

UNIVERSIDAD MAYOR DE SAN ÁNDRES
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE ESTADÍSTICA



TRABAJO DIRIGIDO

**“CARACTERÍSTICAS Y DETERMINANTES DEL EMPLEO
EN EL SECTOR INFORMAL EN BOLIVIA,
APLICACIÓN DEL MODELO LOGIT MULTINOMIAL”**

**Trabajo de investigación
para obtener el grado de:**

Licenciatura en Estadística

Postulante: Elizabeth Sonia Nina Siñani

Tutor: Msc. Rubén Belmonte Coloma

La Paz, Bolivia Junio 2018

DEDICATORIA

A Dios

A mis padres: Eduardo Nina Mamani y Teresa Siñani Lopez por su constante apoyo en todo momento y sus consejos, a quienes les debo lo que soy, ambos son mi apoyo, mi fortaleza y además esto va dedicado para mi hermanito choquito quien me acompañó en momentos de soledad y a quien le prometí este logro y a mi fiel amigo copito quien estuvo siempre a mi lado acompañándome.

AGRADECIMIENTOS

Mis sinceros agradecimientos

A Dios por la oportunidad de vivir esta experiencia.

A la Universidad Mayor de San Andrés, a los docentes de la carrera de Estadística, por ser mis pilares fundamentales en mi formación profesional.

A Msc. Rubén Belmonte Coloma, por su asesoramiento, orientación profesional y por su apoyo en todo momento.

A la Lic. Paloma Aguilar Agramont y al Lic. Jhoel Machaca Suñavi por guiarme en el tema, por sus comentarios, recomendaciones los cuales enriquecieron notablemente el manuscrito.

A mi tribunal revisor: Msc. Juan Carlos Flores López y Lic. Dindo Valdez Blanco, por compartir sus conocimientos y sugerencias para la conclusión de este trabajo.

Agradezco a todos los profesionales que componen la Unidad de Estadísticas e Indicadores Sociales, del Instituto Nacional de Estadística, que de una u otra forma han colaborado, contribuido o aportado en el desarrollo de este trabajo.

Agradezco a todas mis amigas y todos mis amigos que siempre me apoyaron para culminar mi proyecto y también una persona especial en mi vida, el que más me apoyo en los momentos difíciles, a ti Porfirio Miranda gracias.

Resumen

El desarrollo del presente trabajo, muestra los aportes que tiene la aplicación de una metodología estadística en el área de empleo en el sector informal, para lo cual se cuenta con información de la base de Encuesta de Hogares 2014 del Instituto Nacional de Estadística (INE).

Inicialmente el trabajo presenta un marco introductorio de las definiciones de la informalidad de algunas organizaciones que estudiaron este tema, revisado las distintas definiciones se pasa a construir la variable de estudio (sector por mercado de trabajo), además se encuentran los objetivos que se quiere lograr, la justificación, el desarrollo del Modelo Logit Multinomial para su aplicación práctica a la base de la Encuesta de Hogares 2014, dentro la Institución (INE).

Se utiliza el Modelo Logit Multinomial para determinar los efectos cualitativos y cuantitativos de factores de características sociodemográficas y económicas de los individuos, los cuales pueden determinar que de alguna manera las personas opten por estar ocupados en los distintos sectores de trabajo, y así usar una metodología apropiada sobre informalidad a la realidad del país.

Abstract

The development of this work shows the contributions of the application of a statistical methodology in the area of employment in the informal area, for which we have information from the base of Household Survey 2014, the National Institute of Statistics (INE).

To start with, the work presents an introductory framework of the definitions of informality of some organizations that studied this topic, revised the different definitions to build the study variable (sector by labor market), in addition there are the objectives to be achieved, the justification, the development of the Multinomial Logit Model for its practical application to the base of the 2014 Household Survey within the Institution (INE).

The Multinomial Logit Model is used to determine the qualitative and quantitative effects of sociodemographic and economic characteristics of the individuals, which can determine that somehow people choose to be employed in the different sectors of work, and thus use a methodology appropriate on informality to the reality of the country.

Índice

Capítulo I.....	1
1.1. Aspectos generales.....	1
1.2. Antecedentes	2
1.3. Planteamiento del problema	5
1.4. Justificación	5
1.5. Hipótesis de la investigación.....	5
1.6. Objetivos.....	6
1.6.1. Objetivo general	6
1.6.2. Objetivos específicos.....	6
1.7. Alcances y limitaciones	6
Capítulo II.....	8
2.1. Antecedentes de informalidad.....	8
2.2. Definición operativa de los sectores de trabajo.....	12
Capítulo III.....	17
3.1. Consideraciones generales del Análisis Multivariante	18
3.2. Análisis de variables	23
3.2.1. Medida de correlación para variables cualitativas nominales.....	26
3.2.2. Medida de correlación para variables cualitativas ordinales.....	27
3.2.3. Medida de correlación para variables cuantitativas	28
3.2.4. Medida de correlación entre una variable cuantitativa y la otra cualitativa	29
3.3. Modelo de Probabilidad Lineal.....	30
3.3.1. Problemas en la estimación del Modelo de Probabilidad Lineal.....	30
3.4. Modelos de opción cualitativa	33
3.5. Modelo Logit	33

3.5.1. Tipología de la modelización Logit	36
3.6. Modelo Logit Multinomial	36
3.6.1. Odds - Ratio.....	42
3.6.2. Efectos marginales.....	43
3.6.3. Métodos de selección de variables explicativas.....	44
3.6.4. Valoración de la bondad de ajuste del modelo estimado	45
3.6.5. Medidas de confiabilidad del Modelo	46
3.6.6. Consideraciones en el uso del Modelo Logit Multinomial.....	48
Capítulo IV.	52
4.1. Instrumento de información.....	52
4.1.1. Identificación de las variables del modelo	57
4.2. Análisis descriptivo de los sectores de trabajo.....	59
4.3. Análisis previo de los datos de las variables del modelo	78
Capítulo V.	82
5.1. Aplicación y análisis del Modelo Logit Multinomial.....	82
Capítulo VI.	99
6.1. Conclusiones	99
6.2. Recomendaciones	108
ANEXOS	111
GLOSARIO	137
BIBLIOGRAFÍA	138

Índice de Tablas

<i>Tabla 1. CRONOLOGÍA DE LOS CENSOS NACIONALES DE POBLACIÓN Y VIVIENDA 1950 A 2012, ENCUESTAS 1975 A 2014.....</i>	<i>4</i>
<i>Tabla 2. OPERACIONALIZACIÓN DE LAS VARIABLES INCLUIDAS COMO POSIBLES PREDICTORAS.....</i>	<i>59</i>
<i>Tabla 3. SUPUESTO DE INDEPENDENCIA DE ALTERNATIVAS IRRELEVANTES.....</i>	<i>83</i>
<i>Tabla 4. RESUMEN DE PROCESAMIENTO DE CASOS.....</i>	<i>84</i>
<i>Tabla 5. RESUMEN DE PASOS.....</i>	<i>84</i>
<i>Tabla 6. INFORMACIÓN DE AJUSTE DEL MODELO.....</i>	<i>85</i>
<i>Tabla 7. BONDAD DE AJUSTE.....</i>	<i>86</i>
<i>Tabla 8. PSEUDO R CUADRADO.....</i>	<i>86</i>
<i>Tabla 9. CONTRASTE DE LA RAZÓN DE VEROSIMILITUD.....</i>	<i>87</i>
<i>Tabla 10. ESTIMACIONES DE LOS PARÁMETROS.....</i>	<i>88</i>
<i>Tabla 11. EFECTOS MARGINALES.....</i>	<i>94</i>
<i>Tabla 12. CLASIFICACIÓN.....</i>	<i>97</i>

Índice de Figuras

<i>Figura 1. DEFINICIÓN OPERATIVA DE LOS SECTORES DE TRABAJO.....</i>	<i>15</i>
<i>Figura 2. RESUMEN DE TÉCNICAS ESTADÍSTICAS MULTIVARIANTES, DE ACUERDO AL MÉTODO EMPLEADO.....</i>	<i>20</i>

Índice de Cuadros

<i>Cuadro 1. DISTRIBUCIÓN DE LA MUESTRA DE VIVIENDAS POR DEPARTAMENTO Y ÁREA GEOGRÁFICA, ENCUESTA DE HOGARES 2014</i>	55
<i>Cuadro 2. BOLIVIA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACION OCUPADA POR SECTORES DE TRABAJO, SEGÚN ÁREA, ENCUESTA DE HOGARES 2014</i>	59
<i>Cuadro 3. BOLIVIA URBANA: DISTRIBUCIÓN DE LA POBLACIÓN OCUPADA POR SECTORES DE TRABAJO, SEGÚN ALGUNAS CARACTERISTICAS, ENCUESTA DE HOGARES 2014</i>	62
<i>Cuadro 4. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA POR SECTORES DE TRABAJO, SEGÚN GRUPOS DE EDAD, NIVEL DE EDUCACIÓN Y DEPARTAMENTOS DEL EJE CENTRAL, ENCUESTA DE HOGARES 2014</i>	64
<i>Cuadro 5. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR FORMAL E INFORMAL, SEGÚN CARACTERISTICAS DE LA EMPRESA, ENCUESTA DE HOGARES 2014</i>	65
<i>Cuadro 6. BOLIVIA URBANA: POBLACIÓN OCUPADA DEL SECTOR INFORMAL POR SEXO, SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2014</i>	71
<i>Cuadro 7. BOLIVIA URBANA: POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR SEXO SEGÚN GRUPO OCUPACIONAL, ENCUESTA DE HOGARES 2014</i>	72
<i>Cuadro 8. NÚMERO DE CASOS DE LAS VARIABLES PARA EL MODELO</i>	78
<i>Cuadro 9. PRUEBA DE MULTICOLINEALIDAD ENTRE VARIABLES INDEPENDIENTES DEL MODELO</i>	80
<i>Cuadro 10. ASOCIACIÓN ENTRE LA VARIABLE DEPENDIENTE Y LAS VARIABLES INDEPENDIENTES</i>	81

Índice de Gráficos

<i>Gráfico 1. BOLIVIA URBANA: POBLACIÓN OCUPADA POR SECTORES DE TRABAJO, ENCUESTA DE HOGARES 2014</i>	<i>60</i>
<i>Gráfico 2. BOLIVIA URBANA: DISTRIBUCION PORCENTUAL DE LA POBLACIÓN OCUPADA EN LOS DISTINTOS SECTORES DE TRABAJO, SEGÚN CRITERIOS DE MEDICIÓN, ENCUESTA DE HOGARES 2014</i>	<i>61</i>
<i>Gráfico 3. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR GRUPOS DE EDAD, ENCUESTA DE HOGARES 2007 Y 2014.....</i>	<i>66</i>
<i>Gráfico 4. BOLIVIA URBANA: POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR SEXO, ENCUESTA DE HOGARES 2007 Y 2014</i>	<i>67</i>
<i>Gráfico 5. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL, SEGÚN NIVEL DE EDUCACIÓN, ENCUESTA DE HOGARES 2007 Y 2014.....</i>	<i>68</i>
<i>Gráfico 6. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL, POR GÉNERO, SEGÚN NIVEL EDUCATIVO, ENCUESTA DE HOGARES 2014.....</i>	<i>69</i>
<i>Gráfico 7. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL, SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2007 Y 2014</i>	<i>70</i>
<i>Gráfico 8. BOLIVIA URBANA: PROMEDIO DEL INGRESO LABORAL EN EL SECTOR DE INFORMAL SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2007 Y 2014.....</i>	<i>74</i>
<i>Gráfico 9. BOLIVIA URBANA: INGRESO LABORAL PROMEDIO DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR GÉNERO, SEGÚN CATEGORIA OCUPACIONAL, ENCUESTA DE HOGARES 2014.....</i>	<i>75</i>
<i>Gráfico 10. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACION OCUPADA EN EL SECTOR INFORMAL, POR GÉNERO, SEGÚN LUGAR DE TRABAJO, EH - 2014</i>	<i>76</i>
<i>Gráfico 11. BOLIVIAURBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACION OCUPADA EN EL SECTOR INFORMAL POR AFILIACIÓN AFP, SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2014.....</i>	<i>77</i>

Capítulo I.

1.1. Aspectos generales

Introducción

La estadística con el pasar de los tiempos ha evolucionado constantemente, ha llegado a convertirse en un instrumento muy eficaz, necesario para la toma de decisiones y la evaluación del comportamiento de características en distintos universos de estudio, la necesidad de respaldar con la teoría estadística es mayor, con el fin de mejorar los instrumentos metodológicos, para así obtener resultados de calidad.

El presente trabajo se realizó en el Instituto Nacional de Estadística¹, el cual es enmarcado en el convenio institucional entre la Universidad Mayor de San Andrés a través de la Facultad de Ciencias Puras y Naturales y carreras dependientes, lo que establece un convenio interinstitucional de intercambio de conocimientos científicos, según:

Decreto Supremo N° 26615 de 16 de marzo de 2001, que en el marco del Subsistema de Capacitación Productiva establece que: *“Las entidades públicas podrán admitir pasantías de estudiantes y egresados destacados...”*

Decreto Supremo N° 1321 de 13 de agosto de 2013 que tiene por objeto promover y consolidar la realización de pasantías, proyectos de grado, trabajos dirigidos y tesis de los estudiantes de las universidades públicas autónomas y universidades indígenas, en las entidades públicas.

¹Ver Anexo 1, Memoria Institucional del Instituto Nacional de Estadística

De esta manera y gracias a este acuerdo, el Instituto Nacional de Estadística (INE) y la carrera de Estadística, realizaron un convenio en el formato de trabajo dirigido.

El documento, se enfoca en el análisis de la población ocupada, de 10 años o más (edad tomada para el análisis según el INE, que vive en el área urbana de Bolivia, el motivo por el cual se excluye del análisis a la población rural es que debido a la presencia mayoritaria de economía campesina y la poca existencia de relaciones laborales formales en esta área, no es posible analizar las interacciones y transiciones de trabajadores entre distintos sectores de trabajo. Por el contrario en el área urbana se puede observar cierta dualidad entre trabajadores formales e informales y por tanto significativos flujos de movilidad e interacción entre ambos sectores.

El presente trabajo se divide de la siguiente manera: el capítulo I, contiene los objetivos del trabajo, el alcance y sus limitaciones; el capítulo II, abarca los enfoques teóricos de la informalidad; el capítulo III, expone el análisis multivariante, la metodología estadística empleada, es decir el Modelo Logit Multinomial el cual es muy importante para medir probabilidades especialmente en la situación que interesa; el siguiente capítulo muestra el análisis descriptivo de las variables asociadas con la informalidad; el capítulo V, contiene los resultados de la aplicación del Modelo Logit Multinomial y su respectivo análisis y el último capítulo muestra las conclusiones y recomendaciones.

1.2. Antecedentes

Actividades que desarrolla el INE

El Instituto Nacional de Estadística de Bolivia, inicia su labora partir de la información generada en el Censo Demográfico de 1950, con el que se empezaron a construir herramientas de infraestructura estadística que

coadyuvan con la planificación de las encuestas dirigidas a hogares, en la **tabla 1**, se aprecia un orden cronológico de los censos y las encuestas a hogares implementadas por el INE, describiendo el ámbito de referencia de estudio y los entes involucrados con la organización, desde 1950 al 2014. El Instituto Nacional de Estadística desde 1978 ha venido realizando varias rondas de encuestas a hogares, como ser: Encuesta Permanente de Hogares (EPH), Encuesta Integrada de Hogares (EIH), Encuesta Nacional de Empleo (ENE) y Encuestas y Medición sobre Condiciones de Vida (MECOVI).

Con la iniciativa del Banco Mundial, desde 1999 se inició en la región, el Programa de Mejoramiento de las Encuestas y Medición sobre Condiciones de Vida (MECOVI) con el propósito de recopilar información sobre condiciones de vida de la población boliviana.

Entre los años 2003 y 2004 se ejecutó la Encuesta Continua de Hogares con características similares a una encuesta de Presupuestos Familiares, conservándose los tradicionales contenidos de la encuesta, haciendo énfasis en la temática de ingresos y gastos del hogar.

Para los años 2005 a 2014 se retomó la modalidad de encuestas puntuales de hogares, con la ejecución de la Encuesta de Hogares, que mantuvo como temática de exploración a las condiciones de vida de la población boliviana, a través de la aplicación del cuestionario multitemático que permite la investigación de las características sociodemográficas, salud, educación, empleo, ingresos, gastos, vivienda y servicios básicos de los hogares.

Tabla 1. CRONOLOGÍA DE LOS CENSOS NACIONALES DE POBLACIÓN Y VIVIENDA 1950 A 2012, ENCUESTAS 1975 A 2014

Año	Actividad estadística	Ámbito de referencia	Organizador
1950	Censo Demográfico "VII Censo de Población y I Censo de Vivienda" (2.704.165 Habitantes)	Bolivia	INE (Gobierno de Mamerto Urriolagoitia)
1975	Primera Encuesta Demográfica Nacional	Bolivia	INE- CELADE
1976	Censo Nacional de Población y Vivienda "VIII Censo de Población y II Censo de Vivienda" (4.613.486 Habitantes)	Bolivia	INE (Gobierno de Hugo Banzer Suarez)
1977	Encuesta de empleo en los Centros Mineros	Bolivia	INE
1978	Inicio en la realización de las Encuestas Permanentes de Hogares (1978 - 1985)	Bolivia	INE
1980	Segunda Encuesta Demográfica	Bolivia	INE – CELADE
1988	Encuesta Nacional de Población y Vivienda	Bolivia	INE – UNFPA
1988	Inicio en la realización de la Encuesta Integrada de Hogares (1988 a 1994)	Bolivia	INE - Banco Mundial
1989	Primera Encuesta Nacional de Demografía y Salud	Bolivia	INE – Programa DHS
1991	Encuesta Nacional de Consumo de Energía en el Área Rural de Bolivia	Bolivia Área Rural	INE - Banco Mundial
1992	Censo Nacional de Población y Vivienda "IX Censo de Población y III Censo de Vivienda " 6420792 habitantes.	Bolivia	INE (Gobierno Paz Zamora)
1994	Segunda Encuesta Nacional de Demografía y Salud	Bolivia	INE – Programa OHS
1996	Inicio en la realización de la Encuesta Nacional de Empleo (1996 - 1997)	Ciudades Capitales	INE
1998	Tercera Encuesta Nacional de Demografía y Salud	Bolivia	INE –Programa DHS - Ministerio de Salud
1999- 2002	Inicio en la realización de la Encuesta de Mejoramiento de Condiciones de Vida (MECOVI)	Bolivia	INE - Banco Mundial - Banco Internacional de Desarrollo – CEPAL
2003- 2004	Encuesta Continua de Hogares	Bolivia	INE
2005- 2014	Inicio en la realización de la Encuesta de Hogares	Bolivia	INE – Banco Mundial
2001	Censo Nacional de Población y Vivienda "X Censo de Población y IV Censo de Vivienda" (8.274.325 habitantes)	Bolivia	INE (Gobierno de Jorge Quiroga Ramírez)
2012	Censo de Población y Vivienda	Bolivia	INE (Gobierno de Evo Morales Ayma)

Fuente: Elaboración Propia en Base a la Información del INE

1.3. Planteamiento del problema

Que características determinan la probabilidad de pertenecer a algún sector de trabajo de la población ocupada del área urbana de Bolivia.

1.4. Justificación

El modelo Logit ha sido en las últimas décadas una herramienta de gran utilidad en el análisis estadístico de datos categóricos, el tratamiento de información de datos cualitativos se convierte importante en muchas áreas, con el tiempo se desarrolló una serie de modelos, enunciados a partir del Modelo Logit, entre ellos el Modelo Logit Multinomial. El modelo Logit Multinomial es eficiente para tratar variables de respuesta de tipo cualitativa con más de dos categorías en la variable dependiente.

La realidad de Bolivia según reportes de estudios realizados por la Organización Internacional del Trabajo(OIT, 2002), así como otros estudios, más del 50 por ciento de la población ocupada tienen un trabajo informal, por tanto al no existir estudios que permitan determinar la probabilidad de que las personas trabajen en algún sector de trabajo, por tal motivo se desarrollará un estudio de las características sociodemográficas y económicas de la población ocupada que hacen que opten por algún sector de trabajo en Bolivia a través de la aplicación del Modelo Logit Multinomial.

1.5. Hipótesis de la investigación

Hipótesis

“La aplicación del modelo Logit Multinomial permiten determinar la probabilidad de pertenencia a un sector de trabajo de la población ocupada.”

1.6. Objetivos

1.6.1. Objetivo general

Determinar un modelo estadístico que permita obtener la probabilidad de que las personas trabajen en algún sector de trabajo dependiendo de las características sociodemográficas y económicas de las personas, teniendo en cuenta las variables que se encuentran en la base de la Encuesta de Hogares 2014².

1.6.2. Objetivos específicos

- Adoptar una definición operativa que incluya los tres sectores de trabajo.
- Analizar los principales factores que inciden en la informalidad, a través de cuadros descriptivos, gráficos y obtener pruebas estadísticas de las variables de estudio.
- Interpretar los resultados del modelo Logit Multinomial.
- Obtener estimadores de los parámetros del modelo de las variables de estudio.
- Interpretar los efectos marginales y pruebas validación del ajuste del modelo estimado.
- Analizar los resultados del modelo en la investigación.

1.7. Alcances y limitaciones

El presente trabajo tiene como alcance adoptar una definición operativa de clasificación de los sectores de trabajo, dentro del cual se encuentra el sector informal, formal y doméstico, la cual debe ser adecuada a la realidad del país, ya que Bolivia cuenta con distintos sectores de trabajo, se considera a las

²La base y el cuestionario de la Encuesta de Hogares se puede descargar de la página web www.ine.gob.bo/index.php/banco/base-de-datos-sociales

personas ocupadas del sector informal como prioridad y la aplicación de una metodología estadística, como es el Modelo Logit Multinomial.

La limitación de esta investigación está dado a tomar en cuenta a las personas de 10 años o más, que estén ocupadas, del área urbana definido por el Instituto Nacional de Estadística (UDAPE, 2006), el motivo por el cual se excluye del análisis a la población rural es que debido a la presencia mayoritaria de economía campesina y la poca existencia de relaciones laborales formales en estas áreas, no es posible analizar las interacciones y transiciones de trabajadores entre distintos sectores de trabajo. En cambio en el área urbana, se puede observar cierta dualidad entre trabajadores del sector formal e informal y por tanto significativos flujos de movilidad e interacción entre ambos sectores.

Capítulo II.

Marco teórico

2.1. *Antecedentes de informalidad*

El término “informal” (OIT, 2013) generalmente ha estado asociado con la situación de un trabajador o trabajadora “pobre”, que no tiene contrato o que carece de seguro social. La primera vez que aparece históricamente la dicotomía formal - informal es a principios de los años sesenta como resultado de un estudio llevado a cabo en Ghana por (Hart, 2002), el término “economía informal” adquirió popularidad gracias al estudio elaborado por la Organización Internacional del Trabajo (OIT) en una misión realizada a Kenya en el año 1972, en este estudio, la OIT³ definió la informalidad económica considerando características como: la facilidad de entrada, la operación a baja escala, la propiedad familiar, tecnología adaptadas a procesos, reducido capital financiero, intensidad en el uso del factor trabajo, mercados sin regulación pero competitivos y procesos informales de capacitación o de adquisición de habilidades, este estudio se introduce la noción de “oportunidad de ingreso” formal e informal para estudiar la ocupación en el medio urbano de dicho país, en particular entre los estratos de población de menores ingresos.

Durante muchos años los gobiernos y los economistas suponían que con la combinación adecuada de políticas económicas y de recursos, las economías pobres y tradicionales se transformarían en economías modernas dinámicas, en este proceso se esperaba que el sector tradicional o informal desapareciera en la medida en que el sector moderno o formal crecería y absorbería mayor trabajo. Sin embargo, contrariamente a lo esperado, el sector informal y el empleo informal siguen siendo considerables. Muchos países no han sido

³Para mayor detalle, en el Anexo 2 se encuentra la matriz que utiliza la OIT para la clasificación de: empleo informal, empleo en el sector informal y empleo informal fuera del sector informal (OIT, 2002).

capaces de desarrollar una economía moderna capaz de proporcionar oportunidades de empleo para sus poblaciones en rápido crecimiento. El sector informal sigue siendo una de las fuentes principales de empleo, si no la principal fuente de empleo en muchos países del mundo, está compuesto de pequeñas empresas o de empresas por cuenta propia, con poca o ninguna organización formal o capital, y con empleo informal. En los países de Europa del Este y de Asia Central el sector informal representa una parte importante de la economía.

Tanto en los países en desarrollo como en los países desarrollados, el sector moderno o formal se está transformando en virtud de la globalización y de los cambios en las políticas económicas.

Algunas formas de producción están cambiando, abandonan las grandes fábricas registradas y se reorganizan en unidades especializadas, más descentralizadas y flexibles. El trabajo permanente a tiempo completo está siendo reemplazado por acuerdos contractuales no convencionales o por trabajos informales, algunos ejemplos de formas emergentes de empleo comprenden trabajos a tiempo parcial o el empleo en una parte del año, contratos de trabajo a plazo fijo, trabajo temporal proporcionado por las agencias de empleo o contratos con empresas, el empleo informal, la subcontratación y el trabajo externo o trabajo a domicilio. Además no solo existe una parte importante del empleo en el sector informal y en los empleos informales fuera del sector informal, sino que los contratos y los horarios de trabajo son más complejos.

Las actividades típicas del sector informal (como el trabajo no remunerado en una empresa familiar, el trabajo asalariado ocasional, el trabajo a domicilio y la venta en la calle) son las únicas oportunidades que tienen muchas personas pobres para satisfacer sus necesidades básicas de supervivencia. En los países en los que el seguro de desempleo u otras clases de beneficios sociales son

inexistentes, la única alternativa al desempleo es incorporarse en el empleo en el sector informal. Otros tipos de empleo en el sector informal (como los empleadores de establecimientos manufactureros informales o los trabajadores independientes calificados en pequeñas empresas) a veces pueden proporcionar una mejor remuneración. Estos trabajadores pueden ganar incluso mucho más que los asalariados fijos que se desempeñan en empleos formales, pero incluso para estos trabajadores que se encuentran en mejor posición, con frecuencia, el empleo en el sector informal más que el empleo del sector formal, es la única opción. El sector informal es importante no solo como fuente de empleo sino también en la producción de bienes y servicios.

El empleo informal representa una parte considerable del empleo en la actualidad, incluye a las personas ocupadas que, por ley o en la práctica, no se encuentran sometidas a la legislación laboral nacional, al impuesto sobre la renta o que no tienen derecho a la protección social ni a los beneficios laborales. El empleo informal puede existir tanto en el sector formal como en el sector informal de la economía. El empleo informal abarca todas las situaciones de empleo de los trabajadores: pueden ser empleadores, asalariados, trabajadores por cuenta propia, trabajadores familiares auxiliares (no remunerados) y miembros de cooperativas de productores.

El sector informal y el empleo informal interactúan estrechamente con el sector formal. Como se señaló anteriormente, la economía moderna globalizada depende en forma creciente del empleo informal, en formas como la subcontratación, el trabajo a domicilio, el trabajo por agencias y otros tipos de empleo flexibles o temporales. En términos más generales, los establecimientos del sector formal pueden depender del sector informal y del empleo informal para asegurar flexibilidad y menores costos y generar mayores ingresos para sus propietarios. El sector informal no existe en forma separada del sector

formal; más bien produce para él, comercializa con el sector, distribuye y presta servicios al sector formal.

Algunos resultados de Informalidad en Bolivia

Desde hace más de una década, el sector informal se ha convertido en uno de los sectores más importantes y dinámicos de la economía boliviana, abarcando prácticamente todas las ramas de actividad económica y proporcionando más empleo que cualquier otro sector. Si bien hubo una reducción de la informalidad en términos relativos (porcentajes) entre 1996 y 2006, de 63 por ciento a 58 por ciento respectivamente, en términos absolutos se tiene un incremento aproximadamente en 200 mil personas pasando de 1.2 millones a 1.5 millones de personas (UDAPE, 2006).

La OIT en el año 2004 realizó una nueva clasificación de la población ocupada según su grado de informalidad en distintas regiones a nivel mundial. De acuerdo a este informe (OIT, 2014), los países de África, Asia Meridional y aquellos de la Región Andina tienden a presentar los porcentajes más elevados de empleo informal, en este estudio Bolivia es el país que posee los niveles más altos de informalidad de la región de Latinoamérica con el 67 por ciento, Honduras y Perú con el 56 por ciento.

Otro estudio comparativo (Gasparini L. y Tornarolli L., 2009), indaga sobre el tamaño y características del mercado informal en América Latina y el Caribe, ellos adoptan dos definiciones de informalidad: la *definición legalista* la cual se basa en la protección social del trabajador, y la *definición productiva* que se basa en el tamaño de la empresa, ingresos y productividad. Tomando en cuenta la definición productiva, la economía boliviana una vez más presenta los porcentajes más elevados de informalidad (77 por ciento), por encima de Colombia (71 por ciento) y Perú (69 por ciento).

Para el 2006 los resultados en Bolivia, confirman la existencia de un grado muy reducido de protección social dentro de la población ocupada el cual es medido de acuerdo a la cotización que realiza un trabajador a una AFP (Administración de Fondos de Pensiones), en base a este criterio el 79 por ciento de la población ocupada es clasificada como informal. Si se considera el criterio de identificación tributaria, los niveles de informalidad se reducen, obteniendo que el 65 por ciento de la población ocupada es informal.

A pesar del uso de las distintas definiciones de informalidad, según diversas estimaciones, Bolivia es el país con más altos niveles de informalidad en América Latina.

Adoptar una definición de informalidad no es una tarea fácil, dado que el sector informal es heterogéneo y bastante ambiguo y depende de la dimensión del mercado laboral que se quiera analizar y de su contexto, y por tanto en el caso boliviano también lo es.

2.2. Definición operativa de los sectores de trabajo

Revisado la parte teórica de la informalidad, se utiliza la clasificación que realiza el Instituto Nacional de Estadística (INE), en la Encuesta Nacional de Empleo (ENE) elaborada en 1997(UDAPE, 2006), que define al sector informal como un segmento económico que identifica unidades productivas o negocios de escaso nivel de organización e incipiente uso de capital y tecnología.

Para un mejor entendimiento se revisa los conceptos (INE, 2014) de categoría ocupacional del trabajador, como sigue a continuación:

Obreros, personas que trabaja realizando **esfuerzo físico** para un/a empleador/a público o privado y percibe una remuneración monetaria (salario)

y/o en especie. Generalmente son ocupados en actividades de extracción y explotación de recursos naturales (agricultura, minería), o actividades secundarias y de transformación (industria manufacturera, construcción, etc.).

Empleados, personas que trabaja para un/a empleador/a público o privado y percibe una remuneración monetaria (sueldo) y/o en especie. Comprende la **explotación de conocimientos o habilidades intelectuales**. Trabajan en ocupaciones administrativas, técnicas, control, supervisión, comercio y servicios.

Trabajadores por cuenta Propia, es la persona que tiene su **propia empresa o negocio, sin tener ningún trabajador/a remunerado a su cargo ni depender de un patrón**; vende y/o produce bienes o servicios con ayuda de trabajadores/as familiares o aprendices o sin la ayuda de ellos. Por ejemplo: sastre, carpintero, vendedor ambulante, plomeros, etc.

Patrones o empleadores, se tienen:

Patrones o empleadores que si reciben salario, es la persona que dirige su propia empresa o unidad económica. Cuenta con trabajadores/as asalariados (obreros, empleados). A pesar de ser propietario o socio percibe una remuneración mensual por el trabajo desarrollado en su propio establecimiento económico y a fin de gestión le corresponde una parte de las utilidades generadas por el negocio o empresa. Generalmente las empresas donde trabajan, **llevan estados financieros** y un control de toda la empresa o negocio.

Patrones o empleadores que no reciben salario, es la persona que dirige su propia empresa o unidad económica. Cuenta con trabajadores asalariados (obreros, empleados). La característica principal de los establecimientos

económicos donde trabajan es que los mismos **no llevan estados financieros** y no existe un control riguroso de la empresa o negocio.

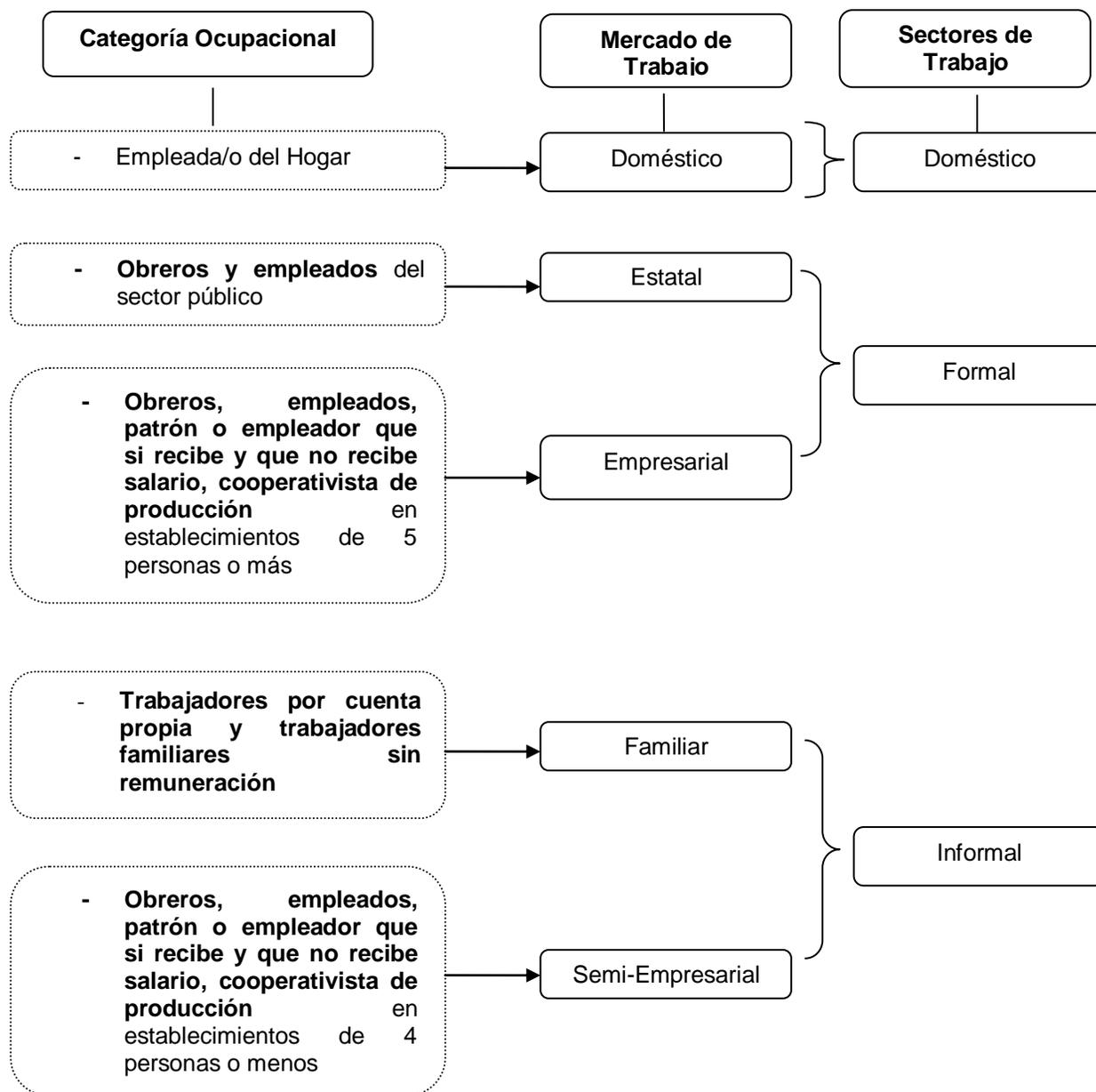
Cooperativista de producción, es la persona que, siendo socia, trabaja activamente en una empresa cooperativa, recibiendo ingresos o asumiendo las pérdidas en su calidad de cooperativista. Por ejemplo: Cooperativistas mineros, auríferos, etc.

Trabajadores familiares o aprendices sin remuneración, personas que realiza alguna actividad sin recibir a cambio remuneración monetaria o en especie, pudiendo ser familiar o no del dueño del lugar donde trabaja. Por ejemplo: Ayudante de taller mecánico, hijos/as de dueños de tiendas que ayudan en la venta, familiares del agricultor campesino, etc.

Empleada/o del hogar, Persona que trabaja en hogares realizando un trabajo de carácter doméstico, recibiendo a cambio un salario en dinero y/o en especie. Por ejemplo: Empleadas domésticas “cama adentro” o “cama afuera”, niñeras, cocineras, mayordomos, jardineros a **contrato fijo**, etc.

De esta forma, la definición operativa adoptada está basada en el tipo de establecimiento y la categoría ocupacional del trabajador, tal como se esquematiza en la siguiente figura:

Figura 1. DEFINICIÓN OPERATIVA DE LOS SECTORES DE TRABAJO



Fuente: Elaboración Propia en Base a la Información del INE – ENE 1997

Tomando en cuenta estos dos criterios de tamaño de la empresa y la categoría ocupacional de un trabajador, se obtiene la siguiente definición operativa a ser utilizada en el presente documento (Figura 1).

Se clasifica primeramente los mercados de trabajo:

- Doméstico: Incluye a las empleadas (os) domésticas (os).
- Estatal: Incluye a la población ocupada que trabaja en la administración pública.
- Familiar: Incluye a los trabajadores por cuenta propia y a los trabajadores familiares sin remuneración.
- Semi-empresarial: Incluye a los ocupados en establecimientos con 4 o menos personas.
- Empresarial: Incluye a los ocupados en establecimientos con más de 5 personas⁴.

De acuerdo a esta clasificación, se tiene a la población ocupada en el sector informal, conformados por los mercados de trabajo familiar y semi-empresarial; el sector formal, lo conforman los ocupados en los mercados de trabajo estatal y empresarial y por último manteniendo el mercado de trabajo doméstico como un sector especial.

La variable que agrupa los tres sectores de trabajo es la variable “mercado de trabajo”, (sector formal, sector informal y sector doméstico). Por tanto esta variable captura información de algunas definiciones que existen de la informalidad, si se tomara en cuenta otra definición y se flexibiliza su aplicabilidad práctica, provocaría que el total de la muestra de la población ocupada se reduzca significativamente, generando problemas de representatividad estadísticas en los resultados a obtener.

⁴La distinción dual que se hace sobre el tamaño de la empresa sirve para fines de la división entre el sector formal e informal, sin embargo se manejan mayores divisiones de este concepto, como por ejemplo, según el INE para el caso boliviano cuatro tamaños de empresas: microempresa (menor a 10 empleados), pequeña empresa (de 10 a 20 empleados), empresa mediana (de 20 a 50 empleados) y gran empresa (más de 50 empleados).

Capítulo III.

Marco Metodológico

Análisis Estadístico

El análisis multivariante es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado, el objetivo del análisis multivariante es proporcionar métodos para el estudio del conjunto de datos multivariantes, ayuda al analista o investigador a tomar decisiones óptimas en el contexto en el que se encuentre, teniendo en cuenta la información disponible por el conjunto de datos analizados, dentro del análisis multivariante se encuentra el Modelo Logit, Logit Multinomial y otros modelos.

Modelo Logit

El Modelo Logit ha sido en las últimas décadas una herramienta de gran utilidad en el análisis estadístico de datos categóricos. La utilización del Modelo Logit en distintas ciencias ocurre a un ritmo desigual e incluso de forma independiente. Los primeros avances de este modelo ocurrieron en: “Epidemiología” los estudios de casos y controles, basados en ratios de probabilidad, propician la rápida incorporación del Modelo Logit, el principal difusor de la regresión logística fue David Roxbee Cox con la publicación de su libro “The Análisis of Binaria Data” (Análisis de Datos Binarios) en el año 1970, la importancia de los trabajos de este autor reside tanto en la utilización que hace del Modelo Logit para formalizar fenómenos binarios (solo dos posibles alternativas) y en la difusión del modelo en aplicaciones econométricas, como en la introducción de transformaciones del modelo, por ejemplo, la creación del Modelo Logit mixto como un modelo alternativo que puede situarse entre el modelo Logit (el cual mantiene su simplicidad) y el Probit (el cual comparte su flexibilidad).

El empuje definitivo para el reconocimiento de Modelo Logit se encuentra en los trabajos del economista americano Daniel McFadden, quien vincula este modelo a la teoría de la elección discreta, abriendo un nuevo campo de trabajo que le hizo merecedor del premio nobel de economía en el año 2000, los trabajos de McFadden se remontan a 1973, cuando trabajaba en California como consultor en un proyecto público en materia de transporte.

La importancia de los modelos de elección discreta radica en la modelización de variables cualitativas, característica que exige la codificación de la variable como paso previo a la modelización, en este proceso los distintos estados de la variable se transforman en códigos para ser tratados utilizando técnicas de regresión, McFadden planteó inicialmente en el caso que los individuos se enfrentan a procesos de decisión dicotómicos, es decir en los que únicamente hay dos posibles alternativas; una de las alternativas implica llevar a cabo una determinada actuación y la otra no.

Con el tiempo se desarrolló una serie de modelos, enunciados a partir del Modelo Logit, entre los que se puede citar son: el Modelo Logit Jerárquico, el Modelo Logit Multinomial, el Mixed Logit.

Actualmente el uso de los modelos de la familia Logit es aplicado en muchas ciencias como la economía, la medicina, la biología, la psicología, entre otras aplicaciones.

3.1. Consideraciones generales del Análisis Multivariante

Una de las ramas de la estadística es el Análisis Multivariante (Salvador Figueras M., 2000), este es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado, donde se tiene variables de tipo cualitativo, cuantitativo o una mezcla de ambos.

Tipos de técnicas multivariantes

Se pueden clasificar en tres grandes grupos (ver **figura 2**):

a. Métodos de dependencia

Suponen que las variables analizadas están divididas en dos grupos: *las variables dependientes y las variables independientes*, el objetivo de estos métodos consiste en determinar si el conjunto de variables independientes, afecta y de qué forma afecta a la variable dependiente.

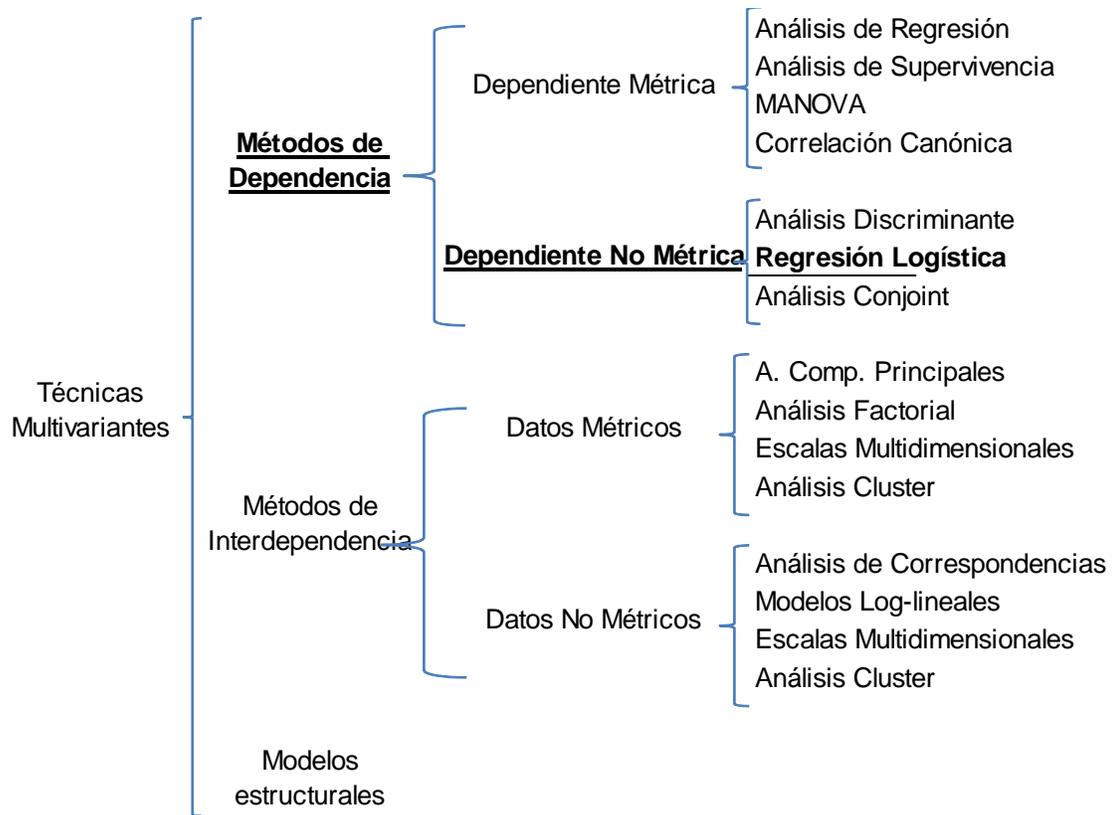
b. Métodos de interdependencia

Estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y porqué.

c. Métodos estructurales

Suponen que las variables están divididas en dos grupos: el de las variables dependientes y las independientes. El objetivo de estos métodos es analizar, no sólo como las variables independientes afectan a la variable dependiente, sino analiza cómo están relacionadas las variables de los dos grupos entre sí.

Figura 2. RESUMEN DE TÉCNICAS ESTADÍSTICAS MULTIVARIANTES, DE ACUERDO AL MÉTODO EMPLEADO



Fuente:(Salvador Figueras M., 2000)

La **figura 2.** Muestra los distintos tipos de técnicas estadísticas multivariantes, tanto para variables de tipo cualitativo como cuantitativo, el método a utilizar en el presente trabajo es el método de dependencia para variables no métricas (variables cualitativas), es decir el método de la regresión logística.

Análisis previo de los datos

Antes de empezar la aplicación de algún método multivariante primeramente se debe hacer un análisis cuidadoso de los datos, el cual conduce a una mejor predicción y a una evaluación más precisa de la dimensionalidad, existen técnicas para examinar las variables individuales como las relaciones entre

ellas mediante gráficos, test y otros, además de evaluar los datos ausentes y aplicar algún método de imputación, identificar los casos atípicos y comprobación de los supuestos si el método lo requiere, para dar solución a los problemas que se puedan presentar al aplicar algún método multivariante (Hair, J. F. y Anderson R. E. y Tatham R. L.; Black W. C., 1999). A continuación se muestran los problemas que se puedan presentar:

Caso de los datos ausentes

Es algo habitual en el análisis multivariante, el hecho de tener ausencia de datos, esto puede ocurrir por muchas razones y en muchas situaciones. Un tipo de proceso de datos ausentes que puede ocurrir en cualquier situación se debe a factores de procedimiento, tales como errores en la entrada de datos que crean códigos inválidos, restricciones de representatividad, fallos al completar el cuestionario o incluso la respuesta que puede dar el encuestado.

Ante la ausencia de datos, se puede aplicar soluciones, entre ellas se tiene: suprimir los datos ausentes, pero existen métodos de imputación, que es el proceso de estimación de valores ausentes basado en valores válidos de otra variable y/o casos de la muestra. El objetivo es emplear relaciones conocidas que pueden identificarse en los valores válidos de la muestra para ayudar en la estimación de valores ausentes.

Los métodos que existen suelen usarse para variables cuantitativas por dos razones: en primer lugar, se pueden hacer estimaciones de los datos ausentes para variables cuantitativas con valores como una media de todos los valores válidos, en segundo lugar, las variables cualitativas requieren una estimación de un valor específico en vez de una estimación en una escala continua.

Los métodos más ampliamente utilizados son: sustitución de caso, sustitución por la media, sustitución por valor constante, imputación por regresión, imputación múltiple, entre otros.

Casos de datos atípicos

Los casos de datos atípicos son observaciones con una combinación única de características identificables que les diferencia claramente de las otras observaciones. Los casos atípicos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y deben ser evaluados por los tipos de información que pueden proporcionar. Cuando son benéficos los casos atípicos, aunque diferentes a la mayor parte de la muestra pueden ser indicativos de las características de la población que se llegarían a descubrir en el curso normal del análisis. Por el contrario, los casos de datos atípicos problemáticos no son representativos de la población y están en contra de los objetivos del análisis, pueden distorsionar seriamente los test estadísticos.

Multicolinealidad

En un modelo Multivariante suele suponerse que entre sus variables (independientes) X_1, X_2, \dots, X_8 , no existe relación lineal entre ellas, es decir que dichas variables no se encuentran correlacionadas, la correlación debe ser relativamente baja, si tal situación no se cumple se dice que el modelo presenta problemas de Multicolinealidad (Perez L. C., 2004), cuando las variables independientes que intervienen en el modelo están muy correlacionadas, conducen a un modelo desprovisto de sentido y por tanto a unos valores de los coeficientes no interpretables, por esta razón debe realizarse un análisis previo de correlación entre las distintas variables independientes.

Si se encuentra que las variables tienen Multicolinealidad, entre las soluciones se tiene: ampliar la muestra, transformar la variable adecuadamente, suprimir algunas variables con justificación estadística u otros métodos.

La matriz de correlaciones es un instrumento que ayuda a detectar la presencia de Multicolinealidad. Valores altos de esta matriz son síntomas de posible dependencia entre las variables implicadas, además de algunos estadísticos para verificar si se tiene valores altos entre estas variables, según el tipo. A continuación se describe algunas medidas de correlación según el tipo de variable.

3.2. Análisis de variables

Tablas de contingencia, el primer paso es el cruce de dos variables categóricas (De la Fuente F. S.), para medir el grado de asociación entre ellas, es construir una tabla de contingencia⁵.

Cuando se desea comparar dos caracteres (X, Y) en una misma población que admiten las modalidades: $X(x_1, x_2, \dots, x_i, \dots, x_k)$, $Y(y_1, y_2, \dots, y_j, \dots, y_m)$, se toma una muestra de tamaño n, representando por n_{ij} el número de elementos de la población que presentan la modalidad x_i de X e y_j de Y.

⁵La tabla de contingencia contiene, en cada casilla el conteo del número de casos en cada una de las combinaciones de valores de ambas variables, se muestran los totales verticales (por columnas), y los totales horizontales (por filas). Estas, por aparecer en los márgenes, se denominan frecuencias marginales. También se muestra el total de casos N, que corresponde a la suma de las frecuencias marginales fila, o a la suma de las frecuencias columna, que son iguales.

$X \backslash y$	y_1	y_2	...	y_j	...	y_m	$\sum_{j=1}^m n_{i\cdot}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	$n_{2\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	$n_{i\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{km}	$n_{k\cdot}$
$\sum_{i=1}^k n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$...	$n_{\cdot j}$...	$n_{\cdot m}$	n

Cada frecuencia observada n_{ij} ($i=1, \dots, k$; $j=1, \dots, m$) de la tabla de contingencia ($k \times m$) hay una frecuencia esperada (e_{ij}) que se obtiene mediante la expresión:

$$e_{ij} = p_{ij} * n = \frac{n_{i\cdot} * n_{\cdot j}}{n}; \quad \text{donde } p_{ij} = \frac{n_{i\cdot}}{n} \times \frac{n_{\cdot j}}{n}$$

Agrupando frecuencias observadas y esperadas en la tabla de contingencia ($k \times m$):

$x \backslash y$	y_1	y_2	...	y_j	...	y_m	$\sum_{j=1}^m n_{i \cdot}$
x_1	n_{11} (e_{11})	n_{12} (e_{12})	...	n_{1j} (e_{1j})	...	n_{1m} (e_{1m})	$n_{1 \cdot}$
x_2	n_{21} (e_{21})	n_{22} (e_{22})	...	n_{2j} (e_{2j})	...	n_{2m} (e_{2m})	$n_{2 \cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_i	n_{i1} (e_{i1})	n_{i2} (e_{i2})	...	n_{ij} (e_{ij})	...	n_{im} (e_{im})	$n_{i \cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_{k1} (e_{k1})	n_{k2} (e_{k2})	...	n_{kj} (e_{kj})	...	n_{km} (e_{km})	$n_{k \cdot}$
$\sum_{i=1}^k n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$...	$n_{\cdot j}$...	$n_{\cdot m}$	n

El estadístico Chi-cuadrado, propuesto por Pearson (1911) que permite contrastar la hipótesis de que las dos variables categóricas son independientes. El estadístico de contraste observado es:

$$\sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \chi_{(k-1)(m-1)}^2$$

Que sigue aproximadamente una Chi-cuadrado con $(k-1) \times (m-1)$ grados de libertad.

Entonces el estadístico χ^2 valdrá cero cuando las variables sean completamente independientes (pues las frecuencias observadas y las esperadas serán iguales) y el valor del estadístico χ^2 será mayor cuando mayor sea la discrepancia entre las frecuencias observadas y las esperadas.

Por tanto, se utilizara la distribución χ^2 para establecer el grado de compatibilidad existente entre el valor del estadístico χ^2 y la hipótesis de independencia. Si los datos son compatibles con la hipótesis de independencia, la probabilidad asociada al estadístico χ^2 será alta (mayor de 0,05). Si esa probabilidad es muy pequeña (menor que 0,05) si los datos son incompatibles con la hipótesis de independencia entonces se concluirá que las variables estudiadas están relacionadas.

Para que las probabilidades de la distribución χ^2 constituyan una buena aproximación a la distribución del estadístico χ^2 conviene que se cumplan algunas condiciones; entre ellas, que las frecuencias esperadas no sean demasiado pequeñas.

La limitación que tiene el estadístico Chi-cuadrado como medida de asociación, es que está acotado inferiormente por cero, pero no tiene cota superior. Por lo tanto es difícil evaluar si su valor es grande o pequeño.

3.2.1. Medida de correlación para variables cualitativas nominales

Como alternativa a efectuar una prueba Chi-cuadrado, se puede simplemente cuantificar el grado de asociación, utilizando alguna medida de asociación adecuada, o puede servir como complemento a la prueba, que sólo concluye si hay o no asociación, pero no dice cuánta asociación. Se dispone de tres medidas, todas basadas en el estadístico Chi-cuadrado.

Coefficiente Phi, definido como:

$$\varphi = \sqrt{\frac{\chi^2}{n}} \quad (\text{Ecuación 3.1})$$

Donde n es el total de observaciones. El coeficiente ϕ es mayor que 0 y aunque es mucho menor que el estadístico Chi-cuadrado, no está acotado superiormente y puede ser mayor que uno.

Coeficiente de Contingencia, que se define como:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (\text{Ecuación 3.2})$$

Este coeficiente toma valores entre 0 y 1, sin embargo difícilmente alcanza el valor 1. Su máximo valor depende del número de filas y columnas.

Coeficiente V de Cramer, definido como:

$$V = \sqrt{\frac{\chi^2}{\chi^2 + (k - 1)}} \quad (\text{Ecuación 3.3})$$

En donde k es el mínimo entre el número de filas y el número de columnas. Este coeficiente está acotado entre 0 y 1, y puede alcanzar ambas cotas, por lo tanto es el mejor de las medidas de asociación, por ser más fácil de interpretar.

3.2.2. Medida de correlación para variables cualitativas ordinales

Estas medidas de correlación ayuda a medir el grado de asociación entre variables con algún orden, se tiene las siguientes medidas:

Coeficiente de Spearman, definido como:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \quad (\text{Ecuación 3.4})$$

En donde d es la diferencia entre los rangos (X menos Y) y “n” es el número de datos. Este coeficiente es una medida de correlación entre dos variables, para calcular “ ρ ”, los datos son ordenados. La interpretación de coeficientes de Spearman es igual que la del coeficiente de correlación de Pearson, oscila entre -1 y +1, indicándonos asociaciones negativas o positivas respectivamente, el valor 0 significa no correlación, pero no independencia.

Coeficiente Gamma, es una medida de asociación simétrica entre dos variables ordinales cuyo valor siempre está comprendido entre -1 y 1. Los valores próximos a 1, en valor absoluto, indican una fuerte relación entre las dos variables, los valores próximos a cero indican que hay poca o ninguna relación entre las dos variables.

Coeficiente Kendall tau-c, medida no paramétrica de asociación para variables ordinales que ignora los empates, el signo del coeficiente indica la dirección de la relación y su valor absoluto indica la magnitud de la misma, de tal modo que los mayores valores absolutos indican relaciones más fuertes. Los valores posibles van de -1 a 1.

3.2.3. Medida de correlación para variables cuantitativas

Para estudiar la relación o comportamiento que puede existir entre dos variables cuantitativas, se utiliza la matriz de correlación, pero primero se va a explicar que es la covarianza, coeficiente simple de correlación lineal que son herramientas estadísticas que nos ayudan a explicar esta relación.

La covarianza entre dos variables: y_1 y y_2 , se denota por la siguiente fórmula:

$$\text{Cov}(y_1, y_2) = E[(y_1 - \mu_1)(y_2 - \mu_2)]$$

Desafortunadamente, el uso de la covarianza como una medida absoluta de dependencia es dificultoso porque su valor depende de la escala de medición.

Es por eso que a primera vista sea difícil saber si una covarianza es grande. Esta dificultad se puede eliminar estandarizando su valor por medio del coeficiente simple de correlación lineal. Su fórmula es:

$$\rho = \frac{\text{Cov}(y_1, y_2)}{\sigma_1 \sigma_2} \quad (\text{Ecuación 3.5})$$

Está relacionado a la covarianza y toma valores en el intervalo $-1 \leq \rho \leq 1$.

La matriz de correlación está conformada por n filas y por n columnas, además es una matriz simétrica. La matriz de correlación explica cómo se encuentran relacionadas cada una de las variables con otra variable. Su diagonal principal siempre contendrá el valor de 1. Si se encuentra un valor 0, indicará que no tiene ninguna relación con esa variable, por lo menos no lineal; es decir, pueda que tenga una relación cuadrática o de otro grado; si la correlación es positiva, esto indica que su proyección de la regresión lineal va a tender a crecer conjuntamente con la contra variable; si la correlación es negativa, esto indica que su proyección de la regresión lineal va a tender a decrecer conjuntamente con la contra variable.

3.2.4. Medida de correlación entre una variable cuantitativa y la otra cualitativa

El **coeficiente de correlación Eta**, sirve para cuantificar el grado de asociación existente entre una variable cuantitativa (medida en escala de intervalo o razón) y una variable categórica (medida en escala nominal u ordinal).

Su mayor utilidad no está precisamente asociada a las tablas de contingencia, pues estas se construyen, utilizando variables categóricas.

Se trata de un coeficiente de correlación que no supone linealidad y cuyo cuadrado puede interpretarse como la proporción de varianza de la variable cuantitativa que esta explicada por la variable categórica.

Verificación de los supuestos del análisis multivariante

La necesidad de comprobar los supuestos estadísticos aumenta en las aplicaciones multivariantes como consecuencia de dos características de este tipo de análisis. En primer lugar, la complejidad de las relaciones, debido al uso habitual de una gran cantidad de variables hace que las distorsiones y los sesgos potenciales sean más potentes cuando se incumplan los supuestos. En segundo lugar, la complejidad del análisis y de los resultados pueden enmascarar los “signos” de las violaciones de los supuestos que son aparentes en el análisis multivariante.

3.3. Modelo de Probabilidad Lineal

En el presente estudio el propósito es analizar la participación de un individuo en el mercado laboral, ya sea formal o informal, con el Modelo de Probabilidad Lineal (MPL) se puede representar de la siguiente manera:

$$Y_i = \beta X_i + \varepsilon_i$$

Donde Y_i , toma dos valores: ejemplo sector formal o sector informal y las variables que tienen las características que hace que la persona se incline por alguno de estos sectores será el conjunto de las variables X_i pero se tiene que la variable respuesta es dicotómica, por tanto se presentan algunos problemas.

3.3.1. Problemas en la estimación del Modelo de Probabilidad Lineal

Al aplicar el Método de Mínimos Cuadrados Ordinales (MCO), para estimar los parámetros del Modelo de Probabilidad Lineal, se presentan los siguientes problemas en la estimación (Gujarati Damodar N., 1995):

✓ **No normalidad de los errores**

Al igual que Y_i, ε , toma dos valores, es decir: $\varepsilon_i = Y_i - \beta X_i$

Entonces se tiene:

$\varepsilon_i = 1 - \beta X_i$, si $Y_i = 1$, lo cual ocurre con probabilidad P_i .

$\varepsilon_i = -\beta X_i$, si $Y_i = 0$, lo cual ocurre con probabilidad $(1 - P_i)$.

Por tanto ε_i sigue una distribución Binomial y no una distribución Normal. La violación del supuesto de normalidad no trae consecuencias sobre los estimadores puntuales de los parámetros, que continúan siendo lineales, insesgados y óptimos. El inconveniente que se presenta es con fines de “inferencia”, ya que los estadísticos usuales para efectuar contrastes de hipótesis no pueden ser utilizados⁶.

✓ **Heterocedasticidad de los errores**

Por tanto los errores ε_i están distribuidos de la siguiente forma:

ε_i	Probabilidad
$-\beta X_i$	$1 - P_i$
$1 - \beta X_i$	P_i

Entonces:

$$V(\varepsilon_i) = E(\varepsilon_i - E(\varepsilon_i))^2 = E(\varepsilon_i^2), \text{ dado que se supone que } E(\varepsilon_i) = 0,$$

Se obtiene:

$$V(\varepsilon_i) = E(\varepsilon_i^2) = (-\beta X_i)^2(1 - P_i) + (1 - \beta X_i)^2 P_i$$

$$V(\varepsilon_i) = \beta X_i(1 - \beta X_i) = P_i(1 - P_i)$$

⁶Además si el tamaño de la muestra aumenta indefinidamente, los estimadores de Mínimos Cuadrados Ordinales (MCO) generalmente tienden a estar normalmente distribuidos. Por tanto, para muestras grandes la metodología usual para la inferencia estadística del MPL sigue el procedimiento habitual bajo el supuesto de normalidad.

La varianza del término de error depende de X_i , generándose problemas de “heterocedasticidad”, provocando que los estimadores de MCO no cumplan con el supuesto de homocedasticidad de varianza (eficiencia)⁷.

✓ **No cumplimiento de** $0 \leq E(Y_i / X_i) \leq 1$

Luego de realizar la regresión, se va a dar la situación en general de que para determinar los valores de $E(Y_i / X_i) = \beta X_i$ puede estar fuera del intervalo $[0,1]$ lo cual es contradictorio con el hecho de que ella representa una probabilidad. Esto constituye una parte fundamental de la limitación del Modelo de Probabilidad Lineal.

A este problema se le pueden dar soluciones:

Se elimina estas observaciones (con la consecuente pérdida de grados de libertad). Se sustituyen valores menores o iguales que cero del Y_i estimado por 0,001 y los valores mayores o iguales que uno por 0,999.

Limitaciones del Modelo de Probabilidad Lineal

El Modelo de Probabilidad Lineal posee problemas a los que de alguna manera se le pueden dar soluciones, dependiendo entre otras cosas del tamaño y la calidad de la muestra con la que se cuenta. No obstante existe un problema de fondo que hace que el Modelo de Probabilidad Lineal no sea el más adecuado a la hora de analizar los modelos de respuesta dicotómica, entre ellos la “*distribución de normalidad de los errores*”.

Al presentarse limitaciones con el Modelo de Probabilidad Lineal, se opta por la aplicación de otro modelo, en el presente trabajo se hará mención al Modelo Logit y su extensión.

⁷Para una posible solución al problema de ineficiencia de los estimadores MCO, está la técnica de Mínimos Cuadrados Ponderados, en que las perturbaciones obtenidas son homocedásticas y de varianza igual a 1.

3.4. Modelos de opción cualitativa

Existen muchos casos en los que el fenómeno que se busca modelar involucra opciones cualitativas, por ejemplo en una elección se vota: “sí” o “no”. Los modelos más conocidos se pueden dividir en:

- **Opción Binaria**, se tiene al Modelo Logit, Modelo Probit.
- **Opción Múltiple**, se tiene las extensiones del Modelo Logit y Probit.

En los modelos de opción binaria, la variable dependiente puede tomar dos valores (sí o no, hombre o mujer, etc.) que se representará por códigos “0” y “1”. Es decir se busca encontrar una relación entre un conjunto de atributos que describen a un individuo y la probabilidad de que el individuo se incline por una opción determinada.

Para estos modelos se necesita una función de distribución acumulada⁸ que cumpla lo siguiente:

Los atributos de las variables X , que normalmente pueden tomar cualquier valor de la recta real, deben de acotarse al intervalo $(0, 1)$.

Incrementos en X deben estar asociados a incrementos o decrementos en la variable respuesta, para todos los valores de X .

La distribución de probabilidad puede representarse por:

$$P_i = F(\alpha + \beta x_i) = F(z_i) \quad (\text{Ecuación 3.6})$$

3.5. Modelo Logit

El Modelo Logit utiliza como función de estimación la función logística (Novales A., 2000), el resultado del modelo es la estimación de la probabilidad de que un nuevo individuo pertenezca a un grupo o al otro.

⁸Una función de distribución acumulada es aquella función de probabilidad que tiene como valor la probabilidad de que el valor observado de una variable x sea menor o igual a una X particular.

El Modelo Logit está basado en la función de probabilidad logística acumulada y está especificado por:

$$P_i = F(z_i) = F(\alpha + \beta x_i) = \frac{1}{1 + e^{-z_i}} = \frac{1}{1 + e^{-(\alpha + \beta x_i)}} \quad (\text{Ecuación 3.7})$$

P_i representa la probabilidad de que un individuo elija una de las dos opciones, dado X_i . Las funciones acumuladas de los Modelos Probit y Logit son parecidas. El Modelo Logit es más comúnmente usado que el Probit por razones computacionales.

Para mostrar cómo se puede estimar el modelo de la ecuación (3.7) se multiplica ambos miembros de la ecuación por $1 + e^{-z_i}$ para obtener:

$$(1 + e^{-z_i})P_i = 1 \quad (\text{Ecuación 3.8})$$

Dividiendo entre P_i y sustrayendo 1 queda:

$$e^{-z_i} = \frac{1}{P_i} - 1 = \frac{1 - P_i}{P_i} \quad (\text{Ecuación 3.9})$$

$$e^{z_i} = \frac{P_i}{1 - P_i} \quad (\text{Ecuación 3.10})$$

Aplicando logaritmo a ambos lados de la ecuación, se tiene:

$$z_i = \ln \left[\frac{P_i}{1 - P_i} \right] = \alpha + \beta x_i \quad (\text{Ecuación 3.11})$$

En esta regresión la variable dependiente es el logaritmo del cociente de probabilidades de que una opción en particular sea escogida. Una característica importante del Modelo Logit es que transforma el problema de predecir

probabilidades (dentro del intervalo 0,1) a un problema de predecir el cociente de probabilidades de la ocurrencia de un evento, a lo largo de la recta real.

La pendiente de la curva logística es mayor en $P=1/2$. Esto implica que cambios en las variables independientes tendrán su efecto máximo sobre la probabilidad de escoger una opción en particular en la mitad de la distribución.

Las pendientes pequeñas cerca de los extremos significan que son necesarios cambios grandes en X_i para obtener un cambio pequeño en la probabilidad.

En el caso de que P_i sea igual a 0 ó a 1, los cocientes de probabilidades, $P_i / (1 - P_i)$ serán iguales a 0, o a infinito; y el logaritmo de los cocientes de probabilidades estará indefinido. Así pues, la aplicación del método de los mínimos cuadrados a la ecuación (3.11) es claramente inapropiada.

Como todo modelo, la regresión sólo intenta reproducir y en algunos casos predecir el comportamiento de un fenómeno real. Los tres modos fundamentales de comportamiento (crecimiento exponencial, búsqueda de un objetivo y estabilidad) son causados por tres estructuras básicas de retroalimentación positiva, retroalimentación negativa y retroalimentación negativa con retardos. Otros patrones más complejos provienen de la interacción no lineal de estas estructuras.

El crecimiento es exponencial al principio y empieza a alentarse hasta que el sistema alcanza un nivel de equilibrio. La forma de la curva se parece a una "S" estirada. Ninguna cantidad física real puede crecer (o decrecer) para siempre, eventualmente, una o más restricciones detienen el crecimiento.

3.5.1. Tipología de la modelización Logit

Existen distintos tipos de modelos Logit en función de las características que presenten las alternativas que definen a la variable dependiente, que es la variable que va a medir el número de grupos existentes en el análisis. Así se tiene:

Logit dicotómico, se utiliza cuando el número de alternativas son dos y son excluyentes entre sí.

Logit de respuesta múltiple, se utiliza cuando el número de alternativas a modelar es superior a dos, entre ellas se encuentran los modelos: Logit Multinomial, el Logit Condicional para alternativas de respuesta no ordenadas, además del Logit con datos ordenados.

3.6. Modelo Logit Multinomial

Los modelos de elección múltiple son aquellos en los que es posible tener más de dos alternativas de respuesta. Normalmente, en este tipo de modelos, las variables explicativas incluidas en la parte determinista o estructural del modelo suelen ser de tres tipos:

Las variables explicativas que son características del individuo, que realiza la elección, por ejemplo, en la decisión de tener un hijo pueden influir sólo las características individuales de la persona.

Las variables explicativas que son características sólo de la elección, por ejemplo, en la elección del medio de transporte utilizado para ir al lugar de trabajo.

Las variables explicativas que recogen características tanto del individuo como de la elección a realizar. Por ejemplo, a la hora de comprar una vivienda puede influir tanto la situación socioeconómica de la persona (nivel de renta, número de hijos, etc.) como las características propias de la vivienda (localización, luminosidad, etc.).

El modelo Logit Multinomial o regresión Logística Multinomial (Greene William H., 2003), es una regresión que generaliza el Modelo Logit, al permitir más de dos resultados discretos. Es decir, utiliza para predecir las probabilidades de los posibles resultados diferentes de una manera categórica (variable dependiente), dado un conjunto de variables independientes (pueden ser variables de tipo cuantitativo o cualitativo). Estos modelos se clasifican en dos grandes grupos, según las alternativas que presenta la variable dependiente (variable con más de 2 categorías), se puedan ordenar (modelos con datos ordenados) o no se puedan ordenar (modelos con datos no ordenados).

El Logit con datos no ordenados que se utiliza cuando las alternativas que presenta la variable dependiente no indican ningún orden, puede a su vez presentar dos modalidades:

Logit Multinomial, se utiliza cuando las variables explicativas del modelo hacen referencia a las observaciones muestrales (individuos), por lo que varían entre observaciones pero no entre alternativas.

Logit Condicional, se utiliza cuando las variables explicativas del modelo hacen referencia a las alternativas, por lo que sus valores varían entre alternativas pudiendo hacerlo o no entre observaciones (individuos).

Por tanto el Modelo Logit Multinomial es apropiado en los casos en que la respuesta no presenta un orden, supone que los datos son casos específicos,

es decir, cada variable independiente tiene un valor único para cada caso. También asume que la variable dependiente no puede perfectamente predecir a partir de las variables independientes para cualquier caso. Al igual que con otros tipos de regresión, no hay la necesidad de que las variables sean independientes, sin embargo, la presencia de Multicolinealidad se supone que es relativamente bajo, como se hace difícil diferenciar entre el impacto de diversas variables si están altamente correlacionados, por tanto en el presente estudio la variable respuesta cuenta con tres categorías sin orden alguno, por tanto se aplica el Modelo Logit Multinomial.

El **Modelo Logit Multinomial** (Greene William H., 2003), se emplea al querer explicar a través de características sociodemográficas y económicas del individuo, bajo ciertas características un individuo optará por trabajar en algún sector de trabajo. Ejemplo cuando la variable respuesta Y_i tiene tres categorías, entonces se codifica a través de números: 1, 2 y 3 correspondiente a cada sector de trabajo y $X_1, X_2, X_3, X_4, X_5, X_6, X_7,$ y X_8 las variables independientes. El Modelo Logit Multinomial que explica los sectores es:

$$\text{Prob}(y_i = j) = \frac{e^{\beta'_j x_i}}{\sum_{k=1}^j e^{\beta'_k x_i}} \quad (\text{Ecuación 3.12})$$

Las ecuaciones estimadas proveen un conjunto de probabilidades para las $J + 1$ alternativas que pueda elegir una persona que haya de tomar una decisión y tenga X_i como características individuales. Antes de proseguir, se debe eliminar una indeterminación que presenta el modelo.

Se define $\beta_j^* = \beta_j + q$ para cualquier vector q , se obtiene exactamente las mismas probabilidades, puesto que todos los términos a los que q afecta se cancelan. Para resolver este problema, se puede normalizar el modelo tomando $\beta_0=0$. Por tanto, las probabilidades resultantes son:

$$\text{Prob}(Y = j) = \frac{e^{\beta'_j x_i}}{\sum_{k=1}^J e^{\beta'_k x_i}}, \text{ para } j = 1, 2, \dots, J \quad (\text{Ecuación 3.13})$$

Entonces para el trabajo se tiene:

Como $J=1,2$ y 3 y se tiene $X_1, X_2, X_3, X_4, X_5, X_6, X_7,$ y X_8

$$\text{Prob}(Y = 1) = \frac{e^{\beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \dots + \beta_{81}X_8}}{1 + e^{\beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \dots + \beta_{81}X_8} + e^{\beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \dots + \beta_{82}X_8}}$$

$$\text{Prob}(Y = 2) = \frac{e^{\beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \dots + \beta_{82}X_8}}{1 + e^{\beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \dots + \beta_{81}X_8} + e^{\beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \dots + \beta_{82}X_8}}$$

Y la probabilidad de elegir la alternativa de referencia se expresa como:

$$\text{Prob}(Y = 0) = \frac{1}{1 + \sum_{k=1}^J e^{\beta'_k x_i}} \quad (\text{Ecuación 3.14})$$

Pero para el trabajo se tiene la categoría base será la última, entonces se tiene:

$$\text{Prob}(Y = 3) = \frac{1}{1 + e^{\beta_{01} + \beta_{11}X_1 + \beta_{21}X_2 + \dots + \beta_{81}X_8} + e^{\beta_{02} + \beta_{12}X_1 + \beta_{22}X_2 + \dots + \beta_{82}X_8}}$$

En este modelo, los coeficientes son difíciles de interpretar, los parámetros β 's son específicos para cada categoría, es decir, por ejemplo si se tiene un caso con 4 opciones ($Y=4$), se obtendrá un conjunto de β 's para cada una de las opciones, en el caso de tener dos variables independientes, se tendría un conjunto de 3 β 's para cada una de las 4 categorías, en total, 12 β 's, de las cuales un conjunto (3 β 's) se anulan para efectos de la normalización.

Derivando la ecuación (3.14), se encuentra que los efectos marginales de los regresores sobre las probabilidades son:

$$\frac{\partial P_j}{\partial x_i} = P_j \left[\beta_j - \sum_k P_k \beta_k \right] \quad (\text{Ecuación 3.15})$$

Esto se puede calcular a partir de los parámetros estimados. El modelo implica que se puede calcular los cocientes:

$$\ln \left[\frac{P_{ij}}{P_{i0}} \right] = \beta_j' X_i \quad (\text{Ecuación 3.16})$$

Desde el punto de vista de la estimación es útil que el cociente de probabilidades no dependa de las otras opciones.

La estimación de los β del Modelo Logit Multinomial se realiza por Máxima Verosimilitud acudiendo a procedimientos iterativos de aproximación numérica, con el método Newton – Raphson, que converge rápidamente, ya que se trata de un método iterativo, empleado en diversos problemas matemáticos, como en la determinación de las raíces de las ecuaciones.

Como ejemplo, se extenderá el modelo al caso de tres opciones (Pindyck R. Rubinfeld D., 1991):

$$\left. \begin{aligned} \ln \left[\frac{P_{i2}}{P_{i1}} \right] &= \alpha_{21} \beta_{21} x_i, \\ \ln \left[\frac{P_{i3}}{P_{i1}} \right] &= \alpha_{31} \beta_{31} x_i, \\ \ln \left[\frac{P_{i3}}{P_{i2}} \right] &= \alpha_{32} \beta_{32} x_i \end{aligned} \right\} (\text{Ecuación 3.17})$$

P_{i1} indica la probabilidad de que el individuo i elija la opción “1”. En cada ecuación se asume que el logaritmo de los cocientes de probabilidades (o “razón de las ventajas” del inglés *Odds ratio*) de una elección relativa a la segunda es una función lineal del atributo X . La única ligadura entre las ecuaciones es que la suma de probabilidades debe ser la unidad.

Para resolver este sistema de ecuaciones, nótese que:

$$\ln \left[\frac{P_{i3}}{P_{i2}} \right] = \ln \left[\frac{P_{i3}}{P_{i1}} \right] + \ln \left[\frac{P_{i1}}{P_{i2}} \right] = \ln \left[\frac{P_{i3}}{P_{i1}} \right] - \ln \left[\frac{P_{i2}}{P_{i1}} \right] = (\alpha_{31} - \alpha_{21}) + (\beta_{31} - \beta_{21}) X_i \quad (\text{Ecuación 3.18})$$

Se tienen así dos ligaduras adicionales:

$$\left. \begin{aligned} \alpha_{32} &= \alpha_{31} - \alpha_{21} \\ \beta_{32} &= \beta_{31} - \beta_{21} \end{aligned} \right\} (\text{Ecuación 3.19})$$

Es más fácil visualizar el Modelo Logit Multinomial si se redefinen todos los parámetros como:

$$\left. \begin{aligned} \alpha_{21} &= \alpha_2 - \alpha_1, \alpha_{31} = \alpha_3 - \alpha_1, \alpha_{32} = \alpha_3 - \alpha_2 \\ \beta_{21} &= \beta_2 - \beta_1, \beta_{31} = \beta_3 - \beta_1, \beta_{32} = \beta_3 - \beta_2 \end{aligned} \right\} (\text{Ecuación 3.20})$$

De esta manera, el sistema de la ecuación (3.17) se puede reescribir así:

$$\left. \begin{aligned} \ln \left[\frac{P_{i2}}{P_{i1}} \right] &= (\alpha_2 - \alpha_1) + (\beta_2 - \beta_1) x_i \\ \ln \left[\frac{P_{i3}}{P_{i1}} \right] &= (\alpha_3 - \alpha_1) + (\beta_3 - \beta_1) x_i \\ \ln \left[\frac{P_{i3}}{P_{i2}} \right] &= (\alpha_3 - \alpha_2) + (\beta_3 - \beta_2) x_i \end{aligned} \right\} (\text{Ecuación 3.21})$$

Como los parámetros de la ecuación anterior se pueden calcular a partir de las dos primeras ecuaciones, no se necesita estimar la tercera ecuación. Si se supone que se tienen suficientes réplicas, se puede usar el Método de Mínimos Cuadrados. Si no es así, deberá usarse el Método de Máxima Verosimilitud⁹.

Una interpretación más clara del significado de los parámetros del modelo puede obtenerse a partir del cálculo de los Odds-Ratio.

⁹Para la demostración por el Método de Máxima Verosimilitud, ver Anexo 3.

3.6.1. Odds - Ratio

Al cociente entre probabilidades de que ocurra un hecho o que se elija la opción “1”, frente a la probabilidad de que no suceda o se elige la opción “0” se denomina Odds Ratio, puede expresarse como:

$$\frac{\text{Pr ob}_{(\text{evento})}}{\text{Pr ob}_{(\text{Noevento})}} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n} \quad (\text{Ecuación 3.22})$$

Los coeficientes estimados ($\beta_0, \beta_1, \dots, \beta_p$) son medidas de los cambios en el ratio de probabilidades, denominado odds ratio. Están expresados en logaritmos, por lo que es necesario re-transformarlos (tomando los valores del antilogaritmo), de tal forma que se evalúe más fácilmente su efecto sobre la probabilidad. Un coeficiente positivo aumenta la probabilidad, lo que significa que el odds incrementa, mientras que un valor negativo disminuye la probabilidad predicha, es decir el odds decrece. Un coeficiente 0 equivale a un valor de 1 lo que no produce cambios en el odds.

En el caso de que el valor de la Odds sea:

Mayor que 1, significa que la opción 1 es más probable que ocurra frente a que no ocurra.

Menor que 1, significa que la opción 1 tiene menor probabilidad que ocurra frente a la alternativa 0.

Es igual a 1, significa que la probabilidad de que ocurra la alternativa 1 es la misma que la de que no ocurra.

La interpretación del modelo puede efectuarse en términos de efectos marginales, dado que el efecto de cada variable explicativa sobre la probabilidad depende tanto del vector de variables explicativas como el conjunto de parámetros del modelo.

3.6.2. Efectos marginales

$$\text{Sea } Z_{ij} = \beta_{1j} + \beta_{2j}X_{2i} + \dots + \beta_{kj}X_{ki} = [1X_{2i} \dots X_{ki}] \begin{bmatrix} \beta_{1j} \\ \beta_{2j} \\ \cdot \\ \beta_{kj} \end{bmatrix} = \beta_j X_i \quad (\text{Ecuación 3.23})$$

En los modelos Multinomiales los parámetros deben interpretarse con mucho cuidado. En principio, se podría pensar que el signo del efecto marginal de una variable sobre la probabilidad de elegir una alternativa solo depende del correspondiente elemento del vector β_j . Sin embargo, se puede establecer que el efecto marginal de la variable X_h es igual a:

$$\frac{\partial P_{ij}}{\partial X_h} = P_{ij} [\beta_{hj} - \bar{\beta}_h] \quad (\text{Ecuación 3.24})$$

Donde $\bar{\beta}_h$ es la media de los parámetros β_{hj} para las alternativas J.

En general el efecto marginal se evalúa en la media de las variables continuas y para valores fijos arbitrarios de las variables dicotómicas.

El efecto marginal depende de P_j , que siempre es positivo, y de la diferencia entre β_{hj} y $\bar{\beta}_h$. Por tanto, el signo dependerá del coeficiente y de la alternativa que se esté analizando. Esta expresión sirve para evaluar los “efectos marginales” de las variables, una ventaja de examinar los efectos marginales es que a diferencia de los parámetros, estos no dependen de la categoría base elegida.

Como interpretación se tiene que, el signo del parámetro, muestra como un aumento en el regresor (variable independiente) afecta de manera positiva o negativa a las probabilidades explicadas.

Los efectos marginales para las variables binarias se calculan en forma diferente debido a que no es posible aplicar para ellas el concepto de derivada. Así, el efecto marginal sobre la probabilidad de cobertura cuando la variable binaria pasa de valer cero a uno, viene dado por:

$$E\Delta P_{ij} = P_{ij}(\beta_h + X\beta) - F(X\beta) \quad (\text{Ecuación 3.25})$$

Donde P es la función de distribución acumulada logística y X tiene un cero en la posición k-esima. Como el efecto marginal depende de los valores que toman cada una de las variables, se lo evalúa usualmente para los valores medios de las variables continuas y para valores fijados arbitrariamente de las variables dummies.

3.6.3. Métodos de selección de variables explicativas

Estadístico de Wald, permite saber cuáles de las variables en el modelo son importantes para determinar variaciones en la variable dependiente “Y” se debe calcular el estadístico de Wald, para lo cual se debe observar la escala de medición de la variable, ya que puede ocurrir dos situaciones: que sea categórica o no categórica.

Si la variable es categórica se tiene $Wald_i = \hat{\beta}_i^T \hat{\beta}_i^{-1} \hat{\beta}_i$, el cual sigue una distribución Chi-cuadrado, con grados de libertad igual al número de parámetros estimados. La importancia del estadístico de Wald radica en que a través de él se puede determinar la significancia del parámetro en el modelo y el peso que tiene, lo que se logra al establecer una sencilla prueba de “hipótesis” en la que se plantea en palabras lo siguiente:

$H_0: \beta_i = 0$. La variable X_i , no es importante para establecer variaciones en el modelo.

$H_1: \beta_i \neq 0$. La variable X_i , si es importante para establecer variaciones en el modelo.

Si se llega a rechazar la hipótesis nula, se observa a qué nivel de confianza se efectúa el rechazo, (comúnmente a nivel de confianza del 95 por ciento).

Si la variable no es categórica, se tiene $Wald_i = \frac{\hat{\beta}_i^2}{\hat{\sigma}_{\beta_i}^2}$, el cual tiene una distribución Chi- cuadrado con un grado de libertad.

Entrada hacia adelante para la selección de variables, este procedimiento “hacia adelante”, inicia el modelo sin términos, es decir por pasos, en cada paso se añade al modelo el término más significativo, hasta que ningún término quede fuera del modelo, cada variable de alguna manera tendrá una contribución estadísticamente significativa, además que estas puedan predecir a la variable dependiente.

3.6.4. Valoración de la bondad de ajuste del modelo estimado

Bondad de ajuste, la prueba de bondad de ajuste, presenta la hipótesis nula el modelo se ajusta adecuadamente a los datos. Por tanto si la hipótesis nula se cumple, **el estadístico Pearson y desviación** tienen una distribución Chi- cuadrado con grados de libertad. Si el valor de la significancia es pequeño menor a 0,05 el modelo no se ajusta adecuadamente a los datos, sin embargo si el valor de la significancia es mayor que 0,05 se acepta la hipótesis nula, además se denota que los supuestos del modelo son consistentes con los datos.

Medidas de bondad de ajuste, en este tipo de modelos no se emplea el R^2 para mostrar la bondad del ajuste, sino que se calcula el incremento de la

verosimilitud, aunque reciben el nombre de R^2 por lo tanto se llaman Pseudos R-cuadrado.

Pseudo R-cuadrado, entre ellas se encuentran los coeficientes de Cox y Snell, R cuadrado de Nagelkerke y McFadden, si el valor se acerca a 1 (V. Pando Fernandez & R. San Martin Fernandez) se dice que existe una relación alta entre la variable “Y” las variables “X”. Aunque estos coeficientes tratan de medir la variabilidad explicada, en general, van a ser muchos más bajos que en la regresión. Entonces se tiene:

$$\text{PseudoR}^2 = 1 - \frac{\log L(\text{completo})}{\log L(\text{reducido})} \quad (\text{Ecuación 3.26})$$

Dónde:

Log L (modelo completo), función de verosimilitud incluyendo todas las X.

Log L (modelo reducido), función de verosimilitud incluyendo solo β_0 .

3.6.5. Medidas de confiabilidad del Modelo

Tabla de clasificación

También llamada matriz de confusión, es una forma sencilla de evaluar el ajuste del modelo de regresión logística, no es tan objetiva pero se usa como indicador de bondad de ajuste.

Se interpreta mediante el porcentaje de objetos o individuos bien clasificados, esto es, aquellos que mediante la probabilidad estimada permanecen en su respectiva categoría, también se interpreta mediante el porcentaje de objetos mal clasificados, esto es, aquellos que mediante la probabilidad estimada se asignan a categorías diferentes del cual fueron observados. Es decir la suma de la diagonal principal es el porcentaje de objetos o individuos bien clasificados mediante el modelo de regresión logística estimado, por tanto lo que se debe

esperar es que este porcentaje sea lo más alto posible, a fin de concluir que el modelo obtenido clasifica bien a los objetos o individuos.

Contrastación de hipótesis

Si el modelo se ha estimado por Máxima Verosimilitud, para contrastar un conjunto de restricciones puede compararse el valor de la función de verosimilitud en el estimador restringido con el valor de dicha función de verosimilitud en el valor obtenido, ignorado las restricciones. El contraste de razón de verosimilitud se basa en el estadístico:

$$-2 \ln \frac{L(\hat{\beta}_R)}{L(\hat{\beta}_{MV})} = -2(L(\hat{\beta}_R) - \ln L(\hat{\beta}_{MV}))$$

Sigue una distribución Chi-cuadrado con q-grados de libertad (el número de restricciones).

En particular, puesto que la significación global del modelo es siempre una hipótesis de interés, debe calcularse siempre el valor L_0 de la función de verosimilitud que se obtiene cuando todos los coeficientes del modelo se hacen iguales a cero, excepto el término independiente, que se estima.

Entonces se tiene:

$$\ln L_0 = n[p \ln p + (1-p)\ln(1-p)]$$

Donde p es la proporción muestral de observaciones con $Y_i=1$. El estadístico $-2(\ln L_0 - \ln L(\hat{\beta}_{MV}))$ se distribuye χ_{k-1}^2 y se puede utilizar como contraste de significación global del modelo.

La función de verosimilitud puede utilizarse asimismo para obtener un estadístico que tiene cierta semejanza con el coeficiente de determinación R^2 :

$$\tilde{R}^2 = 1 - \frac{\ln L_0}{\ln L(\hat{\beta}_{MV})}$$

Donde L_0 es el estadístico antes introducido, y $L(\hat{\beta}_{MV})$ denota el valor maximizado de la función de verosimilitud.

Si los coeficientes β son pocos significativos, la capacidad explicativa del modelo será muy reducida, $L(\hat{\beta}_{MV})$ será similar en valor a L_0 y \tilde{R}^2 próximo a cero, al igual que ocurre con el coeficiente de determinación en el modelo lineal. Por el contrario, cuando mayor sea la capacidad explicativa del modelo, mayor será $L(\hat{\beta}_{MV})$, siempre en exceso de L_0 , y más próximos estará \tilde{R}^2 a 1. Sin embargo, ni llega a valer 1 ni es fácil de interpretar los valores de \tilde{R}^2 entre 0 y 1. Para hipótesis sobre el valor de un coeficiente, puede utilizarse el estadístico de Wald.

3.6.6. Consideraciones en el uso del Modelo Logit Multinomial

Para la utilización de algún método multivariante, según el tipo de método, los datos deben cumplir algunos supuestos de normalidad multivariante. Como el presente trabajo se estudia a variables de tipo cualitativo existen métodos multivariantes como ser el análisis discriminante el cual descansa sobre el cumplimiento estricto de los supuestos de normalidad multivariante y la igualdad de matrices de varianzas y covarianzas entre los grupos, supuestos que no siempre se verifican. En cambio la regresión logística no se enfrenta a estos supuestos tan estrictos, y es mucho más robusta cuando estos supuestos no se cumplen, haciendo muy apropiada su aplicación en muchas situaciones, ya que cuentan con contrastes estadísticos directos, capacidad para incorporar efectos no lineales y permitir una amplia variedad de diagnósticos.

Una de las ventajas del Modelo Logit Multinomial, es que permite el uso de múltiples variables (con características continuas, categóricas nominales, categorías ordinales, etc.). Teniendo en cuenta siempre que la variable dependiente debe ser categórica.

Pero el modelo Logit Multinomial, cuenta con un supuesto importante, el cual debe comprobarse, supuesto conocido como de Independencias Alternativas Irrelevantes (IAI).

Independencia de Alternativas Irrelevante (IAI)

Se refiere a que cada una de las alternativas posibles esta netamente diferenciada y valorada de manera independiente por el que toma la decisión.

Una de las principales desventajas de esta clase de modelos es que se ve afectado por lo que se conoce como el supuesto de Independencia de Alternativas Irrelevantes (IAI), si divido una categoría ya existente en dos muy parecidas, debería esperarse que ambas se repartieran la probabilidad de ser escogida que antes tenía la que ya estaba presente, mientras que el resto de alternativas mantuvieran la misma probabilidad de ser elegidas. No obstante, y de acuerdo con la propiedad de IAI, el modelo Logit Multinomial reasigna las probabilidades de ocurrencia entre el total de categorías existentes, incluyendo la nueva. Por lo mismo, no es apropiado cuando se sabe que se tienen categorías que son sustitutas cercanas (Greene William H., 2003).

De acuerdo con el supuesto de IAI, la aplicación del modelo Multinomial no ordenado logístico supone que el ratio de probabilidades entre dos alternativas no depende de las demás categorías. Basados en este resultado, es evidente que la presencia de una categoría base permite una interpretación directa del signo y magnitud de determinado coeficiente como el efecto que tiene el regresar en cuestión sobre la probabilidad de elegir la j-ésima alternativa respecto a la categoría base. Para esto, basta recordar que los coeficientes de la categoría base han sido normalizados en cero. Mientras las probabilidades predichas serán obtenidas con la categoría base (b), cambiar la base de la categoría podría confundir a algunos, dado que los resultados de los parámetros tienden a ser algo diferentes. Solo habría un cambio en la

parametrización más no en la estimación de las probabilidades predichas, dado que estas serán las mismas.

Test de Independencia de las Alternativas Irrelevantes

En el modelo de regresión Logística Multinomial, las razones de ocurrencia de cada par de valores de la variable dependiente no deben ser afectadas por el resto de alternativas posibles (añadir o eliminar alternativas, no debe modificar el cociente). Se trata del supuesto de Independencia de Alternativas Irrelevantes, que requiere que cada una de las posibles alternativas sea valorada de manera independiente por el que toma la decisión. Si no se cumple este supuesto el modelo Multinomial no tiene lugar pues da lugar a coeficientes incorrectos. Para verificar si la inclusión de determinada categoría afecta la consistencia de nuestros estimados (y con esto los ratios de probabilidad), es posible utilizar la **prueba de Hausman**.

El modelo condicional también tiene como supuesto la Independencia de Alternativas Irrelevantes, mostramos este supuesto en términos del modelo Logit Multinomial¹⁰ (Greene William H., 2003).

Donde el Odds no depende de otras categorías que sean viables. En este sentido, estas categorías alternativas son irrelevantes, lo que significa que al añadir o borrar una categoría esta no afectara la cantidad de Odds en las categorías principales.

$$\frac{\Pr(Y = m/X)}{\Pr(Y = n/X)} = \exp(X(\beta_{m/b} - \beta_{n/b})) \quad (\text{Ecuación 3.27})$$

¹⁰Es habitual en investigación empírica la utilización del modelo Probit Multinomial cuando se vulnera el supuesto de alternativas irrelevantes (En Stata el modelo Probit Multinomial se calcula con el comando mprobit pero cuando se incumple la condición mencionada se puede usar el comando asmprobit. El modelo Probit asumen que los errores son normales por lo tanto los errores de las distintas alternativas están correlacionados y por lo tanto se supone que el modelo Probit no se ve afectado por la vulneración del supuesto. Sin embargo autores como Long y Fresse (2006) argumenta que el modelo Probit Multinomial adolece de los mismos problemas que presenta el modelo Logit Multinomial.

McFadden (1986) mostró que hay tres formas de tratar con la propiedad IAI. Primero, mediante test estadísticos que pueden ser desarrollados para probar si la IAI se mantiene en una aplicación particular, cuando estos test muestran que la IAI se mantiene entonces se pueden emplear los modelos. Segundo, cuando el valor de la escala para una alternativa no depende de las características del conjunto de elección, en este caso la IAI puede ser evitada en el modelo pero a cambio de perder la simplicidad en la estructura del modelo. Tercero, cuando la propiedad de IAI no se cumple en la modelación, entonces se usan otros modelos que evitan esta propiedad pero cuyo costo será la complejidad computacional.

Capítulo IV.

Análisis de los sectores de trabajo

Previamente se realiza el análisis descriptivo de los sectores de trabajo, en especial del sector informal, para luego mostrar por medio de un ejercicio ilustrativo la aplicación y el análisis del Modelo Logit Multinomial en el área de empleo en el sector informal.

4.1. Instrumento de información

El instrumento que recolecta la información para el análisis, es cuestionario, el cual recoge información sobre características demográficas, empleo, educación, entre otras secciones. A continuación se muestra una breve explicación del **Diseño Muestral de la EH-2014** (INE, 2014).

Fuente de información utilizada

Para cumplir con los objetivos de la investigación se tomó en consideración la información de la base de la Encuesta de Hogares 2014 (Instituto Nacional de Estadística), aquí se menciona, algunas de sus características, que son de importancia relevante para la investigación y que permiten comprobar algunas ideas referentes a las características y a los factores determinantes de los sectores de trabajo y como estos benefician o afectan, los criterios de utilización del Modelo Logit Multinomial. Para dicho propósito se analizan variables relacionadas con los sectores de trabajo, en sus aspectos sociodemográficos y económicos.

Se emplea para la Encuesta de Hogares 2014 el marco muestral (INE, 2014) construido en base a la información del Censo Nacional de Población y Vivienda 2012.

Componentes básicos

Población objetivo: la unidad de observación es el **hogar** en la Encuesta de Hogares 2014, el cuestionario de la encuesta explora al objeto persona, delimitándolo por grupos de edad que se ubican en sus diferentes secciones.

Cobertura geográfica: el ámbito geográfico de la Encuesta de Hogares 2014 es el área urbana y rural de todos los departamentos del país: Chuquisaca, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Beni y Pando.

Universo de estudio

La investigación está dirigida al conjunto de hogares establecidos en viviendas particulares ocupadas de las ciudades capitales, resto urbano y área rural de Bolivia, en el periodo definido para la ejecución de la encuesta en la gestión 2014.

Se excluye de la investigación a la población residente en viviendas colectivas tales como hospitales, hoteles, cárceles, recintos militares, hogares de ancianos, unidades educativas, etc.; se incluyen a hogares que, forman un grupo independiente y además residen dentro de estos establecimientos, ejemplo: porteros, cuidadores, etc.

Unidades de observación, análisis y muestreo

La unidad de observación es el Hogar.

Las unidades de análisis son: el hogar como unidad de consumo; los miembros de hogar en cuanto a características sociodemográficas, educación, salud y ocupación y la vivienda en cuanto a características de tamaño, servicios y condiciones físicas.

La unidad de muestreo en su última etapa, es la vivienda particular ocupada, ésta tiene permanencia fija en el tiempo y espacio, característica que la habilita para ser utilizada como unidad de selección en el diseño muestral. La unidad de muestreo, de acuerdo a las áreas de estudio, se conforma de la siguiente manera:

Área Amanzanada

UPM - La unidad primaria de muestreo es el conglomerado de viviendas.

USM - La unidad secundaria de muestreo conforman las viviendas particulares dentro de las UPM seleccionadas.

Área dispersa

UPM.- La unidad primaria de muestreo es el conglomerado.

USM.- Es la comunidad dentro de la UPM seleccionada.

UTM.- La unidad terciaria de muestreo conforman las viviendas particulares dentro de la comunidad seleccionada.

Dominios de estudio

Se garantiza generar estimaciones de indicadores de pobreza, con errores muestrales admisibles, para los siguientes niveles de desagregación:

Nacional

Nacional Urbano

Nacional Rural

Departamental, excepto Beni y Pando, de los cuales se darán estimaciones de manera conjunta.

Para los indicadores de ocupación, los resultados podrán ser desagregados a nivel Nacional y ciudades capitales de los departamentos del eje central (La Paz, Cochabamba y Santa Cruz), y por último la ciudad de El Alto.

Realizar desagregaciones mayores de los resultados estará sujeto a la valoración de los errores muestrales.

Distribución de la muestra

El **cuadro 1** presenta la distribución de la muestra por departamento y área geográfica, de la Encuesta de Hogares 2014:

Cuadro 1. DISTRIBUCIÓN DE LA MUESTRA DE VIVIENDAS POR DEPARTAMENTO Y ÁREA GEOGRÁFICA, ENCUESTA DE HOGARES 2014

Departamento	Número de Viviendas de la muestra			
	Total	Ciudad	Resto Urbano	Rural
BOLIVIA	10.404	6.636	1.392	2.376
Chuquisaca	588	228	48	312
La Paz	1.776	1.272	96	408
El Alto	1.416	1.416	0	0
Cochabamba	1.860	1.392	204	264
Oruro	492	240	48	204
Potosí	468	84	96	288
Tarija	804	300	216	288
Santa Cruz	2.112	1.380	444	288
Beni	420	108	204	108
Pando	468	216	36	216

Fuente: Elaboración Propia en Base a la EH

Definición de las variables

Las características de los trabajadores que influyen en la probabilidad de estar ocupados en el sector formal o informal, estudios sobre el tema de informalidad (Goldberg P. y Nina P., 2003), sugieren incluir las variables edad, sexo, nivel educativo, condición de pareja y otras variables. A continuación se describen las variables (INE, 2014) incluidas en el modelo:

Variables demográficas

Edad, hace referencia a los años cumplidos que tienen la persona hasta el momento de la entrevista.

Sexo, hace referencia a lo biológico o a las características anatómicas sexuales diferentes para hombres y mujeres.

Estado civil, situación de cada persona en relación con las leyes o costumbres conyugales del país. Se releva la situación conyugal de hecho y de derecho o legal de los integrantes del hogar de doce años o más, entre las opciones de respuesta esta: soltera (o), casada (o), conviviente o concubina (o), separada (o), divorciada (o) y viuda (o).

Identificación étnica, con ésta variable se desea saber si en el hogar existe alguna persona de doce años o más, que declara pertenecer a un pueblo originario o indígena bajo un sentido de identidad. Se denomina naciones o pueblos indígenas originarios campesinos a aquel conjunto de personas que descenden de poblaciones asentadas con anterioridad a la conquista y que se encuentran dentro de las actuales fronteras del territorio boliviano.

Variable relacionada a la educación

Años de estudio, hace referencia a la población de acuerdo a sus años de estudio.

Variables indicativas de la condición laboral

Población ocupada, se define a una persona ocupada de siete años y más de edad, que trabajaron por lo menos una hora la semana anterior a la encuesta, o que no trabajaron por estar de vacaciones, por enfermedad o accidente, huelga paro o conflicto laboral; interrupción del trabajo por falta de materiales o clientes, inclemencias del tiempo o suspensión o problemas personales o familiares, pero que tienen empleo al que seguramente volverán. A continuación se describen algunas variables:

Ingreso laboral, hace referencia al ingreso que percibe la persona ocupada.

Horas de trabajo, hace referencia a las horas promedio trabajadas de la persona que tuvo en el día.

Aporte a una AFP, se refiere a la persona ocupada que está afiliada a una Administradora de Fondos de Pensiones (AFP).

4.1.1. Identificación de las variables del modelo

Características de los trabajadores

La encuesta de hogares analiza las características personales que influyen en la probabilidad de trabajar en algún sector de trabajo.

Algunos estudios sugiere que la probabilidad de un trabajador pertenezca al sector informal depende del **sexo** del individuo, y se esperaría que las mujeres sean menos propensas a estar en este sector, en el **cuadro 3** se tiene que la participación femenina en el sector informal ha tendido a aumentar.

La **edad en años** cumplidos de los trabajadores, la literatura sugiere que la informalidad representa la puerta de entrada y de salida del mercado de trabajo, por lo que se espera que las personas más jóvenes y los que tienen edad de jubilación tengan mayor probabilidad de obtener un trabajo en el sector informal, esto se puede corroborar en el **cuadro 3**, e incluso evidencia que un mayor porcentaje de trabajadores adultos trabajan en el sector informal. Por esta razón se espera que el parámetro estimado asociado a esta variable sea positivo: a mayor edad, mayor probabilidad de trabajar en el sector informal.

La **educación**, esta variable se incluye en el modelo como el número de años aprobados de educación, se espera que cuanto mayor sea el nivel educativo,

entonces se reduce la probabilidad de obtener un empleo en el sector informal, de acuerdo con los datos obtenidos en el **cuadro 3**, en el sector informal se presenta mayor porcentaje de trabajadores con menores niveles de escolaridad, por lo que se espera que el parámetro estimado asociado a esta variable tenga signo negativo.

Según algunos documentos es más probable que las **personas con pareja** se desarrollen en actividades del sector informal, pues requieren un trabajo que les permita enfrentar con certeza las obligaciones para con su hogar, esto se puede corroborar con el **cuadro 3**, donde se muestra que en mayor porcentaje de trabajadores del sector informal están casados o conviviendo, es decir ya tienen su hogar establecido.

La **declaración** de una persona ocupada que declara **pertenecer a un pueblo originario o indígena** según el **cuadro 3** se presenta en mayor porcentaje trabajando en el sector informal, por lo que se espera el parámetro estimado asociado a esta variable presente signo positivo en el modelo Logit Multinomial.

Un indicador que permita analizar el sector de trabajo, son las **horas de trabajo**, según con la información del **cuadro 3** se registran mayor horas de trabajo en el sector informal, es decir se espera que el parámetro estimado asociado a esta variable presente un signo positivo en el modelo, en otras palabras a mayor horas de trabajo mayor probabilidad de pertenecer al sector informal.

En la **tabla 2** se observa un resumen de las variables de estudio para el modelo, dentro de las mismas se encuentran variables de tipo cuantitativo como tipo cualitativo, estas últimas son codificadas en variables dummies, ya que el modelo requiere que las variables sean dicotómicas. La variable dependiente

tiene más de dos categorías que son: sector formal, sector informal y sector doméstico.

Tabla 2. OPERACIONALIZACIÓN DE LAS VARIABLES INCLUIDAS COMO POSIBLES PREDICTORAS

Nombre de variable	Variable Dependiente	Alternativas
Informalidad	Sectores de trabajo	1 Formal 2 Informal 3 Doméstico
Nombre de variable	Variable Independiente	Alternativas
s2a_03	Edad	Variable cuantitativa
e	Años de estudio	Variable cuantitativa
ylab	Ingreso laboral mensual	Variable cuantitativa
s6b_23h	Horas promedio de trabajo al día	Variable cuantitativa
sexo	Sexo	0 Mujeres 1 Hombres
est_civil	Condición de pareja	0 Sin Pareja ¹ 1 Con pareja ²
autoident_etn	Identificación étnica	0 No se identifica con ningún pueblo originario o indígena 1 Se identifica con algún pueblo originario o indígena.
s5_58b	Aporte a una AFP	0 Aporta a una AFP 1 No aporta a una AFP

Fuente: Elaboración Propia en Base a la EH-2014

¹ Sin pareja agrupa a las personas solteras(os), viudas(os) y separadas(os)

² Con pareja agrupa a las personas casadas(os) y convivientes/concubina(o).

4.2. Análisis descriptivo de los sectores de trabajo

A continuación se evalúan los resultados de la Encuesta de Hogares 2014, de la población ocupada en los distintos sectores de trabajo, adicionalmente se realiza una desagregación por área geográfica.

Cuadro 2. BOLIVIA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACION OCUPADA POR SECTORES DE TRABAJO, SEGÚN ÁREA, ENCUESTA DE HOGARES 2014

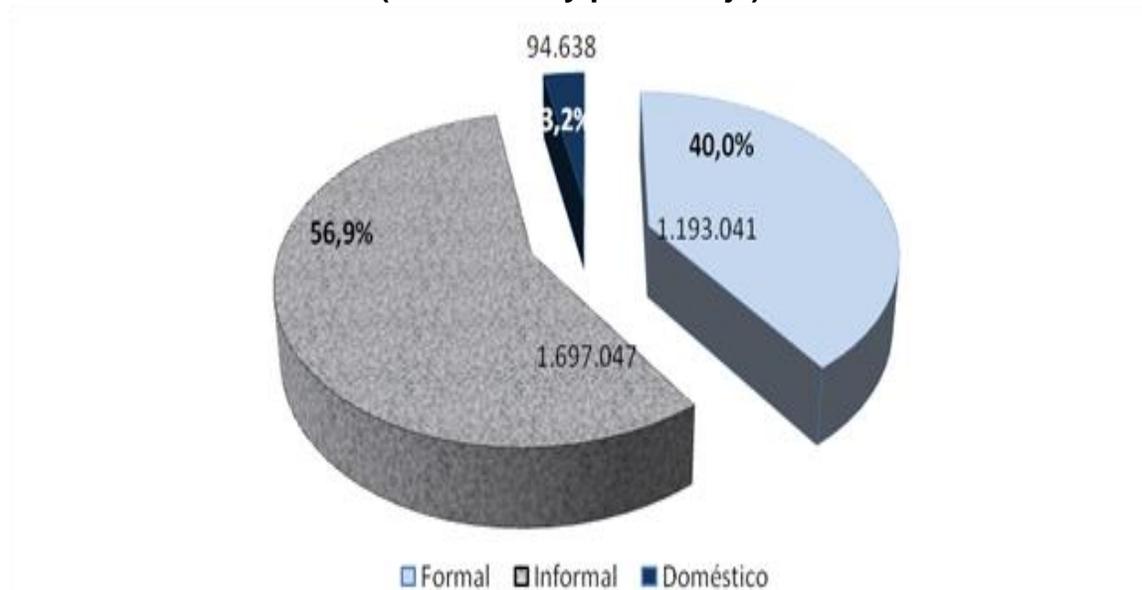
ÁREA	SECTORES DE TRABAJO		
	Formal	Informal	Doméstico
BOLIVIA	33,9	63,6	2,5
Urbana	40,0	56,9	3,2
Rural	18,4	80,9	0,7

Fuente: Elaboración Propia en Base a la EH

Los resultados del **cuadro 2** muestran a la población ocupada a nivel Bolivia que el 63,6 por ciento de los ocupados pertenece al sector informal; el 33,9 por ciento al sector formal y 2,5 por ciento al sector doméstico.

En cuanto a la comparación por área geográfica, según el **gráfico 1**, la población ocupada en el sector informal en el área urbana, para el 2014 fue de 1.697.047 personas, esto representa el 56,9 por ciento; mientras para el sector formal fue de 1.193.041 personas lo que representa al 40 por ciento y por último en el sector doméstico fue de 94.638 personas lo que significa que el 3,2 por ciento pertenece al sector doméstico¹¹.

Gráfico 1. BOLIVIA URBANA: POBLACIÓN OCUPADA POR SECTORES DE TRABAJO, ENCUESTA DE HOGARES 2014
(En número y porcentaje)



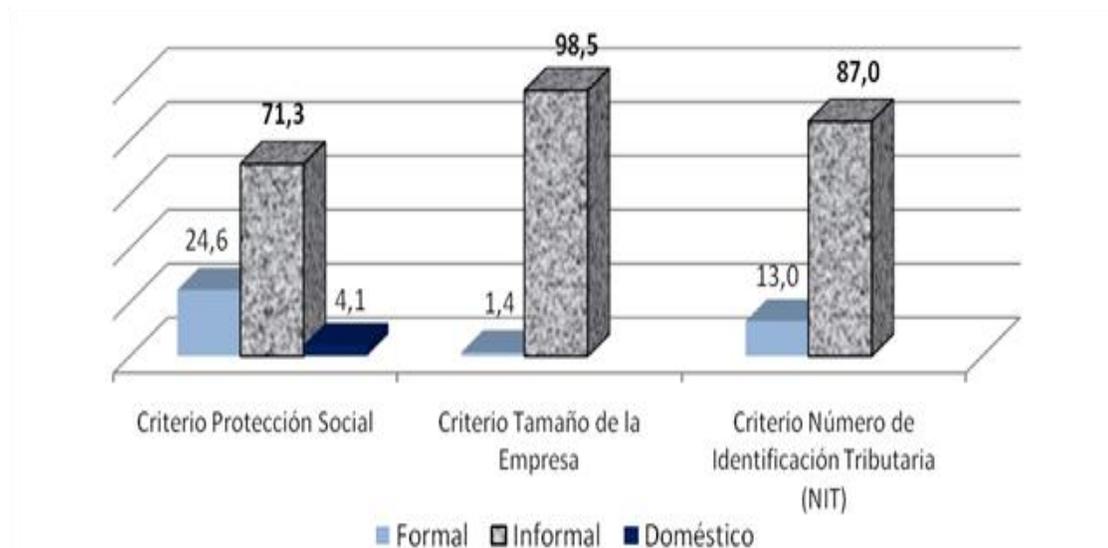
Fuente: Elaboración Propia en Base a la EH

A continuación se muestra los cuadros y gráficos por sectores de trabajo de la población ocupada en el área urbana de Bolivia.

¹¹De la misma forma, se puede concluir que en el área rural se tiene que un 80,9 por ciento de los ocupados pertenece al sector informal; el 18,4 por ciento son del sector formal y el 0,7 por ciento al doméstico. Es necesario mencionar que por las características del área rural, el concepto de informalidad es limitado y no completamente útil.

De acuerdo a las dos definiciones básicas de informalidad, la primera, basada en la protección social del trabajador (“legalista”) y la segunda en función al tamaño de la empresa (“productiva”). Para el caso boliviano, según el **gráfico 2**, en particular aquel referido al aporte a una Administradora de Fondo de Pensiones (AFP), los datos calculados, confirman la existencia de un grado muy reducido de protección social dentro de la población ocupada en Bolivia, en base a este criterio el 71,3 por ciento de la población ocupada en Bolivia podría ser clasificada como informal. Si se refiere a la posesión de un Número de Identificación Tributaria (NIT), que establece una relación formal entre trabajadores o empresas y el estado, con este criterio, los niveles de informalidad aumentan, obteniendo que el 87 por ciento de la población ocupada es informal, es decir ocupados que carecen de identificación tributaria en el lugar donde trabajan. Y en cuanto al criterio de tamaño de la empresa el 98,5 por ciento de la población ocupada es informal, lo cual quiere decir que un alto porcentaje de ocupados trabajan en empresas menores a cinco personas.

Gráfico 2. BOLIVIA URBANA: DISTRIBUCION PORCENTUAL DE LA POBLACIÓN OCUPADA EN LOS DISTINTOS SECTORES DE TRABAJO, SEGÚN CRITERIOS DE MEDICIÓN, ENCUESTA DE HOGARES 2014



Fuente: Elaboración Propia en Base a la EH

Las características demográficas, económicas y de educación, de la población sujeta a investigación son necesarias, puesto que proporcionan información importante respecto del empleo en los distintos sectores de trabajo.

Cuadro 3. BOLIVIA URBANA: DISTRIBUCIÓN DE LA POBLACIÓN OCUPADA POR SECTORES DE TRABAJO, SEGÚN ALGUNAS CARACTERÍSTICAS, ENCUESTA DE HOGARES 2014
(En número y porcentaje)

CARACTERÍSTICAS	SECTORES DE TRABAJO		
	Formal	Informal	Doméstico
Promedio de edad	36	40	31
Promedio de años de estudio	13	10	8
Promedio de horas trabajadas por día	8	9	8
Promedio de ingreso laboral mensual	3.723	3.017	1.474
Sexo			
Mujeres	35,1	57,4	7,5
Hombres	43,5	56,5	0,1*
Estado civil			
Solteros	47,0	48,8	4,2
Casados	39,5	58,4	2,0
Convivientes	35,6	61,0	3,4
Separados	36,5	57,5	6,0
Auto identificación étnica			
Si pertenece a algún pueblo originario o indígena	32,7	64,7	2,7
No pertenece a algún pueblo originario o indígena	42,5	54,2	3,3
Aporte a la AFP			
Si aporta a una AFP	87,6	12,1	0,3*
No aporta a una AFP	24,6	71,3	4,1
Lugar de trabajo			
Otro lugar de Trabajo ¹	4,3	95,7	
Local o Terreno exclusivo	25,7	74,3	

Fuente: Elaboración Propia en Base a la EH

¹Otros lugares de trabajo se refiere a: vivienda particular, puesto móvil, en quiosco o puesto fijo, vehículo, servicio a domicilio, ambulante, ambulante de preventa y otro.

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

Cuando se analiza las características que distinguen a los trabajadores, se observa el **cuadro 3** que para el año 2014, las mujeres representan ligeramente un mayor porcentaje de estar ocupadas en el sector informal en comparación a los hombres.

Es decir la edad promedio de los trabajadores en el sector informal es de 40 años, este dato es mayor en relación al sector formal; por otro lado los años promedio de estudio de la población ocupada en el sector informal y doméstico es menor respecto a los años de estudio de la población ocupada en el sector formal; las horas promedio de trabajo de los trabajadores del sector informal es ligeramente mayor que de las personas que trabajan en el sector formal y doméstico; si se analiza el ingreso mensual de los trabajadores del sector informal este es claramente menor en comparación al ingreso de los trabajadores del sector formal.

En cuanto al estado civil, la población ocupada en el sector informal presenta un porcentaje mayor de solteros, casados, convivientes y separados, que de los demás sectores de trabajo.

La distinción por condición étnica, muestra que todas aquellas personas que declaran pertenecer a algún pueblo originario o indígena, estuvieron ocupadas en actividades del sector informal el cual representa el 64,7 por ciento. Si se analiza el aporte a una AFP, claramente se observa que el mayor porcentaje de personas que realizan el aporte a una AFP se encuentra en el sector formal, lo que representa 87,6 por ciento, en cambio en el sector informal ocurre lo contrario el 71,3 por ciento no aportan a una AFP. Las personas ocupadas en el sector informal trabajan en lugares que no corresponden a un terreno exclusivo esto representa el 95,7 por ciento, lo que corresponde a personas que trabajan en la calle, vehículo, en su vivienda particular y otros lugares.

El **cuadro 4**, muestra que el sector informal se concentra el mayor porcentaje de personas que tiene una edad mayor a 44 años, en cuanto al nivel de educación las personas ocupadas en este sector tienen un nivel de educación primario esto representa el 74,7 por ciento, en cambio en el sector formal las personas ocupadas tienen un nivel superior representado por 64,7 por ciento. Al

observar a la población ocupada de los departamentos del eje central, Cochabamba obtiene el mayor porcentaje de personas ocupadas en el sector informal, seguido de La Paz y por último Santa Cruz, contrariamente ocurre en el sector formal, ya que el departamento de Santa Cruz es donde se concentra mayor población en este sector, seguido de La Paz y por último Cochabamba.

Cuadro 4. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA POR SECTORES DE TRABAJO, SEGÚN GRUPOS DE EDAD, NIVEL DE EDUCACIÓN Y DEPARTAMENTOS DEL EJE CENTRAL, ENCUESTA DE HOGARES 2014

CARACTERÍSTICAS	SECTORES DE TRABAJO		
	Formal	Informal	Doméstico
Grupos de edad			
Menor a 25 años	44,4	50,6	5,0
Entre 25 y 44 años	43,5	53,6	3,0
Mayor a 44 años	31,5	66,1	2,4
Nivel de educación			
Primaria	19,8	74,7	5,5
Secundaria	30,9	64,8	4,2
Superior	64,7	34,8	0,5*
Departamentos del eje central			
La Paz	37,6	60,5	1,9
Cochabamba	36,6	61,0	2,3
Santa Cruz	42,0	53,8	4,3

Fuente: Elaboración Propia en Base a la EH

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

Según las características de la empresa (**cuadro 5**), se tiene la siguiente información: en el sector informal el 93,7 por ciento de las personas ocupadas respondió que en la empresa donde trabajan no cuentan con seguro de salud y tan solo el 6,3 por ciento respondió que sí tienen seguro de salud; por el contrario esto no ocurre en el sector formal, ya que el 51,1 por ciento de las personas que trabajan en este sector respondió que la empresa donde trabajan si cuentan con el seguro de salud y tan solo el 48,9 por ciento respondió que la empresa donde trabajan no cuentan con el seguro de salud.

Otra característica de la empresa es, si esta cuenta con NIT, donde se tiene que: el 77,4 por ciento de las personas respondió que la empresa, negocio o

taller donde trabajan no cuenta con NIT que se encuentran en proceso, mientras el 13,4 por ciento respondió que la empresa lleva NIT (ya sea en régimen general o régimen simplificado) esto ocurre en el sector informal. Por otro parte en el sector formal el 30 por ciento de los ocupados respondió que la empresa donde trabajan si cuenta con NIT y el 45,6 por ciento respondió que no tiene NIT que se encuentra en proceso.

Cuadro 5. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR FORMAL E INFORMAL, SEGÚN CARACTERÍSTICAS DE LA EMPRESA, ENCUESTA DE HOGARES 2014

CARACTERÍSTICAS DE LA EMPRESA	SECTORES DE TRABAJO	
	Formal	Informal
Seguro de Salud de la empresa		
Si	51,1	6,3
No	48,9	93,7
La empresa lleva NIT		
Si en régimen general	24,9	8,6
Si en régimen simplificado	5,0	4,8
No tiene en proceso	45,6	77,4
No sabe	24,5	9,2

Fuente: Elaboración Propia en Base a la EH

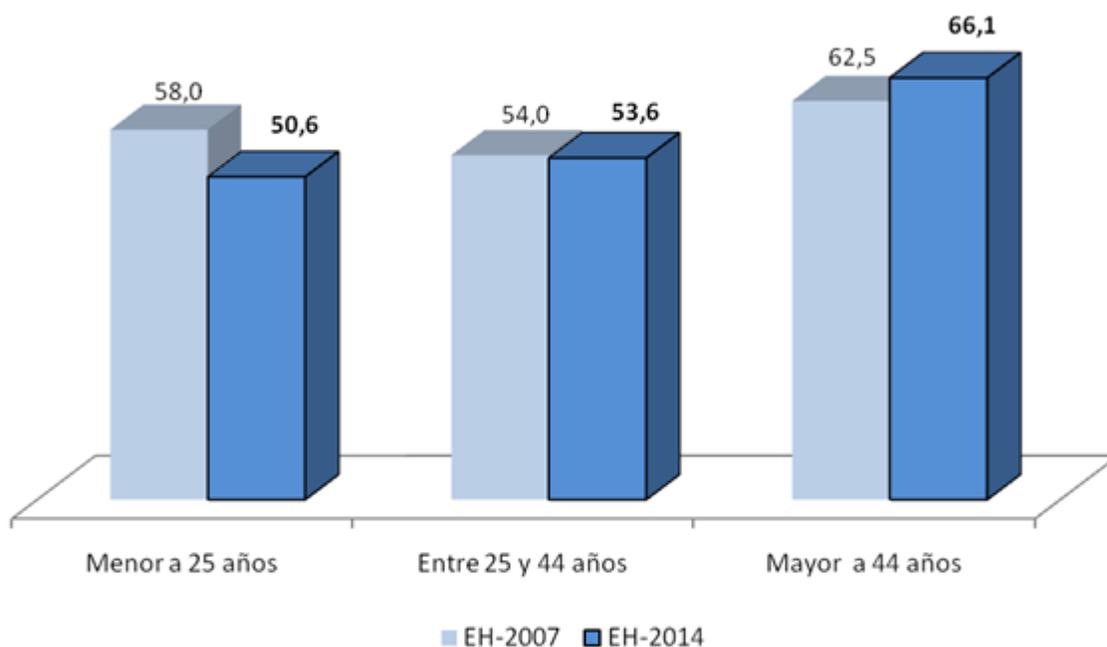
Análisis de la población ocupada en el Sector Informal del área urbana de Bolivia

Edad

En relación a la edad de la población ocupada en el sector informal, según el **gráfico 3** las personas menores a 25 años disminuyeron su participación en este sector de 58 por ciento en el año 2007 a 50,6 por ciento en el año 2014.

En el año 2014 el 53,6 por ciento de la población que trabajaba en este sector tiene una edad entre 25 y 44 años. Sin embargo, la población mayor de 44 años ha aumentado su participación desde el año 2007 al 2014, de 62,5 por ciento a 66,1 por ciento respectivamente.

Gráfico 3. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR GRUPOS DE EDAD, ENCUESTA DE HOGARES 2007 Y 2014

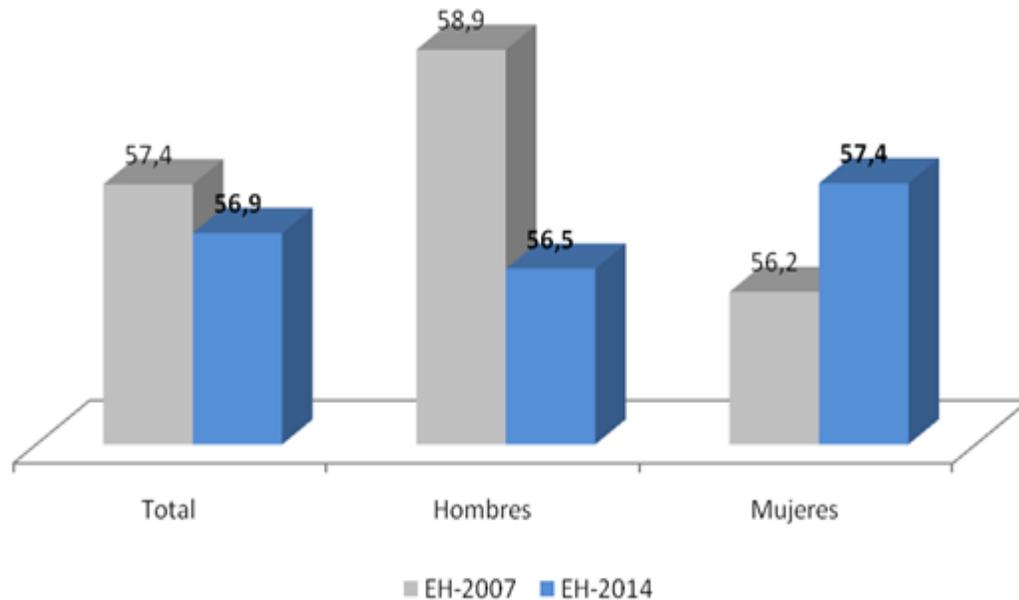


Fuente: Elaboración Propia en Base a la EH – 2007 y EH - 2014

Género

El **gráfico 4** revela que la mayor parte de la población ocupada en el sector informal son mujeres pero su participación aumento mínimamente de 56,2 por ciento en el año 2007 a 57,4 en el 2014. Pero esto no ocurre en los hombres ocupados en este sector, ya que su participación disminuyo de 58,9 por ciento en el año 2007 a 56,5 por ciento en el 2014.

Gráfico 4. BOLIVIA URBANA: POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR SEXO, ENCUESTA DE HOGARES 2007 Y 2014 (En porcentaje)

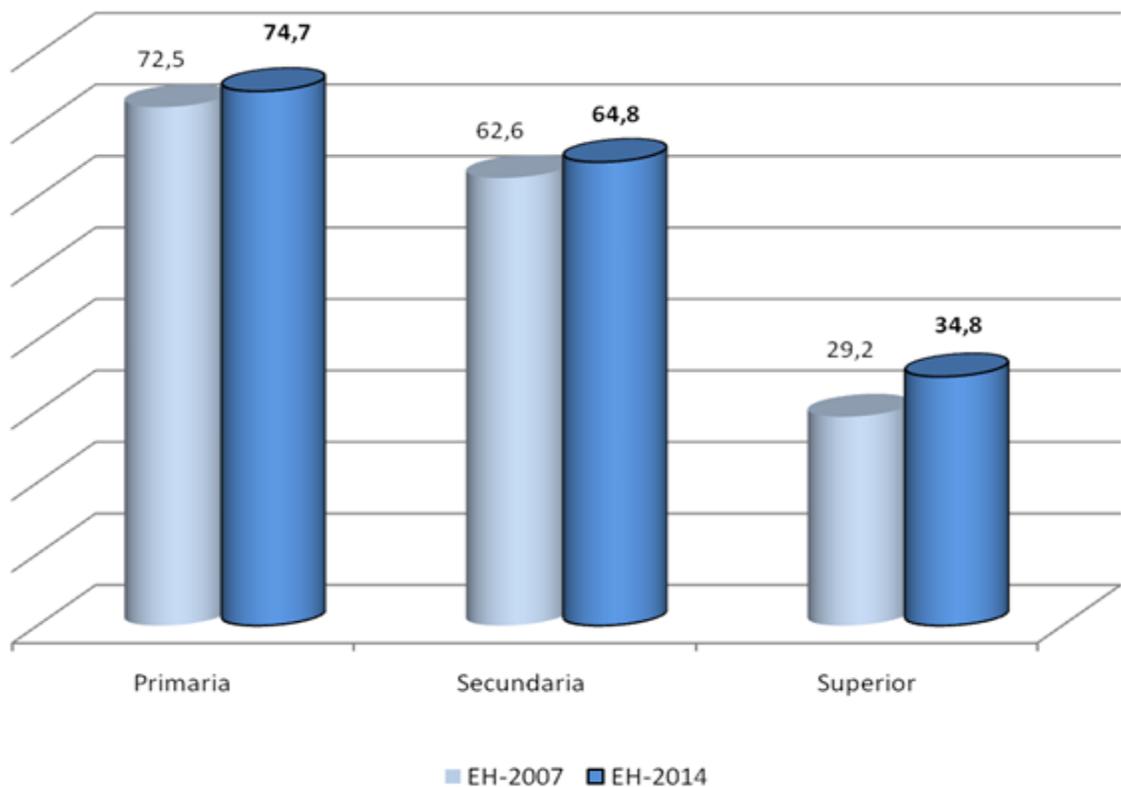


Fuente: Fuente: Elaboración Propia en Base a la EH-2007 y EH-2014

Nivel de Educación

El nivel educativo de las personas ocupadas en el sector informal aumentó (**gráfico 5**), es decir el nivel superior representaba para el año 2007 el 29,2 por ciento, el cual aumentó a 34,8 por ciento en el año 2014. El porcentaje de personas ocupadas con un nivel secundario también aumentó, en el año 2007 fue 62,6 por ciento y para el año 2014 sube a 64,8 por ciento. Sin embargo se mantiene el mayor porcentaje de población ocupada con un nivel primario de 72,5 por ciento en el año 2007 a 74,7 por ciento en el 2014.

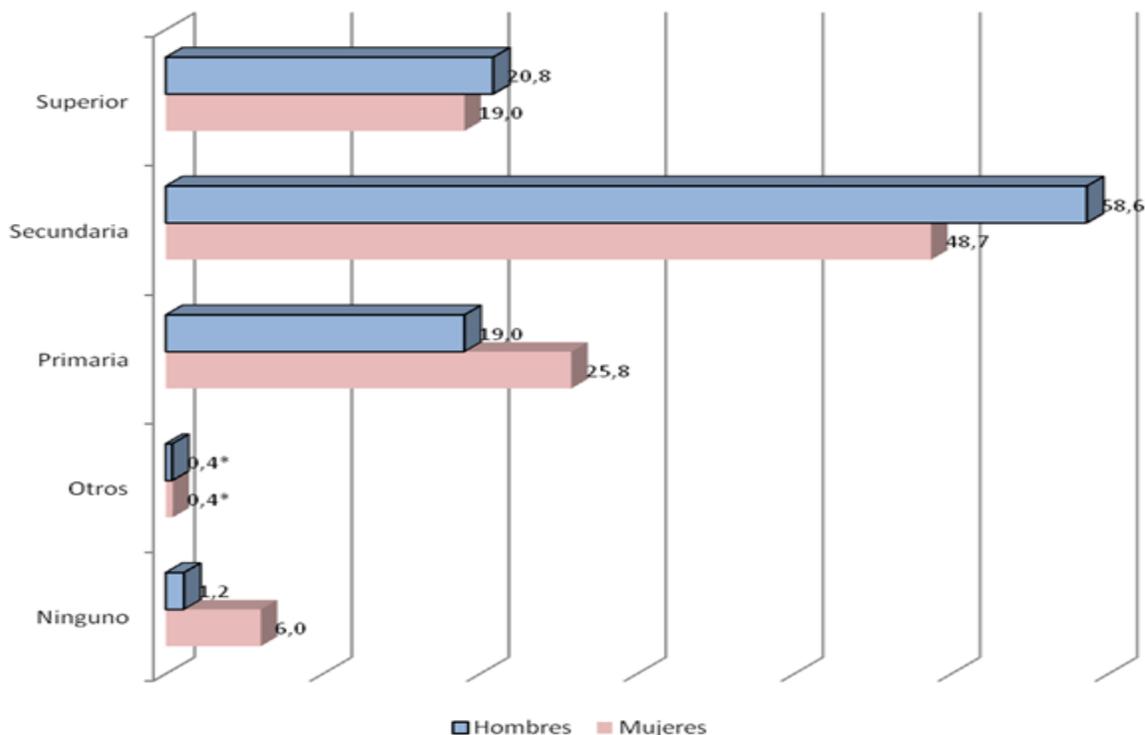
Gráfico 5. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL, SEGÚN NIVEL DE EDUCACIÓN, ENCUESTA DE HOGARES 2007 Y 2014



Fuente: Fuente: Elaboración Propia en Base a la EH-2007 y EH-2014

Si se analiza por género el **gráfico 6** muestra a la población ocupada en el sector informal, para el año 2014 el porcentaje de mujeres sin ningún nivel educativo es mayor respecto de los hombres, esto también ocurre en el nivel educativo primario. Pero el mayor porcentaje se concentra en el nivel educativo secundario, los hombres representan el mayor porcentaje, es decir el 58,6 por ciento y las mujeres tan solo el 48,7 por ciento, en cuanto al nivel superior los hombres representan el 20,8 por ciento, respecto a las mujeres que tan solo representan el 19 por ciento. En resumen las mujeres tienen un nivel educativo bajo (como ser el nivel primario o ningún nivel de estudio) respecto a los hombres, los cuales tienen mayores niveles de educación como ser el nivel secundario o superior.

Gráfico 6. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL, POR GÉNERO, SEGÚN NIVEL EDUCATIVO, ENCUESTA DE HOGARES 2014



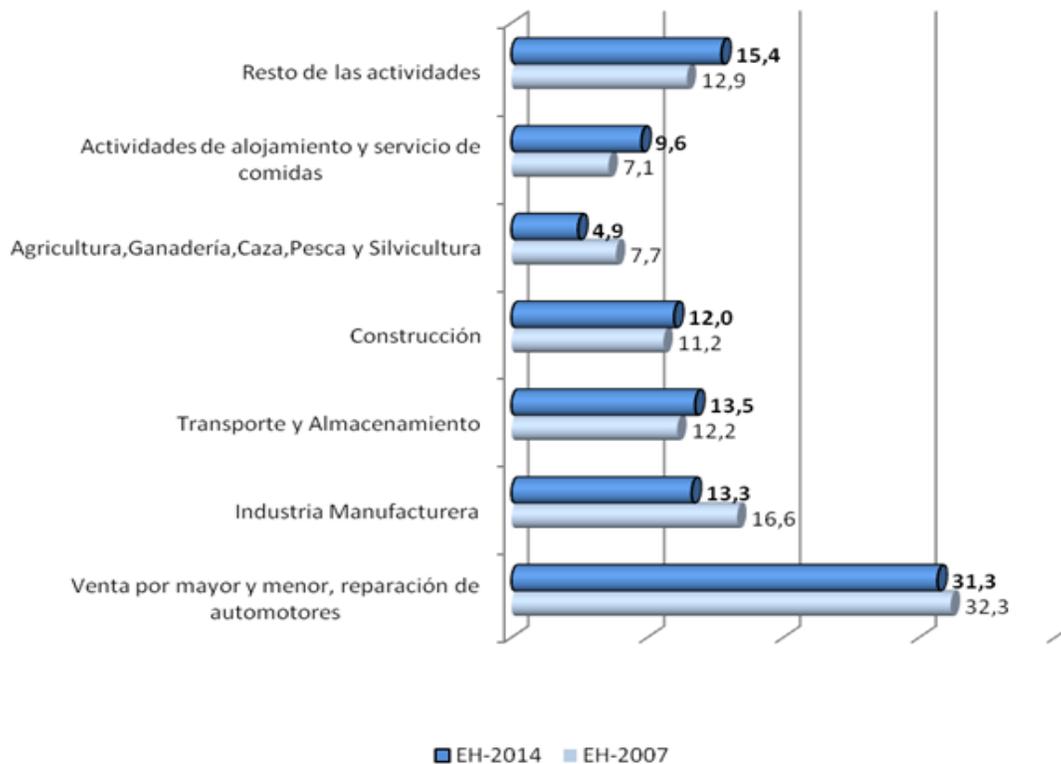
Fuente: Elaboración Propia en Base a la EH

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

Actividad Económica

El **gráfico 7** muestra que para el año 2007 las actividades de venta por mayor y menor; reparación de automotores e industria manufacturera representaban en conjunto el 48,9 por ciento de la población ocupada en el sector informal, para el año 2014 su participación disminuyó a 44,6 por ciento en estas actividades, al tiempo que las actividades de transporte y almacenamiento; y la construcción aumentó su participación en 1,3 y 0,8 puntos porcentuales respectivamente.

Gráfico 7. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL, SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2007 Y 2014



Fuente: Elaboración Propia en Base a la EH-2007 y EH-2014

Resto de las actividades: Para la EH-2007 están: Organismos extraterritoriales; educación; producción y distribución de electricidad, gas y agua; intermediación financiera; administración pública, defensa y seguridad social; hogares privados; explotación de minas y canteras; servicios sociales y de salud; servicios inmobiliarios, empresariales y de alquiler; servicios comunitarios y personales.

Y para la EH-2014 están: Administración Pública, Defensa y Seguridad Social; Actividades de Hogares Privados; Actividades inmobiliarias; Intermediación Financiera y Seguros; Servicios de Educación; Explotación de Minas y Canteras; Actividades artísticas, entretenimiento y recreativas; Informaciones y Comunicaciones; Servicios de Salud y Asistencia Social; Actividades de Servicios Administrativos y de Apoyo; Servicios Profesionales y Técnicos; Otras actividades de servicios.

En el **cuadro 6** se observa a la población ocupada masculina que se distribuye en las actividades de transporte y almacenamiento; construcción; venta por mayor y menor, reparación de automotores, industria manufacturera, que en conjunto representan el 76,3 por ciento de las actividades económicas, mientras que el 46,8 por ciento de las mujeres se emplean en la actividad de: venta por mayor y menor, reparaciones de automotores; seguido de la actividad de alojamiento y servicios de comida que representa el 19,4 por ciento; le sigue la industria manufacturera con el 13,7 por ciento y otras actividades de servicio que representa el 6 por ciento, en conjunto la suma de estas actividades representa el 85,9 por ciento, para las mujeres ocupadas en el sector informal en el año 2014.

Cuadro 6. BOLIVIA URBANA: POBLACIÓN OCUPADA DEL SECTOR INFORMAL POR SEXO, SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2014
(En porcentaje)

Actividad Económica	Sexo		Total
	Hombres	Mujeres	
Total	980.468	716.445	1.696.913
Agricultura, Ganadería, Caza, Pesca y Silvicultura	6,1	3,3	4,9
Explotación de Minas y Canteras	1,1 *	0,0 *	0,6 *
Industria Manufacturera	12,9	13,7	13,3
Suministro de electricidad, gas, vapor y aire acondicionado	0,0 *		0,0 *
Suministro de agua, evacuación de aguas residuos, gestión de desechos	0,1 *		0,0 *
Construcción	20,4	0,4	12,0
Venta por mayor y menor, reparación de automotores	20	46,8	31,3
Transporte y Almacenamiento	23	0,5	13,5
Actividades de alojamiento y servicio de comidas	2,4	19,4	9,6
Informaciones y Comunicaciones	1,4	1,3	1,3
Intermediación Financiera y Seguros	0,3 *	0,4 *	0,3 *
Actividades inmobiliarias	0,3 *	0,3 *	0,3 *
Servicios Profesionales y Técnicos	4,8	3,1	4,1
Actividades de Servicios Administrativos y de Apoyo	1,8	1,4	1,6
Administración Pública, Defensa y Seguridad Social	0,1 *	0,0 *	0,1 *
Servicios de Educación	0,4 *	0,2 *	0,3 *
Servicios de Salud y Asistencia Social	0,9	2,4	1,5
Actividades artísticas, entretenimiento y recreativas	0,9 *	0,5 *	0,7 *
Otras actividades de servicios	3,1	6,0	4,4
Actividades de Hogares Privados	0,0 *	0,3 *	0,1 *

Fuente: Elaboración Propia en Base a la EH

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

Los resultados del **cuadro 7** muestra que para el año 2014 el sector informal empleó al 62,5 por ciento de mujeres ocupadas como trabajadoras de servicio y vendedoras, porcentaje respecto al total; seguido de trabajadoras ocupadas en la construcción, industria manufacturera y otros, cual representa el 15,6 por ciento, estos dos grupos en conjunto representan 78,1 por ciento.

Por otra parte los hombres ocupados en el sector informal, según el grupo ocupacional, se tiene: el más importante los trabajadores de la construcción, industria manufacturera y otros, lo que representa el 34,4 por ciento respecto del total; seguido de los trabajadores operadores de instalaciones y maquinarias que constituye el 24,3 por ciento y los trabajadores de servicio y vendedores que representa el 15,7 por ciento, estos tres grupos en conjunto suman el 74,4 por ciento.

Cuadro 7. BOLIVIA URBANA: POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR SEXO SEGÚN GRUPO OCUPACIONAL, ENCUESTA DE HOGARES 2014 (En porcentaje)

Grupo Ocupacional	Sexo		Total
	Hombres	Mujeres	
Total	980.468	716.445	1.696.913
Directivos Administración Pública y Empresas	0,8 *	1,2 *	1,0 *
Profesionales científicos e intelectuales	5,1	4,7	4,9
Técnicos de Nivel Medio	6,6	3,7	5,3
Empleados de oficina	0,8 *	2,0 *	1,3 *
Trabajadores de Servicio y Vendedores	15,7	62,5	35,4
Trabajadores en Agricultura, Pecuaria, Pesca y otros	5,5	3,3	4,6
Trabajadores de la Construcción, Industria Manufacturera y Otros	34,4	15,6	26,5
Operadores de Instalaciones y Maquinarias	24,3	0,4 *	14,2
Trabajadores No Calificados	6,8	6,5	6,7
Sin especificar		0,1 *	0,02 *

Fuente: Elaboración Propia en Base a la EH

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

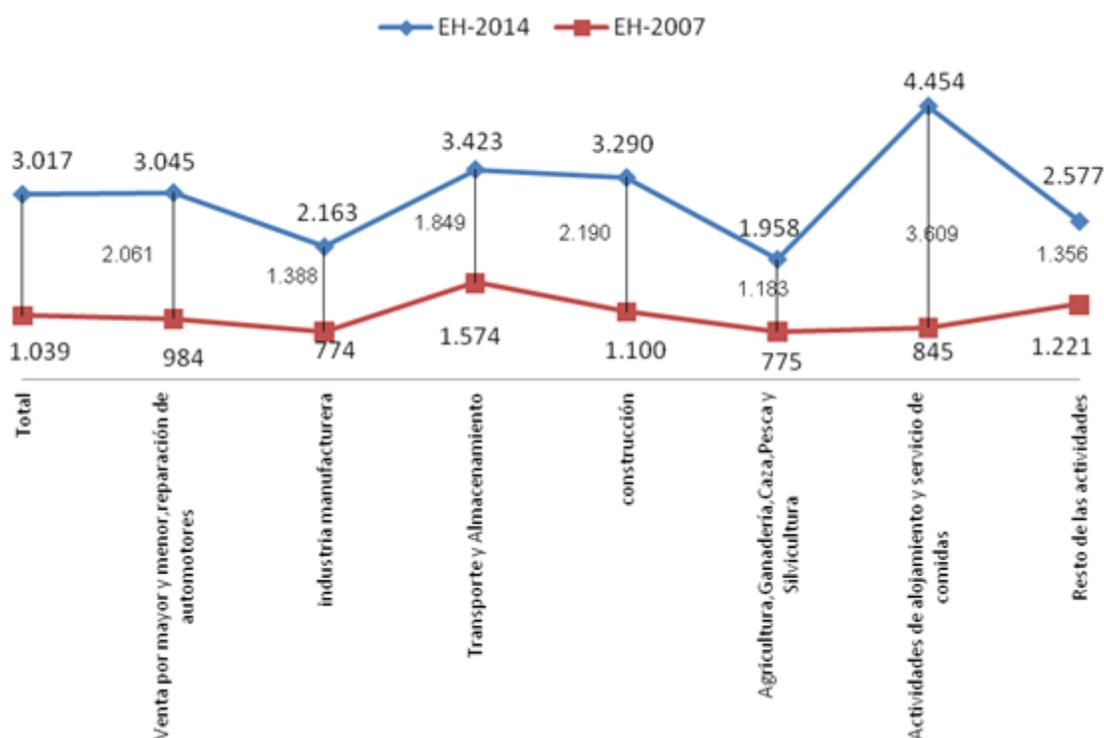
Ingresos

En relación al ingreso promedio mensual de las personas ocupadas en el sector informal, según el **gráfico 8** el promedio declarado de las personas ocupadas para el año 2007 fue de 1.039 Bs y para el año 2014 incrementó a 3.017 Bs.

El ingreso declarado según actividad económica, también fue cambiando en el transcurso de los años, lo que más resalta es la actividad de alojamiento y servicios de comidas, ya que en el año 2007 el ingreso promedio mensual fue de 845 Bs. y para el año 2014 es 4.454 Bs. es decir incremento promedio fue de 3.609 Bs. otro monto que llama la atención, es la actividad de agricultura, ganadería, caza, pesca y silvicultura ya que el año 2007 el ingreso promedio mensual fue 775 Bs y para el 2014 se incrementó a 1.958 Bs. otra actividad donde se observa el incremento es en la actividad de la industria manufacturera ya que para el año 2007 el ingreso promedio mensual fue de 774 Bs. y para el año 2014 incrementó a 2.163 Bs.

Para las demás actividades el ingreso promedio mensual de 2007 a 2014 incrementó, como se observa el gráfico 8, estos datos son para los ocupados en el sector informal.

Gráfico 8. BOLIVIA URBANA: PROMEDIO DEL INGRESO LABORAL EN EL SECTOR DE INFORMAL SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2007 Y 2014



Fuente: Elaboración Propia en Base a la EH-2007 y EH-2014

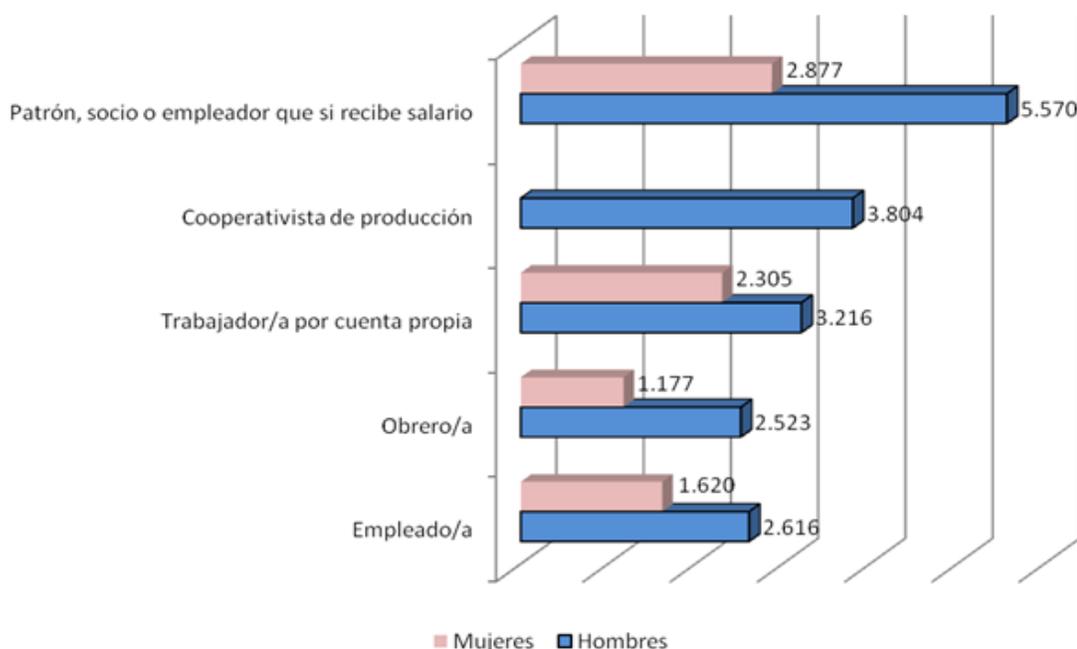
Resto de las actividades: Para la EH-2007 están: Organismos extraterritoriales; educación; producción y distribución de electricidad, gas y agua; intermediación financiera; administración pública, defensa y seguridad social; hogares privados; explotación de minas y canteras; servicios sociales y de salud; servicios inmobiliarios, empresariales y de alquiler; servicios comunitarios y personales.

Y para la EH-2014 están: Administración Pública, Defensa y Seguridad Social; Actividades de Hogares Privados; Actividades inmobiliarias; Intermediación Financiera y Seguros; Servicios de Educación; Explotación de Minas y Canteras; Actividades artísticas, entretenimiento y recreativas; Informaciones y Comunicaciones; Servicios de Salud y Asistencia Social; Actividades de Servicios Administrativos y de Apoyo; Servicios Profesionales y Técnicos; Otras actividades de servicios.

Los datos del **gráfico 9** revelan que para el año 2014, en lo que se refiere a las mujeres trabajadoras, las categorías con menores ingresos es la “Obrera” con ingreso promedio mensual de 1.177 Bs. y la “empleada” con un ingreso promedio mensual de 1.626 Bs. la categoría en la que se encuentra el mayor ingreso promedio es el “Patrón, socio o empleador que si recibe salario” con 2.877 Bs.

En cambio para los hombres, los ingresos promedios mensuales menores se encuentran en la categoría “obrero” con 2.523 Bs. y “empleado” con 2.616 Bs. por el contrario en las categorías “Patrón, Socio Empleador que si recibe salario” y “cooperativista de producción” se encuentran los mayores ingresos promedios mensuales percibidos, con 5.570 Bs. y 3.804 Bs. respectivamente, tal como se muestra en el **gráfico 9**.

Gráfico 9. BOLIVIA URBANA: INGRESO LABORAL PROMEDIO DE LA POBLACIÓN OCUPADA EN EL SECTOR INFORMAL POR GÉNERO, SEGÚN CATEGORIA OCUPACIONAL, ENCUESTA DE HOGARES 2014

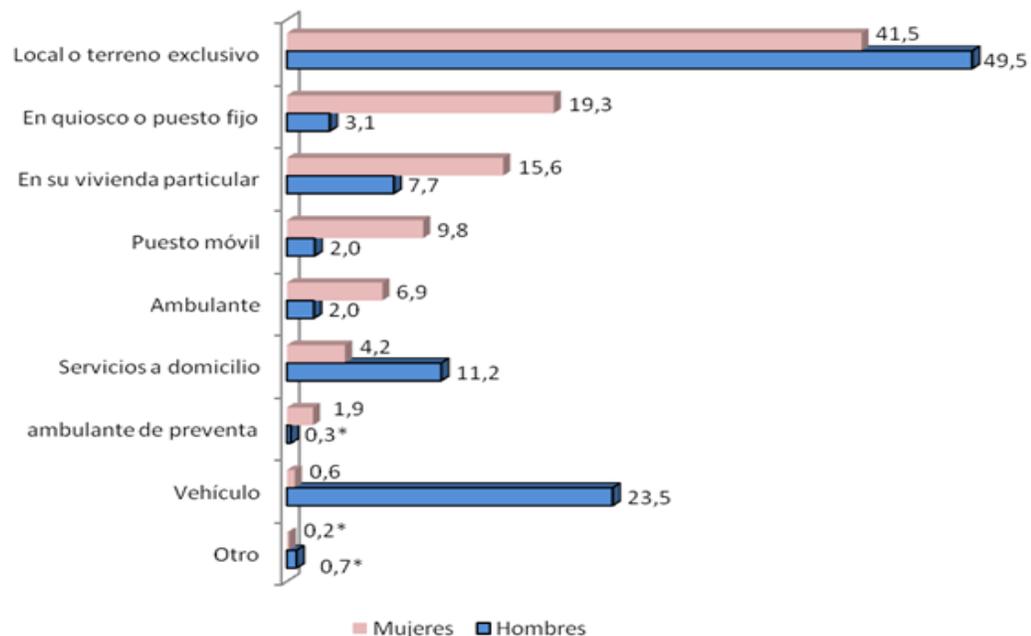


Fuente: Elaboración Propia en Base a la EH

Lugar de trabajo

Los resultados del **gráfico 10**, muestra a las personas ocupadas en el sector informal, lo que indica que existe un mayor porcentaje de mujeres que trabajan en lugares como: puestos móviles¹², quioscos o puesto fijo¹³, es decir lugares precarios, inseguros, lo que representa el 29,1 por ciento, en cambio los hombres que trabajan en este sector, se encuentran en lugares tipo local o terreno exclusivo¹⁴, en vehículos y servicios a domicilio, lo que representa el 84,2 por ciento del total.

Gráfico 10. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACION OCUPADA EN EL SECTOR INFORMAL, POR GÉNERO, SEGÚN LUGAR DE TRABAJO, EH - 2014



Fuente: Elaboración Propia en Base a la EH

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

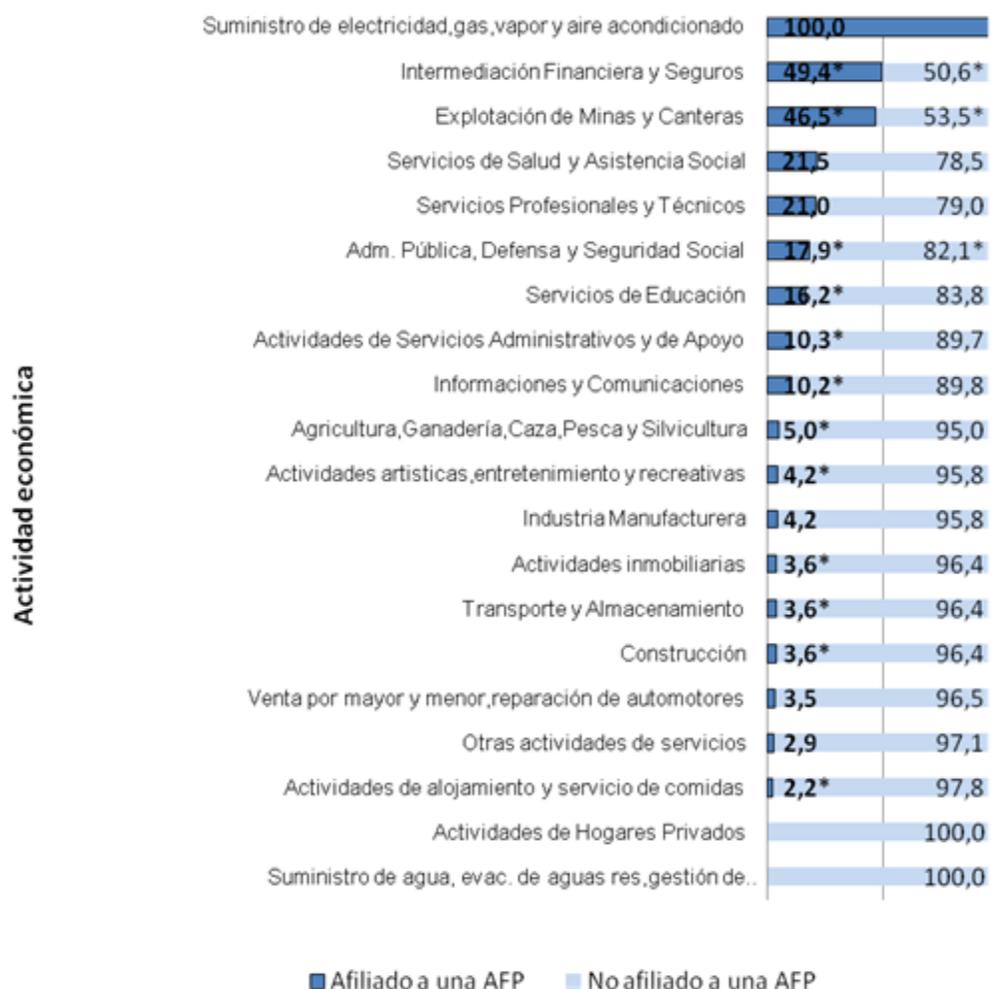
¹²Puesto móvil, son puestos instalados en calles o mercados que pueden ser recogidos al final de la jornada.

¹³Quiosco o puesto fijo, son puestos callejeros de venta o servicios que se encuentran adheridos o empotrados sólidamente en calles, paredes o aceras.

¹⁴Local o terreno exclusivo, es cuando la persona trabaja en un espacio exclusivo de estructura solida para el trabajo, con acceso independiente de la calle, ejemplo: fabricas, consultorio u oficina en un edificio.

A continuación se muestra en el **gráfico 11** a la población ocupada en el sector informal por actividad económica y el aporte que realiza a una AFP, se observa que en la mayor parte de las actividades económicas no están afiliados a una AFP, es decir más del 50 por ciento, solo la actividad económica de suministro de electricidad, gas, vapor y aire acondicionado está afiliada a una AFP en su totalidad.

Gráfico 11. BOLIVIA URBANA: DISTRIBUCIÓN PORCENTUAL DE LA POBLACION OCUPADA EN EL SECTOR INFORMAL POR AFILIACIÓN AFP, SEGÚN ACTIVIDAD ECONÓMICA, ENCUESTA DE HOGARES 2014



Fuente: Elaboración Propia en Base a la EH

* Con coeficiente de variación superior al 20%, emplearlo solo de forma referencial

4.3. Análisis previo de los datos de las variables del modelo

Antes de aplicar algún método multivariante primeramente se debe hacer un análisis cuidadoso de los datos, evaluar datos ausentes y aplicar algún método de imputación si se lo requiere, identificar los casos atípicos y comprobación de los supuestos si el modelo lo requiere, para luego buscar soluciones.

A continuación se muestra un **cuadro 8**, con información si existen datos perdidos, valores mínimos y máximos:

Cuadro 8. NÚMERO DE CASOS DE LAS VARIABLES PARA EL MODELO

Variables	Número de casos	Observaciones	
		Mínimos	Máximos
Edad	2.984.726	10	89
Años de estudio	2.984.726	0	27
Horas promedio de trabajo (día)	2.984.726	1	24
Ingreso promedio laboral (Bs./mensual)	2.984.726	0	265.100
Sexo	2.984.726		
Condición de pareja	2.984.726		
Auto identificación étnica	2.984.726		
Aporte a la AFP	2.984.726		

Fuente: Elaboración Propia en Base a la EH-2014

Observando el cuadro con las variables seleccionadas para el modelo, estas no presentan datos perdidos, pero si existen datos muy pequeños y datos extremos en las variables edad, años de estudio, horas promedio de trabajo e ingreso promedio laboral mensual, los cuales son valores propios de cada sector, revisando los mismos.

Multicolinealidad

Para la aplicación de los modelos de regresión y entre ellos el Modelo Logit Multinomial se debe ver antes la “Multicolinealidad” de las variables independientes X, aplicando pruebas estadísticas para ver si existen correlaciones altas entre las mismas, según el tipo de variables se tienen los

resultados del **cuadro 9**, donde muestra la correlación de las variables independientes, para variables cuantitativas (edad, años de estudio, ingreso y horas de trabajo) se analizó la matriz de correlaciones (**ecuación 3.5**), donde se observa que las correlaciones entre las mismas es relativamente baja (menos de 0,13), para variables cualitativas nominales (sexo, condición de pareja, identificación étnica y aporte a la AFP), se analiza el coeficiente de correlación V de Cramer (**ecuación 3.3**), donde los resultados son valores cercanos a 0,13 por lo que se concluye una correlación relativamente baja, al analizar la significación de cada valor este es menor a 0,05 lo que se afirma que la correlación (mínima) que se ha establecido es muy probable cierta. Otras medidas de correlación como ser: coeficiente de Contingencia y Phi lanzan como resultado valores similares al coeficiente V de Cramer, (ver anexo 4). Por otro lado se tiene la correlación de variables cualitativas (sexo, condición de pareja, identificación étnica y aporte a la AFP) con cuantitativas (previamente se debe codificar las variables edad, años de estudio, ingreso, horas de trabajo en grupos de intervalos), para esto se vio la correlación Eta la cual muestra valores relativamente bajos cercanos a 0,2 y el valor de la significancia de cada variable este es menor a 0,05 lo que indica que la correlación (mínima) que se ha establecido es muy probable cierta, con estas pruebas se concluye que no existe Multicolinealidad¹⁵, es decir la correlación entre las ocho variables independientes del modelo es relativamente baja.

¹⁵Las salidas de correlaciones y otras pruebas estadísticas para probar la Multicolinealidad encuentra en el Anexo 4.

Cuadro 9. PRUEBA DE MULTICOLINEALIDAD ENTRE VARIABLES INDEPENDIENTES DEL MODELO

Variables		Edad	Años de estudio	Horas de trabajo	Ingresos	Sexo	Condición de pareja
Edad	Valor		-0,1384	0,0256	0,0421	0,038	0,267
	Significación					0	0
Años de estudio	Valor			-0,1138	0,1281	0,084	0,08
	Significación					0	0
Horas de trabajo	Valor				0,0739	0,19	0,082
	Significación					0	0
Ingresos	Valor					0,247	0,173
	Significación					0	0
Sexo	Valor						0,136
	Significación						0
Condición de pareja	Valor						
	Significación						
Identificación étnica	Valor						
	Significación						
Aporte a la AFP	Valor						
	Significación						

Fuente: Elaboración Propia en Base a la EH-2014

Para variable de tipo cuantitativo se usó la matriz de correlación.

Para variables de tipo cualitativo nominal, se usó la medida de correlación V de Cramer.

Para variables de tipo cualitativo y cuantitativa se usó la medida de correlación Eta.

Para complementar el análisis se realizó la verificación de correlación de la variable dependiente¹⁶ con cada una de las variables independientes del modelo, por tanto se tiene en el **cuadro 10**: la variable sector de trabajo y la variable edad, donde se tiene una correlación relativamente baja cercana a 0; la variable sector de trabajo con años de estudio tiene una correlación regular; la variable sector de trabajo con las horas de trabajo también presenta una correlación baja; la variable sector de trabajo con la variable ingreso laboral presenta una correlación regular, a estas variables se aplicaron las medidas de correlación Eta, por tanto se concluye que la correlación entre las variables independientes (cuantitativas codificadas en grupos de intervalos) con la variable dependiente, tienen correlaciones relativamente bajas en algunas y en otras regulares, también se observa el valor de la significancia que es menor a 0,05 lo que indica que la correlación se ha establecido, pero esta es mínima.

¹⁶La correlación entre las variables independientes con la variable dependiente se encuentra en el Anexo 5.

Para ver la correlación de las otras cuatro variables de tipo cualitativo (sexo, condición de pareja, identificación étnica y aporte a una AFP), con la variable sector de trabajo, se usó correlación V de Cramer (**ecuación 3.3**), el cual lanza como resultado que 3 variables (sexo, condición de pareja e identificación étnica) no están correlacionadas con la variable sector de trabajo y la que tiene una buena correlación con la variable dependiente es la variable de aporte a una AFP, se analiza el valor de significancia la cual es menor a 0,05 lo que indica que la correlación se ha establecido, aunque para algunas variables esta correlación es mínima.

Cuadro 10. ASOCIACIÓN ENTRE LA VARIABLE DEPENDIENTE Y LAS VARIABLES INDEPENDIENTES

VARIABLES INDEPENDIENTES (X)		VARIABLE DEPENDIENTE "INFORMALIDAD" (Y)
Edad	Valor	0,120
	Significación	0,000
Años de estudio	Valor	0,289
	Significación	0,000
Horas de trabajo	Valor	0,100
	Significación	0,000
Ingresos	Valor	0,277
	Significación	0,000
Sexo	Valor	0,214
	Significación	0,000
Condición de pareja	Valor	0,075
	Significación	0,000
Identificación étnica	Valor	0,092
	Significación	0,000
Aporte a la AFP	Valor	0,553
	Significación	0,000

Fuente: Elaboración Propia en Base a la EH-2014

La prueba de correlación para variable cualitativa con otra cuantitativa es el estadístico Eta.

La prueba de correlación entre variables cualitativas es el estadístico V de Cramer.

El análisis anterior es solo un complemento, ya que el modelo Logit Multinomial cuenta con contrastes estadísticos directos, capacidad para incorporar efectos no lineales y permitir una amplia variedad de diagnósticos. Sin embargo se debe tener cuidado en otros asuntos que se verán en su respectiva aplicación.

Capítulo V.

5.1. *Aplicación y análisis del Modelo Logit Multinomial*

Ahora bien, la decisión de un individuo de participar en algún sector de trabajo, está en función de un conjunto de variables explicativas X (variables independientes), las que pueden ser referidas a las características propias de los individuos, por ejemplo la edad, sexo, años de estudio, etc. El individuo que participa en el mercado laboral, puede elegir entre tres opciones: ocuparse en el sector formal, en el sector informal o en el sector doméstico.

En el presente trabajo se optó por la aplicación del modelo de opción múltiple no ordenado como es el modelo Logit Multinomial descrito en el **capítulo III**, se usa los programas SPSS y STATA¹⁷. Se trabaja con la base de la Encuesta de Hogares 2014, se tiene la variable categórica: **sectores de trabajo** la cual tiene tres alternativas de respuestas: sector formal, sector informal y sector doméstico y las variables explicativas (variables independientes) que se incluyen en el modelo son:

X_1 : Edad

X_2 : Años de estudio

X_3 : Sexo

X_4 : Condición de pareja

X_5 : Identificación étnica

X_6 : Ingreso laboral

X_7 : Horas promedio de trabajo

X_8 : Aporte a una AFP

¹⁷Las salidas de las tablas de este capítulo se encuentran en el Anexo 6.

Antes de aplicar el modelo Logit Multinomial, se verifica el supuesto IAI, el cual es clave para la estimación del modelo, la verificación de este supuesto se realizó en el programa de STATA, como la variable dependiente consta de tres alternativas, una es elegida como categoría base (sector doméstico), se tiene los resultados de la **tabla 3** (prueba de Hausman), donde los sectores de trabajo (formal e informal), son relevantes para el modelo, se aprueba H_0 , es decir que se evidencia la independencia de alternativas irrelevantes para el modelo con un 95 por ciento de nivel de confianza. Es decir cada alternativa de respuesta de la variable dependiente es valorada de manera independiente por cada persona que toma la decisión de optar por algún sector de trabajo, con lo cual no se rechaza el supuesto de IAI, y por tanto es apropiado asumir la distribución logística.

Tabla 3. SUPUESTO DE INDEPENDENCIA DE ALTERNATIVAS IRRELEVANTES

Prueba de Hausman				
Prueba de Independencia de Alternativas Irrelevantes				
Ho: Las respuestas son independientes de otras alternativas				
Categorías	Chi-cuadrado	Grados de Libertad	P>Chi-cuadrado	Evidencia
Formal	0,000	16	0,653	Se cumple IAI
Informal	0,000	16	1,000	Se cumple IAI

Fuente: Elaboración Propia en Base a la EH-2014

La **tabla 4** muestra el resumen de frecuencias de los datos de las variables dicotómicas, en particular los resultados acerca de la preferencia por algún sector de trabajo. De acuerdo con esta tabla, la categoría base es el sector doméstico. En resumen el sector informal comprende el 56,9 por ciento de los casos, el sector formal comprende el 40 por ciento y el sector doméstico representado por 3,2 por ciento. También muestra el número de casos y su porcentaje respectivo de las variables independientes de tipo cualitativo como ser: sexo, condición de pareja, identificación con algún pueblo originario o indígena y el aporte a una AFP.

Tabla 4. RESUMEN DE PROCESAMIENTO DE CASOS

		N	Porcentaje marginal
Informalidad	Sector formal	1.193.041	40,0
	Sector informal	1.697.047	56,9
	Sector domestico	94.638	3,2
Sexo	Mujeres	1.248.140	41,8
	Hombres	1.736.586	58,2
Condición de pareja	Sin pareja	1.106.996	37,1
	Con pareja	1.877.730	62,9
Identificación con algún pueblo	No se identifica	2.223.300	74,5
	Si se identifica	761.426	25,5
Aporte a la AFP	Aporta a una AFP	727.958	24,4
	No aporta a una AFP	2.256.768	75,6
Válido		2.984.726	100,0

Fuente: Elaboración Propia en Base a la EH-2014

Tabla 5. RESUMEN DE PASOS

Modelo	Efecto(s)	- 2 Logaritmo de la verosimilitud	Pruebas de selección de efecto		
			Chi-cuadrado ^a	Grados de libertad	Significación
0	Interceptación	4742920,560			
1	Edad	4679131,014	63789,546	2	0,000
2	Años de estudio	4372707,729	306423,286	2	0,000
3	Sexo	4215644,440	157063,288	2	0,000
4	Condición de pareja	4213704,443	1939,998	2	0,000
5	Identificación étnica	4208961,801	4742,642	2	0,000
6	Ingreso laboral	4188248,130	20713,671	2	0,000
7	Horas de trabajo	4179976,936	8271,194	2	0,000
8	Aporte a una AFP	3439863,092	740113,845	2	0,000

Fuente: Elaboración Propia en Base a la EH-2014

Método por pasos: Avanzar entrada

^a El Chi-cuadrado para la entrada se basa en la prueba de razón de verosimilitud.

La **tabla 5** arroja un resultado importante el cual indica que, utilizando el procedimiento método por pasos¹⁸, todas las variables independientes X para el modelo son significativas. Es decir, las variables: edad, sexo, años de estudio, condición de pareja, horas promedio de trabajo, ingreso promedio laboral, identificación étnica y aporte a una AFP, son indicadores que pueden predecir el sector de trabajo al cual las personas optaran por ingresar (formal o informal), esto ocurre para un valor de significancia menor a 0,05 lo que

¹⁸Método por pasos (entrada hacia adelante) es un procedimiento que permite ir añadiendo paso a paso aquellas variables que resultan más significativas.

significa que el predictor es significativo, por tanto las variables independientes X son significativas para el modelo.

En la columna - 2 Logaritmo de la verosimilitud, similar a la prueba F en regresión lineal, bajo la hipótesis nula de que todos los coeficientes de cada variable X son cero, es decir:

$H_0: \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$, el estadístico $-2\{L(0) - L(\beta)\}$ es X^2 distribuida, con 2 grados de libertad (J-1), de nuevo se tiene una significancia menor a 0,05, por tanto cada variables es significativa para el modelo.

Tabla 6. INFORMACIÓN DE AJUSTE DEL MODELO

Modelo	Criterios de ajuste de modelo	Contraste de la razón de verosimilitud		
	-2 Logaritmo de la verosimilitud	Chi-cuadrado	Grados de libertad	Significación
Sólo interceptación	4742920,560			
Final	3439863,092	1303057,469	16	0,000

Fuente: Elaboración Propia en Base a la EH-2014

Esta **tabla 6** se refiere al estadístico -2 Logaritmo de la verosimilitud de significancia global de los parámetros, como el valor de la significancia es menor a 0,05, entonces se concluye que existe una relación entre la variable dependiente con las variables independientes X, lo cual implica que los motivos para incorporarse en algún sector de trabajo guardan un vínculo con los atributos personales del trabajador como la edad, sexo, años de estudio, condición de pareja, identificación étnica y atributos del empleo como ingreso mensual, horas de trabajo y aporte a una AFP.

Tabla 7. BONDAD DE AJUSTE

Estadísticos	Chi-cuadrado	Grados de libertad	Significación
Pearson	11403070,691	23134	0,061
Desviación	3439713,360	23134	1,000

Fuente: Elaboración Propia en Base a la EH-2014

La **tabla 7** de bondad de ajuste, presenta dos estadísticos, bajo la hipótesis nula el modelo se ajusta adecuadamente los datos, si la hipótesis nula se cumple, el estadístico Pearson y Desviación tiene una distribución Chi-cuadrada con los grados de libertad desplegados en la tabla. Si el valor de la significancia es pequeño (menor que 0,05) el modelo no ajusta adecuadamente los datos. Los resultados muestran el valor de la significancia es mayor que 0,05 denotando que los supuestos del modelo son consistentes con los datos, por tanto el modelo se ajusta adecuadamente a los datos.

Tabla 8. PSEUDO R CUADRADO

Pseudo R cuadrado	
Cox y Snell	0,354
Nagelkerke	0,444
McFadden	0,274

Fuente: Elaboración Propia en Base a la EH-2014

La **tabla 8** muestra los indicadores R^2 (ecuación 3.26)(V. Pando Fernandez & R. San Martin Fernandez) que no están muy cercanos al valor cero, indicando al conjunto de variables independientes significativas y que existe una buena relación entre las variables independientes con la dependiente, aunque estos indicadores miden la variabilidad explicada y en general son mucho más bajas que en la regresión, por tanto el 44,4 por ciento de la variación de la variable dependiente sectores de trabajo puede ser explicada por las variables independientes del modelo, en particular con las variables edad, años de estudio, sexo, relación de pareja, identificación étnica, ingreso laboral promedio, horas promedio de trabajo y aporte a una AFP. Los coeficientes pseudo R-cuadrado permiten estimar la capacidad explicativa del modelo, sin embargo, se recomienda usarlo únicamente como comparativo entre dos modelos, el de

mayor R^2 se preferirá frente a los demás (los datos se corrieron con varias rutinas por pasos hacia adelante y sin pasos) y el que se presenta en la **tabla 8** es el que obtuvo el mayor R^2 .

Tabla 9. CONTRASTE DE LA RAZÓN DE VEROSIMILITUD

Efecto	Criterios de ajuste de modelo	Contraste de la razón de verosimilitud		
	-2 Logaritmo de la verosimilitud del modelo reducido	Chi-cuadrado	Grados de libertad	Significación
Interceptación	3569349,298	129486,206	2	0,000
Edad	3531823,697	91960,605	2	0,000
Años de estudio	3467281,833	27418,742	2	0,000
Sexo	3568820,989	128957,898	2	0,000
Condición de pareja	3442648,418	2785,326	2	0,000
Identificación étnica	3445553,036	5689,944	2	0,000
Ingreso laboral	3452098,851	12235,760	2	0,000
Horas de trabajo	3442321,498	2458,407	2	0,000
Aporte a la AFP	4180008,480	740113,845	2	0,000

Fuente: Elaboración Propia en Base a la EH-2014

El estadístico de Chi-cuadrado es la diferencia de los logaritmos de la verosimilitud -2 entre el modelo final y el modelo reducido. El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de dicho efecto son 0.

La **tabla 9** muestra el resultado de la prueba de cocientes de verosimilitud; esta prueba es usada frecuentemente y compara un modelo con constantes específicas, en lugar de compararlo a un modelo donde todos los β son cero. Así, se compara el modelo completo con un modelo reducido, en este caso, el modelo reducido es aquel que omite el efecto de las 8 variables (edad, años de estudio, sexo, condición de pareja, identificación étnica, ingreso promedio laboral, horas promedio de trabajo y aporte a una AFP). Si la significancia de la prueba es pequeña (menor que 0,05) el efecto contribuye al modelo. En este caso resultó que el efecto de las 8 variables X, si contribuyen significativamente al modelo. Sin embargo a diferencia de la regresión lineal el cual representa una relación lineal entre una variable cuantitativa y una o más variables independientes, es difícil desarrollar una función útil que permita describir la relación entre variables independientes y la respuesta obtenida en modelos de opción múltiple (variable claramente discreta).

Tabla 10. ESTIMACIONES DE LOS PARÁMETROS

Sectores de trabajo	Variables	β	Wald	Significación	Exp(β)
Sector formal	Constante	2,866	9005,067	0,000	
	Edad (X_1)	0,000	2,508	0,113	1,000
	Años de estudio(X_2)	0,091	15431,272	0,000	1,095
	Sexo(X_3)	4,523	30701,266	0,000	92,124
	Condición de pareja(X_4)	0,065	73,564	0,000	1,067
	Identificación étnica(X_5)	0,374	1730,664	0,000	1,453
	Ingreso laboral(X_6)	0,000	8290,851	0,000	1,000
	Horas de trabajo(X_7)	-0,026	416,307	0,000	0,975
	Aporte a la AFP(X_8)	-3,401	19643,312	0,000	0,033
Sector informal	Constante	-0,147	23,998	0,000	
	Edad(X_1)	0,034	15462,462	0,000	1,035
	Años de estudio(X_2)	0,059	6810,145	0,000	1,060
	Sexo(X_3)	4,193	26554,576	0,000	66,252
	Condición de pareja(X_4)	0,214	891,803	0,000	1,238
	Identificación étnica(X_5)	0,533	3860,155	0,000	1,704
	Ingreso laboral(X_6)	0,000	8136,007	0,000	1,000
	Horas de trabajo(X_7)	0,001	1,174	0,000	1,001
	Aporte a la AFP(X_8)	-0,351	205,737	0,000	0,704

Fuente: Elaboración Propia en Base a la EH-2014
La categoría de referencia es: sector doméstico

La **tabla 10** contiene las variables que ingresan en cada modelo como predictoras que determinan el sector formal e informal, muestra los coeficientes estimados (en la columna encabezada por β) y los estadísticos asociados al modelo que predice la probabilidad de pertenecer al sector formal o informal dado el sector de referencia “doméstico”, en función a características individuales de las personas ocupadas como: edad, años de estudio, ingreso laboral, horas de trabajo, sexo, condición de pareja, identificación étnica y aporte a la AFP, estas últimas cuatro variables cualitativas codificadas como dummy.

Además se muestra la prueba de hipótesis sobre los coeficientes, para muestras grandes, la prueba de que un cociente es 0 es basada en el estadístico de Wald que es similar al valor t en regresión múltiple, tiene una distribución Chi-cuadrada, cuando una variable tiene un grado de libertad, el estadístico de Wald es el cuadrado del ratio del coeficiente entre su error

estándar, para variables categóricas, el estadístico de Wald tiene sus grados de libertad equivalentes a uno menos el número de categorías.

Por ejemplo, el coeficiente de años de estudio es 0,091 y su error estándar es 0,001, el estadístico de Wald es $(0,091/0,001)$ o 15431,272, esto es en el grupo del sector formal dado doméstico. El nivel de significancia para el estadístico de Wald es mostrada en la columna de significación, en nuestro caso todas las variables obtenidas en ambos modelos resultan ser significativamente, es decir los valores de β son diferentes de cero, el valor de la columna p-valor es menor al valor crítico prefijado de 0,05 excepto para la variable edad en el modelo del sector formal.

Desafortunadamente, el estadístico de Wald posee una propiedad indeseable para la variable edad en el modelo sector formal. Cuando el valor absoluto del coeficiente de regresión llega a ser demasiado grande, el error estándar también lo es, esto produce que el estadístico de Wald, sea muy pequeño, por tal motivo no estaríamos rechazando la hipótesis nula de que el coeficiente es 0, cuando en realidad si deberíamos. De allí que cuando se tiene un coeficiente grande, uno no debería de confiar en el estadístico de Wald para prueba de hipótesis, en vez de ello, se debería de construir un modelo con la variable y otro sin la variable y basar la prueba de hipótesis en el cambio del logaritmo de la función de verosimilitud.

Los parámetros estimados con coeficientes negativos disminuyen la verosimilitud de esa categoría de respuesta con respecto a la categoría de referencia, en cambio parámetros con coeficientes positivos incrementan la verosimilitud de esa categoría de respuesta.

A continuación se muestran las dos ecuaciones (3.13), para el modelo formal e informal:

$$P(Y = \text{Formal}) = \frac{e^{2,9+0,0X_1+0,1X_2+4,5X_3+0,1X_4+0,4X_5+0,0X_6+0,0X_7-3,4X_8}}{1 + e^{2,9+0,0X_1+0,1X_2+4,5X_3+0,1X_4+0,4X_5+0,0X_6+0,0X_7-3,4X_8} + e^{0,1+0,0X_1+0,1X_2+4,2X_3+0,2X_4+0,5X_5+0,0X_6+0,0X_7-0,4X_8}}$$

$$P(Y = \text{Informal}) = \frac{e^{0,1+0,0X_1+0,1X_2+4,2X_3+0,2X_4+0,5X_5+0,0X_6+0,0X_7-0,4X_8}}{1 + e^{2,9+0,0X_1+0,1X_2+4,5X_3+0,1X_4+0,4X_5+0,0X_6+0,0X_7-3,4X_8} + e^{0,1+0,0X_1+0,1X_2+4,2X_3+0,2X_4+0,5X_5+0,0X_6+0,0X_7-0,4X_8}}$$

Y la ecuación (3.14) para la categoría de referencia es:

$$P(Y = \text{Doméstico}) = \frac{1}{1 + e^{2,9+0,0X_1+0,1X_2+4,5X_3+0,1X_4+0,4X_5+0,0X_6+0,0X_7-3,4X_8} + e^{0,1+0,0X_1+0,1X_2+4,2X_3+0,2X_4+0,5X_5+0,0X_6+0,0X_7-0,4X_8}}$$

Donde las variables independientes para cada modelo son:

X_1 : edad

X_2 : años de estudio

X_3 : sexo

X_4 : condición de pareja

X_5 : identificación étnica

X_6 : ingreso laboral

X_7 : horas de trabajo

X_8 : Aporte a la AFP.

La **tabla 10** muestra dos estimaciones de Logit binario, estimando las chances de cada alternativa (sector formal y sector informal) versus la categoría base (sector doméstico). Los parámetros estimados β en cada regresión (sector formal e informal), representan los betas para cada modelo, cabe mencionar que los modelos parecen ser efectivos al predecir la preferencia de las personas que optaron por los sectores formal e informal.

Parámetros β con valores positivos para cada modelo indican que factores contribuyen para que la persona opte por trabajar en algún sector de trabajo (formal – informal) y parámetros estimados con valores negativos indican que

factores no contribuyen para que las personas opten por algún sector de trabajo (formal - informal).

Así, la probabilidad de que la persona escoja el sector formal dado el sector doméstico, tiene sus propios β 's estimados y sus $\exp(\beta)$, pero una interpretación más clara del significado de los parámetros puede obtenerse a partir del cálculo de sus $\exp(\beta)$, el cual es igual a los odds ratio (ecuación 3.22), a continuación una explicación de los mismos:

Para el **modelo 1** (Sector Formal dado Sector Doméstico), la variable aporte a una AFP que fue considerada como categórica (si aporta a una AFP) como de referencia, se puede observar que el valor es negativo $\beta(-3,401)$ para la categoría "no aporta a una AFP", indicando que disminuye la posibilidad de trabajar en el sector formal en comparación a las personas que aportan a una AFP, esto lo confirma el valor de $\exp(\beta)$ que es menor que 1 (0,033), este último cuantifica esta relación e indica que en el sector formal es menos probable que trabajen personas que no realizan su aporte a una AFP en comparación a los que si realizan el aporte. Así también se tiene para la variable horas de trabajo, donde se observa el valor negativo de $\beta(-0,026)$, es decir que a más horas de trabajo disminuye la posibilidad de trabajar en el sector formal, para estas dos variables con valores negativos en sus parámetros estimados (β), disminuye la posibilidad de trabajar en el sector formal, bajo estas características, aunque estas situaciones en su mayoría no podría ocurrir ya que en el sector formal las horas de trabajo y el aporte a una AFP está definido por el sector, es decir personas que trabaje más horas y que no aporte a una AFP, no podría darse..

Entre los valores positivos de β y mayores a 1 del $\exp(\beta)$ se puede observar en la variable años de estudio, donde se tiene que pertenecer al sector formal es más probable que pase en las personas que tienen más años de estudio que en personas de menos años de estudio ($\exp(\beta)=1,095$).

Otra variable que presenta valor positivo para β es el sexo de la persona, también se observa que el $\exp(\beta)$ es superior a 1, es decir los hombres tienen mayor probabilidad de trabajar en el sector formal respecto a las mujeres.

La variable condición de pareja, que presenta valor positivo para β , donde se observa valores de $\exp(\beta)=1,067$ superior a 1, es decir personas que tienen pareja aumenta la posibilidad de trabajar en el sector formal que en personas que no tienen pareja.

La variable identificación étnica es una variable que influye en la decisión de trabajar en el sector formal, es decir las personas que declararon identificarse con algún pueblo originario tiene más posibilidad de trabajar en este sector ($\exp(\beta)=1,453$) en comparación con las personas que declararon no identificarse con algún pueblo.

La variable ingreso también presenta valor positivo en su β , lo que indica a mayor ingreso mayor aumenta la probabilidad de trabajar en el sector formal, se debe tener en cuenta que el análisis de los odds ratios de las variables del modelo 1 es en comparación con la categoría de referencia (sector doméstico).

Para el **modelo 2** (Sector Informal dado Sector Doméstico), la variable aporte a una AFP que fue considerada como categórica (aporta a una AFP) como de referencia, se puede observar que el valor negativo de $\beta(-0,351)$ para la categoría “no aporta a una AFP”, indica que disminuye la posibilidad de trabajar en el sector informal en comparación a las personas que aportan a una AFP, esto confirma el $\exp(\beta)=0,704$ es menor que 1, este último cuantifica esta relación e indica que en el sector informal es menos probable que trabajen personas que no realizan su aporte a una AFP en comparación a los que si realizan el aporte.

Entre los valores positivos de β se puede observar en la variable edad de la persona el valor de $\exp(\beta)=1,035$ superior a la unidad, es decir a medida que aumenta la edad de una persona, se incrementa la probabilidad de pertenecer al sector informal.

En la variable años de estudio, se tiene que pertenecer al sector informal aumenta la probabilidad en personas que tienen algunos años de estudio que en personas con menos años de estudio el valor de $\exp(\beta)=1,060$.

Otra variable que presenta valor positivo para β es el sexo de la persona, también se observa que el $\exp(\beta)$ es superior a 1, es decir los hombres tienen mayor probabilidad de trabajar en el sector informal que las mujeres.

La variable condición de pareja presenta valor positivo para β , donde se observa valores de $\exp(\beta)=1,238$ superior a 1, es decir personas que tienen pareja aumenta la probabilidad de trabajar en el sector informal que en personas que no tienen pareja.

La variable identificación étnica que influye en la decisión de trabajar en el sector informal, es decir las personas que declaran identificarse con algún pueblo originario aumenta la probabilidad de trabajar en este sector ($\exp(\beta)=1,704$) en comparación con las personas que declararon no identificarse con algún pueblo originario.

La variable ingreso también es un variable que influye en la decisión de trabajar en el sector informal, es decir a mayor ingreso mayor probabilidad de trabajar en el sector informal en 1 vez más en comparación a percibir menores ingresos.

La variable horas de trabajo también es una variable que influye en la decisión de trabajar en el sector informal, es decir por un aumento de las horas de

trabajo aumenta la posibilidad de trabajar en este sector en 1,001 veces más en comparación si no aumentara las horas de trabajo. El análisis de los odds ratios de las variables del modelo 2 es en comparación con la categoría de referencia (sector doméstico).

Los parámetros estimados en cada regresión es representado por los betas para cada modelo y por sí solo no tienen interpretación solo se ve cuantas veces más aporta la variable con el valor de $\exp(\beta)$, para encontrar una interpretación intuitiva de los coeficientes, es importante calcular los efectos marginales, los cuales se calculan después de haber estimado los parámetros del modelo Logit Multinomial, si el signo es positivo indicara que un aumento en la variable independiente, aumenta la propensión a pertenecer a algún sector de trabajo, una ventaja es que no dependen de la categoría base.

A continuación se presenta la estimación de los efectos marginales de cada una de las variables explicativas sobre la probabilidad de pertenecer a cada categoría de sector de trabajo.

Tabla 11. EFECTOS MARGINALES

Variabes Independientes	Formal	Informal	Doméstico
Edad	-0,0080647	0,0081326	-0,0000679
Años de estudio	0,0077729	-0,0075338	-0,0002391
Sexo	0,0932922	-0,0539945	-0,0392977
Condición de pareja	-0,0358123	0,0363303	-0,0005180
Identificación étnica	-0,0375097	0,0389188	-0,0014092
Ingreso	0,0000000	-0,0000000	-0,0000000
Horas promedio de trabajo	-0,0064760	0,0064437	0,0000323
Aporte a la AFP	-0,6276421	0,6235550	0,0040871

Fuente: Elaboración Propia en Base a la EH-2014

La **tabla 11**, indica que un año de más de estudio, aumenta la probabilidad de pertenecer al **sector formal** en 0,7 puntos porcentuales, mientras que para el sector informal y doméstico disminuye en 0,7 y 0,02 puntos porcentuales respectivamente, lo que resulta obvio, es decir que para una persona con más

años de estudio aumenta la probabilidad de trabajar en el sector formal, en comparación con una persona con menos años de estudio.

Otra característica de la persona que hace que aumente la probabilidad de trabajar en el **sector formal** es el sexo, es decir los hombres tienen mayor probabilidad de trabajar en este sector, aumenta la probabilidad en 9,3 puntos porcentuales mientras en el sector informal y doméstico disminuye en 5,4 y 3,9 puntos porcentuales respectivamente, los hombres tienen más probabilidad de trabajar en el sector formal que las mujeres, se podría describir que para el sector formal las mujeres en edad fértil representan un gasto más, ya que por ley tienen derechos a beneficios como días de permiso cuando se encuentra embarazadas, en cambio los hombres no gozan de este beneficio y esto hace que los hombres tengan más probabilidad de ser aceptados en el sector formal.

Una característica del **sector formal** es el ingreso que pueda brindar este, el cual hace que aumente la probabilidad en la decisión de la persona para pertenecer al sector formal, es decir el ingreso en el sector formal es mensual un monto definido según la rama de actividad en cambio en el sector informal y doméstico el ingreso puede ser variante en el mes, por tanto para estos sectores disminuye la posibilidad de trabajar.

Para el **sector informal** la edad de la persona es un factor determinante para optar por este sector, es decir un año más de edad de la persona, aumenta la probabilidad de trabajar en el sector informal en 0,8 puntos porcentuales, en cambio para el sector formal y doméstico disminuye en 0,8 y 0,01 puntos porcentuales respectivamente, en otras palabras las personas adultas en edad de jubilación aumenta la probabilidad de estar ocupados en el sector informal, esto puede deberse a la fuerza de trabajo que pueden ejercer según la actividad económica.

Las personas que tienen pareja tienen mayor probabilidad de estar ocupados en el **sector informal**, es decir aumenta la probabilidad en 3,6 puntos porcentuales, en cambio para el sector formal disminuye en 3,6 puntos porcentuales, es decir las personas con pareja, tiene mayores responsabilidades para con su familia por tanto la única oportunidad de trabajo es en el sector informal, el cual servirá para satisfacer sus necesidades básicas de supervivencia, frente al desempleo.

Si la persona declara identificarse con algún pueblo originario o indígena aumenta la probabilidad en 3,8 puntos porcentuales para trabajar en el sector informal, en cambio para los sectores formal y doméstico disminuye en 3,7 y 0,1 puntos porcentuales respectivamente.

Una de las características propia de cada sector de trabajo es las horas de trabajo, pero para el sector informal no es así, ya que esa decisión de trabajar más o menos horas depende de la persona, así que por una hora de incremento de trabajo, aumenta la probabilidad en 0,6 puntos porcentuales de pertenecer al sector informal, con lo que se puede decir que si la persona está dispuesta a trabajar más horas es por tener más ingresos que satisfagan sus necesidades básicas de supervivencia, un ejemplo de una vendedora cuenta propia que vende en su kiosco obtendrá más ingresos si esta se queda más horas vendiendo, en cambio en el sector formal disminuye en 0,6 puntos porcentuales.

Otra características del sector informal es el aporte a una AFP, ya que en este sector no es obligatorio como en el sector formal, es decir, si una persona no aporta a una AFP aumenta la probabilidad de trabajar en este sector en 62,3 puntos porcentuales y disminuye la probabilidad de pertenecer al sector formal en 62,7 puntos porcentuales, esto se debe a que en el sector informal el aporte a una AFP no es obligatorio esto llama la atención de las personas a la hora de

decidir optar por este sector de trabajo, ya que el aporte significa realizar un gasto más y una disminución en su ingreso laboral.

Para el sector doméstico, el hecho de no aportar a una AFP aumenta la probabilidad de trabajar en el sector en 0,4 puntos porcentuales, esto se debe a que en este sector no es obligatorio el aporte a una AFP y disminuye la probabilidad de pertenecer al sector formal en 62,7 puntos porcentuales.

Bajo estas características de la persona y del sector, mencionadas anteriormente, se tiene las siguientes probabilidades de trabajar en cada sector de trabajo:

Probabilidad cuando $Y = 0 = \text{formal} = 0,40688738$

Probabilidad cuando $Y = 1 = \text{informal} = 0,5897640$

Probabilidad cuando $Y = 2 = \text{doméstico} = 0,00334862$

Tabla 12. CLASIFICACIÓN

Observado	Pronosticado			
	Sector formal	Sector informal	Sector doméstico	Porcentaje correcto
Sector formal	642.787	55.0254	0	53,9
Sector informal	92.831	1.604.216	0	94,5
Sector domestico	1.821	92.817	0	0,0
Porcentaje global	24,7	75,3	0,0	75,3

Fuente: Elaboración Propia en Base a la EH-2014

La **tabla 12** de clasificación, es un camino para determinar que también nuestro modelo ajusta los datos, es comparar las predicciones con los resultados observados, los datos de la diagonal principal son predicciones correctas, los datos fuera de la diagonal principal son predicciones incorrectas, por tanto se tiene que 642.782 personas que escogieron el sector formal fueron correctamente clasificados por el modelo, similarmente, 1.604.216 personas fueron correctamente clasificados en el sector informal. Los datos fuera de la diagonal de la tabla nos indican cuantas personas fueron incorrectamente

clasificadas. En conjunto el 53,9 por ciento de los casos se clasificaron en el sector formal, de las personas del informal el 94,5 por ciento fueron clasificadas correctamente. Del total el 75,3 por ciento de las 2.247.003 personas tomadas para ser analizados fueron correctamente clasificados.

El modelo, pues resulta excelente identificando a los trabajadores que prefieren en sector informal. Sin embargo, no logra clasificar a los trabajadores que prefieren el sector doméstico. Aquí la solución pareciera indicar, a primera vista, que se necesita realizar una investigación mayor para encontrar otro(s) predictor(es) pero, indagando un poco más, la explicación parece ser otra y es la homogeneidad de la muestra, es decir, muy pocos trabajadores trabajan en el sector doméstico de modo que el modelo no los clasifica correctamente.

Capítulo VI.

Conclusiones y Recomendaciones

6.1. Conclusiones

El documento responde los objetivos planteados, se adopta una definición operativa que clasifica los sectores de trabajo, donde se encuentran el sector informal, formal y doméstico, previamente se revisó una serie de documentos del tema de informalidad, para así definir los sectores de trabajo, a partir de esta variable se realiza el análisis respectivo.

Con la aplicación del modelo Logit Multinomial, se determina que el modelo ajusta adecuadamente a los datos de la base de la Encuesta de Hogares 2014, se obtienen las características (variables) más significativas que hacen que las personas opten por alguno de estos sectores de trabajo, estas características llegan a ser de tipo sociodemográficas y económicas.

Los parámetros estimados en cada modelo es representado por los betas, los cuales por sí solo no tienen interpretación solo se ve cuantas veces más aporta la variable con el valor de $\exp(\beta)$. Por tanto se tiene:

Para el **modelo 1** (Sector Formal dado Sector Doméstico), la variable aporte a una AFP que fue considerada como categórica (si aporta a una AFP) como de referencia, se tuvo un valor negativo de $\beta(-3,401)$ para la categoría “no aporta a una AFP”, este indica que disminuye la posibilidad de trabajar en el sector formal en comparación a las personas que si aportan a una AFP, lo que confirma que el $\exp(\beta)= 0,033$ es menor que 1, además indica que en el sector formal es menos probable que trabajen personas que no realizan su aporte a una AFP en comparación a los que si realizan el aporte a una AFP. Así también

se tiene para la variable horas de trabajo, un valor negativo de $\beta = -0,026$, es decir que a más horas de trabajo disminuye la posibilidad de trabajar en el sector formal, para estas dos variables con valores negativos en sus β , disminuye la posibilidad de trabajar en el sector formal. Pero las horas de trabajo y el aporte a una AFP están definidos por el sector formal, así que no puede darse las situaciones que trabajen más horas y que no aporte a una AFP.

Entre los valores positivos de β y mayores a 1 del $\exp(\beta)=1,095$ se tiene la variable años de estudio, donde pertenecer al sector formal es más probable que pase en las personas que tienen más años de estudio que en personas de menos años de estudio.

Otra variable que presenta valor positivo para β es el sexo de la persona, el valor de la $\exp(\beta)$ es superior a 1, es decir los hombres tienen mayor probabilidad de trabajar en el sector formal que las mujeres.

La otra variable que presenta valor positivo para β es condición de pareja, el valor de la $\exp(\beta)$ es superior a 1, es decir personas que tienen pareja tiene más posibilidad de trabajar en el sector formal ($\exp(\beta)=1,067$), que en personas que no tienen pareja.

La variable identificación étnica influye de alguna manera en la decisión de trabajar en el sector formal, es decir las personas que declararon identificarse con algún pueblo originario tiene más posibilidad de trabajar en este sector ($\exp(\beta)=1,453$) en comparación con las personas que declararon no identificarse con ningún pueblo originario. Y por último la variable ingreso es una variable que influye en la decisión de trabajar en el sector formal, es decir a mayor ingreso mayor probabilidad de pertenecer a este sector.

El análisis de $\exp(\beta)$ = **odds ratios** de las variables independiente del modelo 1 es en comparación con la categoría de referencia (sector doméstico).

Para el **modelo 2** (Sector Informal dado Sector Doméstico), la variable aporte a una AFP que fue considerada como categórica (si aporta a una AFP) como de referencia, tiene un valor negativo de $\beta(-0,351)$ para la categoría “no aporta a una AFP”, lo que indica que disminuye la posibilidad de trabajar en el sector informal en comparación a las personas que aportan a una AFP, esto confirma que el valor de $\exp(\beta)=0,704$ es menor que 1 e indica que en el sector informal es menos probable que trabajen personas que no realizan su aporte a una AFP en comparación a los que si realizan el aporte a una AFP.

Entre los valores positivos de β y mayores a 1 del $\exp(\beta)=1,035$ se observa en la variable edad de la persona, es decir a medida que aumenta la edad de la persona, existe mayor probabilidad de pertenecer al sector informal.

Para la variable años de estudio, se tiene que pertenecer al sector informal es más probable que pase en las personas que tienen algunos años de estudio ($\exp(\beta)=1,060$) que en personas que tienen menos años de estudio.

Otra variable que presenta valor positivo para β es el sexo de la persona, el valor de $\exp(\beta)$ es superior a 1, es decir los hombres tienen mayor probabilidad de trabajar en el sector informal que las mujeres.

La variable condición de pareja presenta valor positivo para β el valor de $\exp(\beta)=1,238$ superior a 1, indica que las personas que tienen pareja tiene más posibilidad de trabajar en el sector informal que en personas que no tienen pareja.

La variable identificación étnica influye en la decisión de trabajar en el sector informal, es decir las personas que declararon identificarse con algún pueblo originario aumenta la posibilidad de trabajar en este sector en 1,704 veces más en comparación con las personas que declararon no se identificarse con ningún pueblo originario.

La variable ingreso también aumenta la posibilidad de trabajar en el sector informal en 1 vez más, es decir al incrementar el ingreso, también aumenta la probabilidad de estar ocupado en este sector en comparación a un menor ingreso percibido.

La variable horas de trabajo aumenta la posibilidad de trabajar en el sector informal en 1,001 veces más, es decir por un aumento de trabajo en este sector aumenta también la probabilidad en comparación a que no aumentara las horas de trabajo.

En **resumen** los resultados de los dos modelos (formal - informal) dado el sector de referencia (doméstico), del conjunto de atributos personales, la variable sexo resulto relevante en la explicación de los motivos de la búsqueda desde una condición de ocupado, es decir los hombres buscan incorporarse en el sector formal e informal en comparación a las mujeres.

En cuanto a la edad, esta variable es relevante en la determinación de trabajar en el sector informal, a mayor edad mayor probabilidad de trabajar en el sector informal.

Si se refiere a los años de estudio esta variable es relevante en la determinación de trabajar en el sector formal e informal.

Para las personas que tienen pareja (casadas o convivientes) es un factor determinante al momento de optar por trabajar en el sector formal o informal en comparación a las personas que no tienen pareja.

Si la persona declara identificarse con algún pueblo originario o indígena, este factor es determinante, propio de la persona que elige trabajar en el sector formal o informal.

El ingreso es determinante para poder optar por alguno de estos sectores de trabajo, el hecho está que según la actividad económica en la que trabaje la persona llegara a percibir un ingreso bajo o alto, así como diario o mensual, el cual le será útil para cubrir sus gastos necesarios para su subsistencia.

Otra característica propia de cada sector es el aporte a una AFP, es decir tanto para el sector formal e informal, el aporte es un factor determinante a la hora de optar por alguno de estos sectores de trabajo.

Las horas de trabajo es un factor determinante propio de la persona que elige trabajar en el sector informal, es decir la persona optara trabajar más horas ya que esto significara mayores ingresos.

De todo lo anterior se lograron identificar diferencias en el comportamiento de los sectores de trabajo de las personas, debido a los atributos personales como sexo, edad, condición de pareja, años de estudio e identificación con algún pueblo originario, son factores determinantes a la hora de decidir incorporarse en algún sector de trabajo. Con relación a los atributos del sector de trabajo, las variables que influyen a la hora de decidir por algún sector son el ingreso, las horas promedio de trabajo y la aportación a una AFP.

Pero al analizar el valor de los odds ratio solo se determinó que variables son factores que determinan la decisión para optar por algún sector y en cuanto es más probable que suceda, además se debe recordar que estas están en comparación del sector de referencia (sector doméstico), en cambio al analizar los efectos marginales de cada una de las variables independientes X sobre pertenecer a cada categoría de Y, con el cual se tiene un mejor análisis.

Características del sector formal

Los efectos marginales del modelo Logit Multinomial, revelaron que las variables: edad, condición de pareja e identificación étnica de la persona no son importantes en la determinación de trabajar en el sector formal, lo cual coincide con los cuadros y gráficos descriptivos del capítulo IV, donde indica un porcentaje menor de personas que trabajan en el sector formal con una edad promedio menor que de las personas del sector informal, al igual que las personas con pareja se tiene un menor porcentaje trabajando en el sector formal y las personas que declararon pertenecer o identificarse con algún pueblo originario, representan un menor porcentaje trabajando en este sector.

En cuanto a la característica propia de este sector como las horas de trabajo y el aporte a una AFP, hace que disminuya la probabilidad de trabajar en el sector formal, es decir, las horas de trabajo de este sector ya están establecidas que es de 8 horas día, por lo que a la persona no le llamaría la atención para trabajar en este sector y no le alcanzaría el tiempo para poder tomar una segunda actividad. La otra característica propia del sector es el aporte a la AFP, de hecho en el sector formal (población de la administración pública y establecimientos con más de 5 personas), el aporte a una AFP es obligatorio, por lo que a la persona que opte por trabajar en este sector no le quedará otra que aportar, estas características del sector disminuyen la probabilidad de trabajar en el sector formal, uno porque en este sector los horarios de trabajo ya están definidos y lo segundo es que el aporte a una AFP es obligatorio, y esto

significa una disminución en el ingreso laboral, este análisis también coincide con las salidas de los cuadros y gráficos del capítulo IV, donde se tiene la diferencia de las horas de trabajo entre sectores, en el sector formal las horas de trabajo es de 8 hora por día en promedio y el 87,6 por ciento de los ocupados en el sector formal realizan el aporte a una AFP.

Mientras que las variables **años de estudio, sexo e ingreso**, estas mantienen una relación directa con la probabilidad de pertenecer al sector formal, es decir a mayor años de estudio mayor probabilidad de pertenecer al sector formal, al igual que para los hombres, los cuales tienen más probabilidad de trabajar en el sector formal que las mujeres, y el ingreso que pueda ofrecer este sector también aumenta la probabilidad de pertenecer al sector formal, es decir que para el sector formal las mujeres en edad fértil visto desde este punto, ellas representan al sector un gasto más, por ejemplo por ley tienen beneficios como al seguro y días de permiso cuando se encuentra embarazadas, en cambio los hombres no gozan de algunos de estos beneficios y esto hace que los hombres tengan más probabilidad de ser aceptados en el sector formal, el ingreso que se pueda percibir en este sector aumenta la probabilidad de pertenecer a este sector. Este análisis coincide con los cuadros en el capítulo IV donde se tiene que la mayor parte de las personas que pertenecen al sector formal posee un grado de escolaridad alto, el cual supera la educación secundaria, lo que convierte en una ventaja para estos, ya que aumenta la probabilidad de emplearse en el sector formal al igual que una parte de la población masculina pertenece al sector formal y el ingreso en promedio en el sector formal es un poco más alto en comparación a los demás sectores.

Características del sector Informal

Los efectos marginales del modelo Logit Multinomial, revelaron que las variables años de estudio, sexo de la persona no son importantes en la determinación de trabajar en el sector informal, lo cual coincide con los cuadros

y gráficos descriptivos del capítulo IV, la mayor parte de las personas que pertenecen al sector informal posee un grado de escolaridad bajo, el cual no supera la educación básica ni secundaria, lo que convierte en una desventaja para estos, ya que disminuye la probabilidad de emplearse en el sector formal.

Mientras que las variables: **edad, condición de pareja, identificación étnica** mantienen una relación directa con la probabilidad de pertenecer al sector informal, estos resultados también coinciden con el análisis de los cuadros descriptivos ya que la mayor parte de la población que trabaja en el sector informal tiene una edad mayor a los 44 años, se tiene mayor porcentaje de personas con pareja y la declaración de identificación con algún pueblo originario de las personas trabajando en el sector informal, de lo anterior se puede decir que las personas mayores de 44 años ya no cuentan con la fuerza suficiente para trabajar y esto hace que vean al sector informal como la única opción de trabajo y mucho más para las personas en edad de jubilación, al igual que las personas que tienen pareja, ya que las mismas cuentan con responsabilidades y además complementar los ingresos de su hogar, tienen la necesidad de trabajar sin importar el sector.

Las variables **horas de trabajo y el aporte a una AFP**, son características propias del sector formal, pero no así del sector informal, ya que trabajar menos o más horas y aportar o no a una AFP es decisión propia de la persona, en este sector. Por tanto a más horas de trabajo aumenta la probabilidad de trabajar en el sector informal. En contraste, en el sector informal (establecimientos con menos de 5 personas y trabajadores por cuenta propia y trabajadores familiares sin remuneración) en su mayoría no es obligatorio el aporte a una AFP, si la persona no quiere hacer un gasto más, la persona optará por trabajar en el sector informal, por tanto esta variable también es un factor determinante a la hora de elegir este sector, estos resultados coinciden con los datos de los cuadros y gráficos estadísticos en el capítulo IV, según el sector informal en

promedio se trabajan más horas en comparación al sector formal y en el sector informal en la mayoría de las actividades económicas no realizan el aporte a una AFP.

Por tanto el sector informal puede ser visto como válvula de escape para las personas con menos años de estudio, es decir vista como una posibilidad de trabajo frente al desempleo.

Otro de los objetivos planteados en este documento fue analizar los cuadros y gráficos descriptivos. Desde este punto de vista, también se puede ver las características que tienen estos sectores formal, informal y doméstico como ser de carácter demográfico y económico. Para el 2014, entre la población mayor de 44 años y las mujeres tuvieron mayor participación en el sector informal, además la población femenina tiene un nivel educativo más bajo con relación a los hombres, la mayor parte de ellas tienen un lugar de trabajo precario, venden en las calles, como ambulantes o venden en quioscos o puesto fijo; si se observa por departamento, el mayor nivel de informalidad está en el departamento de Cochabamba, seguido de La Paz y Santa Cruz de la Sierra.

Con respecto a las características propias del sector de trabajo, en el sector informal no ofrece a los ocupados el seguro de salud, tampoco cuentan con el NIT (Número de Identificación Tributaria).

También se advierte que la mayoría de las actividades económicas el personal que trabaja no están afiliadas para el aporte a una AFP, esto ocurre en el sector informal.

En el sector informal en promedio se trabaja más horas en comparación al del sector formal.

Los años promedio de estudio también es clara entre sectores, el sector que tiene trabajadores con menos años de estudio es el doméstico y el sector informal, en cambio el sector formal cuenta con trabajadores con mayores años de estudio.

Referente a los ingresos percibidos por los trabajadores de cada sector, en promedio, el sector menos remunerado es el doméstico y el sector informal, contrariamente en el sector formal en promedio es el que mejores ingresos percibe.

La aplicación del modelo Logit Multinomial en los datos de la base de la EH-2014 y el análisis descriptivo que se realizó, ambos se complementan, además se muestra que el modelo tiene la capacidad de predecir que las personas bajo estas características opten por algún sector de trabajo.

Lo que más resalta entre los resultados son los años de estudio de las personas, es decir mientras una persona tenga mayores años de estudio, mayor nivel de estudio, este optará por un trabajo en el sector formal; en cambio, una persona con un nivel educativo bajo le quedara trabajar en el sector informal, para satisfacer sus necesidades básicas en su hogar.

6.2. Recomendaciones

Se recomienda hacer el análisis del modelo Logit Multinomial por año utilizando las bases de las Encuestas de Hogares, manteniendo las mismas variables de estudio, para así realizar la comparación correspondiente, analizar los factores que determinan en la decisión de optar por trabajar en algún sector de trabajo.

Optar por el análisis de otras variables para el modelo Logit Multinomial, además de realizar el análisis descriptivo.

Analizar el estudio con otras bases, como ser la base de la Encuesta Continua de Empleo, ya que dicha encuesta tienen características propias del empleo.

Aplicar otro método multivariante como ser, el análisis discriminante, para clasificar a las personas ocupadas en cada sector de trabajo, respecto a las variables independientes, para luego estudiar las diferencias que existen de cada grupo.

Se recomienda además analizar otros modelos para variables de estas características, como ser el modelo Logit anidado o el Probit Multinomial, para así compararlo con el Modelo Logit Multinomial, manteniendo las mismas variables para los modelos, para así analizar los cambios en los parámetros si es que hubiesen, también realizar sus pruebas de validación propias en cada modelo, y así ver cual modelo se ajusta mejor a los datos.

ANEXO

Índice de anexo

<i>Anexo 1. Memoria institucional del Instituto Nacional de Estadística</i>	<i>111</i>
<i>Anexo 2. Definición de informalidad propuesta por la Organización Internacional del Trabajo</i>	<i>112</i>
<i>Anexo 3. Estimación por el método de Máxima Verosimilitud.....</i>	<i>114</i>
<i>Anexo 4. Prueba de Multicolinealidad de las variables del modelo.....</i>	<i>116</i>
<i>Anexo 5. Pruebas estadísticas de asociación entre la variable dependiente y las variables independientes</i>	<i>126</i>
<i>Anexo 6. Salidas del modelo Logit Multinomial</i>	<i>129</i>
<i>Anexo 7. Suficiencia Académica del Trabajo Dirigido</i>	<i>133</i>
<i>Anexo 8. Sintaxis del Modelo Logit Multinomial</i>	<i>136</i>

ANEXOS

Anexo 1. Memoria institucional del Instituto Nacional de Estadística

El origen del Instituto Nacional de Estadística se remonta el año 1863, cuando se creó una sección en el Ministerio de Hacienda llamada Mesa Estadística. El año 1869 se denominó Oficina Nacional de Inmigración, Estadística y Propaganda Geográfica. Pasando luego a ser conocida como Oficina de Estadística y Presupuestos dentro del Ministerio. El 14 de Enero de 1936, en la Presidencia de José Luís Tejada, se fundó la Dirección General de Estadística y Censos, el 30 de abril de 1970 en el Instituto Nacional de Estadística (INE).

La Ley del Sistema Nacional de Información Estadística (SNIE) – Decreto Ley N° 14100, de 5 de noviembre de 1976, confiere al INE la responsabilidad de dirigir, planificar, ejecutar, controlar y coordinar las actividades estadísticas del sistema; promover el uso de registros administrativos, tanto en oficinas públicas como privadas para obtener datos estadísticos; además de capacitar recursos humanos y creando conciencia estadística nacional. El INE es el Órgano Ejecutivo del Sistema Nacional de Información Estadística de Bolivia, tiene las funciones de: relevar, clasificar, codificar, compilar y difundir, con carácter oficial, la información estadística del país.

El Decreto Ley N° 14100 es el instrumento legal que norma el funcionamiento actual del INE.

Desde 1978, el Instituto Nacional de Estadística ha venido realizando varias rondas de encuestas a hogares: Encuesta Permanente de Hogares (EPH), Encuesta Integrada de Hogares (EIH), Encuesta Nacional de Empleo y Encuesta Continua de Hogares del Programa MECOVI, entre los años 2003 y 2004 se ejecutó la Encuesta Continua de Hogares con características similares a una encuesta de Presupuestos Familiares, conservándose los tradicionales contenidos de la encuesta, haciendo énfasis en la temática de ingresos y gastos del hogar; para los años 2005 y 2006 se retomó la modalidad de encuestas puntuales de hogares, con la ejecución de la Encuesta de Hogares, que mantuvo como temática de exploración a las condiciones de vida de la población boliviana, a través de la aplicación del cuestionario multitemático que permite la investigación de las características generales, salud, educación, empleo, ingresos, gastos, vivienda y servicios básicos de los hogares.

Anexo 2. Definición de informalidad propuesta por la Organización Internacional del Trabajo

El **cuadro A1** que se muestra a continuación, proporciona el marco propuesto por la Organización Internacional del Trabajo (Herrera J., 2004) que hace posible desglosar el total de empleo de acuerdo con dos dimensiones: el tipo de unidad de producción (filas de la matriz) y el tipo de puesto de trabajo (columnas de la matriz). En la matriz, las unidades de producción están agrupadas en función del tipo, y se distingue entre empresas del sector formal, empresas del sector informal y hogares. Según los criterios adoptados por la 15ª Conferencia Internacional de Estadísticos del Trabajo (CIET), las empresas informales se definen como empresas privadas no constituidas como sociedad por debajo de un cierto tamaño en términos de empleo y/o no registradas en virtud de las leyes comerciales o relativas a las empresas. Los hogares considerados unidades de producción son hogares que producen bienes para su propio uso final (Por ejemplo, granjas de subsistencia, autoconstrucción), así como los que emplean trabajadores domésticos remunerados (criados, lavanderas, jardineras, vigilantes, conductores, etc.)

Cuadro A.1. MATRIZ: UN MARCO CONCEPTUAL PARA LA ECONOMÍA INFORMAL

Unidades de producción por tipo	Puestos de trabajo según la situación								
	Trabajadores independientes		Empleadores		Trabajadores familiares auxiliares	Asalariados		Miembros de cooperativas de productores	
	Informal	Formal	Informal	Formal	Informal	Informal	Formal	Informal	Formal
Empresas del sector formal					1	2			
Empresas del sector informal	3		4		5	6		8	
Hogares	9					1			

Fuente: Haussmanns, (2001). *Informal sector and informal employment: elements of a conceptual framework*. Documento presentado a la quinta reunión del Grupo de Expertos sobre estadísticas del sector informal (Grupo Delhi). Nueva Delhi, 19-21 de septiembre de 2001.

Nota: Según la definición de la Decimoquinta Conferencia Internacional de Estadísticos del Trabajo en 1993; (b) Hogares que producen bienes para su propio uso final y hogares que emplean a trabajadores domésticos.

Las casillas oscuras hacen referencia a los puestos de trabajo que, por definición, no existen en el tipo de unidad de producción que se trate.

Las casillas con líneas hacen referencia a los puestos de trabajo que existen en el tipo de unidad de producción de que se trate pero no son relevantes para los intereses de este estudio.

Las casillas blancas son el punto central de este estudio; se refieren a los tipos de puestos de trabajo que representan los diferentes segmentos de la economía informal.

Casilla 1 y 5: Trabajadores familiares auxiliares: sin contrato ni protección jurídica o social que se derive del propio puesto, en empresas formales (casilla 1) o empresas informales (casilla 5). (Los trabajadores informales con contrato de trabajo, sueldo, protección social, etc., se considerarían trabajadores en el empleo formal.)

Casillas 2, 6 y 10: Trabajadores que tienen puestos de trabajo informales, ya se encuentren en empresas formales (casilla 2) o informales (casilla 6) o se trate de trabajadores domésticos remunerados empleados en hogares (casilla 10).

Casillas 3 y 4: Trabajadores independientes (casilla 3) y empleadores (casilla 4) que poseen sus propias empresas informales. El carácter informal de sus trabajos se desprende directamente de las características de las empresas que poseen.

Casilla 7: Trabajadores empleados en empresas informales pero en puestos formales. (Ello puede ocurrir, por ejemplo, cuando las empresas se definen como informales basándose únicamente en el criterio del tamaño).

Casilla 8: Miembros de cooperativas de productores informales.

Casilla 9: Productores de bienes para su propio uso final en sus hogares (por ejemplo, granjas de subsistencia).

Anexo 3. Estimación por el método de Máxima Verosimilitud

El método de Máxima Verosimilitud (en inglés *Method of Maximum Likelihood*) es un procedimiento que permite estimar los parámetros de un modelo probabilístico de tal manera que sean los más probables a partir de los datos obtenidos. Por ello es útil al comparar diferentes modelos, incluyendo o no variables en el mismo; la forma adecuada de comparar estas diferencias es mediante cocientes, y es por ello que aparecen en los resultados de los programas, el término *likelihood ratio* (Greene W., 1993)

El método de Máxima Verosimilitud se utiliza, por ejemplo, para estimar los coeficientes de un modelo logístico de regresión (Logit). Sea P la probabilidad de que ocurra un determinado suceso, $P = f(\beta x)$, entonces la probabilidad de que el suceso no ocurra será $1-P$. Por tanto, en los sujetos en los que ocurrió el suceso la probabilidad vendrá dada por $P(x_i)$ mientras que para un sujeto en el que no ocurre el suceso, se calcula como $1-P(x_i)$, siendo ambas expresiones función de β .

Si la muestra es aleatoria y las observaciones son independientes entre sí, la probabilidad de que un sujeto de la muestra experimente el suceso es independiente de lo que ocurra a cualquier otro, por lo que la probabilidad de todo el evento se calcula como el producto de las probabilidades individuales y ese producto es la función verosimilitud, que es función únicamente de los coeficientes (β 's). La función de verosimilitud se maximiza derivando e igualando a cero.

Para el caso del modelo Logit Multinomial la función logaritmo de verosimilitud se puede obtener definiendo, para cada individuo, $d_{ij} = 1$ si el individuo i escoge la opción j , y "0" si no, para los $J+1$ posibles resultados. Así, para cada i , sólo uno de los d_{ij} es igual a la unidad.

La función logaritmo de la verosimilitud es:

$$\ln L = \sum_i \sum_{j=0}^J d_{ij} \ln \text{Prob}(y_i = j)$$

Derivando con respecto a los parámetros β 's:

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_i (d_{ij} - P_{ij}) x_i, \text{ Para } j = 1, 2, \dots, J.$$

La matriz de las segundas derivadas (matriz Hessiana) tiene J^2 bloques, cada uno de tamaño $k \times k$:

La j -ésima diagonal es:

$$\sum_i -P_{ij}(1 - P_{ij})x_i x_i'$$

Fuera de la diagonal se tiene:

$$\sum_i (P_{ij}P_{ik})x_i x_i'$$

Como la matriz Hessiana no involucra el término d_{ij} , estos son los valores esperados.

En el cálculo de la función de verosimilitud interviene el producto de las probabilidades individuales, por lo que se toman logaritmos. Es por ello que en los resultados de los programas de computación se tiene el término *Log-likelihood* (Logaritmo de Verosimilitud). Al tratarse de productos de probabilidades, la función de verosimilitud es menor que 1 y por tanto su logaritmo es negativo.

Anexo 4. Prueba de Multicolinealidad de las variables del modelo.

Pruebas estadísticas para variables cuantitativas

Matriz de correlaciones entre las variables cuantitativas

	s2a_03	e	s6b_23h	ylab
s2a_03	1.0000			
e	-0.1384	1.0000		
s6b_23h	0.0256	-0.1138	1.0000	
ylab	0.0421	0.1281	0.0739	1.0000

Pruebas estadísticas para variables cualitativas

Entre la variable sexo con condición de pareja

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,136	0,000
	V de Cramer	0,136	0,000
	Coeficiente de contingencia	0,135	0,000
N de casos válidos		2984726	

Entre la variable sexo con identificación étnica

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,002	0,001
	V de Cramer	0,002	0,001
	Coeficiente de contingencia	0,002	0,001
N de casos válidos		2984726	

Entre la variables sexo con AFP

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	-0,017	0,000
	V de Cramer	0,017	0,000
	Coeficiente de contingencia	0,017	0,000
N de casos válidos		2984726	

Entre la variable condición de pareja con identificación étnica

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,137	0,000
	V de Cramer	0,137	0,000
	Coeficiente de contingencia	0,136	0,000
N de casos válidos		2984726	

Entre la variable condición de pareja con AFP

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	-0,030	0,000
	V de Cramer	0,030	0,000
	Coeficiente de contingencia	0,030	0,000
N de casos válidos		2984726	

Entre la variable identificación étnica con AFP

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,071	0,000
	V de Cramer	0,071	0,000
	Coeficiente de contingencia	0,071	0,000
N de casos válidos		2984726	

Pruebas estadísticas para variables cualitativas y cuantitativas

Entre la variable edad (en grupos) con sexo

Medidas direccionales

			Valor
Nominal por intervalo	Eta	grupos de edad dependiente	0,038
		sexo dependiente	0,041

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,036	0,001	-65,254	0,000
	Tau-c de Kendall	-0,039	0,001	-65,254	0,000
	Gamma	-0,065	0,001	-65,254	0,000
	Correlación de Spearman	-0,038	0,001	-65,064	0,000 ^c
Intervalo por intervalo	R de persona	-0,038	0,001	-66,394	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable edad (en grupos) con condición de pareja

Medidas direccionales

			Valor
Nominal por intervalo	Eta	grupos de edad dependiente	0,267
		Condición de pareja dependiente	0,326

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,245	0,001	440,408	0,000
	Tau-c de Kendall	0,263	0,001	440,408	0,000
	Gamma	0,426	0,001	440,408	0,000
	Correlación de Spearman	0,257	0,001	459,989	0,000 ^c
Intervalo por intervalo	R de persona	0,267	0,001	479,413	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable edad (en grupos) con identificación étnica

Medidas direccionales

			Valor
Nominal por intervalo	Eta	grupos de edad dependiente	0,079
		Se identifica con algún pueblo dependiente	0,081

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,074	0,001	136,600	0,000
	Tau-c de Kendall	0,072	0,001	136,600	0,000
	Gamma	0,152	0,001	136,600	0,000
	Correlación de Spearman	0,078	0,001	134,584	0,000 ^c
Intervalo por intervalo	R de persona	0,079	0,001	136,533	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable edad (en grupos) con AFP

Medidas direccionales

			Valor
Nominal por intervalo	Eta	grupos de edad dependiente	0,082
		Aporte a la AFP dependiente	0,137

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,072	0,001	- 140,401	0,000
	Tau-c de Kendall	-0,069	0,000	- 140,401	0,000
	Gamma	-0,153	0,001	- 140,401	0,000
	Correlación de Spearman	-0,076	0,001	- 131,047	0,000 ^c
Intervalo por intervalo	R de persona	-0,082	0,001	- 141,318	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable años de estudio (en grupos) con sexo

Medidas direccionales

			Valor
Nominal por intervalo	Éta	a_est_g dependiente sexo dependiente	0,084 0,154

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,062	0,001	117,862	0,000
	Tau-c de Kendall	0,077	0,001	117,862	0,000
	Gamma	0,099	0,001	117,862	0,000
	Correlación de Spearman	0,069	0,001	120,232	0,000 ^c
Intervalo por intervalo	R de persona	0,084	0,001	145,976	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable años de estudio (en grupos) con condición de pareja

Medidas direccionales

			Valor
Nominal por intervalo	Éta	a_est_g dependiente Condición de pareja dependiente	0,080 0,090

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,073	0,001	- 141,810	0,000
	Tau-c de Kendall	-0,089	0,001	- 141,810	0,000
	Gamma	-0,120	0,001	- 141,810	0,000
	Correlación de Spearman	-0,082	0,001	- 141,925	0,000 ^c
Intervalo por intervalo	R de persona	-0,080	0,001	- 138,093	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable años de estudio (en grupos) con identificación étnica

Medidas direccionales

		Valor
Nominal por intervalo	Eta a_est_g dependiente	0,179
	Se identifica con algún pueblo dependiente	0,185

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,160	0,001	- 313,671	0,000
	Tau-c de Kendall	-0,176	0,001	- 313,671	0,000
	Gamma	-0,286	0,001	- 313,671	0,000
	Correlación de Spearman	-0,179	0,001	- 314,246	0,000 ^c
Intervalo por intervalo	R de persona	-0,179	0,001	- 314,404	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable años de estudio (en grupos) con AFP

Medidas direccionales

		Valor
Nominal por intervalo	Eta a_est_g dependiente	0,338
	Aporte a la AFP dependiente	0,453

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,329	0,000	- 618,979	0,000
	Tau-c de Kendall	-0,356	0,001	- 618,979	0,000
	Gamma	-0,589	0,001	- 618,979	0,000
	Correlación de Spearman	-0,368	0,001	- 682,752	0,000 ^c
Intervalo por intervalo	R de persona	-0,338	0,001	- 619,529	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable horas de trabajo (en grupos) con sexo

Medidas direccionales

			Valor
Nominal por intervalo	Eta	horas de trabajo por grupos dependiente	0,190
		sexo dependiente	0,225

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,173	0,001	316,031	0,000
	Tau-c de Kendall	0,195	0,001	316,031	0,000
	Gamma	0,298	0,001	316,031	0,000
	Correlación de Spearman	0,183	0,001	321,209	0,000 ^c
Intervalo por intervalo	R de persona	0,190	0,001	333,573	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable horas de trabajo (en grupos) con condición de pareja

Medidas direccionales

			Valor
Nominal por intervalo	Eta	horas de trabajo por grupos dependiente	0,082
		Condición de pareja dependiente	0,084

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,078	0,001	144,069	0,000
	Tau-c de Kendall	0,087	0,001	144,069	0,000
	Gamma	0,141	0,001	144,069	0,000
	Correlación de Spearman	0,083	0,001	143,780	0,000 ^c
Intervalo por intervalo	R de persona	0,082	0,001	142,507	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable horas de trabajo (en grupos) con identificación étnica

Medidas direccionales

			Valor
Nominal por intervalo	Eta	horas de trabajo por grupos dependiente	0,030
		Se identifica con algún pueblo dependiente	0,075

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,032	0,001	56,058	0,000
	Tau-c de Kendall	0,031	0,001	56,058	0,000
	Gamma	0,063	0,001	56,058	0,000
	Correlación de Spearman	0,033	0,001	57,698	0,000 ^c
Intervalo por intervalo	R de persona	0,030	0,001	51,427	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable horas de trabajo (en grupos) con AFP

Medidas direccionales

			Valor
Nominal por intervalo	Eta	horas de trabajo por grupos dependiente	0,102
		Aporte a la AFP dependiente	0,192

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,104	0,001	205,075	0,000
	Tau-c de Kendall	0,103	0,001	205,075	0,000
	Gamma	0,209	0,001	205,075	0,000
	Correlación de Spearman	0,110	0,001	191,918	0,000 ^c
Intervalo por intervalo	R de persona	0,102	0,001	176,911	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable ingreso (en grupos) con sexo

Medidas direccionales

			Valor
Nominal por intervalo	Eta	ingreso por grupos dependiente sexo dependiente	0,247 0,261

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,233	0,001	439,805	0,000
	Tau-c de Kendall	0,265	0,001	439,805	0,000
	Gamma	0,395	0,001	439,805	0,000
	Correlación de Spearman	0,247	0,001	440,253	0,000 ^c
Intervalo por intervalo	R de persona	0,247	0,001	440,446	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable ingreso (en grupos) con condición de pareja

Medidas direccionales

			Valor
Nominal por intervalo	Eta	ingreso por grupos dependiente Condición de pareja dependiente	0,173 0,173

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,163	0,001	305,661	0,000
	Tau-c de Kendall	0,182	0,001	305,661	0,000

	Gamma	0,288	0,001	305,661	0,000
	Correlación de Spearman	0,173	0,001	303,628	0,000 ^c
Intervalo por intervalo	R de persona	0,173	0,001	303,617	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable ingreso (en grupos) con identificación étnica

Medidas direccionales

			Valor
Nominal por intervalo	Eta	ingreso por grupos dependiente	0,046
		Se identifica con algún pueblo dependiente	0,046

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,043	0,001	-79,169	0,000
	Tau-c de Kendall	-0,043	0,001	-79,169	0,000
	Gamma	-0,086	0,001	-79,169	0,000
	Correlación de Spearman	-0,046	0,001	-79,230	0,000 ^c
Intervalo por intervalo	R de persona	-0,046	0,001	-79,235	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Entre la variable ingreso (en grupos) con AFP

Medidas direccionales

			Valor
Nominal por intervalo	Eta	ingreso por grupos dependiente	0,311
		Aporte a la AFP dependiente	0,312

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,293	0,000	- 583,741	0,000
	Tau-c de Kendall	-0,291	0,000	- 583,741	0,000
	Gamma	-0,564	0,001	- 583,741	0,000
	Correlación de Spearman	-0,311	0,000	- 564,903	0,000 ^c
Intervalo por intervalo	R de persona	-0,311	0,000	- 564,962	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Anexo 5. Pruebas estadísticas de asociación entre la variable dependiente y las variables independientes

Correlación entre sector de trabajo con edad

Medidas direccionales

			Valor
Nominal por intervalo	Eta	grupos de edad dependiente	0,120
		Informalidad dependiente	0,094

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,083	0,001	154,598	0,000
	Tau-c de Kendall	0,071	0,000	154,598	0,000
	Gamma	0,146	0,001	154,598	0,000
	Correlación de Spearman	0,089	0,001	153,686	0,000 ^c
Intervalo por intervalo	R de persona	0,076	0,001	130,816	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Correlación entre sector de trabajo con años de estudio

Medidas direccionales

			Valor
Nominal por intervalo	Eta	a_est_g dependiente	0,289
		Informalidad dependiente	0,353

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,272	0,000	-564,532	0,000
	Tau-c de Kendall	-0,261	0,000	-564,532	0,000
	Gamma	-0,413	0,001	-564,532	0,000
	Correlación de Spearman	-0,308	0,001	-558,981	0,000 ^c
Intervalo por intervalo	R de persona	-0,285	0,001	-513,075	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Correlación entre sector de trabajo con horas de trabajo

Medidas direccionales

			Valor
Nominal por intervalo	Eta	horas de trabajo por grupos dependiente	0,100
		Informalidad dependiente	0,210

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	0,077	0,001	145,260	0,000
	Tau-c de Kendall	0,067	0,000	145,260	0,000
	Gamma	0,129	0,001	145,260	0,000
	Correlación de Spearman	0,082	0,001	141,787	0,000 ^c
Intervalo por intervalo	R de persona	0,062	0,001	107,644	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Correlación entre sector de trabajo con ingreso

Medidas direccionales

			Valor
Nominal por intervalo	Eta	ingreso por grupos dependiente	0,277
		Informalidad dependiente	0,282

Medidas simétricas

		Valor	Error estándar asintótico ^a	Aprox. S ^b	Aprox. Sig.
Ordinal por ordinal	Tau-b de Kendall	-0,253	0,000	-502,417	0,000
	Tau-c de Kendall	-0,222	0,000	-502,417	0,000
	Gamma	-0,419	0,001	-502,417	0,000
	Correlación de Spearman	-0,272	0,001	-488,156	0,000 ^c
Intervalo por intervalo	R de persona	-0,276	0,001	-496,356	0,000 ^c
N de casos válidos		2984726			

^a. No se supone la hipótesis nula.

^b. Utilización del error estándar asintótico que asume la hipótesis nula.

^c. Se basa en aproximación normal.

Correlación entre sector de trabajo con sexo

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	,214	0,000
	V de Cramer	,214	0,000
	Coeficiente de contingencia	,210	0,000
N de casos válidos		2984726	

Correlación entre sector de trabajo con condición de pareja

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,075	0,000
	V de Cramer	0,075	0,000
	Coeficiente de contingencia	0,075	0,000
N de casos válidos		2984726	

Correlación entre sector de trabajo con identificación étnica

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,092	0,000
	V de Cramer	0,092	0,000
	Coeficiente de contingencia	0,092	0,000
N de casos válidos		2984726	

Correlación entre sector de trabajo con aporte a una AFP

Medidas simétricas

		Valor	Aprox. Sig.
Nominal por Nominal	Phi	0,553	0,000
	V de Cramer	0,553	0,000
	Coeficiente de contingencia	0,484	0,000
N de casos válidos		2984726	

Anexo 6. Salidas del modelo Logit Multinomial

Salidas del programa SPSS

Resumen de procesamiento de casos

Resumen de procesamiento de casos			
		N	Porcentaje marginal
Informalidad	Sector formal	1193041	40,0%
	Sector informal	1697047	56,9%
	Sector domestico	94638	3,2%
Sexo	Mujeres	1248140	41,8%
	Hombres	1736586	58,2%
Condición de pareja	sin pareja	1106996	37,1%
	con pareja	1877730	62,9%
Se identifica con algún pueblo	No se identifica	2223300	74,5%
	Si se identifica	761426	25,5%
Aporte a la AFP	Aporta a una AFP	727958	24,4%
	no aporta a una AFP	2256768	75,6%
Válido		2984726	100,0%
Perdidos		0	
Total		2984726	
Subpoblación		11576 ^a	

^a. La variable dependiente sólo tiene un valor observado en 11553 (99,8%) subpoblaciones.

Resumen de los pasos

Modelo	Acción	Efecto(s)	Criterios de ajuste de modelo			efecto	
			AIC	BIC	Logaritmo de la verosimilitud -2	Chi-cuadrado ^a	gl
0	Especificado	Interceptación	4742924,6	4742950,4	4742920,56		
1	Especificado	Edad	4679139	4679190,7	4679131,01	63789,55	2
2	Especificado	Años de estudio	4372719,7	4372797,2	4372707,73	306423,3	2
3	Especificado	Sexo	4215660,4	4215763,7	4215644,44	157063,3	2
4	Especificado	Condición de pareja	4213724,4	4213853,5	4213704,44	1939,998	2
5	Especificado	Identificación étnica	4208985,8	4209140,7	4208961,8	4742,642	2
6	Especificado	Ingreso laboral	4188276,1	4188456,9	4188248,13	20713,67	2
7	Especificado	Horas de trabajo	4180008,9	4180215,5	4179976,94	8271,194	2
8	Especificado	Aporte a una AFP	3439899,1	3440131,5	3439863,09	740113,8	2

Método por pasos: Avanzar entrada

^a. El Chi-cuadrado para la entrada se basa en la prueba de razón de verosimilitud.

Información de ajuste de los modelos

Modelo	Criterios de ajuste de modelo			Contraste de la razón de verosimilitud		
	AIC	BIC	Logaritmo de la verosimilitud - 2	Chi-cuadrado	gl	Sig.
Sólo interceptación	4742924,560	4742950,378	4742920,560			
Final	3439899,092	3440131,454	3439863,092	1303057,469	16	0,000

Bondad de Ajuste

	Chi-cuadrado	gl	Sig.
Pearson	11403070,691	23134	0,061
Desvianza	3439713,360	23134	1,000

Pseudo R cuadrado

Cox y Snell	0,354
Nagelkerke	0,444
McFadden	0,274

Contraste de la razón de verosimilitud

Efecto	Criterios de ajuste de modelo			Contraste de la razón de verosimilitud		
	AIC de modelo reducido	BIC de modelo reducido	Logaritmo de la verosimilitud -2 de modelo reducido	Chi-cuadrado	gl	Sig.
Interceptación	3569381,3	3569587,8	3569349,298	129486,206	2	0
Edad	3531855,7	3532062,2	3531823,697	91960,605	2	0
Años de estudio	3467313,8	3467520,4	3467281,833	27418,742	2	0
Sexo	3568853	3569059,5	3568820,989	128957,898	2	0
Condición de pareja	3442680,4	3442887	3442648,418	2785,326	2	0
Identificación étnica	3445585	3445791,6	3445553,036	5689,944	2	0
Ingreso laboral	3452130,9	3452337,4	3452098,851	12235,76	2	0
Horas de trabajo	3442353,5	3442560	3442321,498	2458,407	2	0
Aporte a la AFP	4180008,9	4180215,5	4179976,936	740113,845	2	0

El estadístico de Chi-cuadrado es la diferencia de los logaritmos de la verosimilitud -2 entre el modelo final y el modelo reducido. El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de dicho efecto son 0.

Estimaciones de parámetro

Informalidad ^a		B	Error estándar	Wald	gl	Sig.	Exp(B)
Sector formal	Interceptación	2,866	0,03	9005,067	1	0,000	
	Edad	0	0	2,508	1	0,113	1
	Años de estudio	0,091	0,001	15431,272	1	0,000	1,095
	Sexo	4,523	0,026	30701,266	1	0,000	92,124
	Condición de pareja	0,065	0,008	73,564	1	0,000	1,067
	Identificación étnica	0,374	0,009	1730,664	1	0,000	1,453
	Ingreso laboral	0	0	8290,851	1	0,000	1
	Horas de trabajo	-0,026	0,001	416,307	1	0,000	0,975
	Aporte a la AFP	-3,401	0,024	19643,312	1	0,000	0,033
Sector informal	Interceptación	-0,147	0,03	23,998	1	0,000	
	Edad	0,034	0	15462,462	1	0,000	1,035
	Años de estudio	0,059	0,001	6810,145	1	0,000	1,06
	Sexo	4,193	0,026	26554,576	1	0,000	66,252
	Condición de pareja	0,214	0,007	891,803	1	0,000	1,238
	Identificación étnica	0,533	0,009	3860,155	1	0,000	1,704
	Ingreso laboral	0	0	8136,007	1	0,000	1
	Horas de trabajo	0,001	0,001	1,174	1	0,000	1,001
	Aporte a la AFP	-0,351	0,024	205,737	1	0,000	0,704

^a . La categoría de referencia es: Sector doméstico.

Tabla de Clasificación

Observado	Pronosticado			
	Sector formal	Sector informal	Sector domestico	Porcentaje correcto
Sector formal	642787	550254	0	53,9%
Sector informal	92831	1604216	0	94,5%
Sector domestico	1821	92817	0	0,0%
Porcentaje global	24,7%	75,3%	0,0%	75,3%

Salidas Programa STATA

Efectos marginales

Categoría sector formal

Marginal effects after mlogit

y = Pr(informalidad==sector_formal) (predict, p outcome (0))
= .40688738

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
s2a_03	-.0080647	.00056	-14.32	0.000	-.009169 -.006961	38.1313
e	.0077729	.00113	6.87	0.000	.005556 .00999	11.387
sexo*	.0932922	.01386	6.73	0.000	.066133 .120451	.581824
est_ci~1*	-.0358123	.01548	-2.31	0.021	-.066162 -.005462	.629113
identi~n*	-.0375097	.01651	-2.27	0.023	-.069873 -.005147	.255108
ylab	1.10e-06	.00000	0.95	0.344	-1.2e-06 3.4e-06	3250.34
s6b_23h	-.006476	.00258	-2.51	0.012	-.011542 -.00141	8.42966
afp*	-.6276421	.01105	-56.80	0.000	-.6493 -.605984	.756106

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Categoría sector informal

Marginal effects after mlogit

y = Pr(informalidad==sector_informal) (predict, p outcome (1))
= .589764

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
s2a_03	.0081326	.00056	14.43	0.000	.007028 .009237	38.1313
e	-.0075338	.00114	-6.60	0.000	-.009771 -.005297	11.387
sexo*	-.0539945	.01441	-3.75	0.000	-.082238 -.025751	.581824
est_ci~1*	.0363303	.01543	2.35	0.019	.006089 .066572	.629113
identi~n*	.0389188	.01651	2.36	0.018	.006558 .071279	.255108
ylab	-1.74e-07	.00000	-0.15	0.880	-2.4e-06 2.1e-06	3250.34
s6b_23h	.0064437	.00258	2.50	0.013	.001386 .011502	8.42966
afp*	.623555	.01107	56.35	0.000	.601865 .645245	.756106

(*) dy/dx is for discrete change of dummy variable from 0 to 1

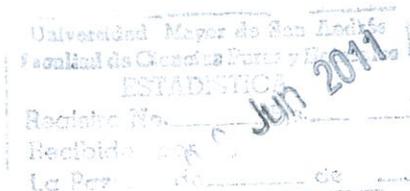
Categoría sector doméstico

Marginal effects after mlogit

y = Pr(informalidad==sector_domestico) (predict, p outcome (2))
= .00334862

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
s2a_03	-.0000679	.00002	-2.74	0.006	-.000117 -.000019	38.1313
e	-.0002391	.00008	-2.94	0.003	-.000398 -.00008	11.387
sexo*	-.0392977	.005	-7.86	0.000	-.049095 -.0295	.581824
est_ci~1*	-.000518	.00049	-1.06	0.289	-.001475 .000439	.629113
identi~n*	-.0014092	.00063	-2.22	0.026	-.002652 -.000167	.255108
ylab	-9.24e-07	.00000	-3.04	0.002	-1.5e-06 -3.3e-07	3250.34
s6b_23h	.0000323	.00007	0.49	0.621	-.000096 .000161	8.42966
afp*	.0040871	.0012	3.41	0.001	.001738 .006436	.756106

(*) dy/dx is for discrete change of dummy variable from 0 to 1



La Paz, 09 de junio de 2011
CITE: INE-RRHH-CAP-209/11

Señor
Ramiro Coa Clemente
Director a.i. de la Carrera de Estadística
Facultad Ciencias Puras y Naturales
UNIVERSIDAD MAYOR DE SAN ANDRES
Presente

Ref.: **Suficiencia Académica de la Pasante Elizabeth Sonia Nina Siñani**

Le mi consideración:

En referencia a la pasantía en la modalidad de Trabajo Dirigido realizada por la señorita **Elizabeth Sonia Nina Siñani** en la Unidad de Estadísticas e Indicadores Sociales de la Dirección de y a solicitud de las misma, nos permitimos informar que de acuerdo a los términos de referencia del trabajo dirigido, previa revisión del contenido de la investigación y aplicación pertinente, así como de los diferentes aspectos que contribuyen al trabajo realizado, se concluye que se ha cumplido con las formalidades de fondo y de forma que exige la pasantía y dada la conformidad de la tutora institucional Lic. Paloma Aguilar Agramont, Profesional en Pobreza y Bienestar, no tenemos inconveniente en expedir la presente carta de **Suficiencia Académica** a favor de la estudiante **Elizabeth Sonia Nina Siñani** con C.I. 6122655 LP.

Asimismo, se adjunta el Informe Final del Trabajo Dirigido presentado por las pasantes e informe de conformidad del tutor institucional sobre la pasantía que fue realizada en nuestra Institución durante el periodo de seis meses, a partir del 25 de octubre de 2010 al 26 de abril de 2011 en horarios establecidos según compromiso suscrito y nota de ampliación de la pasantía.

A tiempo de agradecer el apoyo de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales de la Universidad Mayor de San Andrés, en cuanto al respaldo a los universitarios que realizan sus pasantías en nuestra institución, saludo a usted atentamente.

Adj.: Lo mencionado
c.c.: VU/Arch. RR.HH./File
JChM/dsb

Jaime Choque Manzanedo
RESPONSABLE DE RECURSOS HUMANOS a.i.
INSTITUTO NACIONAL DE ESTADISTICA

La Paz - Bolivia

• Calle Carrasco N° 1391 - Miraflores • Telf. (591-2) 2222333 • Fax (591-2) 2222885
• www.ine.gob.bo • ceninf@ine.gob.bo

INFORME

INE/DEIS /INF N° 147/11



Jimmy Soria Galvarro
Director de Estadísticas e Indicadores Sociales a.i.

Blanca Franco 
Responsable de la Unidad de Estadísticas e Indicadores Sociales a.i

CC: Jaime Choque
Responsable de Recursos Humanos a.i.

Elizabet Sonia Nina Siñani
Pasante de la Unidad de Estadísticas e Indicadores Sociales

De: Paloma Aguilar Agramont 
Profesional en Pobreza y Bienestar

Fecha: 12 de mayo de 2011

Ref.: Revisión de Borrador Final y Entrega del Documento Final del Trabajo Dirigido de Elizabeth Nina Siñani

Señor Director

El borrador del Trabajo Dirigido titulado “Características y Determinantes del Empleo en el Sector Informal en Bolivia, Aplicación del Modelo Logit Multinomial”, fue revisado y realicé las siguientes observaciones para que la Pasante Elizabeth Nina Siñani las tome en cuenta para su documento final:

- La revisión fue exhaustiva y minuciosa, es un documento muy rico en investigación y tiene grandes aportes para nuestra institución. Es importante mencionar que la corrección se hizo en el documento impreso que fue entregado por Elizabeth Nina, el cual se encuentra adjunto a dicho informe.
- En cuanto al formato del texto del documento se sugiere que éste sea justificado. En la mayoría de las páginas, el texto está justificado, sin embargo en la Pág. 4 el texto está a la derecha y se puede ver lo mismo en la Pág. 6. Se debe revisar todo el documento para uniformizar a este formato.

- En cuanto al formato de los cuadros:
 - o Éstos deben contener un mismo formato en la numeración. Actualmente la pasante enumeró de la siguiente manera: Cuadro I.1, Cuadro I.2, etc. Se sugiere realizar la siguiente enumeración: Cuadro N° 1, Cuadro N° 2, etc.
 - o Cada cuadro debe contener líneas y márgenes facilitando la visualización de las filas y las columnas que contienen títulos y resultados.
- En cuanto al análisis de los cuadros, se sugiere realizar un análisis haciendo notar los resultados más importantes, no es aconsejable dar mucho preámbulo.
- Se sugiere realizar cambios en el orden de los subtítulos y a la vez, adicionar algunos en los Capítulos 1, 2 y 3 como se señala en el documento adjunto. Es importante mencionar que el contenido de todos los capítulos son claros y están bien desarrollados.
- Es importante que los títulos y subtítulos tengan un mismo formato, por lo cual se sugiere que todas las palabras comiencen con su primera letra en mayúscula.
- En cuanto a la numeración de las fórmulas ésta contiene después del paréntesis un punto final. Se sugiere quitar el punto a todas las fórmulas del marco teórico.
- La regla de ortografía dice que las mayúsculas pueden o no llevar acento, pero si una palabra lleva acento, todas deben llevar acento o viceversa. Se sugiere poner acento a todas las mayúsculas cuando corresponda.
- Se realizó la revisión de la sintaxis de los programas estadísticos, tanto en SPSS como en STATA. Los programas corren sin problema alguno en ambos paquetes y están bien elaborados, éstos se encuentran adjuntos al presente informe.

Por ultimo, las observaciones y sugerencias realizadas en este informe y en el documento adjunto son de forma y no así de contenido o de fondo. Es importante mencionar que el documento adjunto con las correcciones escritas fue transferido a Elizabeth Nina Siñani quien ya realizó los cambios para obtener el Documento Final de su Trabajo Dirigido, el cual se encuentra adjunto al informe y doy conformidad de que todas las sugerencias y observaciones fueron tomadas en cuenta y se realizaron los cambios de manera pertinente.

Por lo tanto, doy conformidad de que el Documento Final de Trabajo Dirigido de Elizabeth Sonia Nina Siñani con C.I. 6122655 L.P. cumple con todos los términos de referencia establecidos:

- Elaboración de un documento final del trabajo realizado, que contenga: análisis de cuadros, gráficos y pruebas estadísticas del empleo en el sector informal urbano en Bolivia. Además, la aplicación de un modelo estadístico que se ajuste a los datos con su respectivo análisis de los resultados.
- Sintaxis de todos los procedimientos realizados en SPSS y/o STATA.

Es cuanto tengo a bien informar

Anexo 8. Sintaxis del Modelo Logit Multinomial

Sintaxis en programa SPSS.

```
NOMREG Informalidad (BASE=2 ORDER=ASCENDING) WITH s2a_03 e sexo
est_civil identificacion_etn ylab s6b_23h afp
/CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0)
PCONVERGE(0.000001)
SINGULAR(0.00000001)
/MODEL=| FORWARD= s2a_03 e sexo
est_civil identificacion_etn ylab s6b_23h afp
/STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR)
REMOVALMETHOD(LR)
/INTERCEPT=INCLUDE
/PRINT=CLASSTABLE FIT PARAMETER SUMMARY LRT CPS STEP MFI IC.
```

Sintaxis en programa STATA

```
mlogit informalidad s2a_03 e sexo est_civil identificacion_etn ylab s6b_23h afp [fw=factor], b(2)
```

Efectos Marginales

```
mlogit informalidad s2a_03 e sexo est_civil identificacion_etn ylab s6b_23h afp [fw=factor], b(2)
vce(cl upm)
mfx, predict (p outcome (0))
mfx, predict (p outcome (1))
mfx, predict (p outcome (2))
```

GLOSARIO

INE: Instituto Nacional de Estadística
EPH: Encuesta Permanente de Hogares
EIH: Encuesta Integrada de Hogares
ENE: Encuesta Nacional de Empleo
MECOVI: Encuestas y Medición sobre Condiciones de Vida
CNPV: Censo Nacional de Población y Vivienda
IDRC: International Development Research Centre (Centro Internacional de Investigación de desarrollo)
CELADE: Centro Latinoamericano y Caribeño de Demografía
CEPAL: Comisión Económica para América Latina y el Caribe
UNFPA: Fondo de Población de las Naciones Unidas
OIT: Organización Internacional de Estadística
UDAPE: Unidad de Análisis de Políticas Económicas y Sociales
CIET: Conferencia Internacional de Estadísticos del Trabajo
SNIE: Sistema Nacional de Información Estadística
MPL: Modelo de Probabilidad Lineal
MCO: Mínimos Cuadrados Ordinales
NIT: Número de Identificación Tributaria
AFP: Administradora de Fondo de Pensiones
UPM: Unidad Primaria de Muestreo
USM: Unidad Secundaria de Muestreo
UTM: Unidad Terciaria de Muestreo

BIBLIOGRAFÍA

- De la Fuente F. S. (s.f.). *http://www.fuenterebollo.com/Aeronautica*. Obtenido de <http://www.fuenterebollo.com/Aeronautica>:
<http://www.fuenterebollo.com/Aeronautica>
- Gasparini L. y Tornarolli L. (2009). "*Labor informality in Latin America and the Caribbean: Patterns and Trend from Household Survey Microdata*". Obtenido de <http://www.scielo.org.co/pdf/dys/n63/n63a02.pdf>:
<http://www.scielo.org.co/pdf/dys/n63/n63a02.pdf>
- Goldberg P. y Nina P. (2003). "The response of the informal sector to trade liberalization". "*The response of the informal sector to trade liberalization*". Working Paper No. 9443. Cambridge: National Bureau of Economic Research (NBER).
- Greene W. (1993). *Econometric Analysis*. Printice Hall, cap 21, pág. 667.
- Greene William H. (2003). "Econometric Analysis". En G. W. H., "*Econometric Analysis*" (págs. 719 -728, Cap. 21). United State of America, Cuarta Edición.
- Greene, William H. (1999). "Análisis Econométrico". En W. H. Greene, "*Análisis Econométrico*" (pág. 672). Editorial Prentice may, Tercera Edición.
- Gujarati Damodar N. (1995). "Econometría". En G. D. N., "*Econometría*" (pág. 195). Mexico: Editorial Mc Graw Hill, Cuarta Edición.
- Hair, J. F. y Anderson R. E. y Tatham R. L.; Black W. C. (1999). "*Análisis Multivariante de Datos*". Madrid: Prentice Hall, Quinta Edición.
- Hart, K. (2002). "*El trabajo decente y la economía informal-OIT*".
- Herrera J., R. S. (2004). "*El Sector Informal en Colombia y demás Países de la Comunidad Andina*". Bogotá.
- INE. (2014). "Diseño Muestral EH-2014".
- INE. (2014). Manual del Encuestador EH-2014. La Paz - Bolivia.
- Keith, Hart. (1971). "*Trabajo decente y la economía informal*".
- Novalés A. (2000). "Econometría". En N. A., "*Econometría*" (págs. 540 - 548). Madrid: Mc Graw Hill. Interamericana de España, S.A.U.

- OIT. (2002). *Informe "El trabajo Decente y la Economía Informal"*. Ginebra.
- OIT. (2013). "La medición de la informalidad: Manual estadístico sobre el sector informal y el empleo informal". Ginebra, Suiza.
- OIT. (2014). *Informe "Panorama Laboral 2014- América Latina y el Caribe"*.
- Perez L. C. (2004). *"Técnicas de Análisis Multivariante de Datos, Aplicaciones en SPSS"*. Madrid: Pearson Educación S.A.
- Pindyck R. Rubinfeld D. (1991). *Econometric models and economic forecast*. New York 3° ed, cap 10, pág 270.
- Salvador Figueras M. (2000).
<https://ciberconta.unizar.es/leccion/anamul/inicio.html>. Recuperado el 20 de 10 de 2014, de <https://ciberconta.unizar.es/leccion/anamul/inicio.html>:
<https://ciberconta.unizar.es/leccion/anamul/inicio.html>
- UDAPE. (2006). *Informe Especial*.
- V. Pando Fernandez & R. San Martin Fernandez. (s.f.).
http://researchgate.net/publication/40836593_Regresion_logistica_miltino_mial.)Obtenido de
http://researchgate.net/publication/40836593_Regresion_logistica_miltino_mial.
- Instituto Nacional de Estadística. (s.f.). *INE*. Obtenido de sitio Web de INE:
<http://www.ine.gob.bo/index.php/banco/base-de-datos-sociales>