eocupa,b(0)

Iteration 0:   log likelihood = -117353,21
Iteration 1:   log likelihood = -117353,21

Multinomial logistic regression                   Number of obs   =     147635
                                                  LR chi2(0)      =      -0,00
                                                  Prob > chi2     =          .
Log likelihood = -117353,21                       Pseudo R2       =    -0,0000

------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
       _cons |    ,723193   ,0056884   127,13   0,000     ,7120439    ,7343421
-------------+----------------------------------------------------------------
desempleados |
       _cons |  -1,861011   ,0127233  -146,27   0,000    -1,885948   -1,836074
------------------------------------------------------------------------------


 mlogit eocupa area,b(0)

Iteration 0:   log likelihood = -117353,21
Iteration 1:   log likelihood = -112490,27
Iteration 2:   log likelihood = -112451,73
Iteration 3:   log likelihood = -112451,72
Iteration 4:   log likelihood = -112451,72

Multinomial logistic regression                   Number of obs   =     147635
                                                  LR chi2(2)      =    9802,98
                                                  Prob > chi2     =     0,0000
Log likelihood = -112451,72                       Pseudo R2       =     0,0418

------------------------------------------------------------------------------
      eocupa |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |  -1,127067   ,0122522   -91,99   0,000    -1,151081   -1,103053
       _cons |   1,386506   ,0098135   141,29   0,000     1,367272     1,40574
-------------+----------------------------------------------------------------
desempleados |
        area |  -,0805128   ,0278447    -2,89   0,004    -,1350875   -,0259381
       _cons |   -1,80393   ,0233451   -77,27   0,000    -1,849685   -1,758174
------------------------------------------------------------------------------
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

1.2 MODELO GLOBAL

```
. mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0) rrr

Iteration 0:   log likelihood = -117353,21
Iteration 1:   log likelihood = -103971,17
Iteration 2:   log likelihood = -102402,49
Iteration 3:   log likelihood = -87839,194  (backed up)
Iteration 4:   log likelihood =  -70516,86
Iteration 5:   log likelihood = -59450,677
Iteration 6:   log likelihood =  -48593,39
Iteration 7:   log likelihood =  -47292,56
Iteration 8:   log likelihood =  -46660,56
Iteration 9:   log likelihood = -46306,919
Iteration 10:  log likelihood = -46136,702
Iteration 11:  log likelihood = -46074,785
Iteration 12:  log likelihood = -46057,809
Iteration 13:  log likelihood = -46054,534
Iteration 14:  log likelihood = -46053,902
Iteration 15:  log likelihood = -46053,753
Iteration 16:  log likelihood = -46053,723
Iteration 17:  log likelihood = -46053,716
Iteration 18:  log likelihood = -46053,715
Iteration 19:  log likelihood = -46053,715
Iteration 20:  log likelihood = -46053,715
Iteration 21:  log likelihood = -46053,715
Multinomial logistic regression                   Number of obs   =     147635
                                                  LR chi2(22)     =  142598,98
                                                  Prob > chi2     =     0,0000
Log likelihood = -46053,715                       Pseudo R2       =     0,6076

------------------------------------------------------------------------------
      eocupa |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |   ,8313316   ,0155735    -9,86   0,000     ,8013615    ,8624224
         p02 |   1,149956   ,0224658     7,15   0,000     1,106756    1,194842
         p03 |   1,032739   ,0030406    10,94   0,000     1,026796    1,038715
       edad2 |   ,9996862    ,000034    -9,23   0,000     ,9996196    ,9997529
       escol |   1,074178   ,0024112    31,88   0,000     1,069462    1,078914
         p45 |   1,009271   ,0009094    10,24   0,000      1,00749    1,011055
         p50 |   ,5937315   ,0248853   -12,44   0,000     ,5469068    ,6445651
         p75 |   1,284828   ,0498449     6,46   0,000     1,190755    1,386332
     ingreso |   ,0406469   ,0008703  -149,59   0,000     ,0389765    ,0423889
      sierra |   ,8331127   ,0362374    -4,20   0,000     ,7650319     ,907252
       costa |   ,8670742   ,0379259    -3,26   0,001     ,7958378    ,9446869
-------------+----------------------------------------------------------------
desempleados |
        area |   2,223647   ,3248445     5,47   0,000     1,670002    2,960839
         p02 |   2,106569   ,3016665     5,20   0,000     1,591041    2,789139
         p03 |   ,9504415   ,0239715    -2,02   0,044     ,9046005    ,9986055
       edad2 |   1,000603   ,0003227     1,87   0,062     ,9999708    1,001236
       escol |   1,118367   ,0181778     6,88   0,000       1,0833    1,154568
         p45 |   1,01e-09   5,70e-07    -0,04   0,971            0           .
         p50 |   3,81e-14   2,31e-10    -0,01   0,996            0           .
         p75 |   ,9780559   ,2094971    -0,10   0,917     ,6427447    1,488295
     ingreso |   ,0187868   ,0007167  -104,19   0,000     ,0174333    ,0202453
      sierra |    1,12331    ,356191     0,37   0,714     ,6033812    2,091258
       costa |   1,505755   ,4788475     1,29   0,198     ,8073512    2,808318
------------------------------------------------------------------------------
end of do-file
```

1.3 Modelo año 2007

```
mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0) rrr

Iteration 0:   log likelihood = -28531,332
Iteration 1:   log likelihood =  -26355,87
Iteration 2:   log likelihood =  -21002,11
Iteration 3:   log likelihood = -20078,633
Iteration 4:   log likelihood = -15916,971
Iteration 5:   log likelihood = -12631,543
Iteration 6:   log likelihood = -12175,683
Iteration 7:   log likelihood = -12055,113
Iteration 8:   log likelihood = -11979,806
Iteration 9:   log likelihood = -11948,537
Iteration 10:  log likelihood = -11935,851
Iteration 11:  log likelihood = -11932,611
Iteration 12:  log likelihood = -11932,027
Iteration 13:  log likelihood = -11931,901
Iteration 14:  log likelihood = -11931,871
Iteration 15:  log likelihood = -11931,865
Iteration 16:  log likelihood = -11931,864
Iteration 17:  log likelihood = -11931,864
Iteration 18:  log likelihood = -11931,864
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      36633
                                                  LR chi2(22)     =   33198,94
                                                  Prob > chi2     =     0,0000
Log likelihood = -11931,864                       Pseudo R2       =     0,5818

------------------------------------------------------------------------------
      eocupa |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |   ,8261263   ,0308303    -5,12   0,000     ,7678571    ,8888174
         p02 |   1,424858   ,0551247     9,15   0,000      1,32081    1,537102
         p03 |   1,032022   ,0059916     5,43   0,000     1,020345    1,043833
       edad2 |   ,9996932   ,0000672    -4,57   0,000     ,9995616    ,9998249
       escol |   1,081698   ,0047571    17,86   0,000     1,072414    1,091062
         p45 |   1,006122   ,0017499     3,51   0,000     1,002698    1,009557
         p50 |   ,5498073   ,0369622    -8,90   0,000     ,4819325    ,6272415
         p75 |   1,166176   ,0911772     1,97   0,049      1,00049    1,359299
     ingreso |   ,0642224   ,0024178   -72,92   0,000     ,0596541    ,0691405
      sierra |   ,9156708   ,0760808    -1,06   0,289     ,7780637    1,077615
       costa |   1,000017   ,0837756     0,00   1,000      ,848592    1,178464
-------------+----------------------------------------------------------------
desempleados |
        area |   2,893992   ,8150034     3,77   0,000     1,666413    5,025878
         p02 |   4,171906   1,181874     5,04   0,000     2,394388    7,268995
         p03 |   1,017258    ,044187     0,39   0,694     ,9342375    1,107657
       edad2 |   ,9996919   ,0005619    -0,55   0,584     ,9985912    1,000794
       escol |   1,104641   ,0348026     3,16   0,002     1,038492    1,175003
         p45 |   8,92e-09   4,03e-06    -0,04   0,967            0           .
         p50 |   2,98e-13   1,40e-09    -0,01   0,995            0           .
         p75 |   1,285516   ,5043469     0,64   0,522     ,5958342     2,77351
     ingreso |   ,0253087     ,00182   -51,13   0,000     ,0219815    ,0291394
      sierra |   1,254346   ,7694166     0,37   0,712     ,3769552    4,173928
       costa |   1,220333   ,7530538     0,32   0,747     ,3640922    4,090207
------------------------------------------------------------------------------

.
end of do-file
```

1.4 Modelo año 2008

```
 mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0) rrr

Iteration 0:   log likelihood =  -29184,17
Iteration 1:   log likelihood = -25514,295
Iteration 2:   log likelihood = -24218,219
Iteration 3:   log likelihood = -22998,047
Iteration 4:   log likelihood = -18068,531
Iteration 5:   log likelihood =  -17025,93
Iteration 6:   log likelihood = -13123,663
Iteration 7:   log likelihood = -11944,036
Iteration 8:   log likelihood =  -11778,49
Iteration 9:   log likelihood = -11690,525
Iteration 10:  log likelihood =   -11633,6
Iteration 11:  log likelihood = -11610,316
Iteration 12:  log likelihood = -11603,408
Iteration 13:  log likelihood = -11601,965
Iteration 14:  log likelihood = -11601,683
Iteration 15:  log likelihood = -11601,622
Iteration 16:  log likelihood = -11601,609
Iteration 17:  log likelihood = -11601,607
Iteration 18:  log likelihood = -11601,607
Iteration 19:  log likelihood = -11601,607
Iteration 20:  log likelihood = -11601,607
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      36347
                                                  LR chi2(22)     =   35165,13
                                                  Prob > chi2     =     0,0000
Log likelihood = -11601,607                       Pseudo R2       =     0,6025

------------------------------------------------------------------------------
      eocupa |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |   ,8330517   ,0310858    -4,89   0,000     ,7742993    ,8962621
         p02 |   1,066323    ,041597     1,65   0,100     ,9878337     1,15105
         p03 |   1,029632   ,0059637     5,04   0,000      1,01801    1,041387
       edad2 |   ,9996943   ,0000667    -4,58   0,000     ,9995636     ,999825
       escol |   1,074584   ,0048109    16,07   0,000     1,065196    1,084055
         p45 |   1,010566   ,0018004     5,90   0,000     1,007043      1,0141
         p50 |   ,5870074    ,048575    -6,44   0,000     ,4991218     ,690368
         p75 |   1,251679   ,1001615     2,81   0,005     1,069986    1,464224
     ingreso |   ,0452641   ,0018952   -73,93   0,000     ,0416979    ,0491352
      sierra |   ,8812075   ,0746045    -1,49   0,135     ,7464728    1,040261
       costa |   ,8542637   ,0729267    -1,85   0,065     ,7226477    1,009851
-------------+----------------------------------------------------------------
desempleados |
        area |   2,239921   ,6774024     2,67   0,008     1,238256    4,051867
         p02 |   1,453018   ,4320673     1,26   0,209     ,8112582    2,602454
         p03 |   ,9083561   ,0417211    -2,09   0,036     ,8301568    ,9939216
       edad2 |    1,00102   ,0005646     1,81   0,071     ,9999138    1,002127
       escol |   1,084966   ,0366553     2,41   0,016      1,01545    1,159241
         p45 |   8,37e-09   3,62e-06    -0,04   0,966            0           .
         p50 |   7,40e-13   2,11e-09    -0,01   0,992            0           .
         p75 |   1,332766   ,6019205     0,64   0,525     ,5499508    3,229864
     ingreso |   ,0211887   ,0015794   -51,71   0,000     ,0183086    ,0245218
      sierra |   1,061091   ,5633059     0,11   0,911     ,3748621    3,003541
       costa |   1,445293   ,7830628     0,68   0,497     ,4997736    4,179638
------------------------------------------------------------------------------

.
end of do-file
```

1.5 Modelo año 2009

```
mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0) rrr

Iteration 0:   log likelihood = -28931,194
Iteration 1:   log likelihood = -23991,979
Iteration 2:   log likelihood = -18632,831  (backed up)
Iteration 3:   log likelihood = -17924,244
Iteration 4:   log likelihood = -15113,753
Iteration 5:   log likelihood = -11193,524
Iteration 6:   log likelihood = -10559,799
Iteration 7:   log likelihood = -10397,847
Iteration 8:   log likelihood = -10298,292
Iteration 9:   log likelihood = -10247,533
Iteration 10:  log likelihood = -10231,588
Iteration 11:  log likelihood = -10226,994
Iteration 12:  log likelihood = -10225,867
Iteration 13:  log likelihood = -10225,656
Iteration 14:  log likelihood = -10225,622
Iteration 15:  log likelihood = -10225,615
Iteration 16:  log likelihood = -10225,613
Iteration 17:  log likelihood = -10225,613
Iteration 18:  log likelihood = -10225,612
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      36463
                                                  LR chi2(22)     =   37411,16
                                                  Prob > chi2     =     0,0000
Log likelihood = -10225,612                       Pseudo R2       =     0,6466

------------------------------------------------------------------------------
      eocupa |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |   ,8821142   ,0351281    -3,15   0,002     ,8158826    ,9537223
         p02 |    1,12304    ,047053     2,77   0,006     1,034503    1,219155
         p03 |    1,03738   ,0064638     5,89   0,000     1,024788    1,050126
       edad2 |   ,9995831   ,0000715    -5,83   0,000     ,9994429    ,9997233
       escol |   1,076817   ,0052091    15,30   0,000     1,066656    1,087076
         p45 |   1,013607   ,0019827     6,91   0,000     1,009728      1,0175
         p50 |   ,5582049   ,0542611    -6,00   0,000     ,4613722    ,6753608
         p75 |   1,200955   ,0960935     2,29   0,022     1,026641    1,404866
     ingreso |   ,0238505   ,0012041   -74,00   0,000     ,0216035    ,0263312
      sierra |    ,831114   ,0824497    -1,86   0,062     ,6842552    1,009492
       costa |   ,8242132   ,0820776    -1,94   0,052     ,6780697    1,001855
-------------+----------------------------------------------------------------
desempleados |
        area |   1,524045   ,4814034     1,33   0,182      ,820591    2,830535
         p02 |    1,95653   ,5893309     2,23   0,026     1,084162    3,530847
         p03 |   ,8467084     ,05948    -2,37   0,018     ,7377994     ,971694
       edad2 |    1,00223   ,0009341     2,39   0,017     1,000401    1,004063
       escol |   1,133712    ,037014     3,84   0,000     1,063438     1,20863
         p45 |   1,22e-08   4,32e-06    -0,05   0,959            0    1,4e+294
         p50 |   2,17e-13   7,26e-10    -0,01   0,993            0           .
         p75 |    ,570757   ,2404858    -1,33   0,183     ,2499203     1,30347
     ingreso |   ,0115189       ,001   -51,42   0,000     ,0097167    ,0136555
      sierra |    ,918895   ,7419913    -0,10   0,917     ,1887712     4,47297
       costa |   1,310728   1,054546     0,34   0,737     ,2708188    6,343755
------------------------------------------------------------------------------

end of do-file
```

1.6 Modelo año 2010

```
mlogit eocupa area p02 p03 edad2 escol p45 p50 p75 ingreso sierra costa,base(0) rrr

Iteration 0:   log likelihood = -30580,881
Iteration 1:   log likelihood = -26704,268
Iteration 2:   log likelihood = -21124,464  (backed up)
Iteration 3:   log likelihood = -19010,808
Iteration 4:   log likelihood = -14842,641
Iteration 5:   log likelihood = -11911,397
Iteration 6:   log likelihood = -11558,787
Iteration 7:   log likelihood = -11424,426
Iteration 8:   log likelihood = -11346,539
Iteration 9:   log likelihood = -11300,999
Iteration 10:  log likelihood = -11287,122
Iteration 11:  log likelihood = -11281,189
Iteration 12:  log likelihood = -11280,174
Iteration 13:  log likelihood = -11280,014
Iteration 14:  log likelihood = -11279,989
Iteration 15:  log likelihood = -11279,984
Iteration 16:  log likelihood = -11279,982
Iteration 17:  log likelihood = -11279,982
Iteration 18:  log likelihood = -11279,982
Iteration 19:  log likelihood = -11279,982
```

ANEXOS 1 Y 2 Estimación del Modelo Logístico Multinomial

```
Multinomial logistic regression                   Number of obs   =      38192
                                                  LR chi2(22)     =   38601,80
                                                  Prob > chi2     =     0,0000
Log likelihood = -11279,982                       Pseudo R2       =     0,6311

------------------------------------------------------------------------------
      eocupa |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
oplenos      |  (base outcome)
-------------+----------------------------------------------------------------
subempleados |
        area |   ,8632621   ,0323236    -3,93   0,000     ,8021778    ,9289979
         p02 |   1,031652   ,0404065     0,80   0,426       ,95542    1,113966
         p03 |   1,043125   ,0063332     6,95   0,000     1,030786    1,055612
       edad2 |   ,9996265    ,000071    -5,26   0,000     ,9994874    ,9997656
       escol |   1,076671    ,004871    16,33   0,000     1,067166     1,08626
         p45 |   1,010967   ,0018744     5,88   0,000       1,0073    1,014648
         p50 |   ,4737515   ,0495094    -7,15   0,000     ,3860076    ,5814407
         p75 |   1,198566   ,0928528     2,34   0,019     1,029721    1,395097
     ingreso |   ,0179973   ,0009438   -76,61   0,000     ,0162394    ,0199456
      sierra |   ,6563252   ,0580289    -4,76   0,000     ,5518998    ,7805091
       costa |   ,6963283   ,0617172    -4,08   0,000     ,5852886    ,8284343
-------------+----------------------------------------------------------------
desempleados |
        area |   3,087634   1,077943     3,23   0,001     1,557606    6,120601
         p02 |   1,982638   ,6920965     1,96   0,050     1,000247    3,929883
         p03 |   ,8934154   ,0678541    -1,48   0,138     ,7698488    1,036815
       edad2 |   1,001329   ,0009655     1,38   0,168     ,9994384    1,003223
       escol |   1,133921   ,0459425     3,10   0,002     1,047357    1,227638
         p45 |   6,00e-10   4,40e-07    -0,03   0,977            0           .
         p50 |   3,71e-14   2,77e-10    -0,00   0,997            0           .
         p75 |   1,265223   ,7966981     0,37   0,709     ,3682767    4,346703
     ingreso |   ,0083331   ,0008124   -49,11   0,000     ,0068838    ,0100876
      sierra |   1,067677    1,03596     0,07   0,946     ,1594133    7,150807
       costa |   1,340431   1,308942     0,30   0,764     ,1977147    9,087617
------------------------------------------------------------------------------

.
end of do-file
```

2. Predicción de las categorias de la variable independiente

```
predict p0-p2 if e(sample), p

gen peocupa=.
replace peocupa = 0 if p0 == max(p0,p1,p2) & e(sample)
replace peocupa = 1 if p1 == max(p0,p1,p2) & e(sample)
replace peocupa = 2 if p2 == max(p0,p1,p2) & e(sample)
```

3. Tabla de Clasificación

```
tab peocupa eocupa
```

ANEXO 3:

Eocupa		N válido (según lista) No ponderados	Ponderados
Oplenos	área	45899	45899.000
	2. sexo	45899	45899.000
subempleados	área	94598	94598.000
	2. sexo	94598	94598.000
desempleados	área	7138	7138.000
	2. sexo	7138	7138.000
Total	área	147635	147635.000
	2. sexo	147635	147635.000

Tabla 12: Tamaños Muestrales de cada grupo

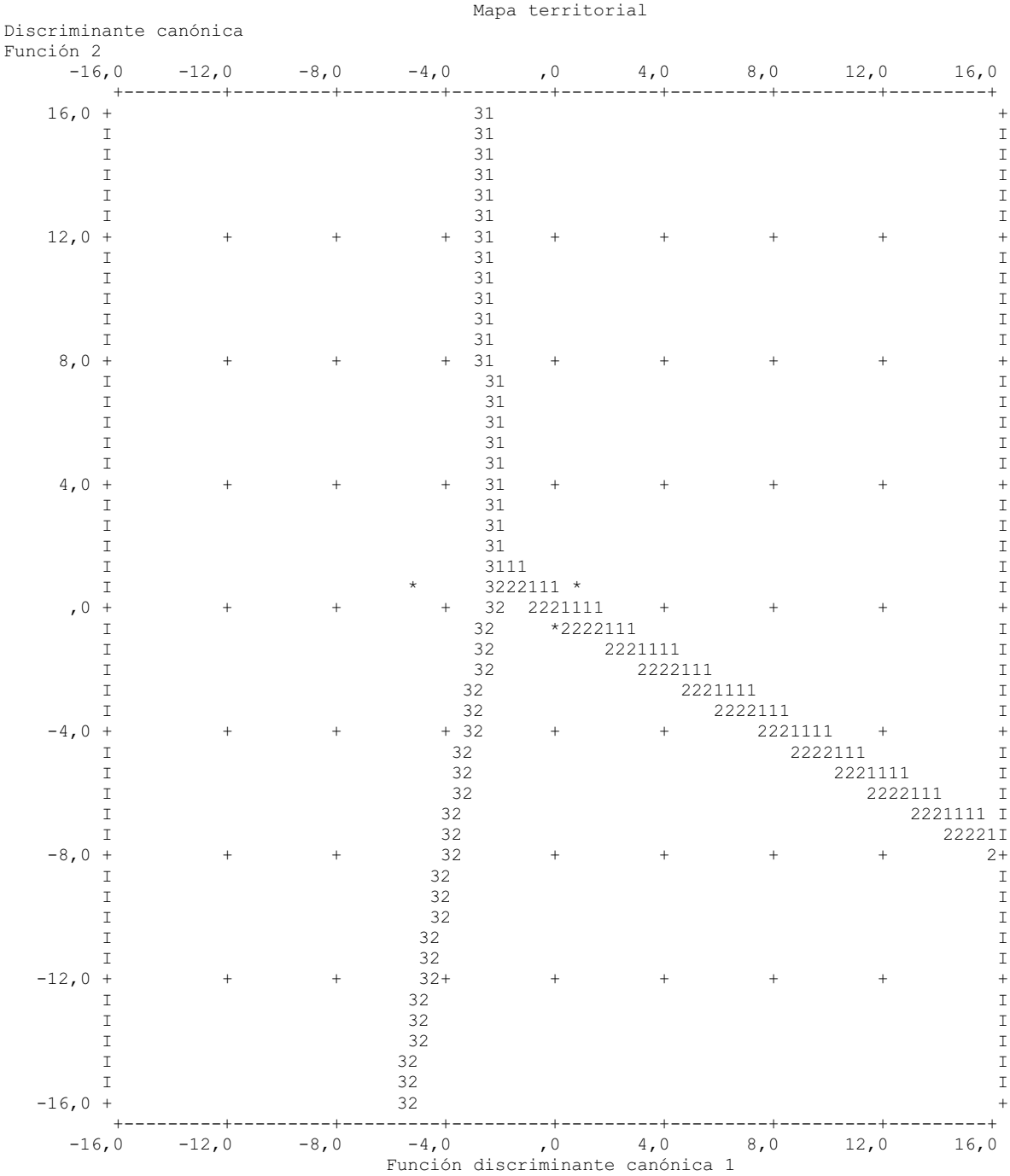
Paso		Tolerancia	F para salir	Lambda de Wilks
11	50. número de trabajos	,989	69819,357	,561
	ingreso	,834	27240,773	,395
	escol	,727	1209,359	,293
	45. cuàntos años trabaja	,509	933,698	,292
	area	,782	914,823	,292
	costa	,175	142,067	,289
	2. sexo	,833	154,841	,289
	75. recibió el bono de desarrollo humano	,831	152,525	,289
	edad2	,040	53,141	,288
	3. edad	,040	49,298	,288
	sierra	,176	16,625	,288

Tabla 13: Variables en el análisis

Comparaciones de grupos por paresa,b,c,d,e,f,g,h,i,j,k

Paso	eocupa		oplenos	subempleados	desempleados
11	oplenos	F		5806,689	20412,677
		Sig.		,000	,000
	subempleados	F	5806,689		18345,826
		Sig.	,000		,000
	desempleados	F	20412,677	18345,826	
		Sig.	,000	,000	

Tabla 15: Comparaciones de Grupos por Pares



Símbolos usados en el mapa territorial

Símbol	Grupo	Etiqu
1	0	oplenos
2	1	subempleados
3	2	desempleados
*		Indica un centroide de grupo

Gráfico 16: Mapa Territorial

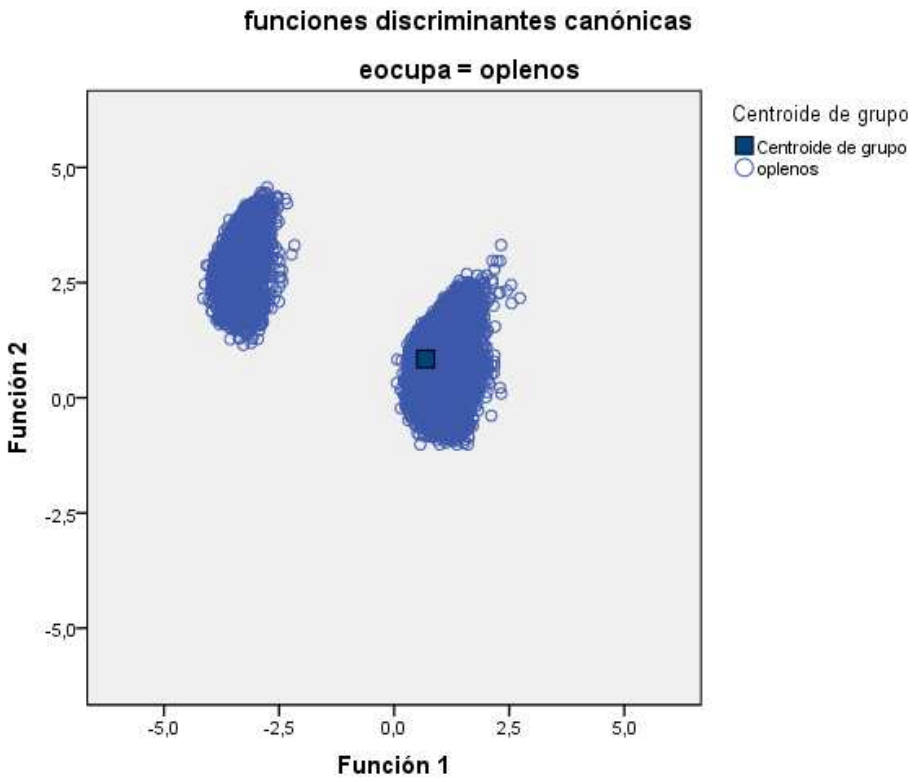


Gráfico 17: Ocupado Plenos

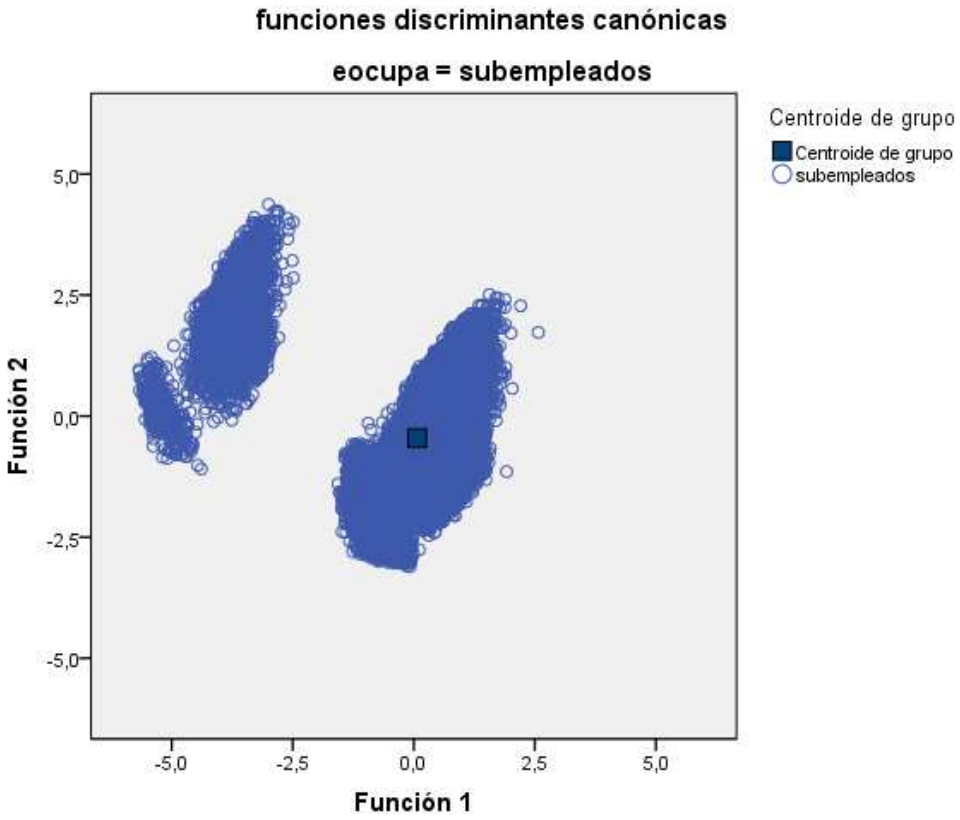


Gráfico 18: Subempleados

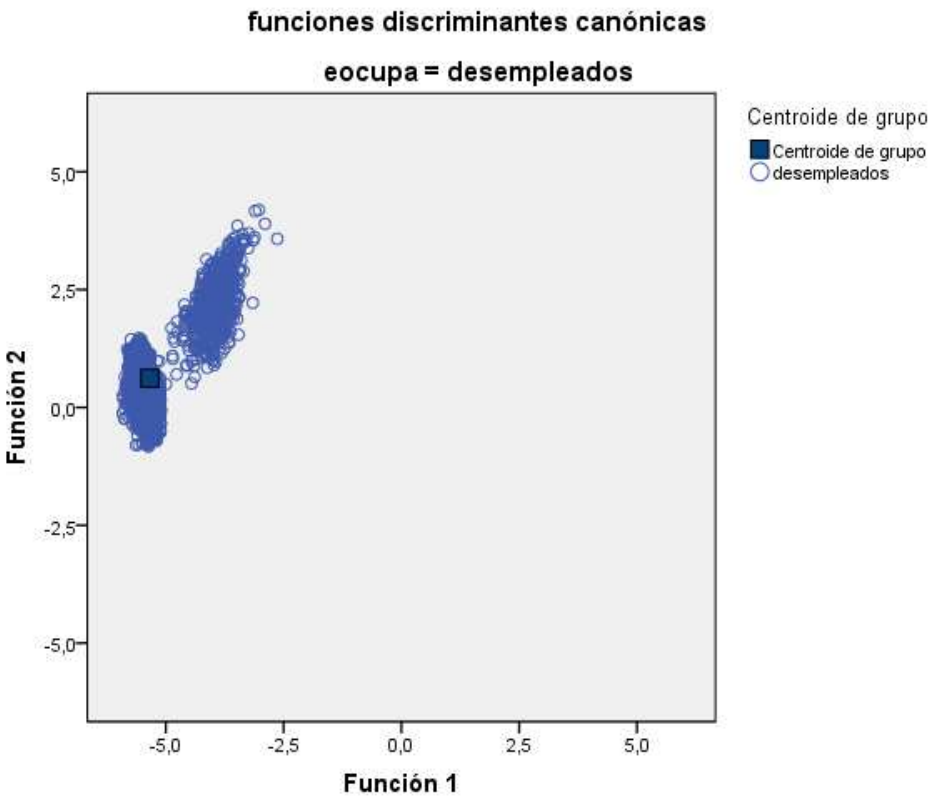


Gráfico 19: Desempleados

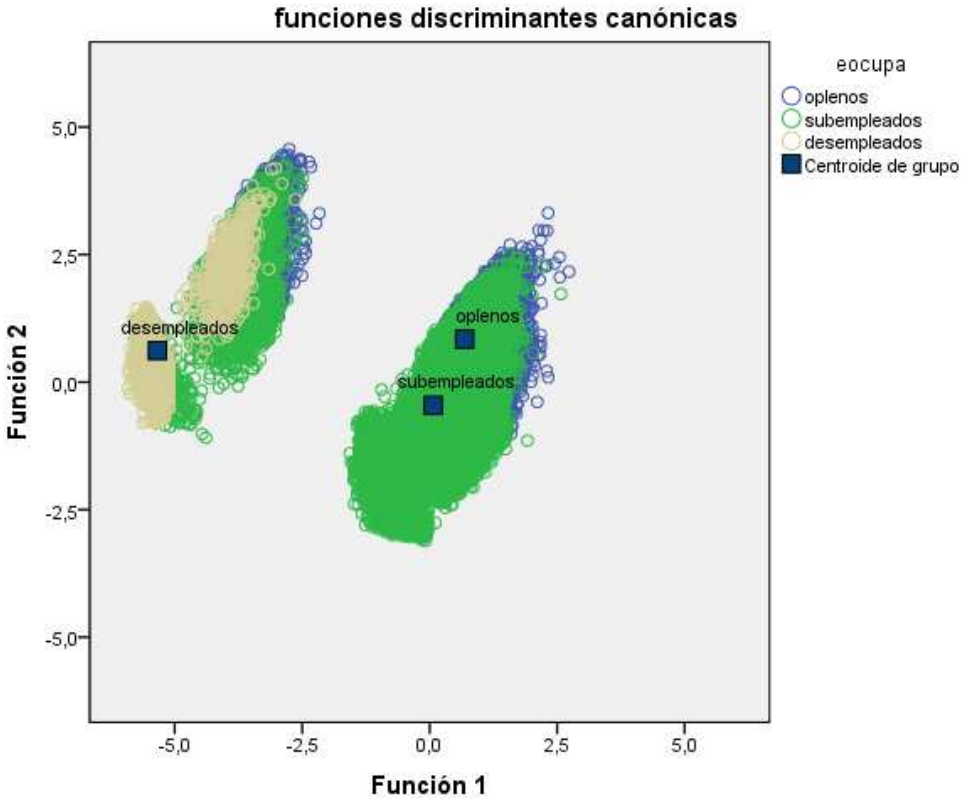


Gráfico 20: Estados de Ocupación

Anexo 4:

METODOLOGÍA DE LA ENCUESTA DE EMPLEO Y DESEMPLEO DEL INEC

El Instituto Nacional de Estadística y Censos (INEC), organismo rector de la producción estadística en el Ecuador y la Unidad Estadísticas de Hogares, al poner a disposición de los usuarios la presente publicación, cumple con el propósito fundamental de satisfacer oportunamente las necesidades de información sobre el Mercado Laboral del Ecuador.
La publicación contiene información sobre las principales variables del mercado laboral: empleo, desempleo y subempleo, la clasificación de la mano de obra ocupada según segmentos de mercado formal, informal y servicio doméstico, a más de una cuarta categoría de no clasificados y datos de la población económicamente activa e inactiva. Tiene un corte de edad para la Población Económicamente Activa (PEA) de 10 años y más, no obstante, la investigación se realiza a toda la población de cinco años de edad en adelante.

La Vigésima Octava Ronda del SIEH-ENEMDU tiene cobertura para el área urbana y rural del territorio continental del Ecuador, por lo que las cifras se presentan en el ámbito nacional, por áreas, por regiones naturales, principales ciudades (Quito, Guayaquil, Cuenca, Machala y Ambato) y a nivel provincial en sierra, costa y como un solo dominio la amazonia. Adicionalmente, la publicación contiene un resumen analítico de variables propias del mercado laboral y sobre la movilidad de la mano de obra, (Matrices de Transición Laboral), proponiendo temas que seguramente merecen profundizarse ya que la intención del documento no es agotarlos sino más bien provocar el interés colectivo sobre los mismos.

Es importante indicar que desde la Décima Sexta Ronda de la ENEMDU (junio-2007), se incorporaron cambios metodológicos tanto en el diseño del formulario de campo como en los algoritmos de cálculo de algunas variables y resultados, sustentados en una necesaria actualización conceptual que permitió la identificación de la investigación con las recomendaciones de la XVI Conferencia
de Estadísticos del Trabajo CIET, así como armonizar con las recomendaciones y acuerdos de los expertos gubernamentales de la Comunidad Andina de Naciones CAN, con la finalidad de lograr la mejor y más precisa comparabilidad regional e internacional.

1. ASPECTOS METODOLÓGICOS

Objetivos

- Visualizar el perfil social, demográfico y económico de la población del país, a través de variables de carácter general como: sexo, edad, nivel de instrucción, condición de actividad, etc.

- Proporcionar información actualizada y periódica sobre la población económicamente activa PEA, con sus principales características y difundir datos e indicadores fidedignos sobre el Mercado Laboral Ecuatoriano, relacionados con el empleo, desempleo y subempleo.

Ámbito de Estudio

El levantamiento de la información se realiza básicamente de acuerdo a la muestra que básicamente tiene tres diseños:

1. Nacional Urbano y Rural que se desarrolla anualmente en diciembre; además de las estimaciones a nivel Nacional Total, Nacional Urbano y Nacional Rural, permite dar estimaciones a nivel provincial en las regiones Costa y Sierra, en las áreas urbano y rural.
2. Nacional Urbano, que se desarrolla trimestralmente. Los dominios de estimación aquí son: 5 ciudades (Quito, Guayaquil, Cuenca, Machala, Ambato), Resto Costa Urbano, Resto Sierra Urbano, Resto Amazonía Urbana; esta encuesta permite dar estimaciones a nivel nacional urbano y en cada domino.
3. 5 ciudades: Quito, Guayaquil, Cuenca, Machala, Ambato, que se desarrolla mensualmente.

La encuesta está dirigida a los hogares y las personas que en ellos residen, establecidos en las viviendas particulares.

1.2 DEFINICIONES

Vivienda

Vivienda es el espacio delimitado por paredes y techo, de cualquier material de construcción, con **entrada independiente**, destinado para ser habitado por una o más personas; la misma que aún cuando no haya sido construida originalmente para tales fines, esté destinada a ser utilizada como vivienda.

Hogar

Es la unidad social conformada por una persona o grupo de personas que se asocian para compartir el alojamiento y la alimentación. Es decir, hogar es el conjunto de personas que residen habitualmente en la misma vivienda o en parte de ella (viven bajo el mismo techo), que están unidas o no por lazos de parentesco, y que cocinan en común para todos sus miembros (comen de la misma olla).

Población Total

Comprende a la población en general; es decir, a los mayores y menores de 10 años.

Población en Edad de Trabajar, PET

Comprende a todas las personas de 10 años y más.

Población Económicamente Inactiva

Son todas aquellas personas de 10 años y más que no están ocupadas, tampoco buscan trabajo y no estaban disponibles para trabajar, es decir no pertenecen a la PEA.

2. DISEÑO MUESTRAL

En esta investigación la muestra está formada por todas las viviendas particulares dentro del territorio ecuatoriano, incluido las zonas “no delimitadas”. Se excluyen: la Región- Insular, la población en viviendas colectivas, viviendas flotantes y población indigente.

2.1. MARCO MUESTRAL

Se basa en los resultados definitivos y la cartografía del VI Censo de Población y V de Vivienda del 2001 (CPV-2001) el mismo que sirvió de base para la construcción de la Muestra Maestra que ha implementado el INEC, para la Encuesta Nacional de Empleo y Desempleo Urbano (ENEMDU) se aplica una submuestra de la Muestra Maestra.

La **Unidad Primaria de Muestreo (UPM)** para la muestra maestra se define como un conjunto de sectores censales donde se tiene aproximadamente 450 viviendas en el área urbana y 320 viviendas en el área rural , mismas que fueron seleccionadas con probabilidad proporcional a su tamaño, el número de UPM´s en la muestra maestra es de 2.094; La unidad secundaria de muestreo son los sectores censales, se seleccionó un sector por cada UPM con

probabilidad proporcional al tamaño (PPT) , el número de sectores censales es 7.409, la unidad última de muestreo son las viviendas, en cada ronda de la ENEMDU se seleccionan 12 viviendas ocupadas por sector.

El INEC, utiliza el Marco Maestro y el esquema de rotación de viviendas establecido para las encuestas trimestrales de Empleo, Desempleo y Subempleo, el cual se implantó a partir de septiembre del 2003.

2.2. ESTRATIFICACIÓN

La estratificación de la muestra maestra consiste en formar grupos de UPM's, homogéneos al interior y heterogéneos entre sí, la misma que es muy importante para aumentar la eficiencia de la muestra. El primer nivel de estratificación generalmente corresponde a los dominios geográficos de análisis. Para la ENEMDU se tienen 8 dominios geográficos a nivel nacional urbano: Quito, Guayaquil, Cuenca, Machala, Ambato, Resto Sierra Urbano, Resto Costa Urbano y la región Amazónica Urbana, y 11 dominios geográficos a nivel nacional total, a más de los 8 dominios urbanos se añaden los siguientes 3 dominios rurales: Costa Rural, Sierra Rural y Amazonía Rural.

Además de los estratos geográficos es necesario estratificar el marco con criterio socioeconómico para lo cual se ajustó un modelo lineal con los datos de la ECV 1999 para predecir el consumo e imputar promedios a nivel de sector censal. Las variables del modelo son las siguientes:

- Nivel de escolaridad en la población de 24 años y más (el promedio de años de instrucción formal recibidos)
- Porcentaje de población de 24 años y más con instrucción superior
- Porcentaje de hogares que disponen de servicio higiénico de uso exclusivo
- Porcentaje de viviendas con agua potable dentro de la vivienda
- Porcentaje de hogares con vivienda propia

Muestra Nacional Urbano – Rural

TABLA 1: Muestras anuales, 32 dominios provinciales urbano y rural por número de sectores.

PROVINCIAS	dic-03	dic-04	dic-05	dic-06	dic-07	dic-08	dic-09
AZUAY	82	82	82	82	96	96	96
BOLIVAR	80	80	80	80	80	80	80
CAÑAR	80	80	80	80	80	80	80
CARCHI	80	80	80	80	80	80	80
COTOPAXI	80	80	80	80	80	80	80
CHIMBORAZO	80	80	80	80	80	80	80
EL ORO	123	123	123	123	137	137	137
ESMERALDAS	120	120	120	120	120	120	120
GUAYAS	192	192	192	192	193	193	177
IMBABURA	80	80	80	80	80	80	80
LOJA	80	80	80	80	80	80	80
LOS RIOS	120	120	120	120	120	120	120
MANABI	120	120	120	120	120	120	120
PICHINCHA	150	150	150	150	150	150	139
TUNGURAHUA	80	80	80	80	103	103	103
AMAZONÍA	80	80	80	80	80	80	80
SANTO DOMINGO*[1]							80
SANTA ELENA*							80
ZONAS NO DELIMITADAS	6	6	6	6	6	6	2
TOTAL	**1633**	**1633**	**1633**	**1633**	**1685**	**1685**	**1814**

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya

Un sector censal es una agrupación de viviendas definido por el INEC en la cartografía del censo 2001; tiene aproximadamente 150 viviendas en el área urbana y 80 en área rural. Cada sector de la muestra contiene cuatro paneles de 3 viviendas cada uno, es decir 12 viviendas por sector; además, para mejorar la cobertura se asigna una vivienda de reemplazo por panel, es decir 4 reemplazos por sector. Este diseño de panel de muestra permite encajar las muestras: las más pequeñas están incluidas en las más grandes, es decir la muestra mensual está incluida en la trimestral y ésta en la anual.En septiembre 2007 se incrementaron 52 sectores en las encuestas mensual y trimestral, 14 en Machala, 14 en Cuenca, 23 en Ambato y 1 en Guayaquil. En diciembre 2009 se incrementaron 129 sectores en la encuesta anual para dar representatividad a las provincias de Santo Domingo de los Tsáchilas y Santa Elena. Todo esto

[1]*Los aumentos de estas provincias obedecen a tener representatividad en valores relativos, más no como valor absoluto, ya que aún no se define los límites exactos de la jurisdicción.

responde a la división político administrativa que además incluye La Concordia a la provincia de Esmeraldas (con esto restan sectores en zonas no delimitadas).

Además de los 32 dominios existen 11 dominios agregados conformados por las 5 ciudades más los restos urbanos y por regiones rurales, como se ve en la tabla 2.

TABLA 2: Muestras anuales, 11 dominios por número de sectores. Muestra Nacional Urbano

DOMINIOS	dic-03	dic-04	dic-05	dic-06	dic-07	dic-08	dic-09
QUITO	76	76	76	76	76	76	76
GUAYAQUIL	91	91	91	91	92	92	92
CUENCA	40	40	40	40	54	54	54
MACHALA	40	40	40	40	54	54	54
AMBATO	31	31	31	31	54	54	54
RESTO SIERRA URBANO	325	325	325	325	325	325	354
RESTO COSTA URBANO	248	248	248	248	248	248	276
AMAZONIA URBANO	40	40	40	40	40	40	40
SIERRA RURAL	400	400	400	400	400	400	440
COSTA RURAL	302	302	302	302	302	302	334
AMAZONIA RURAL	40	40	40	40	40	40	40
Total	**1633**	**1633**	**1633**	**1633**	**1685**	**1685**	**1814**

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya

El tamaño muestral para los 8 dominios urbanos de la encuesta trimestral se presenta en la tabla 3, la mayor novedad es que ingresó Ambato como dominio en septiembre 2007. Como ya se indicó se incrementó la muestra para mejorar la representatividad estadística en estas ciudades.

TABLA 3: Diseño muestral con 8 dominios, muestra trimestral

DOMINIOS	2003	2004	2005	2006	2007	2008	2009	2010
QUITO	76	76	76	76	76	76	76	76
GUAYAQUIL	91	91	91	91	92	92	92	92
CUENCA	40	40	40	40	54	54	54	54
MACHALA	40	40	40	40	54	54	54	54
AMBATO	9	9	9	9	54	54	54	54
RESTO SIERRA URBANO	87	87	87	87	87	87	87	87
RESTO COSTA URBANO	116	116	116	116	116	116	116	116
AMAZONIA URBANO	40	40	40	40	40	40	40	40
Total	**499**	**499**	**499**	**499**	**573**	**573**	**573**	**573**

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya

TABLA 4: Diseño muestral con 5 dominios, muestra mensual
Muestra 5 ciudades

Las encuestas mensuales están incluidas en las trimestrales y las anuales, por tanto los tamaños muestrales se mantienen como se muestra en la tabla 4.

DOMINIOS	2003	2004	2005	2006	2007	2008	2009	2010
QUITO	76	76	76	76	76	76	76	76
GUAYAQUIL	91	91	91	91	92	92	92	92
CUENCA	40	40	40	40	54	54	54	54
MACHALA	40	40	40	40	54	54	54	54
AMBATO	9	9	9	9	54	54	54	54
Total	**256**	**256**	**256**	**256**	**330**	**330**	**330**	**330**

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya

3. Rotación de la Muestra a través del Tiempo

La ENEMDU tiene periodicidad trimestral (Nacional – Urbano) y Mensual (5 ciudades) por tanto, es importante determinar como la muestra puede ser rotada (sustituida) de un mes a otro. Un objetivo de la encuesta trimestral es medir cambios en el empleo y desempleo entre trimestres y meses. Entonces se recomendó una superposición parcial entre las muestras de trimestres y meses.

Dada la importancia de las estimaciones de diferencias trimestrales y anuales en los indicadores de empleo, desempleo y subempleo, se puede considerar un esquema de rotación recomendada por la Organización Internacional del

Trabajo (OIT) para encuestas de empleo trimestrales, llamada 2-2-2. Bajo este esquema, un panel de viviendas seleccionadas en una submuestra (réplica) de 25% se mantiene en la muestra durante dos trimestres consecutivos, seguido de un descanso de dos trimestres, y finalmente entran en la muestra por dos últimos trimestres. Este esquema asegura una superposición de 50% entre las muestras de trimestres consecutivos, como también del mismo trimestre en dos años consecutivos. Comparando las muestras para dos años enteros consecutivos, hay una superposición del 55%.

Este esquema de rotación se continuará utilizando para la ENEMDU trimestral y se incorpora un esquema de superposición parcial del 50% entre meses de un periodo intratrimestral.

Se considera el refrescamiento de la muestra cada 2 años, realizándose este procedimiento por última vez en Septiembre 2009.

Anexo 5

ORGANIZACIÓN INTERNACIONAL DEL TRABAJO OIT

La **Organización Internacional del Trabajo** es un organismo especializado de las Naciones Unidas cuyo mandato fundamental es la promoción de la justicia social y el mejoramiento de las condiciones de vida y de trabajo de los trabajadores, mediante el respeto de los derechos humanos fundamentales y laborales internacionalmente reconocidos. La OIT fue creada en 1919, por el Tratado de Versalles, junto con la Sociedad de las Naciones, como resultado del proceso de paz posterior a la Primera Guerra Mundial. La Organización, desde sus inicios pugnó por la reforma social requerida por las situaciones que emanaban de la revolución industrial. Estas condiciones afectaban de manera severa a los trabajadores y, en particular, a los niños y a las mujeres.

En 1946, la OIT se convirtió en el primer organismo especializado asociado a la recién creada Organización de las Naciones Unidas. En 1969, al cumplir 50 años de fundación, la Organización fue galardonada con el **Premio Nobel de la Paz**.

La Organización Internacional del Trabajo se funda en el principio –inscrito en su **Constitución**- de que la paz universal y permanente sólo puede basarse en la justicia social. En este contexto, la OIT se compromete a:

Establecer **Normas Internacionales del Trabajo**, bajo la forma de **convenios** y **recomendaciones**, por las que se fijan condiciones mínimas en materia de derechos laborales. Estas normas están respaldadas por un sistema singular de control de su aplicación. Las normas fundamentales son las relativas a:

- Libertad sindical.
- Derecho de sindicación.
- Derecho de negociación colectiva.
- Abolición el trabajo forzoso.
- Igualdad de oportunidades y de trato.
- Edad mínima de admisión al empleo o trabajo y abolición del trabajo infantil.

Existen otras normas, por las que se regulan condiciones que abarcan las cuestiones relacionadas con el empleo, las condiciones de trabajo, la protección y el diálogo social.

Prestar asistencia técnica a los mandantes de la Organización y ejecutar programas de cooperación técnica internacional para ayudar a los países a formular y llevar a la práctica políticas y programas para promover los derechos fundamentales en el trabajo, mejorar las condiciones de trabajo y de vida y

fomentar las oportunidades de empleo. La asistencia técnica se realiza principalmente en los siguientes campos:

- Políticas de empleo.
- Desarrollo gerencial de micro y pequeñas empresas.
- Cooperativas.
- Administración del trabajo.
- Legislación del trabajo y relaciones laborales.
- Condiciones de trabajo.
- Seguridad y salud en el trabajo.
- Seguridad social.
- Formación y rehabilitación profesional.
- Estadísticas laborales.
- Erradicación del trabajo infantil.
- Género e igualdad de oportunidades.
- Promoción del diálogo social.

Apoyar a las organizaciones de trabajadores y de empleadores, facilitarles formación y asesoramiento técnico y fomentar el diálogo social tripartito. La OIT ayuda a los gobiernos y a las organizaciones de empleadores y de trabajadores a establecer relaciones laborales armoniosas y a adaptar las leyes laborales a las cambiantes necesidades económicas, en el marco de los principios de la Organización.

Realizar actividades de formación, educación, investigación y publicaciones que contribuyan al progreso laboral. La OIT es el mayor centro mundial de recursos de información, análisis y orientación sobre el mundo del trabajo. La investigación acompaña y refuerza todas las actividades prácticas de la Organización y la OIT es considerada universalmente como una fuente autorizada de información estadística. Además, tiene dos instituciones que fomentan la investigación y la capacitación: el Instituto Internacional de Estudios Laborales y el Centro Internacional de Formación de Turín.

ANEXO 6: CLASIFICACIÓN DE LA POBLACIÓN

NACIONAL URBANO				
	dic-07	dic-08	dic-09	dic-10
POBLACION TOTAL	**9,066,209**	**9,202,348**	**9,356,685**	**9,475,345**
Población Menor de 10 años	1,731,368	1,683,167	1,600,208	1,516,402
Población en Edad de Trabajar (PET)	7,334,841	7,519,181	7,756,476	7,958,943
Población Económicamente Activa (PEA)	4,293,138	4,383,512	4,431,196	4,342,647
Ocupados	4,032,738	4,063,084	4,079,905	4,077,357
Ocupados Plenos	1,830,651	1,910,806	1,720,711	1,980,367
Subempleados	2,156,361	2,138,289	2,236,764	2,046,559
Visibles	444,387	406,067	457,432	410,401
Otras formas	1,711,974	1,732,222	1,779,332	1,636,158
Ocupados No clasificados	45,727	13,990	122,430	50,431
Desocupados	**260,400**	**320,427**	**351,291**	**265,290**
Desempleo Abierto	170,148	228,680	266,961	195,923
Desempleo Oculto	90,252	91,748	84,330	69,367
Cesantes (*)	161,242	198,341	242,338	184,231
Trabajadores Nuevos (*)	99,158	122,087	108,954	81,058
Población Económicamente Inactiva (PEI)	**3,041,703**	**3,135,669**	**3,325,280**	**3,616,296**

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración Paola Troya

NACIONAL RURAL				
	dic-07	dic-08	dic-09	dic-10
POBLACION TOTAL	**4,616,093**	**4,676,355**	**4,724,375**	**4,804,340**
Población Menor de 10 años	1,049,787	982,628	907,205	877,360
Población en Edad de Trabajar (PET)	3,566,306	3,693,727	3,817,171	3,926,980
Población Económicamente Activa (PEA)	2,254,970	2,152,799	2,253,915	2,192,593
Ocupados	2,187,961	2,087,449	2,173,035	2,131,684
Ocupados Plenos	409,011	427,836	387,094	459,900
Subempleados	1,759,172	1,658,121	1,779,039	1,667,820
Visibles	340,938	255,536	290,210	244,087
Otras formas	1,418,234	1,402,585	1,488,830	1,423,733
Ocupados No clasificados	19,778	1,491	6,903	3,963
Desocupados	**67,009**	**65,350**	**80,879**	**60,909**
Desempleo Abierto	25,467	26,527	30,523	29,488
Desempleo Oculto	41,542	38,823	50,357	31,421
Cesantes (*)	32,395	29,103	36,767	32,417
Trabajadores Nuevos (*)	34,614	36,247	44,112	28,492
Población Económicamente Inactiva (PEI)	**1,311,335**	**1,540,929**	**1,563,256**	**1,734,387**

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya

ANEXO 7:

EVOLUCIÓN DEL MERCADO LABORAL EN EL ECUADOR				
NACIONAL URBANO				
	dic-07	dic-08	dic-09	dic-10
Tasa de Participación Bruta	47,35%	47,63%	47,36%	45,83%
Tasa de Participación Global	58,53%	58,30%	57,13%	54,56%
Tasa de Ocupación Bruta	54,98%	54,04%	52,60%	51,23%
Tasa de Ocupación Global	93,93%	92,69%	92,07%	93,89%
Tasa de Ocupados Plenos	42,64%	43,59%	38,83%	45,60%
Tasa de Ocupados no Clasificados	1,07%	0,32%	2,76%	1,16%
Tasa de Subempleo Bruta	50,23%	48,78%	50,48%	47,13%
Tasa de Subempleo Global	53,47%	52,63%	54,82%	50,19%
Tasa de Subempleo Visible	10,35%	9,26%	10,32%	9,45%
Tasa de Otras formas de Subempleo	39,88%	39,52%	40,15%	37,68%
Tasa de Desempleo	6,07%	7,31%	7,93%	6,11%
Tasa de Desempleo Abierto	3,96%	5,22%	6,02%	4,51%
Tasa de Desempleo Oculto	2,10%	2,09%	1,90%	1,60%
Tasa de Subutilización Bruta	56,29%	56,09%	58,41%	53,24%

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya

EVOLUCIÓN DEL MERCADO LABORAL EN EL ECUADOR				
RURAL				
	dic-07	dic-08	dic-09	dic-10
Tasa de Participación Bruta	48,85%	46,04%	47,71%	45,64%
Tasa de Participación Global	63,23%	58,28%	59,05%	55,83%
Tasa de Ocupación Bruta	61,35%	56,51%	56,93%	54,28%
Tasa de Ocupación Global	97,03%	96,96%	96,41%	97,22%
Tasa de Ocupados Plenos	18,14%	19,87%	17,17%	20,98%
Tasa de Ocupados no Clasificados	0,88%	0,07%	0,31%	0,18%
Tasa de Subempleo Bruta	78,01%	77,02%	78,93%	76,07%
Tasa de Subempleo Global	80,40%	79,43%	81,87%	78,24%
Tasa de Subempleo Visible	15,12%	11,87%	12,88%	11,13%
Tasa de Otras formas de Subempleo	62,89%	65,15%	66,06%	64,93%
Tasa de Desempleo	2,97%	3,04%	3,59%	2,78%
Tasa de Desempleo Abierto	1,13%	1,23%	1,35%	1,34%
Tasa de Desempleo Oculto	1,84%	1,80%	2,23%	1,43%
Tasa de Subutilización Bruta	80,98%	80,06%	82,52%	78,84%

Fuente: Instituto Nacional de Estadísticas y Censos.
Elaboración: Paola Troya