

REVISTA DE ESTADÍSTICA Y METODOLOGÍAS

Número 4

Abril 2018



instituto nacional de estadística y censos

REVISTA DE
**ESTADÍSTICA Y
METODOLOGÍAS**

Número 4

Revista de Estadística y Metodologías

AUTORIDADES

Reinaldo Cervantes
Director Ejecutivo

Roberto Castillo A.
Subdirector General

Cristina Fabara
Coordinadora General Técnica de Innovación en Métricas y Análisis de la Información

Markus Nabernegg
Coordinador General Técnico de Producción Estadística

Magaly Paredes
Coordinadora General Técnica de Planificación, Normativas y Calidad Estadística

Giovanni Coronel
Coordinador General Administrativo Financiero

Marx Baquero
Director de Comunicación Social

COMITÉ EDITORIAL DE LA REVISTA

Cristina Fabara
Editora en Jefe

Natalia Garzón
Directora de Estudios y Análisis de Información Subrogante

Lorena Moreno
Directora de Innovación en Métricas y Metodologías

Revisión de redacción y estilo
Coordinación General Técnica de Innovación en Métricas y Análisis de la Información

Esteban Palacios J.
Diseño y diagramación

Propiedad Institucional
Instituto Nacional de Estadística y Censos
Administración Central (Quito)
Juan Larrea N15-36 y José Riofrío,
Teléfonos: (02) 2544 326
2544 561 Fax: (02) 2509 836
Casilla postal: 17-15-135
correo-e: inec@inec.gob.ec

Presentación

El Instituto Nacional de Estadística y Censos (INEC) en su rol de incentivar la investigación científica a través de la generación de estudios sobre metodologías y análisis de la información estadística, pone a disposición de la ciudadanía la Revista de Estadística y Metodologías, línea editorial del INEC donde los investigadores del INEC difunden los avances metodológicos que realiza la institución a manera de artículo de investigación.

La revista tiene como objetivo, actualizar y proponer metodologías para la construcción de indicadores, empalme de series, diseño muestral de operaciones estadísticas, generación de protocolos para el aprovechamiento estadístico de los registros administrativos, entre otros. Incluye artículos que documentan una metodología oficial del Instituto así como propuestas metodológicas no oficiales para su respectiva discusión y mejoramiento.

En el cuarto volumen se pone a consideración de la ciudadanía los siguientes documentos:

- La Escala de Experiencia de Inseguridad Alimentaria FIES en Ecuador: discusión metodológica. El documento hace una revisión del marco conceptual y metodológico propuesto para el monitoreo del indicador 2.1.2 de los ODS en Ecuador el cual mide la prevalencia de inseguridad alimentaria moderada o severa basada en la escala Food Insecurity Experience Scale (FIES) desarrollada por la Oficina de las Naciones Unidas para la Agricultura y la Alimentación (FAO); y presenta la aplicación para el Ecuador realizada en la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU).
- Elaboración de estadísticas de vacantes publicadas en internet. Este artículo documenta el proceso de generar estadísticas de oferta laboral a partir de la identificación de vacantes, a partir de los anuncios de empleos que se publican en el Internet, usando una herramienta de web scraping.
- Encuesta basada en registros administrativos para el análisis de la creación y destrucción de plazas de trabajo. Una propuesta metodológica. Este documento presenta la conceptualización desarrollada para la producción estadística en base a sistemas de registros administrativos (Wallgren y Wallgren 2014) y evidencia el diseño de la estructura de la encuesta y su proceso de compilación.
- Propuesta de Modelo Conceptual para identificar patrones de hechos estadísticos desde texto libre. El documento desarrolla una propuesta de modelo conceptual para la extracción y reconocimiento de textos escritos en lenguaje natural, y se enfoca en realizar una descomposición sistemática del texto de la pregunta abierta de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) que indaga sobre actividades económicas de empresas, cuyo contenido requiere ser comparado con la terminología del clasificador internacional.

La Revista de Estadística y Metodologías permite apoyar la misión institucional de revisión metodológica, definición de medidas e indicadores, y análisis de información estadística, al presentar las metodologías y ejercicios estadísticas planteados en las actividades técnicas de la Institución.

MSc. Reinaldo Cervantes
Director Ejecutivo del INEC

ÍNDICE

DE CONTENIDOS

La Escala de Experiencia de Inseguridad Alimentaria FIES en Ecuador: discusión metodológica

Lorena Moreno, Juan Carlos Serrano, Melody Serrano, Diego Villacreses, Juan José Viteri 7

Elaboración de estadísticas de vacantes publicadas en internet. Una experiencia en Ecuador

Diego Benítez, Sebastián Lucero, Ana M. Pazmiño 33

Encuesta basada en registros administrativos para el análisis de la creación y destrucción de plazas de trabajo. Una propuesta metodológica

Diego Benítez, Boris Espinoza, Ana Grijalva 45

Propuesta de Modelo Conceptual para identificar patrones de hechos estadísticos desde texto libre

Juan Carlos Delgado Loyola 61

La Escala de Experiencia de Inseguridad Alimentaria

FIES en Ecuador: discusión metodológica

Lorena Moreno

Juan Carlos Serrano

Melody Serrano

Diego Villacreses

Juan José Viteri¹

RESUMEN

El presente documento describe el marco conceptual y metodológico propuesto para el monitoreo del indicador 2.1.2 de los ODS en Ecuador el cual mide la prevalencia de inseguridad alimentaria moderada o severa basada en la escala *Food Insecurity Experience Scale* (FIES) desarrollada por la Oficina de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Este indicador nutre a la meta ODS 2.1 la cual establece que para el año 2030 se garantice el acceso de toda la población a alimentos inocuos, nutritivos y suficientes de manera permanente; y es parte del objetivo 2 que se plantea la necesidad de poner fin al hambre, conseguir la seguridad alimentaria, la mejora de la nutrición, y la promoción de la agricultura sostenible. A nivel mundial FIES se levanta en la Encuesta Mundial Gallup (GWP) en más de 150 países; sin embargo, el objetivo de FAO es que cada país pueda emplear encuestas propias para el seguimiento y monitoreo. En este contexto, se seleccionó la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) para primeros ejercicios de levantamiento para la réplica de la FIES en el contexto ecuatoriano. Producto de este trabajo, se presenta una primera discusión metodológica que será importante insumo para la posterior definición de una línea base para el monitoreo del indicador ODS en mención.

Palabras clave: escala de experiencia de inseguridad alimentaria, modelo de Rasch, teoría de respuesta al ítem

Descargo de responsabilidad: Las opiniones e interpretaciones expresadas en este documento pertenecen a los autores y no reflejan el punto de vista oficial del Instituto Nacional de Estadística y Censos (INEC). El INEC no garantiza la exactitud de los datos que figuran en el documento

1 Autores en orden alfabético

1. Introducción

En 1947 la Comisión de Derechos Humanos de la ONU instauró el derecho a los alimentos, como el primer esfuerzo por combatir el hambre en la humanidad. Desde entonces se han dado grandes avances en la lucha contra el hambre; pero, a pesar de estos esfuerzos, para el año 2015 aún se reportaron 795 millones de personas subalimentadas, mayormente en países en desarrollo (FAO, FIDA, & WFP, 2015).

Asegurar la ingesta de alimentos en los individuos no es suficiente para contribuir con su bienestar y por esa razón a lo largo de los años se forjó el concepto de la seguridad alimentaria. El concepto más actual establece que los individuos deben tener acceso económico y físico a alimentos nutritivos e ino cuos en todo momento (FAO, 2006).

El concepto de seguridad alimentaria hace referencia a cuatro grandes componentes o dimensiones: la disponibilidad de alimentos, el acceso a ellos, el aprovechamiento biológico de los mismos y la estabilidad en el tiempo de los tres componentes previos. En el año 2000 los países miembros de Naciones Unidas, Ecuador entre ellos, se comprometieron en la lucha contra el hambre al establecer como una de las metas el “reducir a la mitad la proporción de personas que padecen de hambre para el año 2015”. La meta, que corresponde al objetivo uno de los Objetivos de Desarrollo del Milenio, fue alcanzada casi a nivel mundial, sobresaliendo los avances de países de América Latina.

Ante estos resultados, los miembros de Naciones Unidas fijaron en la Agenda 2030 para el Desarrollo Sostenible un nuevo objetivo que consiste en “Poner fin al hambre y asegurar el acceso de todas las personas a una alimentación sana y suficiente para todo el año.”

El seguimiento de los avances de este objetivo se hace a través del indicador que mide la “Prevalencia entre la población con inseguridad alimentaria moderada o severa basada en la Escala de Experiencia de Inseguridad Alimentaria (FIES)”. Según la FIES, la severidad de la condición de inseguridad alimentaria de un individuo, hogar o población se evalúa como un rasgo latente (no observable), donde la escala mide la falta de “acceso” a alimentos en el hogar (FAO, 2016a).

La estimación de la FIES se la realiza mediante los modelos de Teoría de Respuesta del Ítem (IRT por sus siglas en inglés), que son un conjunto de procedimientos estadísticos que permiten relacionar las respuestas de personas con la variable latente de Inseguridad Alimentaria. El modelo de Rasch asume que dos de los parámetros del modelo general de IRT para variables dicotómicas son constantes, y se estima mediante máxima verosimilitud condicional permitiendo así conocer el nivel de Inseguridad Alimentaria que sufre una persona (FAO, 2016a).

La presente investigación analiza la aplicación de la FIES en el Ecuador. Para ello se emplean los resultados de la sección 14A de la ENEMDU de diciembre 2017. Las preguntas correspondientes tienen un periodo de referencia de un año, lo que las hace comparables en esa característica a las preguntas propuestas en la FIES. Sin embargo, en cuanto a su contenido existen algunas diferencias que pueden o no limitar su comparación con el estándar internacional. Además, como se detallará en el presente estudio, pueden existir limitaciones en la aplicación de Rasch con FIES en Ecuador. Esto puede limitar la interpretación y comparabilidad de los datos. Por estas razones, posteriormente explicadas a detalle, la presentación de cifras tiene como objetivo únicamente realizar una discusión metodológica, y de ninguna forma, reflejar una cifra oficial.

El documento está organizado de la siguiente forma: la segunda sección presenta los antecedentes, la tercera justifica el estudio, la cuarta sección resume las principales leyes y marcos legales relacionados a la temática de la seguridad alimentaria, la quinta sección explica el marco conceptual, mientras que la sexta sección explica en detalle el marco metodológico con énfasis en el modelo de Rasch; la sección siete profundiza en los resultados obtenidos para el Ecuador empleando la metodología propuesta por FAO (2016a). Finalmente la sección de anexos detalla algunos insumos extras para la comprensión del ejercicio.

2. Antecedentes

La preocupación por la seguridad alimentaria se discute a nivel global en la primera Conferencia Mundial de la Alimentación, realizada en noviembre de 1974. En este evento, la FAO reunió en Roma a los representantes de 135 países para dar una respuesta a la crisis alimentaria experimentada desde 1972,

que provocó fluctuaciones en los suministros y los precios de los alimentos (FAO, 1975). Los gobiernos participantes firmaron la Declaración Universal sobre la Erradicación del Hambre y la Malnutrición, proclamando que “todos los hombres, mujeres y niños tienen el derecho inalienable a no padecer de hambre y malnutrición a fin de poder desarrollarse plenamente y conservar sus facultades físicas y mentales” (UN, 1975, p. 8).

En esta conferencia se destaca el papel de la producción y distribución de los alimentos en el bienestar de los pueblos; además, se plantea la necesidad de establecer un sistema mundial de seguridad alimentaria que pueda garantizar la “disponibilidad adecuada y precios razonables de los alimentos en todo momento”, sin que se vean afectados por las fluctuaciones periódicas del clima o de las presiones políticas y económicas (UN, 1975, p. 10). Además, los países participantes se comprometieron a erradicar el hambre, la inseguridad alimentaria y la malnutrición en un plazo de 10 años, a través de políticas y programas de cooperación internacional, sin embargo, esta meta no fue alcanzada (FAO, 1999).

Años más tarde, con la finalidad de renovar este compromiso mundial, la FAO realiza en noviembre de 1996 la Cumbre Mundial sobre la Alimentación (CMA), donde reúnen en Roma a 10 mil personas representantes de 185 países y de la Comunidad Europea. Este encuentro generó dos acuerdos globales, la Declaración de Roma y el Plan de acción de la cumbre, donde se reafirmó el derecho de toda persona a tener acceso a alimentos sanos y nutritivos; y se prometió realizar un esfuerzo conjunto para conseguir seguridad alimentaria para todos y erradicar el hambre en todos los países, proponiendo como meta reducir el número de personas desnutridas a la mitad hasta el año 2015 (FAO, 1999).

Cuatro años después, la Organización de las Naciones Unidas (ONU) reunió en Nueva York a 189 líderes de sus estados miembros en la llamada Cumbre del Milenio, realizada en septiembre del año 2000. Esta reunión concluyó con la aprobación de la Declaración del Milenio, en la cual se establecieron ocho Objetivos de Desarrollo del Milenio (ODM), expresados en 21 metas a cumplirse hasta el 2015. La meta 1.C. del ODM, reafirmó el compromiso mundial de la lucha contra la pobreza y el hambre, proponiendo reducir a la mitad –entre 1990 y 2015– la proporción de personas que padecen hambre (ONU, 2015a).

Para evaluar y vigilar el progreso de ambos acuerdos, la FAO organiza en Roma el Simposio Científico Internacional sobre medición y evaluación de la carencia de alimentos y la desnutrición en junio de 2002, con la participación de 8 expertos miembros del Comité Asesor Científico y de 141 expertos provenientes de 25 países² que cuenten con un punto de vista científico de las metodologías para medir el hambre. En esta reunión se destacó la necesidad de contar con “mediciones precisas, fiables y oportunas de la prevalencia del hambre y la malnutrición, la inseguridad alimentaria y la vulnerabilidad y de la forma en que varían en el curso del tiempo” (Vercueil et al., 2002, p. 73).

En los años subsiguientes, los esfuerzos se enfocaron en monitorear la consecución tanto del objetivo de la CMA como de la meta 1.C de los ODM, siendo esta última menos ambiciosa que la primera³. De acuerdo a las estimaciones de FAO en promedio entre 2014-2016, existen alrededor de 795 millones de personas subalimentadas⁴ en el mundo (10.9% de la población mundial), de las cuales 780 millones (el 98%) viven en las regiones en desarrollo. Esta cifra, no obstante, implica 148 millones de personas menos a nivel mundial que entre 2005-2007 (pasar de 14.3% al 10.9%) y 216 millones menos que entre 1990-1992 (de 18.6% a 10.9%). Dándose un descenso más pronunciado en las regiones en desarrollo⁵ (FAO, FIDA, & WFP, 2015).

Al finalizar el 2015, también concluyó el periodo de evaluación de estos dos acuerdos mundiales, obteniendo resultados positivos pero que no logran poner fin al hambre en el mundo. De los 129 países en desarrollo evaluados a partir de 1990, 73 alcanzaron la meta 1.C de los ODM, reduciendo a la mitad la proporción de personas subalimentadas para el 2015, o bien, reduciendo –o manteniendo– esta proporción por debajo del 5%; esto implica que la meta 1.C. casi se alcanza a nivel mundial. Sin embargo, de esos

2 Estados Unidos, Tailandia, Italia, Yugoslavia, Indonesia, Reino Unido, Filipinas, China, India, Suiza, Países Bajos, Colombia, Portugal, Nueva Zelanda, Bélgica, Turquía, Alemania, Canadá, Suecia, Francia, Kenia, España, Bangladesh, Grecia y Hungría.

3 La prevalencia y número de personas subalimentadas han sido adoptadas como indicadores para el seguimiento de la meta 1.C de los ODM y la meta de la CMA respectivamente, como se detalla en el Anexo 2 del informe presentado por (FAO et al., 2015).

4 Se entiende a la subalimentación como el estado de incapacidad para adquirir alimentos suficientes con una duración mínima de un año. Implica un nivel de ingesta de alimentos insuficiente para satisfacer las necesidades de energía alimentaria (FAO et al., 2015).

5 La FAO utiliza el trienio 1990-92 como punto de partida y el trienio 2014-16 como punto de llegada –período promedio de tres años centrado en 2015–.

73 países, solo 29 alcanzaron la meta más exigente de la CMA de 1996 –reducir a la mitad el número de personas que padecen hambre para el 2015–; resultado que se ubica lejos del cumplimiento a nivel mundial (FAO et al., 2015).

América Latina fue una de las regiones que más progresaron en la reducción del hambre y subalimentación en ese periodo –en especial los países de América del Sur–, ya que logró cumplir ambos acuerdos. Además, se dio una reducción del 53,8% en el número de personas subalimentadas –pasando de 58 millones a 27 millones de personas–, y por otro lado, experimentó una caída significativa en la prevalencia de la subalimentación, pasando de 13,9% en 1990-92 a menos del 5% en 2014-16. Ecuador, por su parte, logró reducir en 44% la prevalencia de subalimentación –al pasar de 19,4% en 1990-92 a 10,9% en 2014-16–, acercándose a cumplir la meta 1.C en los próximos años si mantiene la tendencia; y además, una reducción del 12,3% en el número de personas subalimentadas, pasando de 2 a 1,8 millones de personas, progreso aún insuficiente para alcanzar la meta de la CMA (FAO et al., 2015).

Estos resultados revelaron que aunque se ha dado una reducción del hambre a nivel mundial, aún prevalece la inseguridad alimentaria –especialmente en países en vías de desarrollo–. Además, existe todavía incertidumbre sobre el número exacto de hogares que se ven afectados, la ubicación de los mismos y el grado de severidad que enfrentan (Comité Científico de la ELCSA, 2012). A pesar de que desde hace muchos años se cuenta con métodos para medir la suficiencia de alimentos y la nutrición, sigue siendo difícil la medición oportuna y precisa en toda la población (Vos, 2016).

Teniendo esto en mente, la FAO lanzó el proyecto *Las Voces del Hambre* (VoH) en 2013, con el objetivo de llegar a proporcionar información actualizada sobre la inseguridad alimentaria a nivel global, que sea relevante para las decisiones políticas. Para lograrlo, este proyecto propuso un nuevo estándar para estimar la prevalencia del fenómeno: la FIES, una escala basada en la experiencia de los hogares que utiliza las respuestas de las personas a preguntas sobre su acceso a una alimentación adecuada (en cantidad y calidad). Esta iniciativa tiene como objetivo contar con información confiable, que sea comparable entre países y que tenga mayor frecuencia de recolección, al ser de bajo costo comparado a otros indicadores (FAO, 2016a).

A partir del 2014, la FAO comienza el levantamiento de la FIES en un módulo de inseguridad alimentaria incorporado en la Encuesta Mundial Gallup (GWP). La información es recogida en más de 150 países, con una muestra nacionalmente representativa de la población de 15 años o más (FAO, 2016a). Paralelamente, algunos países ya habían comenzado a recoger regularmente datos de seguridad alimentaria en sus encuestas nacionales, obteniendo información compatible con la FIES, por ejemplo: Estados Unidos (desde 1995), Brasil (desde 2004), México (desde 2008), Guatemala (desde 2011), entre otros. Siguiendo esa línea, Ecuador incluyó la Escala Latinoamericana y Caribeña de Seguridad Alimentaria ELCSA⁶ en su ENEMDU de diciembre de 2016, con la intención de obtener la información de forma periódica.

En septiembre de 2015, los líderes mundiales de los 193 estados miembros de las Naciones Unidas se reunieron en Nueva York para acordar un nuevo conjunto de objetivos globales, como parte de la agenda de desarrollo sostenible a cumplirse hasta el 2030. En este encuentro, los estados participantes se comprometieron por trabajar arduamente en los próximos 15 años a fin de lograr los 17 *Objetivos de Desarrollo Sostenible* (ODS) y las 169 metas establecidas en la nueva agenda, reconociendo que son de crítica importancia para la humanidad y el planeta (ONU, 2015b).

Los nuevos objetivos planteados abarcan aquellos temas que quedaron inconclusos en los ODM; como por ejemplo, la decisión de terminar con el hambre, todas las formas de malnutrición y lograr la seguridad alimentaria para todos, de forma prioritaria (ODS2). De forma específica, el objetivo establece “*Poner fin al hambre, lograr la seguridad alimentaria y la mejora de la nutrición y promover la agricultura sostenible*” (ONU, 2015b).

Se cuenta con ocho metas alineadas a este objetivo, cinco se relacionan a resultados y tres se relacionan a los medios de ejecución (FAO, 2016b). Dentro de este grupo, la primera meta de resultados plantea “*Para el año 2030, poner fin al hambre y garantizar el acceso de todas las personas [...] a alimentos inocuos, nutritivos y suficientes durante todo el año*”, poniendo una particular atención a las personas en situación de vulnerabilidad (ONU, 2015b).

⁶ La ELCSA es una escala previa a la FIES que también se basa en la experiencia de los hogares y que recepta información compatible con la FIES. La primera vez que fue incluida en la ENEMDU fue en marzo de 2010 para un estudio puntual.

Uno de los dos indicadores establecidos para medir los avances en esta dirección es la prevalencia entre la población de inseguridad alimentaria moderada o severa basada en la FIES. Gracias a la incorporación de una escala de experiencia compatible con la FIES en la ENEMDU, Ecuador está en la capacidad de reportar este indicador.

3. Justificación

A nivel mundial, los indicadores relacionados a salud, nutrición y alimentación eran medidos a través de varios índices, de los cuales se destacan: Índice de Nutrición (2000) Wiesmann, Índice de Hambre (2001) e Índice Global del Hambre (2006). Estas aproximaciones, a diferencia de la FIES, no miden la situación de acceso de los hogares a alimentos de calidad (FAO, 2011).

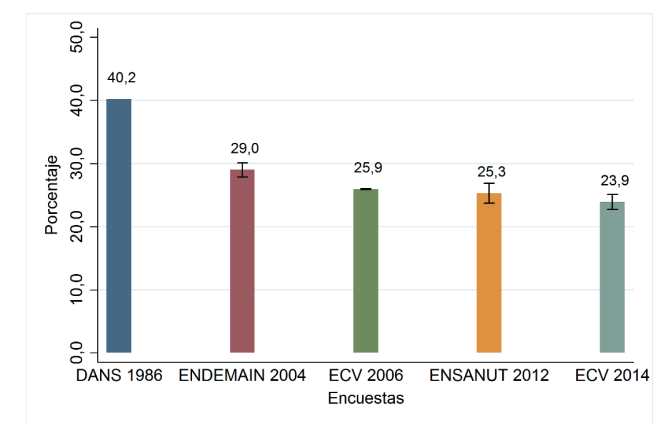
Con miras al cumplimiento de la Agenda 2030, la FIES, de acuerdo con el Plan de Desarrollo Estadístico (PDE), se presentó como el mecanismo idóneo para medir la dimensión de acceso de la seguridad alimentaria y a través del cual se plantea dar seguimiento a la meta 2.1 de desarrollo sostenible (INEC, 2017). Y es que la falta de acceso a alimentos y de control sobre los mismos, son frecuentemente la causa de la inseguridad alimentaria. Se puede tratar de un problema de índole física, en caso de que los alimentos no están disponibles en la cantidad suficiente en lugares aislados –o con falta de infraestructura–; o de índole económica cuando las personas no cuentan con el dinero suficiente para alimentarse con regularidad y calidad. En Latinoamérica, el segundo caso es el más común (MCDS/FAO, 2010).

Dentro del Ecuador, han existido varias fuentes de información para medir y evidenciar el estado nutricional y de alimentación de la población. Inicialmente se implementó para el año de 1986 un Diagnóstico de la Situación Alimentaria Nutricional y de Salud de la Población Ecuatoriana (DANS) (CONADE, 1988). Para el año 2004 se realizó la Encuesta Demográfica y de Salud Materna e Infantil (ENDEMAIN), la cual contó con cuatro ediciones (CEPAR, 2004). Posteriormente en la UT

Salud y Nutrición (ENSANUT) del año 2012 se colectaron datos referentes a salud alimentaria y nutrición.

No obstante, los principales datos recopilados por las mencionadas encuestas se trasladan a indicadores antropométricos, entre los que constan: peso, talla, longitud e índice de masa corporal (Freire et al., 2014), enfocados más bien a la utilización biológica de los alimentos. Otra de las fuentes de información que evidencia la situación de las condiciones de salud y nutrición de la población, es la Encuesta de Condiciones de Vida (ECV) última ronda 2013-2014, la cual presenta dentro de sus principales indicadores relacionados a la nutrición: desnutrición global; desnutrición crónica y desnutrición infantil. (INEC, 2014). En el gráfico 1 se incluyen la evolución de la desnutrición crónica infantil según diferentes fuentes de información tomando como referencia los patrones de crecimiento de referencia WHO 2006.

Gráfico 1: Histórico de Desnutrición Crónica Infantil en Ecuador (menores de 5 años)



Fuente: DANS, ENDEMAIN, ECV, ENSANUT
Elaboración: Autores

Las medidas antropométricas mencionadas anteriormente proveen información útil para realizar evaluaciones de crecimiento y para constatar el estado nutricional de niño, niñas y adolescentes (Freire et al., 2014); sin embargo, debido a sus costos, la periodicidad con la que se levanta la información y el hecho de que se centran principalmente en resultados antropométricos, presentan limitantes para medir y monitorear el acceso a alimentos de calidad (FAO, 2016a).

Uno de los indicadores que más se aproxima a medir la dimensión de acceso a alimentos de una población es la pobreza extrema por consumo. Según Sen (2000) las personas que cuentan con pocos activos productivos o un bajo poder adquisitivo, son los que tienen mayor probabilidad de enfrentar una mayor dificultad de acceso a una cantidad de alimentos adecuada.

Sin embargo, al ser una proxy no refleja los componentes más correlacionados al fenómeno de inseguridad alimentaria

Tomando en cuenta lo antes mencionado, los indicadores basados en la experiencia en comparación con otros indicadores, destacan por su facilidad de gestionar la información, bajo costo relativo, periodicidad y la posibilidad de comparar los resultados entre países (FAO, 2016a). En este contexto, dada la necesidad de encontrar una medida para la inseguridad alimentaria que se pueda levantar con una encuesta de mayor periodicidad como la ENEMDU, se incorporó una versión adaptada de la ELCSA⁷, escala armonizada para Latinoamérica y el Caribe (LAC) como una medida válida del nivel de inseguridad alimentaria en los hogares. Más adelante se profundizará en los conceptos y metodologías detrás de la escala empleada para medir la FIES. A continuación se enmarca el contexto legal.

4. Marco Legal

El derecho nacional reconoce el concepto de seguridad alimentaria; por ejemplo en la vigente Constitución, el artículo 3, garantiza sin discriminación alguna el derecho a la alimentación. Así mismo el Capítulo Segundo, Derechos del buen vivir, sección primera, art. 13 establece que “las personas y colectividades tienen derecho al acceso seguro y permanente a alimentos sanos, suficientes y nutritivos; preferentemente producidos a nivel local y en correspondencia con sus diversas identidades y tradiciones culturales”.

Para garantizar el pleno cumplimiento de este derecho se establece, en el art. 281 de la Constitución, a la soberanía alimentaria como un objetivo estratégico y de responsabilidad del Estado, el cual a través de la aplicación de instrumentos de política económica (art. 284) y comercial (art. 304) asegurará la soberanía alimentaria de forma permanente.

Adicionalmente, el Estado es el encargado de promover el acceso equitativo a los factores de producción, mediante: “el desarrollo de políticas de fomento a la producción nacional en todos los sectores...”, así como: “brindar a los agricultores y a las comunidades rurales apoyo para la conservación y restauración

7 Las diferencias entre la ELCSA y la adaptación de la ELCSA empleada en la ENEMDU se detallan en secciones posteriores.

de suelos, así como para el desarrollo de prácticas agrícolas que protejan y promuevan la soberanía alimentaria”, tal y como lo señalan los artículos 334 y 410 de la Constitución respectivamente.

El Art. 3 de la Constitución, en los numerales 1 y 5, establece que son deberes primordiales del Estado:

- Garantizar sin discriminación alguna el efectivo goce de los derechos establecidos en la Constitución y en los instrumentos internacionales, en particular la educación, la salud, la alimentación, la seguridad social y el agua para sus habitantes.
- Planificar el desarrollo nacional, erradicar la pobreza, promover el desarrollo sustentable y la redistribución equitativa de los recursos y la riqueza, para acceder al Buen Vivir.

Adicionalmente, el actual Plan Nacional de Desarrollo, el plan “Toda una Vida” 2017-2021 se instituye como el marco político rector de la planificación para el desarrollo en el Ecuador y constituye un guía dentro de la cual sus objetivos y metas establecen la importancia de garantizar la seguridad y soberanía alimentaria. En el plan, se establece dentro del Eje 2 “Economía al servicio de la sociedad”, objetivo 6 “Desarrollar las capacidades productivas y del entorno, para lograr la soberanía alimentaria y el Buen Vivir Rural”, esto a través de políticas direccionadas a combatir toda forma de malnutrición y desnutrición dentro del marco de la seguridad y soberanía alimentaria. Estos conceptos se presentan como ejes transversales dentro del plan para garantizar el cumplimiento de varios derechos estipulados en los diferentes objetivos del Plan.

Por otro lado, uno de los mecanismos establecidos para que el Estado cumpla con el objetivo estratégico de soberanía alimentaria, es la Ley Orgánica del Régimen de la Soberanía Alimentaria (LORSA)⁸. En la cual, de acuerdo al art. 1, se establece el régimen para garantizar “la autosuficiencia de alimentos sanos, nutritivos y culturalmente apropiados de forma permanente”, a través de la producción sustentable y sostenible de alimentos con incentivos para el uso productivo de la tierra y el consumo de alimentos sanos.

8 Ley Orgánica del Régimen de la Soberanía Alimentaria (2011): Asamblea Nacional de la República del Ecuador.

Para el pleno ejercicio de las competencias constitucionales entre los diferentes niveles de gobierno, el Código Orgánico de Organización Territorial Autonomía y Descentralización (COOTAD) establece en su art. 134 que los gobiernos autónomos descentralizados regionales son los encargados del fomento de la seguridad alimentaria, a través de la aplicación de las disposiciones constitucionales y legales para garantizar la soberanía alimentaria. Así mismo, con la implementación de políticas públicas que comprenden (MCP, 2011):

- a) Brindar asistencia técnica;
- b) Implementar coordinadamente con otros niveles de gobierno la producción sustentable de alimentos;
- c) Planificar y construir infraestructura adecuada;
- d) Fomentar el acceso de los ciudadanos a alimentos suficientes y sanos.

Bajo el marco legal antes mencionado surgieron en los últimos años varios programas emblemáticos que nacieron con el objetivo de ayudar a alcanzar la soberanía alimentaria en el Ecuador, por ejemplo:

Bajo el marco legal antes mencionado surgieron en los últimos años varios programas emblemáticos con el objetivo de ayudar a alcanzar la soberanía alimentaria en el Ecuador. A través de estos programas se ha ido marcado una hoja de ruta, a pesar de que algunos de ellos ya no están en vigencia, como por ejemplo el programa Aliméntate Ecuador⁹ (2003-2011), el cual tenía como objetivo reducir problemas alimentarios presentes en el país, en base a dos sub-programas: Protección Alimentaria y Alimentación Nutricional (SIISE, 2010a). De igual forma, el programa de Complementación Alimentaria (2002-2010), tuvo como objetivo combatir la desnutrición en las niñas y niños pequeños, en especial los de sectores vulnerables (SIISE, 2010c).

Por otro lado, dentro de los programas que permanecen vigentes se encuentran el programa Alimentación Escolar, cuyo propósito es el de brindar de manera gratuita servicios de alimentación escolar, con el objetivo de reducir la brecha en el acceso a la universalización de la educación y al mejoramiento de su calidad y eficiencia y que a la vez, mejore el estado nutricional de los estudiantes de instituciones públicas

9 Registro oficial No. 603-24 de junio del 2002.

(SIISE, 2010b). El Bono de Desarrollo Humano (BDH) que busca combatir la pobreza a través de la entrega de compensaciones monetarias directas (MIES, 2017).

Finalmente, el programa más reciente en entrar en vigencia se denomina Misión Ternura. Este programa se presenta como la continuación de la Estrategia Acción Nutrición a raíz de la conformación de la Secretaría Técnica “Plan Toda una Vida”, en reemplazo del Ministerio Coordinador de Desarrollo Social. El objetivo de este programa es: “impulsar el desarrollo infantil integral, a través del diseño e implementación de mecanismos de política pública y coordinación interinstitucional, nacional y local”. La población objetivo, son las niñas y niños menores de 5 años, con un mayor énfasis en los menores de 2 años (1000 primeros días).

Varias de las leyes mencionadas anteriormente siguen las líneas dispuestas en diferentes acuerdos internacionales. Por ejemplo, el Ecuador fue una de las naciones que suscribió los Objetivos de Desarrollo del Milenio, mismos que formaron parte del Plan Nacional para el Buen Vivir 2013-2017, en donde, entre otras cosas, se hacía énfasis en la reducción del hambre (SENPLADES, 2015). También, Ecuador al ser parte de la Comunidad de Estados Latinoamericanos y Caribeños (CELAC) forma parte del Plan para la Seguridad Alimentaria, Nutrición y Erradicación del Hambre de la CELAC 2025. El objetivo de este plan es el mejoramiento de la calidad de vida de las personas, reduciendo la pobreza y garantizando la seguridad alimentaria y la nutrición, con enfoque de género y respetando la diversidad de hábitos alimentarios (CELAC, 2015).

Adicionalmente el Ecuador como miembro de la Organización de las Naciones Unidas (ONU) fue parte del acuerdo establecido en septiembre del 2015, en donde se da inicio a la Agenda 2030 para el Desarrollo Sostenible. Como ya se mencionó, el segundo objetivo propone a mejorar las condiciones de vida de la población, a través de la implementación de metas que generen cambios en el sistema mundial de agricultura y alimentación. Cuenta con ocho metas, de las cuales, la meta 2.1 es la que se enfoca de manera más específica a resolver el problema de la inseguridad alimentaria. Esta meta cuenta a su vez con dos indicadores:

1. Prevalencia de la subalimentación
2. Prevalencia de la inseguridad alimentaria moderada o grave en la población, según la Escala de Experiencia de Inseguridad Alimentaria (FIES).

Sobre el segundo indicador se presenta en las siguientes secciones un desarrollo conceptual y metodológico contextualizado a nuestro país en base a la propuesta internacional de FAO.

5. Marco Conceptual

La seguridad alimentaria

El concepto de seguridad alimentaria ha evolucionado a través de los años. El primer acercamiento conceptual se generó en los años setenta, en la *Conferencia Mundial de la Alimentación* organizada por la FAO en 1974. En la *Declaración Universal sobre la Erradicación del Hambre y la Malnutrición*. Los gobiernos participantes enunciaron que era responsabilidad de toda la comunidad internacional “asegurar la disponibilidad en todo momento de suministros mundiales adecuados de alimentos básicos mediante reservas adecuadas, incluidas las reservas de emergencia” (UN, 1975, p. 10). En ese sentido, se reconoce la importancia de que los gobiernos trabajen por “una mayor producción de alimentos y una distribución más equitativa y eficiente de los alimentos entre los países y dentro de los países” (UN, 1975, p. 8).

Hasta ese momento, se ve a la inseguridad alimentaria exclusivamente como un problema que proviene desde la oferta –o disponibilidad– de alimentos; una visión influenciada por las crisis alimentarias post guerras mundiales que trajeron tanto escasez de alimentos como fluctuaciones en los precios de los mismos (Napoli, 2010). Por ello, en esta conferencia se establecieron como objetivos principales: garantizar la disponibilidad de alimentos, a través de un aumento de la producción y el uso de excedentes, mejorar el consumo y la distribución de alimentos, y garantizar la estabilidad de precios a nivel nacional e internacional (UN, 1975).

Con la llegada de los años ochenta, la producción de alimentos incrementó significativamente, no obstante, contrario a los pronósticos de la conferencia, el problema del hambre no desapareció. Es aquí cuando se comienza a comprender que un adecuado suministro de alimentos no es la única arista de la seguridad alimentaria (Napoli, 2010). En 1981, Amartya Sen comienza su libro *Pobreza y Hambrunas* puntualizando que la “inanición es la característica de que algunas personas no tengan suficientes alimentos

para comer. No es la característica de que no hayan suficientes alimentos para comer” (Sen, 1981, p. 11), añadiendo que si bien la última puede causar la primera, ésta es apenas una de las posibles causas.

A partir de ese momento se comienzan a considerar las barreras económicas y físicas para acceder a los alimentos. Los aspectos físicos incluyen el estado de las carreteras y demás vías para transportar los alimentos, mientras que los aspectos económicos se relacionan al poder adquisitivo de los diferentes grupos sociales. Sin embargo, el acceso solo puede prevenir el hambre si se acompaña de estabilidad (Napoli, 2010).

Es por eso que en 1983 la FAO incluyó el acceso a la producción y la estabilidad de precios en el concepto, señalando en uno de sus reportes que "el objetivo último de la seguridad alimentaria mundial debe ser garantizar que todas las personas en todo momento tengan tanto acceso físico y económico a los alimentos básicos que necesitan. [...] [Buscando] garantizar la producción de suministros alimentarios adecuados; maximizar la estabilidad en el flujo de suministros; y asegurar el acceso a los suministros disponibles por parte de aquellos que los necesitan" (WFS, 1983, p. 8).

El último elemento importante que se incorporó en la definición de seguridad alimentaria es la calidad de los alimentos suministrados, en base a la idea de que no solo deben satisfacer las necesidades proteínicas y de energía del organismo sino que también deben aportar el equilibrio nutricional necesario para una vida sana y activa, reconociendo las preferencias, hábitos, tradiciones y tipos de alimentos socialmente aceptables (Napoli, 2010).

La definición que considera todos estos aspectos se concretó en la *Cumbre Mundial sobre la Alimentación* llevada a cabo en 1996, donde se declara que:

Existe seguridad alimentaria cuando todas las personas tienen en todo momento acceso físico y económico a suficientes alimentos inocuos y nutritivos para satisfacer sus necesidades alimenticias y sus preferencias en cuanto a los alimentos a fin de llevar una vida activa y sana (FAO, 2006).

Hasta la fecha, esta es la definición que ha adquirido la mayor aceptación (Jones, Ngure, Pelto, & Young, 2013) y hace referencia a cuatro grandes componentes o dimensiones del fenómeno: la disponibilidad de alimentos, el acceso a los alimentos, la utilización biológica de los alimentos y la estabilidad en el tiempo

de los mismos, conocidos también como los cuatro pilares de la seguridad alimentaria (Tabla 1).

Tabla 1: Dimensiones de la Seguridad alimentaria

Dimensión	Definición
Disponibilidad	Disponibilidad de cantidades suficientes de alimentos de calidad apropiada, proporcionados por la producción a nivel local o nacional, importaciones, almacenamientos y ayuda alimentaria, que se encuentra presente en una determinada área.
Acceso	Acceso a los recursos adecuados para poder adquirir y/o producir alimentos apropiados para una alimentación nutritiva. Estos recursos pueden ser alimentos disponibles en el mercado, así como también los insumos agrícolas, medios de producción, conocimiento, tierra, agua, tecnología, entre otros, sobre los cuales las personas tienen derecho; dados los arreglos legales, políticos, económicos y sociales de la comunidad en la que viven.
Utilización	Utilización biológica de los alimentos, a través de una dieta adecuada, agua potable, sanidad y atención médica, que permita lograr un estado de bienestar nutricional, donde se satisfagan todas las necesidades fisiológicas de la persona. Una inadecuada utilización biológica puede tener consecuencias como la desnutrición y/o la malnutrición, problemas que pueden traer repercusiones a largo plazo o permanentes –especialmente si sucede en la niñez–.
Estabilidad	Una persona, hogar o población, debe poder acceder a alimentos adecuados en todo momento. Factores externos como crisis repentinas o acontecimientos cíclicos, no deberían poner en riesgo el acceso ni la disponibilidad de alimentos a la población. Cobran importancia la infraestructura de almacenamiento a nivel nacional y/o local, así como la disponibilidad de recursos alimenticios e insumos de contingencia. Se relaciona tanto a la disponibilidad como al acceso a alimentos.

Fuente: FAO (2016b) y MCDS/FAO (2010)
Elaboración: Autores

Hoy en día no se cuenta con un indicador que, por sí solo, pueda explicar todas las dimensiones de la inseguridad alimentaria, por lo cual la discusión se ha enfocado en definir un conjunto de indicadores que permitan medir los distintos aspectos del fenómeno (FAO, 2016a). La propuesta global de FAO se encuentra resumida en el Anexo 1 y los indicadores más relevantes por dimensión que han sido reportados por Ecuador se pueden ver en el Anexo 2.

La gran ventaja de contar con este conjunto de indicadores, complementarios entre sí, es que logra visibilizar las cuatro dimensiones de la Seguridad Alimentaria, aunque históricamente se han encontrado dificultades en el seguimiento de muchos de ellos (FAO; FIDA & PMA, 2013). Por otro lado, la desventaja es que se enfocan en medir causas y consecuencias (resultados), y no alcanzan a medir el fenómeno de manera directa (Comité Científico de la ELCSA, 2012).

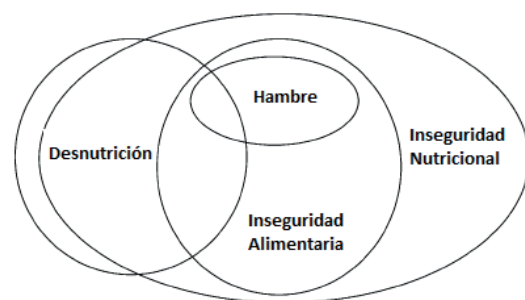
La Inseguridad Alimentaria y su medición en base a escalas

La ausencia de seguridad alimentaria se traduce en inseguridad alimentaria, y de la misma forma puede ser medida en cada una de sus dimensiones. En este apartado se hace un breve resumen de la relación entre inseguridad alimentaria y conceptos que se traslapan con la misma; y además se explica una de las propuestas de su medición en relación a la dimensión de acceso.

La inseguridad alimentaria ha tenido numerosas y variadas definiciones, debido en parte a su naturaleza multidimensional y multisectorial, lo que ha ocasionado que se confunda muchas veces con conceptos similares como el hambre, la desnutrición y la inseguridad nutricional (Jones et al., 2013). En la Figura 1, se puede observar cómo estos conceptos se traslapan y relacionan entre sí. A continuación se

define con claridad cada concepto de tal forma que se pueda establecer qué es lo que se mide y qué no se mide en el presente ejercicio.

Figura 1: Conceptos traslapados con la inseguridad alimentaria



Fuente: (Jones et al., 2013)
Elaboración: Autores

Inseguridad alimentaria: Ausencia de una o más de las condiciones establecidas para la seguridad alimentaria en la definición instaurada desde la CMA de 1996 (Jones et al., 2013). Esto es, la “situación que se da cuando las personas carecen de acceso seguro a una cantidad suficiente de alimentos inocuos y nutritivos para su crecimiento y desarrollo normal así como para llevar una vida activa y sana”. Se determina entonces cuatro posibles causas de la inseguridad alimentaria: no disponibilidad de alimentos, poder adquisitivo insuficiente, una distribución inapropiada o uso inadecuado de los alimentos en el interior del hogar (FAO et al., 2015, p. 58).

Inseguridad nutricional: A veces, la inseguridad nutricional se utiliza indistintamente con la inseguridad alimentaria, pero esta definición es más amplia. La inseguridad alimentaria es condición necesaria pero no suficiente para la seguridad nutricional, ya que ésta también considera las prácticas de atención y cuidado, salud, higiene y saneamiento, además de la suficiencia de la dieta (Jones et al., 2013). Se define a la seguridad nutricional como la situación que se da cuando se dispone de “acceso seguro a una dieta suficientemente nutritiva combinado con un entorno salubre y servicios sanitarios y de atención de salud adecuados, a fin de que todos los miembros de la familia puedan llevar una vida sana y activa” (FAO et al., 2015, p. 58)

Desnutrición: Se la considera una forma de inseguridad alimentaria extrema, donde la “ingesta calórica es inferior al mínimo requerimiento de energía dietética” (Jones et al., 2013, p. 482). La FAO define la desnutrición como el “resultado de la

subalimentación, o de absorción y/o uso biológico deficientes de los nutrientes consumidos como resultado de repetidas enfermedades infecciosas”. Este concepto comprende “la insuficiencia ponderal en relación a la edad, la estatura demasiado baja para la edad (retraso del crecimiento), la delgadez peligrosa en relación con la estatura (emaciación) y el déficit de vitaminas y minerales (malnutrición por carencia de micronutrientes)” (FAO et al., 2015, p. 58).

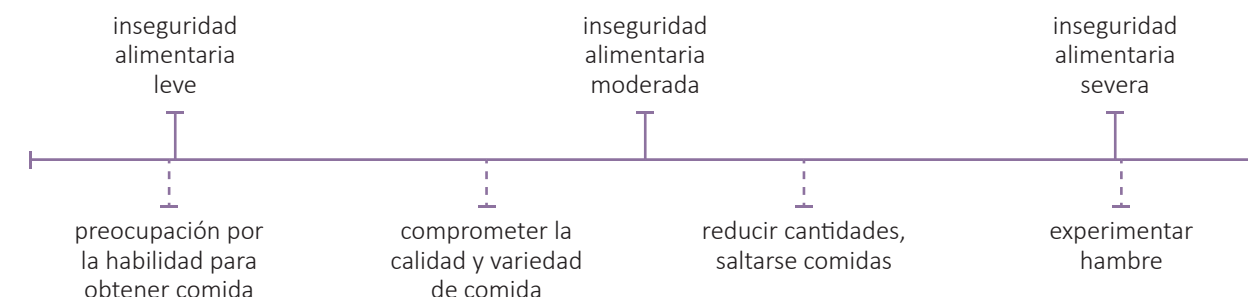
Hambre: De acuerdo a (Jones et al., 2013, p. 482), la definición más explícita de hambre se encuentra en un informe de 1990 del American Institute of Nutrition (ahora ASN), donde se la define como la “sensación incómoda o dolorosa causada por la falta de alimento” y una “recurrente e involuntaria falta de acceso a los alimentos”. En el informe de la FAO de 2015 sobre el estado de la inseguridad alimentaria en el mundo, consideran al hambre como sinónimo de subalimentación crónica. En el mismo informe se define la subalimentación como el “estado, con una duración de al menos un año, de incapacidad para adquirir alimentos suficientes, que se define como un nivel de ingesta de alimentos insuficiente para satisfacer las necesidades de energía alimentaria” (FAO et al., 2015, p. 58).

Escala de Experiencia de Inseguridad Alimentaria (FIES)

La FIES, planteada en 2013 por el proyecto VoH, es una escala basada en la experiencia de la inseguridad alimentaria de los individuos u hogares que permite calcular indicadores *válidos y fiables* de la prevalencia de la inseguridad alimentaria (FAO, 2016a). Se trata de una métrica que permite calibrar el alcance y la profundidad del acceso limitado a los alimentos a nivel de país, permitiendo comparaciones internacionales (Ballard, Kepple, & Cafiero, 2013).

El origen de estas mediciones basadas en la experiencia de los individuos proviene de una investigación etnográfica llevada a cabo en Estados Unidos, que estuvo enfocada en entender la experiencia del hambre entre las personas. Este estudio identificó que se trata de un proceso que comienza por una sensación de ansiedad por tener suficiente comida, seguido de cambios en la dieta para hacer que los recursos alimenticios limitados duren (reducción en calidad), y por último, una disminución del consumo de alimentos en el hogar (reducción en cantidad) (FAO, 2016a). Por esta razón, las diferentes experiencias –y sus severidades respectivas– suelen asociarse a diferentes niveles de inseguridad alimentaria (Gráfico 2).

Gráfico 2: Las experiencias de inseguridad alimentaria y los niveles de seguridad asociadas



Fuente: FAO, 2016a
Elaboración: Autores

El Módulo de la Encuesta de Seguridad Alimentaria de los Estados Unidos (US HFSSM) –aplicado en ese país desde 1995– se basó en esta construcción teórica, así como en otras escalas y encuestas de inseguridad alimentaria como: escala de experiencia Radimer/Cornell, la *Continuing Survey of Food Intake by Individuals* (CSFII), la *National Health and Nutrition Examination Survey* (NHANES) y el *Comunity Childhood Hunger Identification Project* (CCHIP), aplicadas previamente en los Estados Unidos durante los años de 1988-1995.

A nivel regional, la primera fue la Escala Brasileira de Inseguridad Alimentaria (EBIA) ,levantada en Brasil desde 2004, posteriormente está la Escala Lorenza de Colombia –año 2004–, luego vienen: la Escala Latinoamericana y Caribeña de Inseguridad Alimentaria (ELCSA), y la Escala Mexicana de Seguridad Alimentaria (EMSA), entre las más importantes (FAO, 2016a). La FIES, al igual que sus predecesoras, se sustenta dentro de este marco conceptual.

Todas estas escalas basadas en la experiencia parten de una hipótesis esencial: la severidad de la condición de inseguridad alimentaria de un individuo, hogar o población puede evaluarse como un rasgo latente. Esto es, una característica, o rasgo, que no se puede observar directamente pero que sí puede deducir de la evidencia observable al aplicar algún modelo de la Teoría de Respuesta al Ítem¹⁰ (IRT), específicamente, el Modelo de Rasch (Ballard et al., 2013).

Si se aplica adecuadamente la metodología del modelo de Rasch a la medición de la inseguridad alimentaria, se puede asegurar dos elementos importantes:

¹⁰ Conjunto de métodos enraizados en estadísticas con amplia aplicación a los problemas de medición en los ámbitos de las ciencias humanas y sociales (FAO, 2016).

1. La severidad de la condición de inseguridad alimentaria del encuestado y la severidad asociada a cada una de las experiencias pueden situarse en la misma escala unidimensional. El supuesto básico del Modelo de Rasch asume que la probabilidad de responder afirmativamente a una pregunta es independiente del resto de las preguntas. La escala es unidimensional, y por lo tanto mide una sola dimensión del fenómeno en cuestión, es decir la falta de acceso a alimentos en el hogar.
2. La mayor severidad de la condición de inseguridad alimentaria de un encuestado aumentará la probabilidad de reportar la ocurrencia de experiencias asociadas con la inseguridad alimentaria (FAO, 2016a). A continuación, se profundizará en la teoría detrás del modelo IRT para estimar la FIES.

6. Marco Metodológico

A continuación se presenta el IRT en términos generales; y, detalla al Modelo de Rasch y su aplicación específica para la medición de Inseguridad Alimentaria según (FAO, 2016a).

a. Detalles generales de la Teoría de Respuesta al Ítem

La IRT se refiere a un conjunto de procedimientos estadísticos que buscan modelar la relación entre una variable latente con las respuestas dadas por las personas entrevistadas a cada pregunta (ítem) de un instrumento (prueba diseñada para medir el rasgo latente) (Hambleton, Robin, & Xing, 2000). Una variable latente se refiere a una característica no observable directamente.

Por ejemplo, la ansiedad podría definirse como sentimientos que varían entre inquietud a incapacitantes ataques de terror. Este tipo de fenómenos por su propia naturaleza no pueden medirse directamente, por ello y para fines de estimación mediante modelos estadísticos se denominarían variables latentes (Ayala, 2009). Se puede estudiar, como variable latente, rasgos psicológicos como ansiedad, impulsividad y depresión; estado de salud o calidad de vida; conocimiento; habilidades (intelectuales e inclusive habilidades naturales como la capacidad de crecer de los vegetales bajo un clima específico) (Hardouin, 2007).

La Teoría de Respuesta al Ítem no es una teoría en el sentido tradicional del término ya que no plantea explicación alguna sobre las razones de las personas encuestadas a responder de la forma en la que lo hacen. Se podría decir que IRT es la teoría de estimación estadística que usa una caracterización de la variable latente de los individuos y los ítems como predictores de las respuestas observadas, tratando a los individuos que responden dichas preguntas como cajas negras. Para estimar esta caracterización los ítems y las personas se

encuentran en el mismo continuum (valores que puede obtener la variable latente). De esta forma los individuos se caracterizan por su posición en la variable latente y los ítems por su localización y otras propiedades que se analizarán más adelante (tres parámetros denominados a , b y c) (Ayala, 2009). Los modelos IRT pueden estudiar ítems con dos categorías (dicotómico) o con múltiples categorías (politómico). Considerando que el Modelo de Rasch trata únicamente con ítems dicotómicos a continuación se detallan únicamente dicha familia de modelos.

b. Modelos IRT para Ítems Dicotómicos

Considerando que los ítems estudiados son dicotómicos, la respuesta puede ser 0 o 1, para fines de comprensión del documento se denominará la respuesta igual a 1 como *respuesta correcta*. En IRT se puede expresar la *probabilidad de respuesta correcta* como una función logística de 3, 2 o 1 parámetros (3PL, 2PL y 1PL respectivamente) (DeMars, 2010); donde el modelo 3PL será el caso general del resto de modelos. Según Hambleton et al (2000) se puede expresar al modelo 3PL como:

$$P_j(X_{ij} = 1 | \theta_i, a_j, \beta_j, c_j) = c_j + (1 - c_j) \frac{e^{Da_j(\theta - \beta_j)}}{1 + e^{Da_j(\theta - \beta_j)}}, j = (1, 2, \dots, k). \quad (1)$$

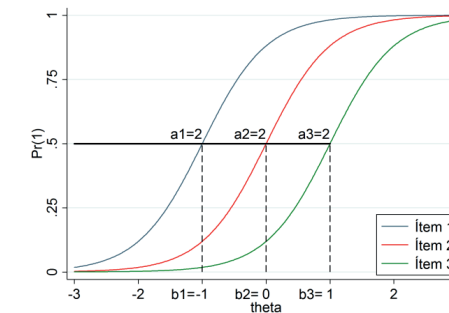
Donde $P_j(X_{ij} = 1 | \theta_i, a_j, \beta_j, c_j)$ es la probabilidad de un individuo de obtener una respuesta afirmativa para el ítem j en función de la variable latente θ , y los parámetros a_j , β_j y c_j . Además, X_{ij} se referirá a la respuesta del individuo i en el ítem j . Para simplificar la notación, cuando sea posible, se expresará esta probabilidad únicamente como $P_j(\theta)$. El parámetro a_j se refiere a la discriminación del ítem j , β_j es el parámetro de dificultad del ítem j , c_j será el parámetro de adivinanza, también conocido como asíntota inferior (DeMars, 2010). Finalmente, D es un parámetro de escalamiento. Esta representación de la relación entre la variable latente y la respuesta correcta del ítem se denomina Función Característica del Ítem o ICC por sus siglas en inglés.

La dificultad (β_j) simboliza que tan difícil es responder correctamente a un ítem. Por ejemplo, obsérvese el Gráfico 3, donde el ítem 3 es *más difícil* que el 1, si un individuo tuviera un valor de θ de 1, esta persona tendría una probabilidad de responder correctamente el ítem 3 del 50%, mientras que tendría una probabilidad cercana al 100% de responder el ítem

1. Formalmente β_j se define como el valor de θ que corresponde a una probabilidad del 50% de responder correctamente al ítem j .

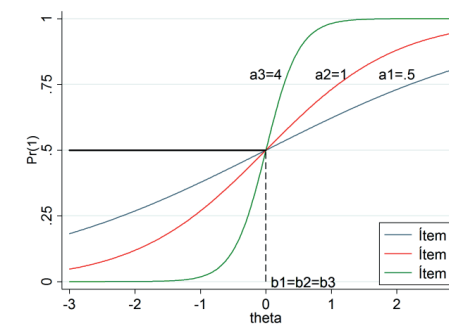
La discriminación (a_j) es la capacidad del ítem de diferenciar entre dos individuos un valor de variable latente diferente, es igual a la pendiente de la ICC. Como se puede observar en el Gráfico 3, el ítem 3 discrimina mejor que el ítem 1. Si bien para los dos ítems cuando un individuo presenta un valor de θ igual a 0 la probabilidad de responder correctamente es 50%, cuando θ es igual a -1 la probabilidad de responder correctamente el ítem 1 es de 40% mientras que para el ítem 3 es cercana al 0% (DeMars, 2010).

Gráfico 3: ICC igual discriminación y parámetro de adivinanza-distinta dificultad



Fuente y elaboración: (Jácome & Restrepo, 2017, p. 19)

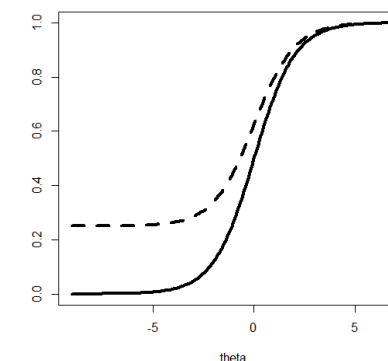
Gráfico 4: ICC igual dificultad y parámetro de adivinanza-distinta discriminación



Fuente y elaboración: (Jácome & Restrepo, 2017, p. 19)

El parámetro de adivinanza c_j es la probabilidad que un examinado con algún valor muy bajo en la variable latente responda correctamente. Se considera que un buen instrumento está diseñado de tal forma que evita que personas con bajos valores en la variable latente logren adivinar la respuesta correcta. Se puede apreciar en el Gráfico 5 que el ítem 3 (línea sólida) tiene un mayor parámetro de adivinanza que el ítem 1 (línea entre cortada).

Gráfico 5: ICC igual discriminación y dificultad, distinto parámetro de adivinanza



Fuente: DeMars, 2010
Elaboración: Autores

El modelo 3P corresponde a una estimación de los 3 parámetros anteriormente mencionados, el modelo 2P asume que $c_j = 1$. El modelo 1P presume que $c_j = 1$ y a_j es igual a alguna constante (Nering & Ostini, 2010, p. 28).

Es también de interés conocer cuántas respuestas correctas obtendrá una persona dado un valor de θ , para responder esta pregunta en IRT se utiliza la Curva Característica de la Prueba o TCC por sus siglas en inglés. La TCC presenta la cantidad de respuestas correctas esperadas dado un valor de θ (DeMars, 2010, p. 22). Matemáticamente la TCC se expresa:

$$TCC(\theta) = \sum_{j=1}^p P_j(\theta) \quad (2)$$

c. Modelo de Rasch

El Modelo de Rasch es un modelo 1PL donde para todo ítem j el parámetro a_j será igual a 1; y, para todo j el parámetro c_j será igual a 0. Los primeros investigadores de los modelos IRT no presentaron atención a los modelos 1PL ya que los consideraban poco realistas. En la perspectiva con la que Rasch desarrolló su modelo, no eran de interés valores de discriminación diferentes entre ítems o parámetros de adivinanza diferentes de 0, ya que se encontraba trabajando con pruebas de inteligencia (Nering & Ostini, 2010). Cuando los supuestos necesarios para la correcta estimación de un modelo Rasch o 1PL se mantienen, estos tienen propiedades estadísticas deseables que no se pueden obtener con modelos 2PL o 3PL. La primera de estas propiedades establece que el puntaje bruto es un estadístico suficiente para la estimación θ , lo que significa que los examinados que tienen el mismo número de respuestas correctas (puntaje bruto) tendrán el mismo valor de θ . Por otra parte en los modelos 2PL y 3PL dos personas con la misma cantidad de respuestas correctas tendrá diferentes puntajes si sus patrones de respuesta correcta fueron diferentes. Además, la proporción de respuestas correctas por ítem es un estadístico suficiente para la estimación de la dificultad, por lo tanto, dos ítems con la misma proporción de respuestas correctas tendrán la misma dificultad (DeMars, 2010).

El modelo de Rasch tiene dos versiones dependiendo de cómo se considere a la variable latente: efectos fijos o aleatorios. Considerando que FAO (2016) utiliza un modelo de efectos fijos se desarrollará únicamente dicha versión.

El modelo de Rasch explica la ocurrencia de una matriz de datos que contienen respuestas dicotómicas de una muestra de n sujetos $\{S_1, S_2, \dots, S_n\}$ que responden a un número fijo de ítems $\{I_1, I_2, \dots, I_k\}$ que intentan medir la variable latente de interés (en este caso la inseguridad alimentaria). Cada individuo S_i tiene asociado un parámetro de θ_i definido en los reales, que define su nivel de inseguridad alimentaria dentro del continuum θ . En adición cada ítem I_j tiene asociado un parámetro de dificultad β_j definido en los reales. Las respuestas de los n individuos a los k ítems se almacenan en una matriz de datos $X_{n \times k}$ con elementos $x_{ij} \in \{0, 1\}$, que representan las respuestas afirmativas ($x_{ij} = 1$) o negativas ($x_{ij} = 0$) de los individuos al ítem I_j . Finalmente el modelo de Rasch asocia una probabilidad para x_{ij} según la siguiente forma funcional de la ICC (Fischer & Molenaar, 1995):

$$P(X_{ij} = x_{ij} | \theta_i, \beta_j) = \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{x_{ij}(\theta_i - \beta_j)}} \quad (3)$$

Los parámetros pueden ser estimados mediante diferentes técnicas, sin embargo el método más recomendado es el método de máximo verosimilitud condicional o CML (Fischer & Molenaar, 1995). La FAO utiliza CML para estimar Rasch (Nord, 2014). Al utilizar CML se condiciona la función de Máxima Verosimilitud a θ_i , lo que implica que la CML ya no dependerá de θ_i simplificando los cálculos. Bajo esta metodología la ecuación (3) se convierte en:

$$P(X_{ij} = 1) = \frac{\xi_i \epsilon_j}{1 + \xi_i \epsilon_j} \quad (4)$$

$$P(X_{ij} = 0) = \frac{1}{1 + \xi_i \epsilon_j} \quad (5)$$

En donde $\xi_i = e^{\theta_i}$ y $\epsilon_j = e^{-\beta_j}$ (Fischer & Molenaar, 1995). En base a estas ecuaciones se puede construir la siguiente función de máximo verosimilitud para el caso general de k ítems y n personas.

$$L_c(\epsilon | r) = \left(\prod_i \gamma_{r_i} \right)^{-1} \prod_j \epsilon_j^{X_{ij}}, \quad (6)$$

Donde:

$$\gamma_{r_v}(\epsilon_1, \dots, \epsilon_k) = \sum_{\mathcal{Y}(r_v)} \prod_{i=1}^k (\epsilon_i)^{y_i}, \quad (7)$$

Y la sumatoria de todas las posibles patrones de respuesta $\mathbf{y} = (y_1, \dots, y_k)$ con

$$\sum_{i=1}^k y_i = r_v. \quad (8)$$

Nótese que la función de máxima verosimilitud está construida para el vector de ítems ϵ de dimensión k y el vector r de todas las notas de todos los individuos r_i . Al maximizar la transformación logarítmica de la ecuación (6), $\ln L_c$ obtenemos el siguiente conjunto de ecuaciones de CML:

$$x_{.j} - \sum_{i=1}^n \frac{\epsilon_i \gamma_{r_i-1}}{\gamma_{r_i}} = 0, \quad (9)$$

Donde $x_{.j} = \sum_{i=1}^n x_{ij}$, para $j = 1, 2, \dots, k$. La solución iterativa de todas estas ecuaciones nos permite estimar los parámetros de los ítems. El método de CML tiene propiedades de convergencia; consistencia para $n \rightarrow \infty$ y k fijo; eficiencia asintótica y con un pérdida de información trivial para $n \rightarrow \infty$ (Fischer & Molenaar, 1995). Nótese que en las ecuaciones de la 6 a la 9 el término ξ_i no está presente, por lo que se puede estimar los parámetros de los ítems independientemente de la variable latente.

d. Supuestos del Modelo

Según Hardouin el IRT se basa en los siguientes tres supuestos fundamentales (Hardouin, 2007):

- Unidimensionalidad: las respuestas observadas a los ítems dependerán únicamente de una variable latente θ .
- Monotonidad: la probabilidad de respuesta correcta es una función monótona no-decreciente de θ . Lo que quiere decir que al aumentar el valor de θ , la probabilidad de respuesta correcta no puede decrecer.
- Independencia local: todos los ítems estudiados serán independientes entre sí al condicionar con θ .

Los modelos presentados asumen unidimensionalidad; sin embargo, existen modelos IRT que permiten modelar espacios multidimensionales cuando las respuestas de las personas se basan en más de una variable latente (Ayala, 2009). La independencia local es un supuesto necesario para realizar la estimación

de los parámetros estudiados (Hardouin, 2007). Al momento de escoger entre los modelos disponibles se incluye un supuesto adicional: los datos observados pueden ser representados por el modelo escogido. Si el supuesto de unicidad no se cumpliera los parámetros se encontrarían estimados incorrectamente. En este punto cabe recalcar una diferencia, no es igual la unidimensionalidad matemática que la conceptual, puede suceder que un grupo de ítems se conceptualice en de varias dimensiones pero matemáticamente se pueda representar en una sola (DeMars, 2010).

e. Evaluación de supuestos

Unidimensionalidad: Valores Propios

Se puede realizar un estudio de los Valores Propios de un Análisis de Componentes Principales (ACP) de los ítems de interés. Considerando que se usan ítems dicotómicos las correlaciones de Pearson presentarán estimaciones incorrectas por lo que se recomienda el uso de correlaciones Tetracóricas en el cálculo de los Valores Propios.

Infit y outfit

Estos estadísticos analizan la bondad de ajuste en base a los residuos, es decir comparan a los valores observados del modelo versus los valores esperados utilizando el modelo de Rasch para evaluar el supuesto de unidimensionalidad (Christensen, Kreiner, & Mesbah, 2012, p. 84).

Los residuos se definen como:

$$r_{ij} = x_{ij} - \hat{\pi}_{s_{ij}},$$

Donde $\hat{\pi}_{s_{ij}}$ es la estimación bajo el modelo de Rasch de la probabilidad de una respuesta afirmativa dada una suma de respuestas afirmativas S para el ítem j .

El índice infit para el ítem j es:

$$INFIT_j = \frac{\sum_{i=1}^N r_{nj}^2 w_i}{\sum_{i=1}^N \hat{\pi}_{ij}(1 - \hat{\pi}_{ij}) \cdot w_i},$$

y el índice outfit es, donde w_i es la ponderación del individuo i .

$$OUTFIT_j = \frac{1}{\sum_{i=1}^N w_i} \times \sum_{i=1}^N \frac{r_{nj}^2 w_i}{\hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}$$

Si los supuestos del modelo de Rasch se cumplen, el valor esperado de estos índices debe ser igual a uno¹¹. Valores del *infit* o *outfit* cercanos a cero o muy lejanos a uno son evidencia de problemas con el ajuste del modelo (Christensen et al., 2012, p. 86). El *outfit* permite detectar *outliers* o respuestas afirmativas incongruentes en los primero o últimos ítems, mientras que el *infit* denota patrones no esperados en ítems contiguos, cercanos al medio de la escala o *inliers* (Hardouin, 2007). No se puede determinar con facilidad cuando estos índices toman valores muy altos o muy bajos para que el modelo sea aceptable. Hay dos razones principales detrás de este fenómeno: la distribución de los estadísticos y la forma en que se estima $\hat{\pi}_{s_{ij}}$. En referencia al primer punto, se puede transformar los residuos estandarizados, pero esto conlleva añadir más supuestos al modelo. Respecto al segundo punto $\hat{\pi}_{s_{ij}}$ depende de dos parámetros, pero solo se puede obtener estimadores consistentes de uno de ellos (Christensen et al., 2012, p. 86). Sin embargo, para la FAO valores de *infit* entre 0,8 y 1,2 son muy buenos; mientras que valores entre 0,7 y 1,3 son utilizables. En el caso del *outfit* la interpretación es similar a los *infits*, pero se debe tomar en cuenta que el indicador es muy sensible a pocas observaciones muy incongruentes (por ejemplo el individuo i tiene un *raw score* de 2 pero contesta el afirmativamente el ítem 8), en general no hay un valor de corte establecido para los *outfits* (Nord, 2014).

Independencia local: correlación residual

Si los ítems no presentaran independencia local otra dimensión podría causar esta dependencia, violando a la vez el supuesto de unidimensionalidad, obteniendo parámetros incorrectamente estimados. Las pruebas de unidimensionalidad solo pueden detectar dimensiones adicionales que influyeran a grandes grupos de ítems, mientras que las pruebas de independencia local pueden detectar dimensiones subyacentes por pares de ítems.

Si los ítems son localmente independientes se encontrarán no correlacionados tras condicionar a θ . Una metodología usada para medir independencia local es mediante la matriz de correlaciones residuales, para ítems dicotómicos. Se calcula el residuo como $e = P(\theta) - X_{ij}$. Se considera una correlación residual como problemática si es mayor a 0,2 (DeMars, 2010).

11 Para ello primero se debe estandarizar el residuo.

f. Aplicación según FAO

La FAO, en la encuesta mundial de Gallup incluye, en el 2014, la FIES. Esta es una encuesta multipropósito que se realiza desde el 2005 e incluye a 150 países (C. Cafiero et al., 2016). La encuesta está dirigida a individuos adultos mayores a 15 años. En base a esta encuesta se crea un estándar internacional o escala global de inseguridad alimentaria que permite la comparabilidad de las escalas de inseguridad alimentaria con aquellas que siguen la metodología de la FAO.

La propiedad de la invariancia de los parámetros de los ítems inherente a los modelos IRT, implica que los parámetros de los ítems deberían ser los mismos para poblaciones diferentes (DeMars, 2010). Esta propiedad permite la comparación entre la escala FIES y la escala del Ecuador. Para que los datos sean comparables se hace un análisis comparativo entre la escala global y la escala nacional y se determinar hasta qué punto las preguntas de una y otra escala son comparables. En el caso del Ecuador se hizo un ejercicio a finales del 2017 cuando se incluye un módulo de inseguridad alimentaria en la ENEMDU de diciembre en la sección 14A con características similares a la FIES. La única diferencia importante es la inclusión de una pregunta adicional. Así la pregunta 8 indaga alguna persona adulta del hogar ha dejado de comer alguna vez durante el día (pregunta que proviene de la ELCSA), y la pregunta 9 que se refiere a si alguna persona del hogar dejó de comer todo el día (pregunta 8 de la escala FIES). La razón de esta pregunta es tratar de rescatar la ELCSA. Para mayor detalle comparativo entre las preguntas del ejercicio mundial y el de Ecuador revisar el Anexo 3.

A parte de estas diferencias, es importante considerar que las poblaciones tienen apreciaciones diferentes del grado de inseguridad alimentaria que representa cada pregunta de la escala, y esto se refleja en las severidades de los ítems que resultan de la aplicación del modelo de Rasch. En adición cada país tiene diferentes diseños muestrales para sus encuestas, y esto puede tener incidencias en la varianzas de los estimadores de inseguridad alimentaria.

El paquete RM.weights (A. C. Cafiero, Nord, & Viviani, 2016) está diseñado para trabajar con todas las cuestiones antes mencionadas, exceptuando el tema de la varianza. En efecto, antes de aplicar el modelo de Rasch se deben normalizar las ponderaciones. Este procedimiento se aplica cuando el software no incluye el diseño muestral dentro de

sus procedimientos. Sin embargo, este mecanismo considera los pesos muestrales pero ignora el diseño muestral. En efecto, la aplicación de este método produce estimadores correctos del error estándar únicamente bajo el supuesto de muestreo aleatorio simple. Si en la muestra hay posibilidad de dependencia en las observaciones, se debe ajustar los pesos estandarizados (Hahs-Vaughn, 2005). Las consecuencias de no hacer estas correcciones pueden ser: sobrestimación del número efectivo de observaciones y subestimación de la variabilidad¹².

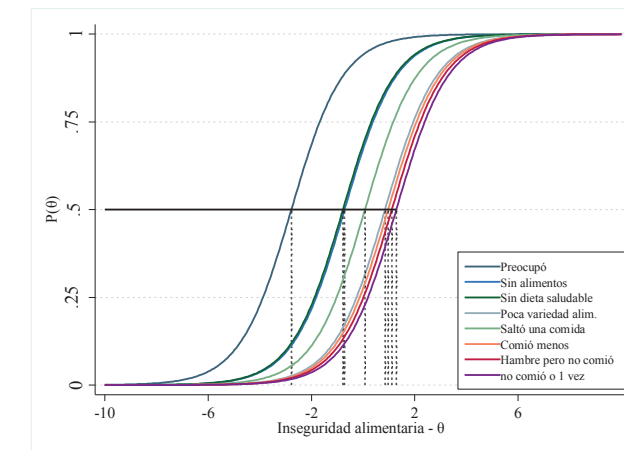
Una vez estandarizados los pesos, eliminados los valores faltantes de la matriz de respuestas y transformadas dichas respuestas en variables dicotómicas de unos y ceros; se puede utilizar la rutina RM.w del paquete RM.weights. Esta rutina utiliza el modelo de Rasch para estimar la severidad de los ítems. La rutina optimiza la función de verosimilitud condicionada, presentada en la sección anterior, con la ayuda del comando base “optim” opción “BFGS”, que hace la optimización en base a un método quasi-Newtoniano¹³. La rutina genera: los scores brutos normalizados, la severidad de los ítems normalizados, sus errores estándar, el infit y el outfit. Cabe recalcar que los parámetros para los valores extremos de los scores brutos, en este caso 0 y 8, no están definidos bajo máximo verosimilitud condicionada (MVC), por lo tanto el comando suma 0,5 al score bruto de cero y resta 0,5 al score bruto de 8 (A. C. Cafiero et al., 2016).

De la sección anterior sabemos que la estimación MVC, permite obtener los estimadores de los ítems β_j . Este es el único insumo necesario para determinar el valor de inseguridad asociado a cada ítem, es decir la estimación del parámetro de interés θ_i . Para ilustrar el proceso analizaremos los resultados de la FIES-Ecuador, para el escenario en el que se combinan las respuestas de las preguntas 8 y 9 de la sección 14A del formulario de la ENEMDU de diciembre 2017. Los β_j estimados se incorporan a la ecuación (3) para construir las ICC, para $x_{ij} = 1$. Las ICC con los datos de Ecuador se muestran a continuación:

12 Canadian Research Data Centre (2018). Toronto RDC events. Toronto, Canadá. Recuperado de: http://sites.utoronto.ca/rdc/files/3_NLSCY_Workshop__Nonresponse_and_Normalized_Weights_and_Pooling_Data_and_Full_Example.pdf ; página 4.

13 La explicación de este método está fuera del espectro del estudio, pero una explicación básica se puede encontrar en: https://en.wikipedia.org/wiki/Broyden%E2%80%93Fletcher%E2%80%93Goldfarb%E2%80%93Shanno_algorithm

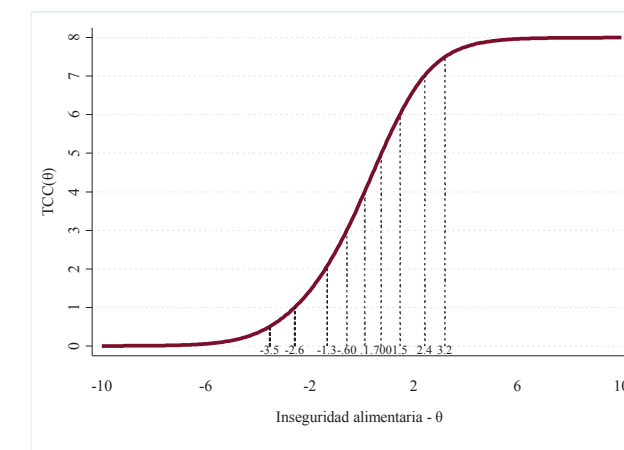
Gráfico 6: curvas ICC para Ecuador



Fuente: ENEMDU 2017
Elaboración: Autores

De la agregación de estas curvas obtenemos la TCC según la fórmula (2). La TCC se presenta en el gráfico siguiente:

Gráfico 7: curva TCC para Ecuador



Fuente: ENEMDU 2017
Elaboración: Autores

La curva TCC relaciona el score bruto y los niveles de inseguridad alimentaria, de tal forma que cada score bruto se asocia a un nivel de inseguridad alimentaria θ . Por ejemplo un *score* bruto de 1 corresponde, según la TCC, a un $\theta = -2,6$.

Para el cálculo de prevalencias, se deben definir niveles de β que se relacionen a cierto nivel de inseguridad alimentaria de interés, y que sirven como umbrales. En el caso de la metodología FIES, los β de interés pertenecen a una escala global calculada

por dicha institución, y corresponden al 4to ítem más severo (...ha comido menos) y al 8vo ítem más severo (dejó de comer una vez al día o todo el día) (C. Cafiero, Viviani, & Nord, 2018).

La escala global resulta de combinar las escalas de inseguridad alimentaria de más de 150 países seleccionados por la FAO (Nord, 2014). Para poder lograr esta combinación entre los distintos ítems, se estandarizan y se analizan cuáles son comunes; con un criterio basado en desviaciones de la mediana de los ítems de todos los países y un nivel de tolerancia de 0,35. Una vez identificados los ítems comunes se obtienen las medianas de la severidad de los ítems estimados y se les escala con la siguiente transformación lineal:

$$\gamma = \sigma_{\beta} / \sigma_{\beta}; \alpha = \mu_{\beta} - \mu_{\beta} \times \gamma$$

$$\check{\beta} = \alpha + \check{\beta} \times \gamma$$

Donde σ es la desviación estándar, μ es la media, $\check{\beta}$ es la mediana de los ítems que se establecieron como comunes, $\check{\beta}$ es el estimador de la severidad del ítem escalado por la desviación estándar de todos los $\check{\beta}$ identificados como comunes. Nótese que $\check{\beta}$ se transforma con los parámetros α y γ respectivamente. De este $\check{\beta}$ se vuelve a tomar la mediana para construir un nuevo $\check{\beta}$. Finalmente para obtener el $\check{\beta}$ de la escala global, se realiza el siguiente procedimiento:

$$\check{\beta} = \frac{(\check{\beta} - \mu_{\check{\beta}})}{\sigma_{\check{\beta}} \times \sqrt{7/8}}$$

Este procedimiento se realiza para cada uno de los ítems. Los valores $\check{\beta}$ de la escala global los procesa la FIES.

Una vez que se tiene los $\check{\beta}$ s de la escala global, es necesario revisar si los ítems de la escala nacional son comparables. El análisis de comparabilidad se puede hacer de dos formas; se puede identificar *a priori* ítems que se consideran no comparables ya sea por diferencias en las formas de preguntar. Por otro lado se puede comparar la escala global con la nacional en base a una transformación lineal, sacar las diferencias de estas escalas y determinar si los ítems son diferentes en base a un nivel de tolerancia (en el caso del Ecuador este es 0,4). Una vez definidos los ítems que sean comunes, se transforma a los $\check{\beta}$ s mediante la siguiente transformación para que sean comparables con los ítems nacionales:

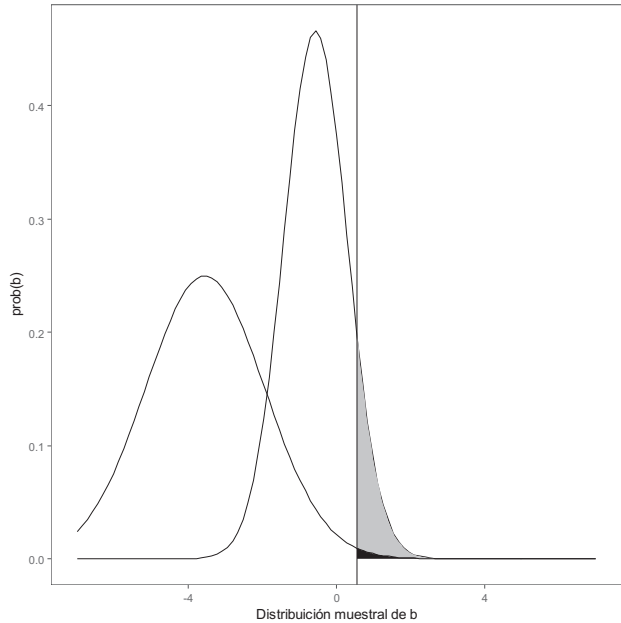
$$\check{\beta} = (\check{\beta} - \kappa) / \lambda$$

$$\lambda = \sigma_{\hat{\beta}}^{\dots} / \sigma_{\hat{\beta}}$$
$$\kappa = \mu_{\hat{\beta}} - \mu_{\hat{\beta}} \times \lambda$$

Donde $\hat{\beta}^{\dots}$ representa un ítem de la escala global definido como común, $\hat{\beta}$ es un ítem nacional definido como común; μ y σ representan al operador media y varianza respectivamente y su subíndice indica de qué variables se obtiene el operador. λ se conoce como *scale*, y κ como *shift*. $\hat{\beta}$ corresponde a los ítems de la escala global comparables con los $\hat{\beta}$ de Ecuador. De estos $\hat{\beta}$ el cuarto ítem y el octavo ítem son los umbrales que se utilizan para para el cálculo de las prevalencias.

La severidad del puntaje bruto es una estimación muestral y por lo tanto tiene asociado un nivel de incertidumbre. Según la FAO, cada $\hat{\theta}_i$ sigue una distribución normal con sus parámetros definidos según la rutina RM.w (C. Cafiero et al., 2018). Cada uno de los umbrales seleccionados de los $\hat{\beta}$ se evalúa dentro de estas distribuciones. Y se calcula la probabilidad de inseguridad alimentaria según el parámetro correspondiente. Para ilustrar, el gráfico 8 contiene las distribuciones de los puntajes brutos correspondientes al 1 y al 4; mientras que la línea vertical corresponde al umbral de inseguridad alimentaria de moderada a severa.

Gráfico 8: Distribuciones de los scores brutos 1 y 4, y umbral de inseguridad alimentaria de moderada a severa



Elaboración: autores

La sumatoria ponderada de las probabilidades de inseguridad alimentaria, según el umbral seleccionado, corresponde a la prevalencia. En el caso del presente estudio las prevalencias de inseguridad alimentaria (I.A.) de moderada a severa y severa se calculan así:

$$Prevalencia\ de\ I.A.\ Moderada\ a\ severa = \frac{\sum_{i=1}^n (prob(I.A.\ Moderada\ a\ severa)_i * w_i)}{\sum_{i=1}^n (w_i)}$$
$$Prevalencia\ de\ I.A.\ Severa = \frac{\sum_{i=1}^n (prob(I.A.\ Severa)_i * w_i)}{\sum_{i=1}^n (w_i)}$$

La rutina del paquete *RM.weights* que calcula las prevalencias, es la rutina *equating.fun*.

7. Resultados en Ecuador

a. Proporciones de las respuestas

Los ítems o preguntas de la FIES fueron formulados con la finalidad de que se pueda determinar la posición de los individuos dentro de una escala de inseguridad alimentaria, a través de resultados obtenidos de las respuestas que los individuos dan, basados en sus experiencias a lo largo del último año.

Los patrones de respuesta de la aplicación de la FIES se resumen en la Tabla 2.

Tabla 2: Proporciones repuestas FIES 2017

Pregunta	Orden	Proporción ponderada
¿Usted u otra persona adulta se preocupó por no tener suficientes alimentos?	1	33,8%
¿Usted u otra persona adulta dejó de comer alimentos saludables y nutritivos?	2	20,0%
¿Usted u otra persona adulta tuvo una alimentación basada en poca variedad de alimentos?	3	20,4%
¿Usted u otra persona adulta en su hogar dejó de desayunar, almorzar o cenar?	4	11,2%
¿Usted u otra persona adulta comió menos de lo que pensaba que debía comer?	5	15,2%
¿Se quedaron sin alimentos?	6	10,6%
¿Usted u otra persona adulta sintió hambre pero no comió?	7	10,0%
¿Usted u otra persona adulta comió una sola vez al día?	8	9,0%
¿Usted u otra persona adulta dejó de comer durante todo un día?	9	4,0%

Fuente: ENEMDU diciembre-2017
Elaboración: Autores

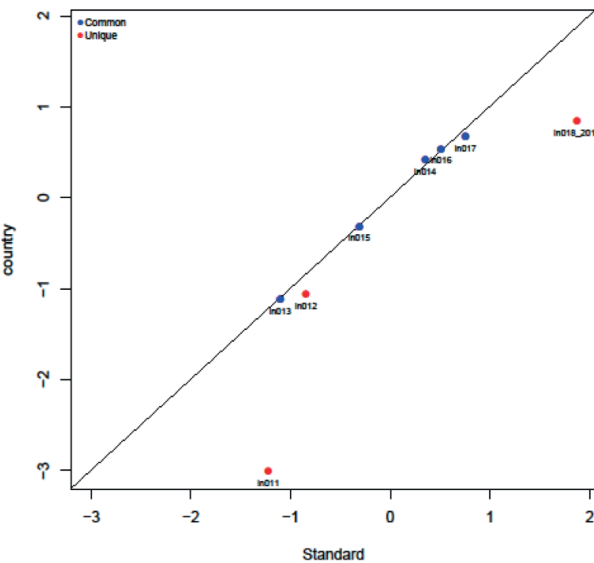
a. Resultados de la escala

El cálculo de las prevalencias requiere de aplicar una rutina que compare la escala FIES global con la del Ecuador, como se explicó en el literal e) de la sección anterior. En el caso del ejercicio del 2017; se determinó que las siguientes preguntas pueden ser no comparables con la FIES: ¿Usted u otra persona adulta se preocupó por no tener suficientes alimentos?; ¿Usted u otra persona adulta dejó de comer alimentos saludables y nutritivos? y ¿Usted o algún adulto solo comió una vez al día o dejó de comer

durante todo un día? Unida con la pregunta ¿Usted u otra persona adulta dejó de comer durante todo un día? Estas preguntas entrarían como *priors* en la rutina de *equating* del paquete *RM.weights*. Los *priors* son aquellos ítems que se definen, a priori, como diferentes de la FIES global FAO. La rutina calculará la prevalencia para estos *priors* extrapolando de la escala global, con parámetros de ajuste obtenidos de la escala nacional.

La primera, tercera y cuarta de estas preguntas son muy diferentes a sus equivalentes de la FIES, ya que están fuera del rango que permite la función *equating*. En el caso de la segunda pregunta, se incluyó por sugerencia de FAO; ya que hay diferencias en la redacción. Sin embargo, este punto está claramente cerca de la línea de igualdad. El gráfico a continuación resumo la severidad de los ítems comparados con la escala global, tomando en cuenta un umbral de 0,4.

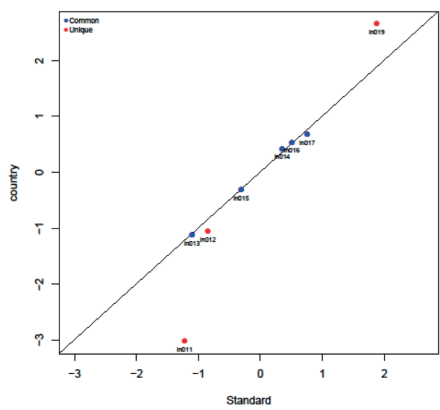
Gráfico 9: *equating plot* de la escala FIES/ELCSA vs. La escala nacional 2017



Fuente: ENEMDU 2017
Elaboración: Autores
Nota: En el gráfico el prefijo in01 indica la codificación de la pregunta, referente a inseguridad alimentaria. Los números restantes corresponden al número de la pregunta. La pregunta 8 tiene un sufijo, e indica que se unen las respuestas de la pregunta 8 y 9.

El ejercicio anterior corresponde a utilizar la FIES como una escala similar a la ELCSA, pero si utilizamos estrictamente las preguntas correspondientes a la escala FIES el gráfico anterior cambia. Como se muestra en el gráfico 10.

Gráfico 10: *equating plot* de la escala FIES vs. La escala nacional 2017



Fuente: ENEMDU 2017
Elaboración: Autores
Nota: En el gráfico el prefijo in01 indica la codificación de la pregunta, referente a inseguridad alimentaria. Los números restantes corresponden al número de la pregunta. La pregunta 8 tiene un sufijo, e indica que se unen las respuestas de la pregunta 8 y 9.

Tabla 3: Prevalencias de Inseguridad alimentaria severa y moderada 2017

Nivel de IA	Año	P.1	P.2	P.8	Versión 1 (pregunta 8 y 9)	Versión 2 (pregunta 9)	Orden	Escenario	Prevalencia
moderada	2017	si	no	no	si	no	FIES	1	14,94%
severa	2017	si	no	no	si	no	FIES	1	5,40%
moderada	2017	si	no	no	no	si	FIES	1	14,76%
severa	2017	si	no	no	no	si	FIES	1	3,34%
Moderada	2017	si	si	si	si	no	FIES	2	14,69%
severa	2017	si	si	si	si	no	FIES	2	3,83%
moderada	2017	si	si	si	no	si	FIES	2	14,81%
severa	2017	si	si	si	no	si	FIES	2	4,35%
moderada	2017	si	no	si	si	no	FIES	3	14,65%
severa	2017	si	no	si	si	no	FIES	3	3,77%
moderada	2017	si	no	si	no	si	FIES	3	14,77%
severa	2017	si	no	si	no	si	FIES	3	4,29%
moderada	2017	no	no	si	si	no	FIES	4	16,72%
severa	2017	no	no	si	si	no	FIES	4	1,97%
moderada	2017	no	no	si	no	si	FIES	4	16,75%
severa	2017	no	no	si	no	si	FIES	4	2,43%
moderada	2017	no	no	no	si	no	FIES	5	17,83%
severa	2017	no	no	no	si	no	FIES	5	3,87%
moderada	2017	no	no	no	no	si	FIES	5	16,85%
severa	2017	no	no	no	no	si	FIES	5	2,66%

Elaboración: Autores
Nota: El sufijo P. hace referencia a los “prior” que se incluye en el comando *equating.fun*.

En definitiva la pregunta 1 y la pregunta 9 parecen alejarse del umbral. Sin embargo la pregunta 9 está más cerca que la combinación de la pregunta 8 y 9, como se muestra en el gráfico anterior. La pregunta 2, no modifica significativamente las prevalencias por lo que se puede prescindir de su inclusión como prior; situación que se ejemplifica en la siguiente tabla.

La razón de mostrar resultados en distintos escenarios, responde a determinar el efecto en las prevalencias de definir distintos *priors*. La escala planteada en la ENEMDU 2017, es muy similar a la FIES y se esperaría poca diferencia entre el estándar global, y la escala en Ecuador. Sin embargo, en el gráfico 10 se observa que las preguntas 1 y 8, están alejadas de la línea de la igualdad. Esto da luces de que hay ciertas categorías de la escala ecuatoriana que no son comparables con la FIES. Al analizar los resultados se verifica que los escenarios del 1 al 3 son muy similares, sin importar la versión. Los últimos dos escenarios, varían un poco siendo el quinto escenario el que más cambio en las prevalencias presenta; sin importar la versión. El último escenario es el menos factible, ya que asume no hay mayor diferencia entre los ítems de la FIES global FAO y los ítems de la escala nacional, a pesar de que el gráfico 10 indica todo lo contrario.

8. Conclusiones

La escala FIES, constituye un mecanismo eficiente, en términos de recursos, para la medición de la dimensión de acceso de la seguridad alimentaria; aunque, hacen falta más ejercicios para determinar una línea base oficial para el Ecuador.

En base a los supuestos de Rasch y los diferentes escenarios de “*equating*” la prevalencia de inseguridad alimentaria moderada a severa en el Ecuador en 2017 puede ir a un rango de 14,65% a 17,83%; mientras que la prevalencia de la inseguridad severa es de 1,97% a 5,4%. Esto evidencia la sensibilidad de la metodología a los criterios selección para la comparación con la FIES global FAO.

La propiedad de la invariancia permite calcular las prevalencias para distintas desagregaciones de la población. Las desagregaciones de los resultados 2017 en grupos poblacionales parecen indicar que la inseguridad alimentaria se acopla a la realidad del país. Los pobres por ingreso y por necesidades básicas insatisfechas-NBI tienen una prevalencia mayor de inseguridad alimentaria que los no pobres. Así mismo las prevalencias de inseguridad alimentaria son mayores en el sector rural, área que es por lo general más vulnerable que la urbana. Al dividir por sexo, no existen diferencias significativas, lo cual merece futura discusión dado que la pregunta es dirigida a jefes de hogar.

Al desagregar por regiones naturales, la Sierra presenta valores de inseguridad alimentaria significativamente más bajos que la región Costa. Esto llama la atención dado que los indicadores de antropometría indican, tradicionalmente, valores menores para la región Costa respecto a la Sierra. Según la última Encuesta de Condiciones de Vida para Ecuador, correspondiente al año 2014, la desnutrición crónica para la región Costa alcanza 18,49% mientras que esta cifra para la Sierra es 29,56%¹⁴. Sin embargo, al ser un ejercicio puntual y con fines metodológicos, solamente representa un ejemplo de los posibles análisis entre las dimensiones de acceso y utilización de la seguridad alimentaria.

9. Limitaciones y recomendaciones

El incumplimiento de algún supuesto de Rasch, unidimensionalidad, monotonicidad o independencia local, resultaría en la incorrecta estimación de los parámetros del modelo. Las estimaciones de las pruebas sugeridas por FAO establecen que dichos supuestos se cumplen. Sin embargo, las pruebas utilizadas para comprobar los supuestos presentan limitaciones metodológicas, por lo que se recomienda utilizar otro grupo de pruebas o implementar modelos para más de un parámetro.

El presente documento no evalúa la invariancia de parámetros. Dependiendo de las características locales, las estimaciones podrían no ser comparables. Esto quiere decir que los resultados pueden no ser comparables para grupos poblacionales del Ecuador, para distintos años en un mismo país, e incluso entre países. Las pruebas específicas requieren de mayor revisión pero como un primer insumo se recomienda construir modelos de regresión de *pooled data*; para determinar si existen características particulares de la población que pongan en duda la invariancia de los parámetros.

Finalmente hay ciertos aspectos que en la aplicación modelo de Rasch potencialmente incidirían en las prevalencias: Nord (2014) asume que la probabilidad de los puntajes brutos con puntaje perfecto deben tener probabilidad de 1. Además, en la metodología para cálculo de prevalencias se asigna probabilidades de acuerdo a los estándares internacionales propuestos por FAO (cuya selección son de cierta forma arbitrarios).

14 Los cálculos se realizaron por los autores, con las bases de la ECV tomadas de: http://www.ecuadorencifras.gob.ec/documentos/web-inec/ECV/ECV_2015/

En adición, no se incluye el efecto de diseño de la muestra en las prevalencias, y como se indicó en secciones anteriores, puede tener un efecto importante en las prevalencias calculadas.

10. Referencias

Ayala, R. J. de. (2009). *The Theory and Practice of Item Response Theory*. Nueva York.

Ballard, T. J., Kepple, A. W., & Cafiero, C. (2013). *The Food Insecurity Experience Scale Development of a Global Standard for Monitoring Hunger Worldwide*.

Cafiero, A. C., Nord, M., & Viviani, S. (2016). Package RM.weights. CRAN, FAO. Recuperado a partir de <https://cran.r-project.org/web/packages/RM.weights/index.html>

Cafiero, C., Nord, M., Viviani, S., Del Grossi, M. E., Ballard, T. J., Kepple, A. W., ... Nwosu, C. (2016). *Voices of the Hungry Technical Report* (Vol. 2016).

Cafiero, C., Viviani, S., & Nord, M. (2018). Food security measurement in a global context: The food insecurity experience scale. *Measurement: Journal of the International Measurement Confederation*. <http://doi.org/10.1016/j.measurement.2017.10.065>

CELAC. (2015). Plan Para La Seguridad Y Erradicación Del Hambre De La Celac 2025. *Plan para la Seguridad Alimentaria, nutrición y erradicación del hambre de la CELAC 2025*, 10.

CEPAR. (2004). Encuesta Demográfica y de Salud Materna e Infantil.

Christensen, K. B., Kreiner, S., & Mesbah, M. (2012). Rasch Models in Health. (K. B. Christensen, S. Kreiner, & M. Mesbah, Eds.) *Rasch Models in Health* (1st ed.). Hoboken, NJ USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/9781118574454>

Comité Científico de la ELCSA. (2012). *Escala Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA)*. Roma.

CONADE. (1988). DIAGNOSTICO DE LA SITUACION ALIMENTARIA NUTRICIONAL Y DE POBLACION ECUATORIANA DANS 1988.pdf. Quito-Ecuador: CONADE y MSP.

DeMars, C. (2010). *Item Response Theory* (1st ed.). New York: Oxford university press.

FAO. (1975). The state of food and agriculture 1974. Recuperado a partir de <http://www.fao.org/docrep/017/f3350e/f3350e.pdf>

FAO. (1999). La Cumbre Mundial sobre la Alimentación.

FAO. (2006). Food security. *Policy Brief*, (2), 1-4. <http://doi.org/10.1016/j.jneb.2010.12.007>

FAO. (2011). Una introducción a los conceptos básicos de la seguridad alimentaria. *La Seguridad Alimentaria: información para la toma de desiciones*, 1-4.

FAO. (2016a). *Métodos para la estimación de índices comparables de prevalencia de la inseguridad alimentaria experimentada por adultos en todo el mundo Reporte Técnico* (Vol. 2016). Recuperado a partir de <http://www.fao.org/3/b-i4830s.pdf>

FAO. (2016b). Monitoring Food Security and Nutrition in Support of the 2030 Agenda for Sustainable Development. *FAO, Rome*. Recuperado a partir de <http://www.fao.org/3/a-i6188e.pdf>

FAO; FIDA & PMA. (2013). *El estado de la inseguridad alimentaria en el mundo 2013*, 1-63.

FAO, FIDA, & WFP. (2015). *El estado de la inseguridad alimentaria en el mundo*. Recuperado a partir de <http://www.fao.org/3/a-i4646s.pdf>

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments, and applications*. (G. H. Fischer & I. W. Molenaar, Eds.) (1st ed.). New York: Springer.Vedag. <http://doi.org/10.1007/978-1-4612-4230-7>

Freire, W., Belmont, P., Ramírez, M., Mendieta, M., Silva, K., Romero, N., ... Monge, R. (2014). *Encuesta Nacional de Salud y Nutrición 2012. Ensanut-Ecu 2012* (Vol. Tomo 1). <http://doi.org/044669>

Hahs-Vaughn, D. L. (2005). A primer for using and understanding weights with national datasets. *Journal of Experimental Education*. <http://doi.org/10.3200/JEXE.73.3.221-248>

Hambleton, R. K., Robin, F., & Xing, D. (2000). Test Data. *Analysis*.

Hardouin, J. B. (2007). Rasch analysis: Estimation and tests with raschtest. *Stata Journal*, 7(1), 22-44. <http://doi.org/The Stata Journal>

INEC. (2014). Compendio de Resultados Encuesta de Vida ECV, sexta ronda 2015. *Inec*, 197. <http://doi.org/10.1007/s13398-014-0173-7.2>

INEC. (2017). Plan de Desarrollo Estadístico para el reporte de los indicadores de los Objetivos de Desarrollo Sostenible, Tomo I: Diagnóstico de la capacidad estadística del Ecuador y estrategias de fortalecimiento estadístico. Quito-Ecuador.

Jácome, F., & Restrepo, M. C. (2017). Validación del uso de la escala CESD aplicada en la Encuesta de Condiciones de Vida 2014. *Revista de Estadística y Metodologías*, 3, 7-28. Recuperado a partir de http://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Revista_Estadistica/Revista_de_Estadistica_y_Metodologias_3.pdf

Jones, A. D., Ngure, F. M., Pelto, G., & Young, S. L. (2013). What Are We Assessing When We Measure Food Security? A Compendium and Review of Current Metrics. *Adv. Nutr*, 4, 481-505. <http://doi.org/10.3945/an.113.004119>

MCDS/FAO.(2010).SeguridadAlimentariayNutricional en El Ecuador. *Organización de las Naciones Unidas para la Alimentación y la Agricultura*, 335. Recuperado a partir de <http://es.scribd.com/doc/100616097/Seguridad-Alimentaria-y-Nutricional-en-El-Ecuador>

MCP. (2011). Código Orgánico de Organización Territorial, Autonomía y Descentralización. *Territorial Organization, Autonomy and Decentralization Code, (COOTAD)*, 81.

MIES. (2017). Ministerio de Inclusión Económica y Social.

Napoli, M. (2010). Towards a Food Insecurity Multidimensional Index (FIMI). Recuperado a partir de <http://www.fao.org/fileadmin/templates/ERP/uni/FIMI.pdf>

Nering, M. L., & Ostini, R. (2010). Handbook of Polytomous Item Response Theory Models. *Psychometrika*. <http://doi.org/10.1007/BF02294473>

Nord, M. (2014). INTRODUCTION TO ITEM RESPONSE THEORY Basic Concepts , Parameters and Statistics. *FAO Report*, 1-20. Recuperado a partir de <http://www.fao.org/economic/ess/ess-fs/voices/en>

ONU. (2015a). Objetivos de Desarrollo del Milenio Informe de 2015. *Naciones Unidas*, 72. <http://doi.org/10.1108/17427370810932141>

ONU. (2015b). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible. Washington: ONU.

Sen, A. (1981). *Poverty and Famines: An Essay on Entitlement and Deprivation*. Recuperado a partir de http://staging.ilo.org/public/libdoc/ilo/1981/81B09_608_engl.pdf

Sen, A. (2000). Desarrollo Y Libertad Amartya_Sen. Pdf.

SENPLADES. (2015). Objetivos del Milenio: Balance Ecuador 2014, 32.

SIISE. (2010a). Programa Aliméntate Ecuador.

SIISE. (2010b). Programa de Alimentación Escolar.

SIISE. (2010c). Programa de Complementación Alimentaria.

UN. (1975). Report of the world food conference (Rome, 5-16 Noviembre 1974). Recuperado a partir de http://ernaehrungsdenkwerkstatt.de/fileadmin/user_upload/EDWText/Abbildungen/Vorlesung_OLT/Surveillance_Nutrition_Resolution_1974_ICN_Infos.pdf

Vercueil, J., Naiken, L., Burlingame, B., Kinlay, D., Mernies, J., Raney, T., ... Willett, W. (2002). Resumen de los debates respecto al Simposio Científico Internacional sobre Medición y Evaluación de la Carencia de Alimentos y de la Desnutrición.

Vos, R. (2016). First things first : food to live well : A new method to estimate undernourishment and food insecurity, (August), 141-151.

WFS. (1983). *Reporte de la 8va sesión del comité*.

11. Anexos

Anexo 1

Indicadores de la seguridad Alimentaria por dimensión FAO

Indicadores de la seguridad Alimentaria	DIMENSIÓN
Suficiencia del suministro medio de energía alimentaria Valor medio de la producción de alimentos	DISPONIBILIDAD
Proporción del suministro de energía alimentaria derivada de cereales, raíces y tubérculos Suministro medio de proteínas Suministro medio de proteínas de origen animal	
Porcentaje del total de carreteras que están asfaltadas Densidad de carreteras Densidad de líneas de ferrocarril	
Producto Interno Bruto (a paridad del poder adquisitivo)	ACCESO
Índice nacional de precios de los alimentos	
Prevalencia de la subalimentación Proporción del gasto de los pobres destinada a alimentos Alcance del déficit de alimentos Prevalencia de la insuficiencia de alimentos	
Coeficiente de dependencia de las importaciones de cereales Porcentaje de las tierras cultivables equipadas para el riego Valor de las importaciones de alimentos respecto de las exportaciones totales de mercancías	
Estabilidad política y ausencia de violencia o terrorismo Volatilidad de los precios nacionales de los alimentos Variabilidad de la producción de alimentos per cápita Variabilidad del suministro de alimentos per cápita	ESTABILIDAD
Acceso a fuentes de agua mejoradas Acceso a servicios de saneamiento mejorados	UTILIZACIÓN
Porcentaje de niños menores de cinco años que padecen emaciación Porcentaje de niños menores de cinco años que padecen retraso de crecimiento Porcentaje de niños menores de cinco años que padecen insuficiencia ponderal Porcentaje de adultos que padecen insuficiencia ponderal Prevalencia de la anemia entre las mujeres embarazadas Prevalencia de la anemia entre los niños menores de cinco años Prevalencia de la carencia de vitamina A en la población Prevalencia de la carencia de yodo en la población	

Fuente: FAO (2016)
Elaboración: Autores

Anexo 2

MEDICIÓN DE LAS DIMENSIONES PARA ECUADOR

DIMENSIÓN	EJEMPLOS DE INDICADORES
DISPONIBILIDAD	- Acceso a riego, distribución de la tierra, superficie y productividad de cultivos, uso de semillas y fertilizantes (ESPAC – INEC) - Concentración de la tierra (ESPAC – INEC, Censo Nacional Agropecuario – INEC SICA MAG) - Producción de alimentos, disponibilidad de energía, grasa o proteína, grado de suficiencia energética o proteica, dependencia de importaciones alimentos (Hojas de balance – FAOSTAT) - Tenencia de la tierra (Censo Nacional Agropecuario – INEC SICA MAG)
ACCESO	- Ingreso laboral medio, ingreso per cápita y miembros del hogar por deciles (ENEMDU - INEC) - Participación del gasto en alimentos sobre el gasto total (ECV - INEC) - Nivel educativo de la población, PEA, tasas de desempleo, pobreza y pobreza extrema por ingresos, por consumo y por NBI (ENEMDU, ECV o CPV – INEC) - Costo y brecha de la canasta básica, IPC (Series históricas - INEC)
UTILIZACIÓN	- Esperanza de vida al nacer (Estimaciones y proyecciones de población - INEC) - Prevalencia de enfermedades: cólera, diabetes, diarrea, hipertensión arterial (Estadísticas epidemiológicas - MSP) - Tasa de mortalidad (infantil, materna y niñez), principales causas muerte (Estadísticas vitales - INEC) - Agua segura, acceso a eliminación de excretas, red de alcantarillado (ENEMDU, ECV o CPV - INEC) - Desnutrición (aguda, crónica y global), sobrepeso, lactancia materna (ENDEMAIN - CEPAR, ECV- INEC) - Consumo medio per cápita de energía, índice de adecuación energética, autoconsumo (ECV - INEC)
ESTABILIDAD	- Variación anual de la productividad de cultivos (ESPAC – INEC) - Índice de disponibilidad de energía, grasa o proteína per cápita, índice de producción per cápita de alimentos, índice global de producción (Hojas de balance – FAOSTAT) - Índice de Gini, Perceptores de ingresos del hogar por deciles de ingreso (ENEMDU – INEC)

Fuente: SIISE-SISSAN
Elaboración: Autores

Anexo 3

Comparación de preguntas ENEMDU 2017 y FIES

Instrucción	ENEMDU-2017-14A	FIES Global-Hogares
	En los últimos 12 meses en su hogar, alguna vez por falta de dinero u otros recursos:	
Preguntas	¿Usted u otra persona adulta se preocupó por no tener suficientes alimentos?	¿Usted u otra persona en su hogar se haya preocupado por no tener suficientes alimentos para comer por falta de dinero u otros recursos?
	¿Usted u otra persona adulta dejó de comer alimentos saludables y nutritivos?	Pensando aún en los últimos 12 meses, ¿hubo alguna vez en que usted u otra persona en su hogar no haya podido comer alimentos saludables y nutritivos por falta de dinero u otros recursos?
	¿Usted u otra persona adulta tuvo una alimentación basada en poca variedad de alimentos?	¿Hubo alguna vez en que usted u otra persona en su hogar haya comido poca variedad de alimentos por falta de dinero u otros recursos?
	¿Usted u otra persona adulta en su hogar dejó de desayunar, almorzar o cenar?	¿Hubo alguna vez en que usted u otra persona en su hogar haya tenido que dejar de desayunar, almorzar o cenar porque no había suficiente dinero u otros recursos para obtener alimentos?
	¿Usted u otra persona adulta comió menos de lo que pensaba que debía comer?	Pensando aún en los últimos 12 meses, ¿hubo alguna vez en que usted u otra persona en su hogar haya comido menos de lo que pensaba que debía comer por falta de dinero u otros recursos?
	¿Se quedaron sin alimentos?	¿Hubo alguna vez en que su hogar se haya quedado sin alimentos por falta de dinero u otros recursos?
	¿Usted u otra persona adulta sintió hambre pero no comió?	¿Hubo alguna vez en que usted u otra persona en su hogar haya sentido hambre pero no comió porque no había suficiente dinero u otros recursos para obtener alimentos?
	¿Usted u otra persona adulta comió una sola vez al día?	¿Hubo alguna vez en que usted u otra persona en su hogar haya dejado de comer todo un día por falta de dinero u otros recursos?
	¿Usted u otra persona adulta dejó de comer durante todo un día?	

Fuentes: INEC y FAO
Elaboración: autores

2

Elaboración de estadísticas de vacantes publicadas en internet

Una experiencia en Ecuador

Diego Benítez⁺

Sebastián Lucero^{*}

Ana M. Pazmiño^{*}

RESUMEN

La presente Nota Metodológica documenta un ejercicio de elaboración de estadísticas de vacantes, a partir de los anuncios de empleos que se publican en el Internet, usando una herramienta de web scraping. Se describen los aspectos más relevantes del proceso; y se muestra, a manera de ejemplo, los principales resultados correspondientes a los meses de septiembre y octubre de 2016.

Palabras clave: webscraping, intenet, vacantes

Descargo de responsabilidad: Las opiniones e interpretaciones expresadas en este documento pertenecen a los autores y no reflejan el punto de vista oficial del Instituto Nacional de Estadística y Censos (INEC). El INEC no garantiza la exactitud de los datos que figuran en el documento.

.....
⁺ Consultores del Banco Mundial. Los autores agradecen el apoyo financiero del *Jobs Umbrella Trust Fund* para la realización de este documento así como también para muchas otras actividades orientadas al fortalecimiento de la producción de estadísticas laborales en Ecuador. Lo expresado en este documento no representa la posición del INEC, del gobierno del Ecuador, ni del Grupo Banco Mundial. Cualquier error es de los autores. Correspondencia a través de dieben79@yahoo.com. Se agradece también la orientación proporcionada por Jeisson Cárdenas para la realización del ejercicio. Con la finalidad de contribuir al debate técnico respecto al mercado laboral, orientado a la formulación de mejores políticas públicas, el Instituto Nacional de Estadísticas y Censos (INEC) y el Banco Mundial (BM) emprendieron un proyecto conjunto denominado “Data, Diagnostics and Institutional Innovations for Jobs in Ecuador” en el que uno de sus componentes es el fortalecimiento de las estadísticas relacionadas con la temática. En este contexto, se decidió explorar la posibilidad de elaborar estadísticas de vacantes a partir de anuncios de empleos que se publican en internet, a manera complementaria de otras estadísticas que cuenta o que está desarrollando el INEC.

1. Introducción

Actualmente, millones de gigabytes de información se suben al internet diariamente. Muchos profesionales como periodistas, investigadores, analistas de datos, agentes de ventas, desarrolladores de software, entre otros, acuden a esta información, y con técnicas de “*copy and paste*” la organizan en hojas de cálculo y forman parte de sus informes y presentaciones. Sin embargo, cuando ésta es muy grande, tal es el caso si la información a consultar tiene fines estadísticos, este trabajo manual puede volverse demasiado tedioso.

Para superar este inconveniente, se desarrollaron técnicas de *web scraping* que permiten extraer información de los sitios web de forma eficiente y automática, y convertirla en formatos más estructurados y fáciles de usar (Castrillo-Fernández, 2015). Los usos del *web scraping* van desde el análisis de comportamiento hasta aplicaciones de economía política, e incluso índices de precios complementarios. Una de sus aplicaciones más destacadas es el estudio de la demanda laboral, como es el caso de Australia, donde su oficina nacional de estadística cuenta con el Índice de Vacantes de Internet (*Internet Vacancy Index*) (Reimsbasch-Kounatze, 2015), el cual muestra mensualmente la variación de vacantes publicadas en las principales bolsas de empleo *online*. A nivel regional, en Colombia, Cárdenas et al. (2015) realizaron una metodología para el análisis de demanda laboral mediante datos de Internet que, pese aún no estar institucionalizada, sentó un precedente y ha sido un elemento clave en el desarrollo del presente trabajo.

Inspirados en la experiencia colombiana, el INEC, con el apoyo del Banco Mundial, impulsó la evaluación de la posibilidad de elaborar estadísticas de vacantes como parte del sistema de información relacionado con el mercado laboral en Ecuador. El presente artículo tiene por objetivo documentar esta experiencia. Para lograr este objetivo, a manera de justificación, primero se explica qué parte de la información acerca de la demanda laboral puede cubrirse con los anuncios de empleo que se publican en internet. Segundo, se hace una breve revisión de otras experiencias en el uso del *web scraping* con fines estadísticos. Tercero, se describe el proceso técnico de extracción de la información en el caso ecuatoriano. Cuarto, se presenta los principales resultados correspondientes a los meses de septiembre y octubre de 2016. Finalmente se exponen las principales conclusiones y recomendaciones.

2. Justificación

Si bien es deseable una situación en la que tanto la oferta como la demanda en el mercado de trabajo se encuentren en un punto de equilibrio óptimo, es decir, dónde todos los empleadores tienen cubiertas sus plazas de trabajo y las personas que desean tener un empleo así lo hacen, esto no siempre sucede así y mucho menos en tiempos dónde la actividad económica en general tiende a disminuir. Uno de estos desequilibrios es el desempleo, mismo que puede ser originado por factores estructurales, friccionales o cíclicos.

Estructuralmente, el desempleo se origina por varios factores, uno de ellos es las brechas entre las habilidades que requieren las empresas y las que pueden ofrecer los trabajadores en un momento dado. Generalmente, los saltos tecnológicos hacen que las habilidades de algunos trabajadores queden obsoletas. Las medidas para corregir el desempleo estructural tardan un largo plazo en hacer efecto. Por su parte, los factores friccionales del desempleo corresponden al tiempo que tarda un trabajador en transitar de un empleo a otro; este tipo de desempleo podría reducirse tanto si empleadores como trabajadores contaran con información adecuada. Finalmente, los factores cíclicos están relacionados con la actividad económica general, y el desempleo ocurre cuando la demanda agregada se contrae a tal punto en que algunas empresas prescinden de sus trabajadores.

La intermediación laboral ayuda a resolver las brechas y distorsiones propias del desempleo friccional; entendida como tal a las instituciones y procesos que facilitan el intercambio de información entre empleadores y trabajadores con el fin de reducir el tiempo que una vacante toma en ser llenada. La proliferación del uso de tecnologías de información y comunicación han permitido la aparición de nuevos esquemas de intermediación laboral más ágiles y versátiles, como es el caso de las bolsas de empleo *online*. Kuhn (2011), provee evidencia sobre el importante rol de las herramientas *online* para emparejar trabajadores y plazas de trabajo.

Por lo expuesto, el estudio de la información que se publica en las bolsas de empleo puede proveer elementos útiles para entender una parte importante del mercado de trabajo. Esto es importante porque potencialmente permite llenar un vacío de información respecto al mercado laboral, que las fuentes tradicionales, enfocadas principalmente a la oferta laboral, no cubren.

Específicamente, en Ecuador la principal fuente de información sobre el mercado laboral es la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), a través de la cual se calculan los principales indicadores laborales vinculados con la condición de actividad como las tasas empleo adecuado, inadecuado y desempleo, entre otros (INEC, 2016). Por el lado de la demanda de trabajo pueden citarse fuentes como los Estudios Mensuales de Opinión Empresarial (EMOE) del Banco Central del Ecuador (BCE), o las Encuestas Industriales y el Laboratorio de Dinámica Laboral y Empresarial (LDLE)¹ del INEC (Benítez, Espinoza, Grijalva, Rivadeneira y Oviedo (2016).²

La información de los anuncios de los empleos que se publican en internet permitiría estudiar una parte del mercado laboral que aún no ha sido explorada por las otras fuentes. Con información sistematizada sobre las ofertas de plazas de trabajo se podría conocer el perfil de profesionales que las empresas requieren, incluyendo detalles como: habilidades, localización de la vacante, salario, experiencia mínima, tipo de contrato, entre otros. Al utilizar como técnica el *web scraping*, la información que se obtiene es oportuna, y su costo de levantamiento es relativamente bajo. En este sentido, Carnevale, et al. (2014) señalan como ventajas el bajo costo para su recolección, el permitir contar con el detalle de ocupaciones, industria y ubicación, el mostrar la demanda de habilidades, el posibilitar realizar monitoreos y notar tendencias, y el ser útil para mejorar la política educativa.

Indudablemente, existen también ciertos reparos a la obtención de estadísticas por esta vía. Muchos de estos problemas se relacionan con la calidad, confiabilidad y representatividad de la información obtenida (Stefanik, 2012). Estas preocupaciones son especialmente pertinentes, dado que las principales herramientas cuantitativas del análisis económico están basadas en el uso de la estadística inferencial. Sin embargo, en un horizonte lejano, hay que considerar que en la medida en que la población va digitalizándose por completo, las técnicas de muestreo pueden incluso llegar a ser obsoletas (Askita & Zimmermann, 2015).

1 El LDLE es un sistema de registros administrativos con fines estadísticos. Esto implica que se nutre de cierta información de empresas que reportan información a administraciones públicas como la oficina de impuestos o el instituto de seguridad social, SRI o IESS en el caso particular del Ecuador.

2 Otras fuentes se están desarrollando en el marco de la cooperación INEC-BM, como son los casos del módulo de demanda laboral de corto plazo anclado al Sistema de Indicadores de la Producción (SIPRO), o el módulo de demanda laboral estructural anclado a las encuestas empresariales; pero ambas se encuentran aún en fase de desarrollo.

3. Otros usos de web scraping con fines académicos

Pese a los problemas que presenta, el *web scraping* se viene usando de manera creciente en diferentes disciplinas. En este sentido, la Economía no constituye una excepción. Según Edelman (2012) la toma sistemática de información de internet ha sido utilizada en varios estudios acerca de: historia del pensamiento económico (Azar, 2007), microeconomía (Bajari & Hortacsu, 2003), finanzas (Antweiler & Murray, 2004), desarrollo económico (Seamans & Zhu, 2010), entre otros.

Azar (2007), examina las tendencias en los tiempos de respuesta de las revistas científicas a los borradores de artículos que reciben, recogiendo información de tiempos de respuesta de diferentes páginas web. Bajari & Hortacsu (2003), toman información de las ventas de monedas coleccionables en eBay para analizar el comportamiento de los postores en las subastas. Antweiler & Murray (2004), analizan las discusiones de los foros *online* de los mercados de valores y encuentran que las discusiones ayudan a predecir precios. Finalmente, Seamans & Zhu (2010) exploran la forma en que los periódicos locales adaptan sus tarifas de suscripción, de publicidad y de anuncios clasificados frente a la entrada de páginas como *Craigslist*³.

Uno de los usos más destacados de la construcción de estadísticas a partir de Internet tiene que ver con la estimación de índices de precios. En la medida en que cada vez más productos son vendidos *online*, el Internet se convirtió en una fuente rica para obtener información de precios. Esto lo supo ver bien Alberto Cavallo que en 2007 presentó su primer desarrollo utilizando precios en línea de Argentina, Chile, Brazil, y Colombia (Cavallo, 2009). Sus exitosos resultados le permitieron expandir su colección de información a más de 50 países, dando inicio a la iniciativa académica Billion Price Project (BPP) que actualmente es manejada por PriceStats. En 2012, finalmente logró demostrar que las cifras oficiales de Argentina estaban siendo manipuladas por lo que su sitio online www.inflacionverdadera.com pasó a ser la fuente más creíble de inflación (Cavallo, 2012).

3 Craigslist es un sitio web de anuncios clasificados con secciones dedicadas al empleo, vivienda, contactos personales, ventas, ítems, servicios, comunidad, conciertos, hojas de vida, y foros de discusión.

La utilización de información tomada a partir de Internet en la investigación y análisis del mercado laboral difiere según los objetivos y enfoques perseguidos. Reimsbasch-Kounatze (2015) expone un par de casos significativos: por ejemplo, en Estados Unidos, la actividad de creación de plazas de trabajo es monitoreada a nivel nacional, regional, estatal y metropolitano a través de los anuncios de trabajo condensados en la forma de un índice compuesto denominado *Help-Wanted Index*.⁴

Otro caso importante está en Australia, en dónde se da seguimiento a la variación mensual del total de anuncios de las tres bolsas de empleo más grandes a través del *Internet Vacancy Index* (IVI). El IVI incluye información de al menos 350 tipos de ocupaciones para todas sus 37 regiones. En la Unión Europea (UE), una fuente potencial para información de vacantes online es el sitio web EURES (European Employment Services), el cual almacena las vacantes laborales de todos los países de la UE en una plataforma estandarizada (Kurekova, et al., 2015). A nivel regional, Colombia es pionero en el uso de vacantes *online* para estimar la demanda laboral. Cárdenas, et al. (2015) presentan una metodología para analizar la demanda laboral mediante la extracción de información *online* de plazas de trabajo vacantes.

4. La experiencia en Ecuador

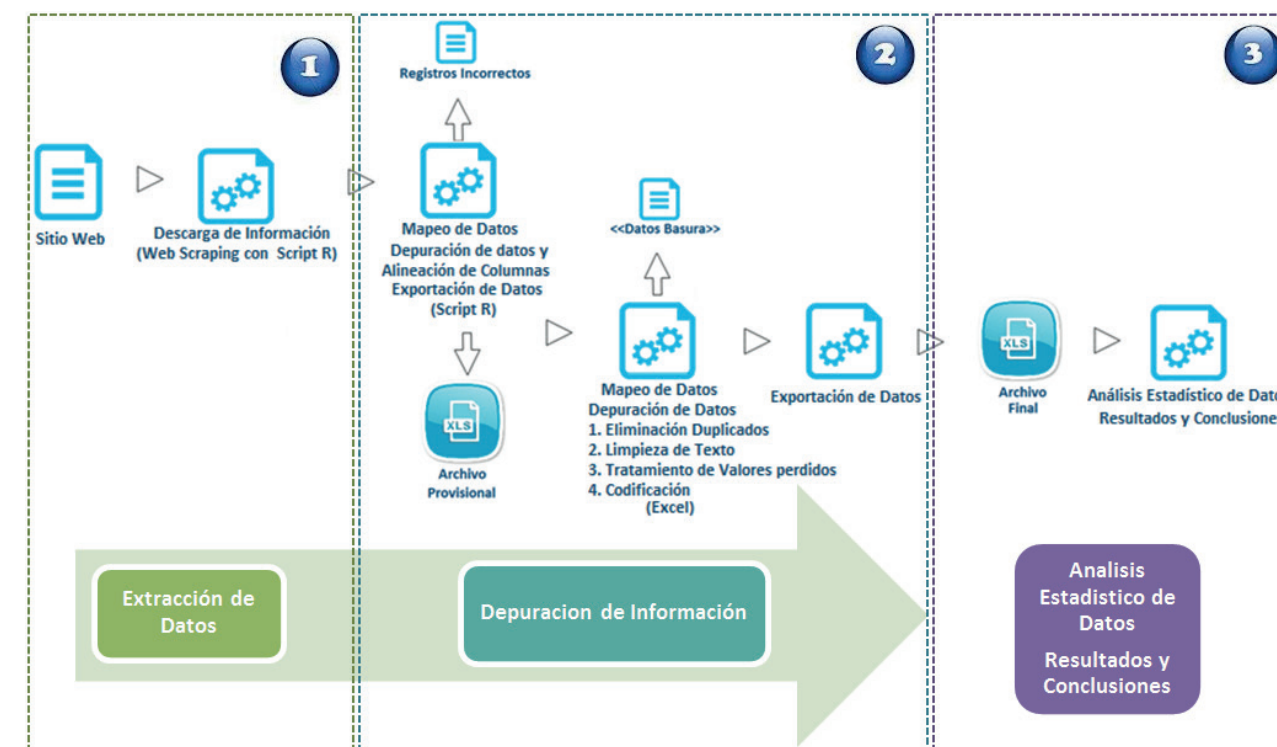
Específicamente, la experiencia que aquí se presenta consistió en elaborar estadísticas de las vacantes que se publican en internet. En Ecuador existen distintas bolsas de empleo online entre las que destacan una pública (Red Socioempleo⁵), y algunas privadas como: CompuTrabajo⁶, Multitrabajos⁷, Porfinempleo⁸, entre otras. De ellas, para este ejercicio se eligió CompuTrabajo, por ser considerada una de las más utilizadas en el medio. En Ecuador, esta bolsa publica mensualmente más de 2 000 anuncios que incluyen campos de información como el nombre de la empresa, nombre del cargo, tipo de contrato, salario, educación mínima, entre otros elementos.

Es probable que un estudio que pretenda abarcar el universo de los anuncios de empleo, requiera incorporar varias de las bolsas disponibles. Sin embargo, resulta difícil armar una base de datos con el total de anuncios, dado que, no se conoce si los anuncios publicados en las distintas páginas web se complementan o se superponen, y en qué medida lo hacen, lo cual constituye un impedimento para combinar la información de dos o más bolsas distintas.

Una solución a este problema podría ser identificar si dos o más anuncios que se han publicado en distintos medios corresponden realmente a la misma vacante. Esto podría hacerse mediante la comparación de los campos de los anuncios, siempre que puedan establecerse patrones que permitan determinar la similitud y sugerir la superposición del dato. Sin embargo, esto no necesariamente es fácil de implementar, para conocer su factibilidad se requiere de un desarrollo tecnológico adicional. Una solución complementaria o alternativa consiste en utilizar encuestas a las empresas para conocer el contexto en el cual ellas difunden sus vacantes disponibles, con lo cual podría saberse, por ejemplo, el porcentaje de empresas que usan un medio u otro. Esta segunda opción debe considerar el costo inherente a un levantamiento de campo, por lo que, para ahorrar costos, podría aprovecharse alguna encuesta en marcha para incluir un conjunto de preguntas en este sentido.

Bajo estas consideraciones, la experiencia que aquí se reporta se centró en elaborar estadísticas de plazas de trabajo publicadas en la página web de CompuTrabajo. El proceso consiste en: i) utilizar herramientas de *web scraping* para descargar periódicamente la información de anuncios de empleo que aquí se publican, ii) depurar esta información para almacenarla en una base de datos estructurada, y finalmente, iii) obtener estadísticas a partir de la misma. El gráfico 1 describe este proceso.

Gráfico 1. Proceso de construcción de la base de información



4.1. Descarga de la información

En el proceso de captura y almacenamiento de la información, *web scraping*, se utilizan algoritmos de interpretación que permiten modelar la estructura de información de las páginas web para poder transformar datos sin estructura en bases de datos estructuradas. Este proceso se puede explicar más fácilmente de la siguiente manera: piénsese en una página web como un conjunto de campos de información ordenados donde cada campo de información puede ser identificado en el código fuente de la página mediante la interpretación del lenguaje HTML, técnicamente, este procedimiento se lo denomina análisis del esquema HTML. En este sentido, conociendo como se presenta la información en la web se puede extraer información de cada campo de información. Este proceso se programa en algoritmos para que sea realizado de manera repetitiva y estandarizada. Sin embargo, se debe tener presente que cualquier cambio en la forma de presentar la información, modifica el esquema HTML de la página y perjudica los códigos de programación *web scraping* y, en consecuencia, se requiere un esfuerzo de supervisión permanente.

La información extraída, es almacenada en una matriz donde las filas corresponden a los anuncios y las columnas a sus características asociadas (campos de información). Dado que los campos de información de CompuTrabajo son semiestructurados, es decir, no guardan siempre el mismo ordenamiento de la información para todos los anuncios, se identificó, en la base descargada, la existencia de atributos que no corresponden con la variable que debería. Para alinear correctamente todos los campos, se cuenta con un algoritmo capaz de corregir estos problemas, sin embargo, algunas observaciones por sus particularidades deben ser alineadas manualmente. Finalmente, posterior a la etapa de depuración, la base de datos se exporta a una hoja de cálculo.

La propuesta busca generar información mensual de vacantes. Para esto, se estableció que la extracción de información se realice cada 10 días con el fin de garantizar que todos los anuncios publicados en el mes de referencia han sido capturados adecuadamente. (Al momento, se cuenta con bases de datos para los meses de septiembre y octubre 2016).

⁴ El índice es elaborado por el *Conference Board*, una organización no gubernamental y *think tank* privado.

⁵ <http://www.socioempleo.gob.ec/>

⁶ <http://www.computrabajo.com.ec/>

⁷ <http://www.multitrabajos.com/>

⁸ <http://www.porfinempleo.com/>

4.2. Depuración de la data

En la depuración de la información extraída, se siguen una serie de pasos fundamentales previo a la obtención de la base de datos a ser analizada. Esto pasos tienen que ver con: 1) Eliminación de registros duplicados, 2) Limpieza de caracteres basura, 3) Tratamiento de valores *missing*, 4) Codificación de variables. A continuación, se describe cada uno de estos pasos:

1. **Eliminación de observaciones duplicadas:** El primer paso en el proceso de limpieza es la eliminación de valores duplicados. Dado que la recolección y almacenamiento de información se realiza cada diez días, al momento de consolidar en una única base se detectó registros repetidos. Para solucionar esto, sobre la base consolidada se utiliza la función ‘Quitar duplicados’ de Microsoft Excel, la cual suprime todas las observaciones que sean duplicadas.
2. **Limpieza del texto:** Una vez que se cuenta con registros únicos, se procede a realizar la limpieza de texto en todos los campos de información, esto es, remover todos los caracteres que no aportan o no guardan coherencia con la descripción de la variable. Por tratarse de extracción de información en lenguaje HTML, al momento de la captura de datos existen *<<datos basura>>*. Algunos de estos caracteres se refieren a signos de puntuación,

conjunciones y palabras no relacionadas. Este procedimiento se lo realiza con técnicas de minería de texto.

3. **Tratamiento de valores perdidos:** El criterio utilizado para llenar los campos vacíos de información fue estrictamente basado en la observación y análisis de la variable FUNCIONES (tareas y competencias requeridas para ocupar la vacante) que, contiene generalmente una amplia descripción de la vacante y permite detectar palabras clave para completar los vacíos de información presentes en el resto de variables.
4. **Codificación:** Para el análisis de la información y facilitar la comparabilidad es necesario codificar ciertos campos. Estos campos se refieren en específico a las variables capturadas del cargo y la localidad. Respecto al cargo, la codificación se hizo mediante el contraste con las ocupaciones registradas a 8 dígitos en la Clasificación Internacional Uniforme de Ocupaciones (CIUO-08). En cuanto a la localidad (que reporta el texto de provincia y ciudad), se separó su texto en “PROVINCIA” y “CIUDAD”; luego, empatando con la División Político Administrativa (DPA) se obtienen los códigos de provincia y cantón.

Finalmente, concluido todo el proceso de depuración, se logra contar con una base de datos lista para ser analizada, obtener resultados y emitir conclusiones. A continuación, se muestra la descripción de la base para cada una de sus variables (Tabla 1):

Tabla 1. Estructura de la base de datos

Variable	Tipo	Descripción
EMPRESA	Alfanumérico	Nombre de la empresa
DESC. OFERTA	Alfanumérico	Nombre de la vacante tal como se reporta
AREA	Alfanumérico	Nombre de la vacante tal como se reporta
CARGO	Alfanumérico	Nombre de la vacante
CODIGO CIUO	Código numérico	Código CIUO correspondiente con el cargo
DESCRIPCION CIUO	Alfanumérico	Nombre de la ocupación según el CIUO-08 a 8 dígitos
JORNADA	Categórica	Jornada de trabajo (tiempo completo, parcial, por horas, otros)
CONTRATO	Categórica	Tipo de contrato (indefinido, determinado, por obra, otros)
SALARIO	Alfanumérico	Salario
DESC EMPRESA	Área de texto	Descripción de la actividad de la empresa
LOCALIDAD	Alfabética	Ubicación de la vacante por cantón-provincia

DPA-PRV	Código numérico	Código de la provincia conforme DPA
PROVINCIA	Alfabética	Nombre de la provincia donde se localiza la vacante
DPA-CNT	Código numérico	Código de la ciudad conforme DPA
CIUDAD	Alfabética	Nombre de la ciudad donde se localiza la vacante
FUNCIONES	Área de texto	Tareas y competencias requeridas para ocupar la vacante
FECHA CONTRATACION	Fecha	Fecha de posible vinculación
FECHA PUBLICACION	Fecha	Fecha de publicación de la vacante
VACANTES	Numérica	Número de puestos de trabajo por vacante
EDUCACION MINIMA	Categórica	Nivel de estudios mínimo para acceder al cargo (bachillerato, universidad, postgrado, otros)
EXPERIENCIA	Numérica	Experiencia en años requerida
DIS. VIAJE	Dicotómico	Disponibilidad para viajar
IDIOMA	Alfabética	Idiomas requeridos para aplicar
SEXO	Categórica	Preferencia de sexo para la vacante (hombre, mujer, ambos sexos, no específica)

Fuente: CompuTrabajo, elaboración de los autores.

5. Resultados

A continuación, se presenta algunas estadísticas obtenidas del ejercicio. El análisis consiste en analizar: distribuciones para el total de observaciones, variaciones entre los meses de septiembre y octubre, y algunas relaciones importantes entre variables.

Del proceso de descarga y depuración se obtiene una base de datos que, entre septiembre y octubre de 2016, consta de 4.385 registros (anuncios) correspondientes a 13.271 vacantes (plazas de trabajo vacías). Esta diferenciación entre anuncios y vacantes se hace porque ocurre con cierta frecuencia que las empresas requieren contratar más de una vacante en un mismo anuncio, algo que puede verse más fácilmente, por ejemplo, en una empresa de limpieza que requiere diez personas para el cargo de asistente de limpieza. La Tabla 2 muestra el número de anuncios y de vacantes que se publicaron en CompuTrabajo en los meses de septiembre y octubre.

Tabla 2. Anuncios y vacantes

2016			
	Septiembre	Octubre	Total
Anuncios	2.227	2.158	4.385
Vacantes	6.491	6.780	13.271

Fuente: CompuTrabajo, elaboración de los autores.

Si bien en octubre hay menos anuncios, existe un aumento en el número de vacantes, debido a que en octubre se tiene más anuncios que ofertan 10 o más plazas de trabajo (y menos anuncios con una sola vacante). La mayoría de anuncios, cerca de un 60%, ofrece una única plaza de trabajo, mientras que, en el otro extremo, hay un 7% de anuncios que presentan 10 o más plazas de trabajo vacantes. La Tabla 3 muestra la frecuencia de los números de vacantes por anuncio en los meses de septiembre y octubre.

Tabla 3. vacantes por anuncio

Número de puestos por anuncio	Sep	Oct	Total	%
1	1.345	1.299	2.644	60,3
2	377	373	750	17,1
3	150	140	290	6,6
4	58	63	121	2,8
5	111	91	202	4,6
6	23	21	44	1,0
7	5	2	7	0,2
8	9	9	18	0,4
9	0	2	2	0,0
10+	149	158	307	7,0
Total	2.227	2.158	4.385	100

Fuente: CompuTrabajo, elaboración de los autores.

A nivel geográfico, en la Tabla 4 se puede observar que la gran mayoría de vacantes se concentra en las tres principales ciudades del país. De esta manera, Quito aglutina un 51% de vacantes, seguido por Guayaquil con un 31% y Cuenca con apenas un 3%. En ambos meses no se aprecia ningún cambio significativo en la clasificación de ciudades.

Tabla 4. Vacantes por ciudad

Ciudad	Sep	Oct	Total	%
Quito	3.389	3.313	6.702	50,7
Guayaquil	1.973	2.113	4.086	30,9
Cuenca	179	203	382	2,9
Ambato	156	85	241	1,8
Machala	83	107	190	1,4
Santo Domingo	79	92	171	1,3
Rumiñahui	60	68	128	1,0
Portoviejo	59	50	109	0,8
Manta	56	31	87	0,7
Riobamba	49	44	93	0,7
Extranjero	47	0	47	0,4
Ibarra	44	58	102	0,8
Otros	317	556	873	6,6
Total	6.491	6.780	13.271	100,0

Fuente: CompuTrabajo, elaboración de los autores.

El salario asociado a la demanda laboral de una empresa es una variable sensible de presentar en las páginas web. Incluso, en algunos anuncios laborales es la empresa quién solicita al postulante que detalle el monto o rango de interés. La información del salario, es uno de los factores determinantes tanto para la oferta como para la demanda laboral de continuar en un proceso de selección.

Respecto a la información de salarios en este ejercicio, estos son presentados en rangos para facilitar su análisis. El primer intervalo se construyó en función del Salario Básico Unificado (SBU) vigente para 2016 que corresponde a US\$ 366 mensuales, en tanto que el último intervalo fue acotado en US\$2.000 y más, dado el reducido número de observaciones para este rango. De la Tabla 5 se desprende que, un 42% de plazas de trabajo no mencionan el salario y prefieren acordarlo con los candidatos al momento del reclutamiento, un 52% de plazas de trabajo ofrecen un salario menor a US\$900 y, llama la atención que cerca de un 10% de vacantes propongan salarios menores al SBU.

En octubre se aprecia una mayor oferta de vacantes con salarios menores a los US\$400. Un análisis con una mayor temporalidad permitirá conocer si esto es una coincidencia u obedece a algún fenómeno particular.

Tabla 5. Distribución de salarios

Rangos salariales	Sep	Oct	Total	%
0-366	554	705	1.259	9,5
366-400	920	1.222	2.142	16,1
400-500	654	590	1.244	9,4
500-600	493	530	1.023	7,7
600-700	321	377	698	5,3
700-800	124	95	219	1,7
800-900	171	79	250	1,9
900-1000	50	27	77	0,6
1000-1250	119	109	228	1,7
1250-1500	22	8	30	0,2
1500-1750	44	30	74	0,6
1750-2000	9	4	13	0,1
2000+	14	7	21	0,2
A convenir	2.855	2.767	5.622	42,4
No especifica	141	230	371	2,8
Total	6.491	6.780	13.271	100,0

Fuente: CompuTrabajo, elaboración de los autores.

Como se puede observar en la Tabla 6, la experiencia laboral es un elemento determinante para acceder a una plaza de trabajo. Apenas el 3% de vacantes no requieren ningún tipo de experiencia para ser satisfechas mientras que para el 85% de los puestos de trabajo se condiciona el tener tiempo de experiencia; el 12,7% de plazas no especifican la necesidad de contar o no con experiencia. De las vacantes que requieren experiencia, la mayoría solicitan tener al menos un año de experiencia (48%), seguido de las que necesitan al menos dos (21%) y tres años de experiencia (10%). Es fácil notar que existe una relación inversa entre años de experiencia y plazas de trabajo en los anuncios.

Tabla 6. Distribución de los años de experiencia

Experiencia	Sep	Oct	Total	%
0	192	164	356	2,7
1	3.063	3.306	6.369	48,0
2	1.351	1.423	2.774	20,9
3	635	734	1.369	10,3
4	147	93	240	1,8
5	209	174	383	2,9
6	12	12	24	0,2
7	4	5	9	0,1

8	27	3	30	0,2
10	10	24	34	0,3
No especifica	841	842	1.683	12,7
Total	6.491	6.780	13.271	100,0

Fuente: CompuTrabajo, elaboración de los autores.

Al analizar en conjunto la relación entre años de experiencia y salario propuesto, se puede observar que existe una correlación positiva. El 59% de las plazas de trabajo ofrecen menos de US\$ 700 y exigen un año de experiencia o menos. Las vacantes que ofertan US\$2.000 o más exigen tener al menos 3 años de experiencia (Tabla 7).

Tabla 7. Rangos salariales y años de experiencia.

Salario vs Exp	0	1	2	3	4	5+	Total %
0-366	0,20%	12,50%	1,10%	0,20%	0,30%	1,00%	15,30%
366-400	1,40%	20,40%	5,00%	0,70%	0,20%	0,20%	27,90%
400-500	1,80%	6,80%	5,50%	2,10%	0,20%	0,20%	16,60%
500-600	0,00%	8,40%	3,10%	1,50%	0,10%	0,10%	13,20%
600-700	0,70%	6,70%	1,00%	1,20%	0,20%	0,00%	9,80%
700-800	0,00%	1,10%	1,20%	0,70%	0,40%	0,40%	3,80%
800-900	0,00%	3,80%	0,90%	0,60%	0,10%	0,40%	5,80%
900-1000	0,40%	0,10%	0,40%	0,10%	0,00%	0,00%	1,00%
1000-1250	0,00%	0,80%	1,30%	0,80%	0,80%	0,20%	3,90%
1250-1500	0,00%	0,10%	0,40%	0,20%	0,00%	0,00%	0,70%
1500-1750	0,00%	0,10%	0,30%	0,50%	0,50%	0,20%	1,60%
1750-2000	0,00%	0,00%	0,00%	0,00%	0,00%	0,30%	0,30%
2000+	0,00%	0,00%	0,00%	0,10%	0,10%	0,20%	0,40%
Total %	4,50%	60,80%	20,20%	8,70%	2,90%	3,20%	100%

Fuente: CompuTrabajo, elaboración de los autores.

En lo que respecta a las características educativas de las vacantes, en la Tabla 8 se observa que en su mayoría (43%) requieren aspirantes que al menos hayan concluido el bachillerato, un 22% exigen título universitario y apenas un 1% demanda profesionales de cuarto nivel. Entre septiembre y octubre la demanda de bachilleres aumentó mientras que la de universitarios bajó en una proporción similar.

Esta mayor demanda de personal menos calificado puede responder a un fenómeno estacional, sin embargo, es todavía una hipótesis a ser comprobada⁹.

⁹ Se debe considerar que en el presente ejercicio el periodo de análisis es muy corto.(2 meses)

Tabla 8. Educación mínima requerida

Educación mínima	Sep	Oct	Total	%
Bachillerato	2.725	3.037	5.762	43,4
Universidad	1.584	1.361	2.945	22,2
Educación Técnica	964	981	1.945	14,7
Educación Básica Secundaria	872	893	1.765	13,3
Primaria	246	444	690	5,2
Postgrado	100	64	164	1,2
Total	6.491	6.780	13.271	100

Fuente: CompuTrabajo, elaboración de los autores.

Tabla 9. Rangos salariales y educación mínima

Salario vs Educación Mínima	Primaria	Básica	Bachillerato	Técnico	Universidad	Postgrado
0-366	23,2%	12,7%	9,8%	9,1%	31,2%	28,6%
366-400	7,2%	34,5%	37,9%	18,8%	8,3%	0,0%
400-500	28,3%	35,6%	14,2%	21,1%	14,9%	0,0%
500-600	40,6%	11,0%	14,9%	16,4%	9,0%	0,0%
600-700	0,0%	0,4%	13,6%	11,2%	6,4%	0,0%
700-800	0,0%	3,4%	1,5%	10,4%	4,0%	7,1%
800-900	0,0%	0,0%	5,7%	5,8%	6,2%	21,4%
900-1000	0,0%	2,1%	0,3%	2,1%	3,0%	0,0%
1000-1250	0,7%	0,2%	1,8%	3,5%	8,3%	14,3%
1250-1500	0,0%	0,0%	0,3%	0,4%	1,8%	0,0%
1500-1750	0,0%	0,0%	0,0%	1,0%	4,5%	7,1%
1750-2000	0,0%	0,0%	0,0%	0,0%	1,1%	0,0%
2000+	0,0%	0,0%	0,0%	0,2%	1,2%	21,4%
Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Fuente: CompuTrabajo, elaboración de los autores.

Además, gracias al ejercicio de codificación realizado sobre los nombres de los cargos ofertados en los anuncios, podemos llegar a conocer los tipos de ocupación más demandados. En la Tabla 10, puede

Asociando el nivel de instrucción requerido con el salario ofertado (Tabla 9), se contrasta que existe una relación positiva entre ambas, en el sentido de que los rangos salariales más altos sólo se alcanzan a partir de cierto nivel de instrucción, por ejemplo, el 64% de vacantes de postgrado prometen salarios mayores a los US\$ 800 algo que sólo llega al 8% en secundaria y menos del 1% en primaria. Sin embargo, esta relación no es tan fuerte ya que se pueden detectar ciertas anomalías interesantes, como por ejemplo: llama muchísimo la atención que de las vacantes que requieren sólo instrucción primaria, el 41% de ellas pague entre US\$500 y \$600, mientras que en las de secundaria, un 83% no alcanza ni los US\$500; así mismo, es cuestionable que el 31% de vacantes de universidad y el 29% de postgrado oferten salarios menores al SBU, siendo estas las incidencias más altas entre todos los niveles de instrucción, incluso mayores que primaria (23%).

apreciarse que las ocupaciones más comunes son aquellas relacionados con el tema de ventas. De las 10 ocupaciones más demandas, 7 tienen que ver con característica previamente mencionada.

Tabla 10. Cargos más demandados

Categoría de ocupación	Sep	Oct	Total	%
Asesor comercial	877	715	1.592	12,0
Agente, ventas	201	631	832	6,3
Vendedor, comercio	578	213	791	6,0
Vendedor por teléfono (Telemarketer)	372	365	737	5,6
Vendedor puerta a puerta	252	341	593	4,5
Cobrador, deudas	260	180	440	3,3
Asistente de contabilidad	132	280	412	3,1
Demostrador (impulsador ventas)	267	114	381	2,9
Cajero	127	156	283	2,1
Ayudante de cocina	179	81	260	2,0
Otros	3.105	3.474	6.579	49,6
No especifica/No es posible codificar	141	230	371	2,8
Total	6.491	6.780	13.271	100,0

Fuente: CompuTrabajo, elaboración de los autores.

Finalmente, el Gráfico 2 muestra un mapa de palabras realizado para la variable “FUNCIONES”. Este gráfico fue realizado gracias al uso de técnicas de *Text mining* que permiten realizar análisis sobre variables de texto. Debido a que con las variables de texto no pueden hacerse tabulaciones ni conteos que se hacen con variables categóricas o numéricas, en ocasiones es oportuno usar otro tipo de técnicas como el mostrado en el Gráfico 2, donde se puede apreciar, de forma general, una nube de palabras que muestra aquellos términos que se repiten con mayor frecuencia. El tamaño del texto se relaciona con la frecuencia y tener una descripción visual del texto que se está analizando. De manera específica, observando el Gráfico 2 podemos tener una idea de que las funciones más demandadas por parte de las empresas están relacionadas con palabras como: clientes, manejo, equipo, área, conocimiento, venta, control, servicio, comercial, etc., funciones que guardan coherencia con las ocupaciones más demandadas.

Gráfico 2: Mapeo de palabras de las funciones solicitadas



Fuente: CompuTrabajo, elaboración de los autores.

6. Conclusiones y recomendaciones

El objetivo de este artículo ha sido presentar la experiencia que se ha tenido en el INEC en la elaboración de estadísticas de puestos de trabajo que se publican en internet. Pese a que se conoce que existen distintas páginas web que realizan este tipo de intermediación laboral, el ejercicio aquí documentado se limita a explotar la información del portal de CompuTrabajo. Aunque se reconoce que potencialmente es importante incorporar diferentes bolsas de empleo al análisis¹⁰, se ha optado dejar esta tarea para futuros desarrollos.

Una de las conclusiones destacables, es que el INEC es capaz de utilizar la información publicada en Internet para elaborar estadísticas. Concretamente, las tablas mostradas en este documento constituyen una prueba de ello. Este tipo de estadística puede obtenerse a un costo relativamente bajo y con una oportunidad muy alta, ya que la información está disponible y se puede descargar día a día. Sin embargo, se sabe que este tipo de información tiene ciertas limitaciones, relacionadas principalmente con la imposibilidad de aplicar métodos estadísticos para realizar inferencias.

10Al incorporar nuevas bolsas de empleo para el análisis se puede visibilizar el panorama global de la demanda laboral y se podría realizar ejercicios como por ejemplo, identificar si las empresas posicionan sus vacantes en un portal web específico o en varios.

Para enfrentar algunas de estas debilidades se ha mencionado como alternativa el incluir en alguna de las encuestas dirigidas a empresas (o quizás diseñar una encuesta para el efecto) un conjunto de preguntas que permitan poner en contexto la información de las vacantes publicadas en internet. Por ejemplo, sería útil conocer qué porcentaje de las empresas usan este tipo de bolsas de empleo para promocionar sus necesidades de mano de obra, y algunos datos adicionales para caracterizar estas prácticas y sus tendencias.

Aún con las debilidades conocidas, es probable que las estadísticas de las vacantes publicadas en internet sean una forma útil y práctica de medir el pulso de la economía. Sin embargo, la evaluación de la relación que puede existir entre el número de vacantes publicadas y la actividad económica es algo que sólo puede hacerse una vez que se cuente con un rango temporal de la información lo suficientemente extenso. De cualquier forma, conocer en tiempo real las demandas de las empresas, respecto del personal a contratar, puede proveer de información útil sobre el tipo de perfiles de trabajadores que se requieren.

Dado que este artículo se basó en el análisis de dos meses de referencia, se recomienda realizar la investigación sobre una temporalidad mayor con la finalidad de poder identificar patrones y fluctuaciones en las variables consideradas.

7. Referencias

Antweiler, W. & Murray, F., 2004. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards.. *Journal of Finance*, 1259-94(59), p. 1.

Askitas, N. & Zimmermann, K. F., 2015. The Internet as a data source for advancement in social sciences. *Institute for the Study of Labor*, Issue 8899, p. 3.

Azar, O. H., 2007. The Slowdown in First-Response Times of Economics Journals: Can it Be Beneficial?. *Economic Inquiry*, 179-87(45), p. 1.

Bajari, P. & Hortacsu, A., 2003. The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions. *RAND Journal of Economics*, 329-55(34), p. 1.

Benítez, D.; Espinoza, B.; Grijalva, A.; Rivadeneira, Ana.; Oviedo, Ana María., 2016. Using administrative

records to study employer-employee dynamics: The case of Ecuador's Laboratory of Labor and Business Dynamics. *Working Paper*, Issue 1, pp. 2-4.

Cárdenas, J., Guataquí, J. C. & Montaña, J., 2015. Metodología para el análisis de demanda laboral mediante datos de Internet: el caso colombiano.

Carnevale, A. P., Jayasundera, T. & Repnikov, D., 2014. Understanding Online Job Ads Data. A Technical Report. *Georgetown University*, p. 17.

Castrillo-Fernández, O., 2015. Web Scraping: Applications and Tools. *European Public Sector Information Platform*, Issue 2015/10, pp. 1-15.

Cavallo, A., 2009. Scraped Data and Sticky Prices: Frequency, Hazards, and Synchronisation. *Harvard University Job Market Paper*, p. 5.

Cavallo, A., 2012. Online and official price indexes: measuring Argentina's inflation. *Massachusetts Institute of Technology*, p. 9.

Edelman, B., 2012. Using Internet Data for Economic Research. *American Economic Association*, 26(2), p. 192.

INEC, Instituto Nacional de Estadísticas y Censos., 2016. Metodología para la medición del empleo en el Ecuador. pp. 8-10.

Kuhn, P. & Mansour, H., 2011. Is Internet Job Search Still Ineffective?. *Institute for the Study of Labor (IZA) DP*, Issue 5955, p. 1.

Kurekova, L. M., Miroslav, B. & Anna, T.-T., 2015. Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. *IZA Journal of Labor Economics*, Volumen 4:18, pp. 2-3.

Reimsbasch-Kounatze, C., 2015. The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies. A preliminary analysis.. *OECD Digital Economy Papers*, Issue 245, p. 15.

Seamans, R. & Zhu, F., 2010. Technology Shocks in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers. *NET Institute Working Paper*, 10(11), p. 1.

Stefanik, M., 2012. Internet job search data as a possible source of information on skills demand (with results for Slovak university graduates). *Building on skills Forecasts-Comparing Methods and Applications*, p. 1.

3

Encuesta basada en registros administrativos para el análisis de la creación y destrucción de plazas de trabajo
Una propuesta metodológica

Diego Benítez*

Boris Espinoza*

Ana Grijalva+

RESUMEN

La utilización de registros administrativos con fines estadísticos requiere de un proceso de depuración, armonización, modificación y validación de los mismos. Como resultado, se obtiene una fuente de información, denominada encuesta basada en registros, orientada a responder preguntas específicas. Esta fuente, al no incluir información sensible o de identificación de las unidades estadísticas, es susceptible de ser publicada para su utilización académica. El presente artículo describe la elaboración de una encuesta basada en registros administrativos para analizar la creación y destrucción de plazas de trabajo. Se plantean las potencialidades y se describen algunos problemas encontrados.

Palabras clave: encuesta por registro, plazas de trabajo

Descargo de responsabilidad: Las opiniones e interpretaciones expresadas en este documento pertenecen a los autores y no reflejan el punto de vista oficial del Instituto Nacional de Estadística y Censos (INEC). El INEC no garantiza la exactitud de los datos que figuran en el documento.

+ Consultores del Banco Mundial. Los autores agradecen el apoyo financiero del *Jobs Umbrella Trust Fund* para la realización de este documento así como también para muchas otras actividades orientadas al fortalecimiento de la producción de estadísticas laborales en Ecuador. Lo expresado en este documento no representa la posición del INEC, del gobierno del Ecuador, ni del Grupo Banco Mundial. Cualquier error corresponde a los autores. Correspondencia a través de dieben79@yahoo.com.

1. Introducción

Desde el año 2014, el Instituto Nacional de Estadística y Censos (INEC), con el apoyo de la Comisión Económica para América Latina y el Caribe (CEPAL), ha desarrollado el Laboratorio de Dinámica Laboral y Empresarial (LDLE). El LDLE es un Sistema de registros administrativos con fines estadísticos (Benítez, Espinoza, Grijalva, Rivadeneira y Oviedo (2016)), según el modelo propuesto por Wallgren y Wallgren (2014). Con la intención de impulsar el fortalecimiento de las estadísticas sobre el mercado laboral, el Banco Mundial auspició la exploración de la información existente en el LDLE, así como el diseño de nuevos indicadores, y la realización de nuevos análisis.

En este contexto, se diseñó un proceso que vincula información de las empresas, sus empleados, y sus actividades para estructurar una fuente de información que permita calcular indicadores relacionados con la dinámica de ocupación de las plazas de trabajo, y realizar otros estudios relacionados con la temática. Según la metodología utilizada, este tipo de fuente recibe el nombre de encuesta basada en registros administrativos (Wallgren y Wallgren 2014).

El presente documento tiene por finalidad describir el proceso de elaboración de esta encuesta basada en registros administrativos, y proponer el cálculo de algunos indicadores acerca de la creación y destrucción de plazas de trabajo. Para lograr este objetivo, primero se describe la estructura de la encuesta y el proceso de su compilación. Segundo, se proponen algunos indicadores de creación y destrucción de plazas. Tercero, se muestran algunos resultados entre los que se aborda la identificación de una singularidad en un segmento de la data. Finalmente, se plantean algunas conclusiones.

2. Estructura y proceso de compilación de la encuesta

Una encuesta basada en registros administrativos (register-based survey) es una fuente de información construida a partir de la vinculación de registros estadísticos, que responde a una necesidad específica de producción estadística o de investigación (Wallgren y Wallgren 2014). Los registros estadísticos son aquellos

que, a partir de un registro administrativo, han pasado por un proceso que los dota de las características necesarias para poder ser utilizados en la elaboración de estadísticas. Los registros estadísticos se vinculan en macro estructuras denominados registros base. Teóricamente, el sistema de registros del INEC se compone de cuatro registros base: los Registros de i) Población, ii) Residencia, iii) Actividades, y iv) Negocios; cada uno de los cuales está caracterizado por la unidad estadística que lo define.

En el caso del Registro de Población la unidad estadística es el individuo; en el caso del Registro de Negocios es la persona jurídica, que por simplificación se denominará simplemente empresa. Y, en el caso del Registro de Actividades, la unidad estadística es la relación laboral existente entre una empresa y un individuo. La fuente del Registro de Población es la información del Registro Civil¹ (RC) complementada con el registro de títulos en niveles de educación superior de la Secretaría Nacional de Educación Superior Ciencia y Tecnología² (SENESCYT). La fuente del Registro de Negocios es, principalmente, el Servicio de Rentas Internas³ (SRI). Y, la fuente del Registro de Actividades es el registro de afiliados al Instituto Ecuatoriano de Seguridad Social (IESS).

1 Institución encargada del registro de los ecuatorianos o residentes en el país. Cuenta con información como la edad, el sexo, el estado civil, el nombre de los padres y el lugar de nacimiento, entre otra.

2 Es una dependencia del poder ejecutivo cuya finalidad es la adopción de las políticas para asegurar la calidad de la educación superior.

3 Oficina nacional de impuestos.

Tabla 1. Distribución de las empresas activas entre 2009 y 2014 según su constitución jurídica

Constitución jurídica	2009	2010	2011	2012	2013	2014
Persona natural obligada a llevar contabilidad	13.673	16.591	21.473	26.232	28.521	30.494
Sociedad con fines de lucro	26.919	29.287	32.970	36.070	37.079	40.207
Sociedad sin fines de lucro	3.141	3.186	3.629	3.952	3.995	4.373
Empresa pública	82	95	139	187	234	245
Institución pública	4.296	4.404	4.506	4.558	4.657	4.744
Economía popular y solidaria	743	831	977	1.064	1.137	1.254
Total	48.854	54.394	63.694	72.063	75.623	81.317

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Dado que la propuesta que aquí se plantea tiene como objetivo el análisis de la dinámica del empleo desde el enfoque de las empresas, el eje de la encuesta basada en registros administrativos⁴ es la información disponible en el Registro de Negocios, que cuenta, entre otras, con variables como las ventas anuales, el número de trabajadores y la rama de actividad económica, durante el periodo comprendido entre los años 2009 y 2014. Por lo tanto, la unidad estadística de la encuesta también será la empresa, considerando como tal a: las sociedades con y sin fines de lucro, las personas naturales obligadas a llevar contabilidad, las instituciones públicas, las empresas públicas, y las entidades productivas registradas bajo la constitución jurídica de economía popular y solidaria⁵. La encuesta cubrirá el mismo periodo que posibilita el registro base (a la fecha, 2009-2014), y la Tabla 1 exhibe el número de empresas o unidades institucionales activas⁶ en cada año.

La encuesta tendrá entonces una estructura de panel, en la que las observaciones-las empresas o unidades institucionales- se presentan en todo, o en parte de, el periodo analizado⁷. Cada observación toma

4 Con la finalidad de guiar al lector, en el resto del documento se llamará únicamente “encuesta” a la “encuesta por registros administrativos”.

5 Así, se excluyen a las personas naturales no obligadas a llevar contabilidad y a las entidades pertenecientes al régimen impositivo simplificado ecuatoriano (RISE).

6 Según la metodología del LDLE, se considera empresas activas a aquellas que registran ventas, empleo y rama de actividad CIU 4 dígitos.

7 Debido a la demografía empresarial, existen entidades que se originan durante el periodo que cubre la encuesta, así como otras que desaparecen; en total, la encuesta contabiliza información para más de 108 mil empresas o unidades institucionales distintas.

del Registro de Negocios las siguientes variables: a) el identificador único de la empresa o unidad institucional; b) el sector al que pertenece-público o privado-; c) la provincia de su domicilio fiscal; d) su rama de actividad-CIU 4-; e) el tipo de constitución jurídica⁸; f) su tamaño-grande, mediana o pequeña-; y f) su volumen de ventas anual.

La estructura de la encuesta se completa al vincular la información de las empresas o unidades institucionales con el Registro de Actividades y, a través de éste, con el Registro de Población. El Registro de Actividades cumple dos funciones esenciales para la encuesta: en primer lugar, a través del RUC, posibilita la cuantificación del número de afiliaciones correspondientes a cada empresa o unidad institucional⁹; y en segundo lugar, a través de la cédula de identidad, permite enlazar cada registro de afiliación¹⁰ con las características del trabajador afiliado, mismas que se encuentran en el registro base de población.

Cabe aquí la introducción de dos de los elementos conceptuales de esta encuesta, y del enfoque que con ella se pretende instrumentar.

8 Se refiere a las categorías de la Tabla 1: sociedad con fines de lucro, empresa pública, etc.

9 El registro base de actividades contiene las afiliaciones registradas para cada mes del año; para esta encuesta en especial, se utilizarán las afiliaciones reportadas en los meses de mayo, ya que es uno de los meses con menor efecto estacional (García, Garzón, Palacios y Puebla (2016)).

10 Nótese que se habla de afiliaciones, puesto que la unidad estadística del registro base de actividades es la relación laboral y no el trabajador o empleado. Al respecto, por ejemplo, un mismo trabajador podría tener más de un empleo -relación laboral- y constar en más de un asiento de afiliación.

Primero, se asume que las afiliaciones contabilizadas para una empresa equivalen al número de *plazas ocupadas* de que esta dispone¹¹. Segundo, se propone caracterizar dichas plazas a través de las cualidades o atributos de los trabajadores que las ocupan; se plantea, por ejemplo, que si la mitad de las plazas de un empresa están ocupadas por trabajadores con una titulación técnica, ese número de plazas requieren, en sí, una titulación técnica. Para la realización de esto último, es necesario enlazar a las plazas con las características de los trabajadores, es decir, vincular lo hasta ahora descrito con la información del Registro de Población.

A la información disponible en el RC, el Registro de Población incorpora los títulos de educación registrados por la SENESCYT, de esta manera se puede identificar si un individuo posee alguna titulación en educación superior y, de ser el caso, conocer el nivel de la misma, así como el área de conocimiento (ver Tabla 4 en Anexos). En caso de que un individuo posea más de una titulación en educación superior, el registro se refiere a la de más alto nivel. La integración de los diferentes registros permite conocer el número de plazas que las empresas llenan con individuos según sus títulos profesionales. De esta forma, a las variables de la encuesta que describen características de la empresa se suman las siguientes, que caracterizan a las plazas ocupadas:

- a. 6 variables que agrupan por el nivel de instrucción al número de plazas de trabajo (ver Tabla 4 en Anexos),
- b. 8 variables que agrupan por área de conocimiento al número de plazas de trabajo (ver Tabla 4 en Anexos),
- c. 3 variables que agrupan por grupos etarios: jóvenes -15 a 24 años-, adultos A-25 a 44 años-, y adultos B-45 años o más- al número de plazas de trabajo
- d. 8 variables que agrupan las carreras técnicas y tecnológicas para cada área de conocimiento al número de plazas de trabajo (plazas del grupo b ocupadas desglosado por trabajadores titulados en carreras técnicas y tecnológicas),
- e. 8 variables que agrupan el grupo etario de 15 a 24 años (jóvenes)desglosado por el nivel de

11 Podrían existir, sin embargo, plazas ocupadas por empleados que no han sido afiliados.

formación técnico y tecnológico y por cada área de conocimiento al número de plazas de trabajo(plazas del grupo b ocupadas por jóvenes y que están titulados en carreras técnicas y tecnológicas),

- f. 3 variables que agrupan por formación técnico/tecnológico desglosadas por los 3 grupos etarios al número de plazas de trabajo. (plazas del grupo c ocupadas por trabajadores titulados en carreras técnicas y tecnológicas).

A continuación se exponen las principales fases del proceso a través del cual se determinó el universo de empresas expuesto en la Tabla 1. Se debe considerar que para determinar este universo se vincula la base de negocios con la base de actividades. Como primer paso, en la Tabla 2, se observan las exclusiones en el registro base de negocios correspondientes a los registros RISE y personas naturales no obligadas a llevar contabilidad¹². Cada registro consiste en una persona, institución o empresa en un año determinado, por lo que, si una empresa aparece durante todo el periodo, le corresponderán 6 registros. Después de este filtro el número de registros de empresas utilizados en la encuesta es de aproximadamente 1 millón 64 mil. De este número deben excluirse aún las empresas no activas (no registran ventas, o empleo o CIU 4 dígitos).

Tabla 2. Registros excluidos. Registro base de negocios 2009-2014

Registros	Valores absolutos
Total inicial	7'555.164
- RISE	-3'026.886
- Persona Natural no Obligada a Llevar Contabilidad	-3'463.530
Residuo	1'064.748

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

En cuanto al registro de actividades, la Tabla 3 muestra las exclusiones de distintos registros que presentaban incongruencias (duplicados, RUC o cédula de identidad nulos, número de días de trabajo al mes superior a 31) o que no son congruentes con el objetivo de la encuesta (afiliados voluntarios, empleados domésticos, trabajadores autónomos o de una entidad RISE).

12 Se excluye estas categorías por fines metodológicos, dado que, por la naturaleza de su actividad, no necesariamente contarán con registros en la base de actividades (afiliados en la seguridad social).

Cada registro equivale a la afiliación de un trabajador en un año y una empresa determinada, por lo que, a un individuo afiliado durante todo el periodo en una misma empresa le corresponderían 6 registros en la base de datos; el total inicial es la suma de todos los registros existentes en el registro base durante el periodo de estudio. Luego de estos filtros, el número de registros de afiliaciones disponibles es de aproximadamente 13 millones 506 mil.

Tabla 3. Registros excluidos. Registro base de actividades 2009-2014*

Registros	Valores absolutos
Total inicial	14'961.341
-Afiliados voluntarios	-306.831
-Registros con RUC nulo	-466.686
-Registros con cédula nula	-37.349
-Registros duplicados	-53.840
-Trabajadores con número de días<=0 ó sueldo <=0	-30.529
-Trabajadores con número de días mayores a 31	-2.613
-Empleados domésticos	-76
-Trabajadores autónomos y sin relación de trabajo	-21.335
-Trabajadores de entidades RISE	-70.216
-Trabajadores por cuenta propia	-465.111
-Residuo	13'506.755

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)
*Mes de referencia: Mayo

La encuesta se consuma con la vinculación de ambos registros depurados (negocios y actividades). En esta fase se da una nueva depuración ya que, por características propias de los registros administrativos, no todos los registros pueden vincularse¹³. Finalmente, la encuesta abarca un número de 395.945 registros útiles, es decir el total de registros para el periodo de análisis (de empresas) distribuidos en cada uno de los periodos de estudio (ver Tabla 1).

13 En este ejercicio el punto de partida es la base de actividades y al vincular con la base de negocios se evidencia que no todas las empresas registran afiliaciones en la seguridad social.

3. Indicadores de creación y destrucción de plazas

Para el efecto de la presente propuesta, se definen como plazas de trabajo a la relación existente entre un individuo y una empresa. Por un lado, la empresa debe constar en el Registro de Negocios, y por el otro, la relación laborar debe constar en el Registro de Actividades, es decir, en el registro de afiliados del IESS. Por lo tanto, los indicadores se referirán a las plazas ocupadas por trabajadores formales, afiliados al seguro general del IESS. Esto implica que se excluyen las plazas ocupadas por trabajadores informales, y otras ocupadas por afiliados a otros sistemas de seguridad social como son el ISSFA y el ISSPOL¹⁴.

Por otro lado, se considera que ha habido una creación de plazas de trabajo en uno de los dos siguientes casos: i) cuando, para una misma empresa, el número de plazas en *t+1* es mayor que en *t*; y ii) cuando aparecen plazas correspondientes a una nueva empresa, es decir una empresa que existe en *t+1* pero no en *t*. Análogamente, se considera que ha habido una destrucción de plazas de trabajo: i) cuando, para una misma empresa, el número de plazas en *t+1* es menor que el número de plazas en *t*, y ii) cuando desaparecen plazas correspondientes a una empresa que cerró, es decir que existía en *t* pero no en *t+1*¹⁵.

Considerando los criterios planteados, siguiendo la propuesta de García, Grijalva, Palacios y Mauricio (2016), para el análisis de la dinámica del empleo registrado en el Ecuador, se calculan tasas de creación, de destrucción y de permanencia de plazas de trabajo, según el criterio que se detalla a continuación:

Tasa bruta de creación de plazas: número de plazas existente en *t* y no existentes en *t-1* sobre denominador (1) o denominador (2).

Tasa bruta de destrucción de plazas: número de plazas inexistente en *t+1* pero existentes en *t* sobre denominador (1) o denominador (2).

Tasa de plazas que permanecen: número de plazas existentes en *t* y en *t+1* sobre denominador (1) o denominador (2).

14 Institutos de seguridad social de las Fuerzas Armadas y de la Policía, respectivamente.

15 Se debe recordar que, en lo que a esta metodología concierne, la “existencia” de una empresa, y análogamente la de una plaza, está determinada por su aparición en los registros como una empresa activa.

Tasa de creación neta de plazas: la diferencia entre la tasa bruta de creación de plazas y la tasa bruta de destrucción de plazas

Dónde:

Denominador (1): es el resultado de la sumatoria de las plazas creadas, destruidas y permanentes en un periodo. Al dividir las tasas anteriores para este denominador se conoce, respecto a la totalidad del número de plazas registradas durante el periodo, la porción que se creó, destruyó y mantuvo.

Denominador (2): es el stock inicial de plazas de un periodo. Al dividir las tasas anteriores para este denominador se conoce, respecto a la totalidad del número de plazas iniciales, la porción que se creó, destruyó y mantuvo.

Gracias a que la información de base para estos cálculos proviene de Registros Administrativos, los indicadores pueden estimarse para subconjuntos detallados de empresas según su tamaño, el sector, la rama de actividad, y por las características de las plazas como son el nivel de formación, el área de conocimiento, el grupo etario, y el sexo. De esta forma, se aspira poder caracterizar la dinámica del empleo para distintos grupos; respondiendo a preguntas como: ¿En qué ramas productivas existe mayor generación de plazas de trabajo?, ¿Qué plazas, según el nivel de instrucción que las ocupa, están sujetas a una tasa neta de creación elevada?, etc. En la siguiente sección se muestran los principales resultados.

4. Principales resultados

El Gráfico 1 muestra las principales estadísticas sobre la evolución del número de empresas y plazas. En el panel A se muestra el número de empresas en el periodo comprendido entre 2009 y 2014, el panel B exhibe la demografía empresarial en el mismo periodo, el panel C contiene la evolución del número de plazas, y el panel D representa la *demografía* de las plazas.

En el panel A se identifica que gradualmente el número de unidades institucionales o empresas activas pasó de 48.862 en el año 2009 a 81.311 en 2014; éste es el resultado de un proceso dinámico de creación y destrucción de empresas que se expone en el panel B,

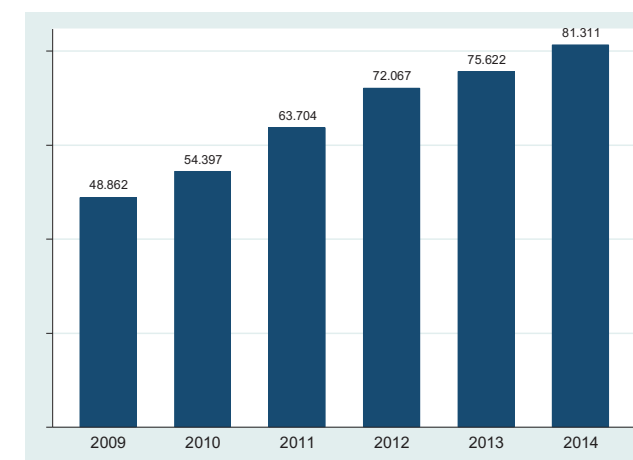
se trata de la demografía empresarial durante periodos interanuales a través de tres magnitudes principales: i) las empresas que se crean corresponden a entidades que no están registradas en el inicio del periodo pero aparecen al final del mismo, ii) las empresas que se mantienen aparecen registradas en ambos extremos del periodo, y iii) las empresas que se destruyen solo aparecen registradas al inicio del periodo. Durante el periodo 2009-2010, por ejemplo, se crearon 10.244 empresas, se mantuvieron 44.153, y se destruyeron 4.709.

Es necesario notar que, en un periodo determinado, el stock inicial es igual a la suma de las empresas que se mantienen y las empresas que cerraron durante ese periodo; por ejemplo, el stock inicial en el periodo 2014-2015 es de 75.622 empresas (67.494 + 8.128). El stock final de un periodo, en cambio, es la sumatoria de las empresas que se mantienen y las empresas que se crearon; por ejemplo, en el periodo 2012-2013, la adición de las empresas que se mantienen -62.853- y las que se crean -12.769- resultan en el stock final de ese periodo, 75.622, que, de hecho, es el stock inicial del siguiente periodo. Por tanto el crecimiento (o decrecimiento, si fuera el caso) en el número de empresas en un periodo dado, está determinado por la diferencia entre el stock final y el stock inicial, que es la misma diferencia que existe entre las empresas que se crean y las empresas que se destruyen.

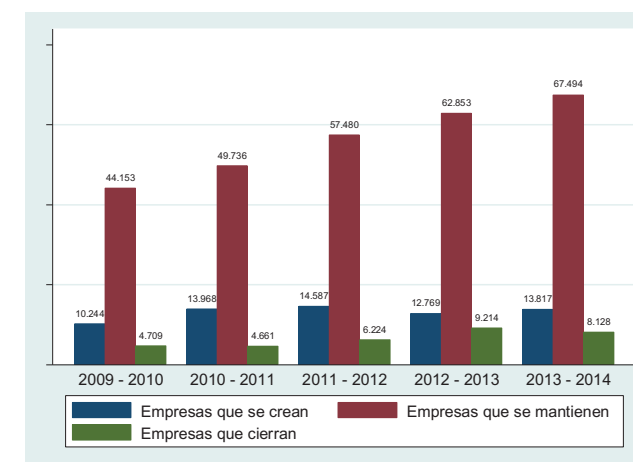
En cuanto a las plazas, en el panel C se observa un aumento gradual del número de plazas de trabajo registradas¹⁶, que pasa desde aproximadamente 1 millón 460 mil, en 2009, hasta superar los 2 millones 200 mil, en 2014. Con una metodología análoga a la descrita en el caso de las empresas, en el panel D puede observarse la demografía de las plazas; de tal modo que entre 2009 y 2010 se crearon aproximadamente 267 mil plazas de trabajo, se destruyeron o cerraron cerca de 140 mil y se mantuvieron alrededor de 1 millón 319 mil; entre 2013 y 2014, en cambio, se crearon cerca de 428 mil plazas, se destruyeron alrededor de 326 mil y se mantuvieron más de 1 millón 780 mil. Resulta de interés el hecho de que, aunque el número de empresas que se crean no tiene un ascenso continuo -panel B-, el número de plazas que se crean sí aumenta en cada periodo -panel D-; observación también puede trasladarse al caso de las empresas y plazas que cierran.

¹⁶ Como se mencionó en la sección anterior, la encuesta basada en registros administrativos reporta las plazas que existirían en los meses de mayo de cada año.

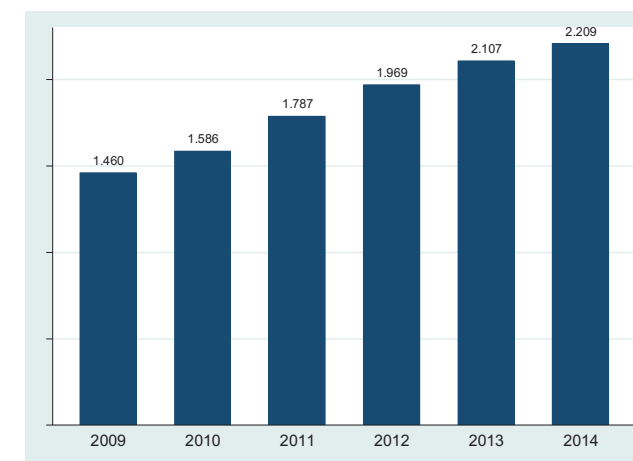
Gráfico 1. Evolución y demografía de empresas y plazas



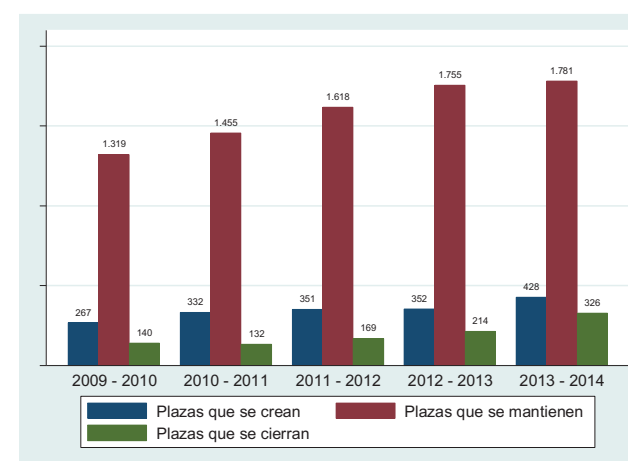
A) Evolución del número de empresas entre 2009 y 2014



B) Demografía empresarial entre 2009 y 2014



C) Evolución del número de plazas entre 2009 y 2014 (miles de plazas)

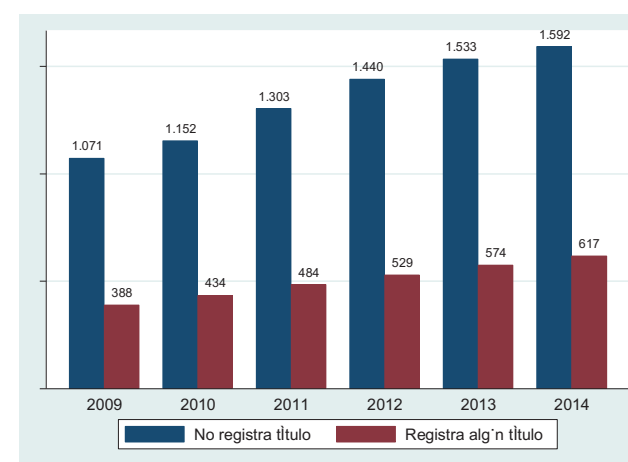


D) Demografía empresarial entre 2009 y 2014

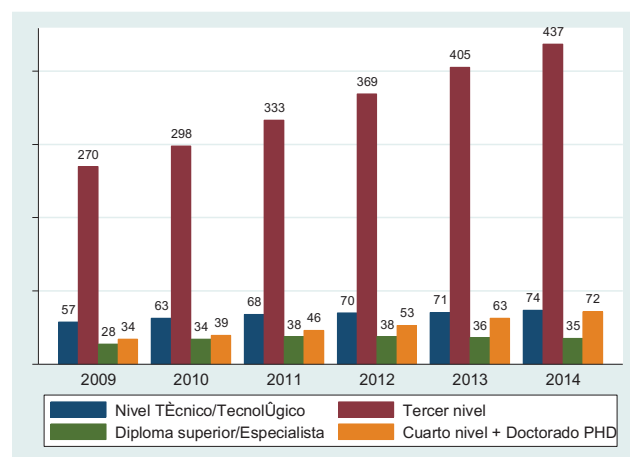
Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

El método de tipificación de las plazas según las características de los trabajadores, resulta de especial interés cuando se trata de analizar las plazas de trabajo según la instrucción de quienes las ocupan. El panel A del Gráfico 2 muestra el número de plazas ocupadas por personas que registran algún título de instrucción superior, que en 2009 fue aproximadamente 388 mil y en 2014 superó las 610 mil plazas de trabajo. Éstas plazas se desagregan en el panel B según el nivel de instrucción registrado, siendo notorio que mayoritariamente se trata de plazas ocupadas por trabajadores con instrucción de tercer nivel, seguidas por las que están ocupadas por trabajadores con nivel de instrucción técnico/tecnológico, que gradualmente ascendieron desde alrededor de 57 mil, en 2009, a cerca de 72 mil, en 2014.

Gráfico 2. Plazas según instrucción de los trabajadores



A) Número de plazas según titulación entre 2009 y 2014 (miles)



B) Número de plazas según nivel de instrucción entre 2009 y 2014 (miles)

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

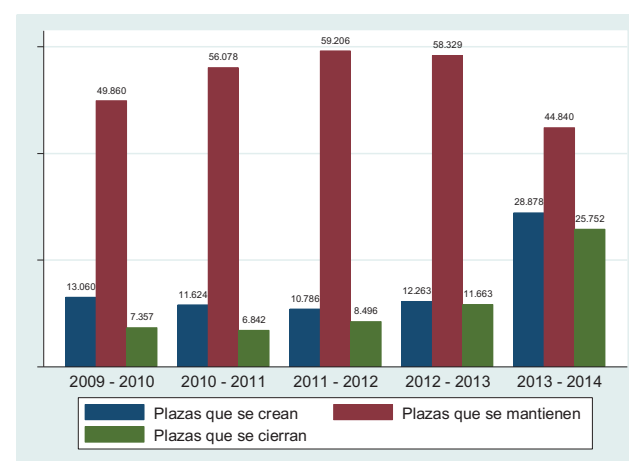
El perfil profesional es precisamente la dimensión escogida en este documento para profundizar en las capacidades de la encuesta para analizar el dinamismo de las plazas, y, también, para exponer una falencia técnica localizada en uno de los registros administrativos de los que se insume el sistema, cuyo examen resulta ilustrativo. A continuación se desarrolla el análisis de creación y destrucción de las plazas ocupadas por trabajadores con un nivel de instrucción en específico: el técnico/tecnológico.

El panel A del Gráfico 3 muestra la demografía de las plazas, que de aquí en más se denominarán como técnico/tecnológicas. Se puede advertir en el gráfico una diferencia sustancial de las magnitudes pertenecientes al periodo 2013-2014 respecto a los periodos previos: tanto el número de plazas que se crean como el de plazas que se destruyen en este periodo, superan por mucho las magnitudes correspondientes a otros periodos. Se encuentran más detalles sobre esta singularidad al aplicar la metodología de demografía a subconjuntos de las plazas técnico/tecnológicas. Una primera aproximación es analizar estas plazas según el área de conocimiento al que pertenezcan las titulaciones de los profesionales que las ocupan; a través de este ejercicio, se logró identificar que según los indicadores de demografía la singularidad afecta, de forma exclusiva, a las plazas ocupadas por profesionales con titulaciones técnico/tecnológicas relacionadas con el área de conocimiento de Educación, cuya demografía se presenta en el panel B del Gráfico 3.

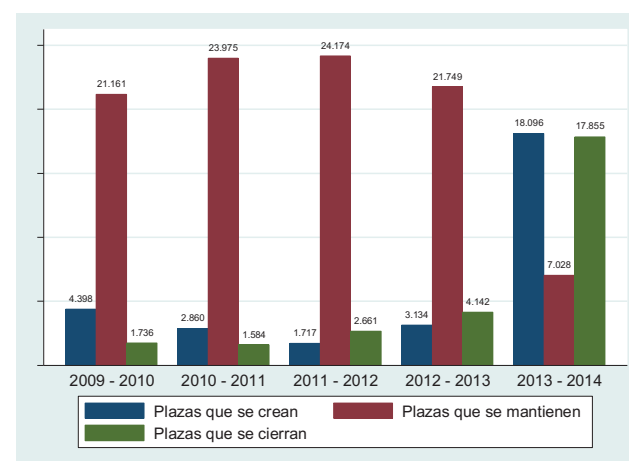
Otro enfoque utilizado fue la observación de la demografía en cada rama de actividad; se detectó así que las empresas que pertenecen a la rama de

actividad de Administración Pública-CIU 4- registran, en el periodo 2013-2014, volúmenes de creación y destrucción particularmente distintos a otros periodos, esto se representa en el panel C del Gráfico 3. Con este antecedente, se verificó la demografía de las plazas técnico/tecnológicas pertenecientes a las empresas con esta rama de actividad, que se expone en el panel D del Gráfico 3 y que replica lo observado en la demografía de las plazas técnico/tecnológicas del área de Educación (panel B).

Gráfico 3. Demografía de las plazas técnicas¹⁷

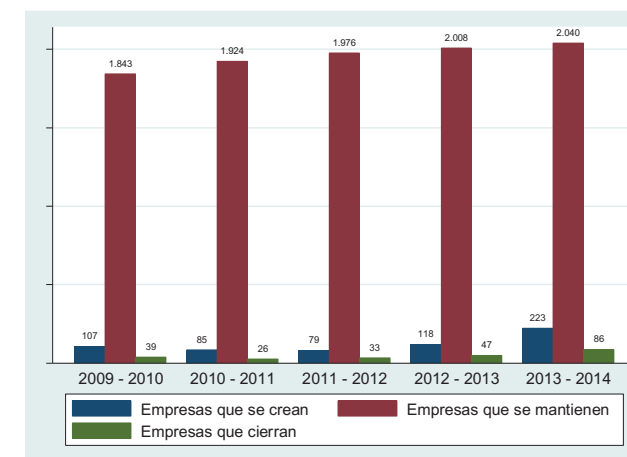


A) Demografía de plazas técnicas entre 2009 y 2014

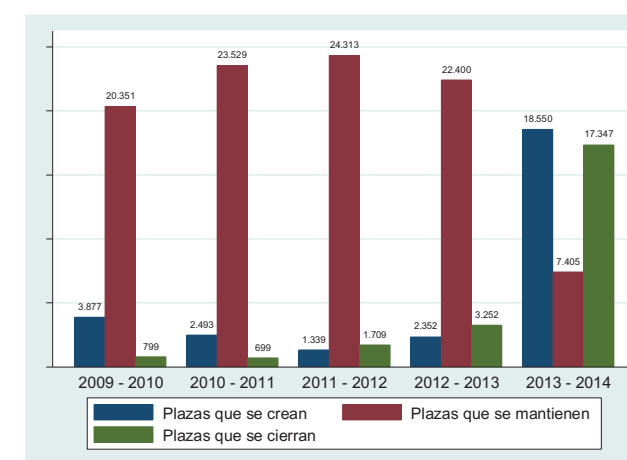


B) Demografía de plazas técnicas del área de conocimiento educación entre 2009 y 2014

¹⁷ El Gráfico 3 se presenta tres enfoques diferentes de análisis de plazas de trabajo respecto al nivel de formación técnico/tecnológico en distintos niveles de desagregación y un enfoque empresarial.



C) Demografía empresarial en la rama de actividad Administración Pública entre 2009 y 2014



D) Demografía de plazas técnicas en la rama de actividad Administración Pública entre 2009 y 2014

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Los hallazgos mencionados condujeron a una revisión de los registros originales, cuya información se incluye en la encuesta basada en registros, específicamente los que corresponden a las empresas clasificadas en la rama de actividad de Administración Pública. Esta revisión permitió identificar una particularidad propia de la naturaleza de los registros administrativos utilizados. Concretamente, se trata de un cambio en la forma de registro de las entidades educativas de la administración pública denominadas “Direcciones provinciales de educación” que, en 2014, pasaron a registrar su fuerza laboral a través de subunidades denominadas “Direcciones distritales de educación”. Es decir, en 2014 aparecen las Direcciones distritales al tiempo que desaparecen las Direcciones provinciales.

Un fenómeno de creación y destrucción de plazas, consistiría en que las nuevas entidades hubieran efectuado un procedimiento de contratación de personal, y en que las anteriores hubieran cerrado, con ellas, sus plazas. No obstante, lo sucedido es, más bien, un cambio en la denominación y estructura organizacional de los empleadores -las Direcciones provinciales-, y con ello también de sus identificadores únicos -RUC-. En tal caso, no existiría el singular fenómeno de creación y destrucción de plazas que se observa en el Gráfico 3, que en realidad se origina en un cambio institucional que da esa apariencia. Este particular no reduce la utilidad de la encuesta para el análisis de plazas, más bien se exhibe como ilustración de las precauciones y consideraciones que deben tomarse en el diseño encuestas basadas en registros, sobre todo en lo que tiene que ver con las peculiaridades de los registros administrativos.

En Anexos se exponen los principales resultados sobre los indicadores de dinámica, que también se calculan para algunos tipos de empresas y plazas, se trata de: la tasa bruta de creación, la tasa bruta de destrucción, la tasa de permanencia, y la tasa de creación neta.

5. Conclusiones

La presente propuesta metodológica, utiliza la conceptualización desarrollada para la producción estadística en base a sistemas de registros administrativos (Wallgren y Wallgren 2014) y desarrolla el diseño y compilación de una encuesta basada en registros; esto con el fin de dar respuesta a un planteamiento investigativo, en este caso, el análisis de creación y destrucción de plazas de trabajo formales por perfil profesional. Se ha expuesto la utilidad de este tipo de herramientas para la elaboración de estadísticas, así como para fines investigativos, sobre todo a causa de su versatilidad y capacidad de desagregación.

La encuesta aborda un universo de entidades determinado por la constitución jurídica de las empresas del registro base de negocios, así como por la metodología de depuración y vinculación con el registro base de actividades. La metodología utilizada plantea la caracterización de las plazas de trabajo ocupadas, primero, reconociendo las propiedades de las empresas o unidades institucionales a las que pertenecen: plazas del sector privado o público,

plazas de empresas pequeñas o grandes, etc.; segundo, asignando a las plazas las características de los trabajadores que las ocupan: plazas técnico-tecnológicas, plazas del área de la educación, plazas para jóvenes, etc. Esta última estrategia, posibilita una aproximación a la constitución de la fuerza laboral en las empresas, al poder reconocer, por ejemplo, las proporciones de plazas caracterizadas por requerir individuos con instrucción superior e, inclusive, con determinado nivel o área de conocimientos; y al enlazar estos resultados con las características propias de las empresas las posibilidades de análisis se amplifican; por ejemplo, relacionar la rama de actividad en la que se enmarcan las empresas con el área de conocimiento más frecuente en sus plazas.

La estructura y el conjunto de variables disponibles en la encuesta basada en registros permiten la elaboración de distintos indicadores acerca de la dinámica de creación y destrucción de empleo. Estos indicadores pueden replicarse para subconjuntos de empresas o plazas, siendo factible realizar análisis para segmentos muy desagregados.

El uso de registros administrativos para fines estadísticos es un ejercicio relativamente reciente; y uno de las principales limitaciones es la calidad de las fuentes primarias. En el presente documento también se ha expuesto cómo, en un grupo de observaciones, se detectaron particularidades que tienen su origen en la estructura institucional de las fuentes de los registros, y que no necesariamente pueden calificarse como errores. No obstante, una vez detectadas, este tipo de circunstancias pueden ser resueltas sin reducir la capacidad estadística de la herramienta.

Este ejercicio evidencia un proceso de integración de registros administrativos de las actividades y negocios que permitió identificar una aproximación de la evolución de las plazas de trabajo en la realidad ecuatoriana, actualmente el Laboratorio de Dinámica Laboral y Empresarial-LDLE continua realizando actividades de depuración y mejoramiento de las bases expuestas en el documento incorporando nuevas fuentes de información.

6. Referencias

Benítez, Diego, Boris Espinoza , Ana Grijalva, Ana Rivadeneira, y Ana María Oviedo. «Using administrative records to study employer-employee dynamics: The case of Ecuador’s Laboratory of Labor and Business Dynamics.» 2016.

García, María Isabel, Ana María Grijalva, Juan Carlos Palacios , y Roxana Mauricio . «Dinámica del empleo registrado en la seguridad social en Ecuador: 2006-2015.» *Cuaderno de trabajo INEC*, nº 2 (2016).

García, María Isabel, Natalia Garzón, Juan Carlos Palacios, y David Puebla. «Desestacionalización del empleo registrado en la Seguridad Social 2009-2014.» *Cuaderno de trabajo INEC* (INEC), 2016.

Garzón, Natalia, Matías Kulfas, Juan Carlos Palacios, y Drichelmo Tamayo. «Evolución del sector manufacturero ecuatoriano 2010-2013. Tipología estáticas y dinámicas de las manufacturas.» *Cuaderno de trabajo INEC*, nº 1 (2016).

INEC. «Ecuador en Cifras, Estadísticas Económicas.» 2016. <http://www.ecuadorencifras.gob.ec/directoriodeempresas/> (último acceso: Mayo de 2016).

INEC, CEPAL. *Panorama laboral y empresarial del Ecuador, 2009-2013*. Quito: INEC, 2015.

Wallgren, Anders, y Britt Wallgren. *Register-based Statistics. Statistical Methods for Administrative Data*. United Kingdom: John Wiley & Sons Ltd, 2014.

7. Anexos

Tabla 4. Categorías de los niveles de formación y áreas de conocimiento de las titulaciones en educación superior

Nivel de formación	
Nivel_0	No registran título de educación superior
Nivel_1	Nivel Técnico/ Tecnológico
Nivel_2	Tercer Nivel
Nivel_3	Diplomado / Especialización
Nivel_5	Cuarto Nivel
Nivel_6	PhD
Área de conocimiento	
area_1	Educación
area_2	Humanidades y Artes
area_3	Ciencias sociales, educación comercial y derecho
area_4	Ciencias
area_5	Ingeniería, industria y construcción
area_6	Agricultura
area_7	Salud y servicios sociales
area_8	Servicios
area_999	Registro título más no su área de conocimiento

*Cada entrada se corresponde con una variable de la encuesta que da cuenta del número de plazas que una empresa dispone en esa categoría (v.g. número de plazas con PhD).

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Tabla 5. Demografía general de plazas

	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014
Stock inicial de plazas	1.459.729	1.586.453	1.787.242	1.969.250	2.107.431
Stock final plazas	1.586.453	1.787.242	1.969.250	2.107.431	2.209.057
<u>Plazas que se crean</u>	267.036	332.431	350.914	352.478	428.082
En empresas que se crean	80.191	103.010	101.267	115.471	192.974
En empresas ya existentes en t	186.845	229.421	249.647	237.007	235.108
<u>Plazas que se destruyen</u>	140.312	131.642	168.906	214.297	326.456
En empresas que cierran	60.257	55.339	61.935	87.350	157.093
En empresas que siguen en t+1	80.055	76.303	106.971	126.947	169.363
<u>Plazas que se mantienen</u>	1.319.417	1.454.811	1.618.336	1.754.953	1.780.975
Suma de entran+salen+perm	1.726.765	1.918.884	2.138.156	2.321.728	2.535.513
<u>En relación al total de plazas en t y en t+1</u>					
Tasa bruta de creación de plazas	15,46%	17,32%	16,41%	15,18%	16,88%
Tasa bruta de destrucción de plazas	8,13%	6,86%	7,90%	9,23%	12,88%
Tasa de creación neta de plazas	7,34%	10,46%	8,51%	5,95%	4,01%
Plazas que permanecen	76,41%	75,82%	75,69%	75,59%	70,24%
<u>En relación al total de plazas en t</u>					
Tasa bruta de creación de plazas	18,29%	20,95%	19,63%	17,90%	20,31%
Tasa bruta de destrucción de plazas	9,61%	8,30%	9,45%	10,88%	15,49%
Tasa de creación neta de plazas	8,68%	12,66%	10,18%	7,02%	4,82%
Plazas que permanecen	90,39%	91,70%	90,55%	89,12%	84,51%

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Tabla 6. Demografía de plazas según titulación superior

	PLAZAS OCUPADAS POR PERSONAS SIN TÍTULO REGISTRADO							PLAZAS OCUPADAS PERSONAS NIVEL TÉCNICO/ TECNOLÓGICO						
	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014			2009-2010	2010-2011	2011-2012	2012-2013	2013-2014		
Stock inicial de plazas	1.071.376	1.152.396	1.303.106	1.440.011	1.533.171			57.217	62.920	67.702	69.992	70.592		
Stock final plazas	1.152.396	1.303.106	1.440.011	1.533.171	1.591.701			62.920	67.702	69.992	70.592	73.718		
Plazas que se crean	205.548	269.263	283.611	274.525	276.570			13.060	11.624	10.786	12.263	28.878		
En empresas que se crean	64.858	85.396	84.501	84.735	100.823			2.450	2.422	2.169	4.166	17.008		
En empresas ya existentes en t	140.690	183.867	199.110	189.790	175.747			10.610	9.202	8.617	8.097	11.870		
Plazas que se destruyen	124.528	118.553	146.706	181.365	218.040			7.357	6.842	8.496	11.663	25.752		
En empresas que cierran	46.232	44.631	50.521	68.718	77.554			2.159	1.609	1.665	2.320	16.656		
En empresas que siguen en t+1	78.296	73.922	96.185	112.647	140.486			5.198	5.233	6.831	9.343	9.096		
Plazas que se mantienen	946.848	1.033.843	1.156.400	1.258.646	1.315.131			49.860	56.078	59.206	58.329	44.840		
Suma de entran+salen+perm	1.276.924	1.421.659	1.586.717	1.714.536	1.809.741			70.277	74.544	78.488	82.255	99.470		
<i>En relación al total de plazas en t y en t+1</i>														
Tasa bruta de creación de plazas	16,10%	18,94%	17,87%	16,01%	15,28%			18,58%	15,59%	13,74%	14,91%	29,03%		
Tasa bruta de destrucción de plazas	9,75%	8,34%	9,25%	10,58%	12,05%			10,47%	9,18%	10,82%	14,18%	25,89%		
Tasa de creación neta de plazas	6,34%	10,60%	8,63%	5,43%	3,23%			8,12%	6,42%	2,92%	0,73%	3,14%		
Plazas que permanecen	74,15%	72,72%	72,88%	73,41%	72,67%			70,95%	75,23%	75,43%	70,91%	45,08%		
<i>En relación al total de plazas en t</i>														
Tasa bruta de creación de plazas	19,19%	23,37%	21,76%	19,06%	18,04%			22,83%	18,47%	15,93%	17,52%	40,91%		
Tasa bruta de destrucción de plazas	11,62%	10,29%	11,26%	12,59%	14,22%			12,86%	10,87%	12,55%	16,66%	36,48%		
Tasa de creación neta de plazas	7,56%	13,08%	10,51%	6,47%	3,82%			9,97%	7,60%	3,38%	0,86%	4,43%		
Plazas que permanecen	88,38%	89,71%	88,74%	87,41%	85,78%			87,14%	89,13%	87,45%	83,34%	63,52%		

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Tabla 7. Demografía de plazas según grupos de edad

	PLAZAS OCUPADAS POR MENORES DE 25 AÑOS							PLAZAS OCUPADAS POR PERSONAS ENTRE 25 Y 45 AÑOS						
	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014			2009-2010	2010-2011	2011-2012	2012-2013	2013-2014		
Stock inicial de plazas	221.339	235.231	271.406	304.155	320.786			844.702	927.006	1.046.185	1.157.773	1.245.548		
Stock final plazas	235.231	271.406	304.155	320.786	326.533			927.006	1.046.185	1.157.773	1.245.548	1.317.429		
Plazas que se crean	60.224	78.483	87.108	82.597	80.877			171.932	205.225	219.017	225.111	262.408		
En empresas que se crean	14.953	20.016	20.715	19.973	21.045			47.902	60.425	58.404	69.313	107.267		
En empresas ya existentes en t	45.271	58.467	66.393	62.624	59.832			124.030	144.800	160.613	155.798	155.141		
Plazas que se destruyen	46.332	42.308	54.359	65.966	75.130			89.628	86.046	107.429	137.336	190.527		
En empresas que cierran	8.976	8.399	11.034	14.550	15.536			35.648	32.820	36.131	52.286	82.040		
En empresas que siguen en t+1	37.356	33.909	43.325	51.416	59.594			53.980	53.226	71.298	85.050	108.487		
Plazas que se mantienen	175.007	192.923	217.047	238.189	245.656			755.074	840.960	938.756	1.020.437	1.055.021		
Suma de entran+salen+perm	281.563	313.714	358.514	386.752	401.663			1.016.634	1.132.231	1.265.202	1.382.884	1.507.956		
<i>En relación al total de plazas en t y en t+1</i>														
Tasa bruta de creación de plazas	21,39%	25,02%	24,30%	21,36%	20,14%			16,91%	18,13%	17,31%	16,28%	17,40%		
Tasa bruta de destrucción de plazas	16,46%	13,49%	15,16%	17,06%	18,70%			8,82%	7,60%	8,49%	9,93%	12,63%		
Tasa de creación neta de plazas	4,93%	11,53%	9,13%	4,30%	1,43%			8,10%	10,53%	8,82%	6,35%	4,77%		
Plazas que permanecen	62,16%	61,50%	60,54%	61,59%	61,16%			74,27%	74,27%	74,20%	73,79%	69,96%		
<i>En relación al total de plazas en t</i>														
Tasa bruta de creación de plazas	27,21%	33,36%	32,10%	27,16%	25,21%			20,35%	22,14%	20,93%	19,44%	21,07%		
Tasa bruta de destrucción de plazas	20,93%	17,99%	20,03%	21,69%	23,42%			10,61%	9,28%	10,27%	11,86%	15,30%		
Tasa de creación neta de plazas	6,28%	15,38%	12,07%	5,47%	1,79%			9,74%	12,86%	10,67%	7,58%	5,77%		
Plazas que permanecen	79,07%	82,01%	79,97%	78,31%	76,58%			89,39%	90,72%	89,73%	88,14%	84,70%		

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Tabla 8. Demografía de plazas según sector

	PLAZAS EN EL SECTOR PRIVADO						PLAZAS EN EL SECTOR PÚBLICO					
	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014		2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	
Stock inicial de plazas	1.052.567	1.132.391	1.292.787	1.439.908	1.531.855		407.162	454.062	494.455	529.342	575.576	
Stock final plazas	1.132.391	1.292.787	1.439.908	1.531.855	1.615.024		454.062	494.455	529.342	575.576	594.033	
Plazas que se crean	206.171	273.029	286.799	276.170	270.065		60.865	59.402	64.115	76.308	158.017	
En empresas que se crean	67.141	91.983	93.257	89.615	96.677		13.050	11.027	8.010	25.856	96.297	
En empresas ya existentes en t	139.030	181.046	193.542	186.555	173.388		47.815	48.375	56.105	50.452	61.720	
Plazas que se destruyen	126.347	112.633	139.678	184.223	186.896		13.965	19.009	29.228	30.074	139.560	
En empresas que cierran	55.210	45.341	50.489	80.697	61.897		5.047	9.998	11.446	6.653	95.196	
En empresas que siguen en t+1	71.137	67.292	89.189	103.526	124.999		8.918	9.011	17.782	23.421	44.364	
Plazas que se mantienen	926.220	1.019.758	1.153.109	1.255.685	1.344.959		393.197	435.053	465.227	499.268	436.016	
Suma de entran+salen+perm	1.258.738	1.405.420	1.579.586	1.716.078	1.801.920		468.027	513.464	558.570	605.650	733.593	
<i>En relación al total de plazas en t y en t+1</i>												
Tasa bruta de creación de plazas	16%	19%	18%	16%	15%		13%	12%	11%	13%	22%	
Tasa bruta de destrucción de plazas	10%	8%	9%	11%	10%		3%	4%	5%	5%	19%	
Tasa de creación neta de plazas	6%	11%	9%	5%	5%		10%	8%	6%	8%	3%	
Plazas que permanecen	74%	73%	73%	73%	75%		84%	85%	83%	82%	59%	
<i>En relación al total de plazas en t</i>												
Tasa bruta de creación de plazas	20%	24%	22%	19%	18%		15%	13%	13%	14%	27%	
Tasa bruta de destrucción de plazas	12%	10%	11%	13%	12%		3%	4%	6%	6%	24%	
Tasa de creación neta de plazas	8%	14%	11%	6%	5%		12%	9%	7%	9%	3%	
Plazas que permanecen	88%	90%	89%	87%	88%		97%	96%	94%	94%	76%	

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

Tabla 9. Demografía de plazas según rama de actividad

	PLAZAS EN MANUFACTURA						PLAZAS EN COMERCIO					
	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014		2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	
Stock inicial de plazas	220.961	245.550	270.867	293.313	296.962		248.839	275.423	317.954	352.733	371.921	
Stock final plazas	245.550	270.867	293.313	296.962	300.456		275.423	317.954	352.733	371.921	403.157	
Plazas que se crean	38.992	39.822	39.426	34.103	27.524		52.395	68.807	66.469	58.534	69.107	
En empresas que se crean	6.572	7.953	10.288	8.840	6.672		19.991	25.719	24.116	19.087	30.629	
En empresas ya existentes en t	32.420	31.869	29.138	25.263	20.852		32.404	43.088	42.353	39.447	38.478	
Plazas que se destruyen	14.403	14.505	16.980	30.454	24.030		25.811	26.276	31.690	39.346	37.871	
En empresas que cierran	3.915	3.201	5.621	11.612	6.003		11.655	12.030	11.849	18.385	13.959	
En empresas que siguen en t+1	10.488	11.304	11.359	18.842	18.027		14.156	14.246	19.841	20.961	23.912	
Plazas que se mantienen	206.558	231.045	253.887	262.859	272.932		223.028	249.147	286.264	313.387	334.050	
Suma de entran+salen+perm	259.953	285.372	310.293	327.416	324.486		301.234	344.230	384.423	411.267	441.028	
<i>En relación al total de plazas en t y en t+1</i>												
Tasa bruta de creación de plazas	15,00%	13,95%	12,71%	10,42%	8,48%		17,39%	19,99%	17,29%	14,23%	15,67%	
Tasa bruta de destrucción de plazas	5,54%	5,08%	5,47%	9,30%	7,41%		8,57%	7,63%	8,24%	9,57%	8,59%	
Tasa de creación neta de plazas	9,46%	8,87%	7,23%	1,11%	1,08%		8,83%	12,36%	9,05%	4,67%	7,08%	
Plazas que permanecen												
<i>En relación al total de plazas en t</i>												
Tasa bruta de creación de plazas	17,65%	16,22%	14,56%	11,63%	9,27%		21,06%	24,98%	20,91%	16,59%	18,58%	
Tasa bruta de destrucción de plazas	6,52%	5,91%	6,27%	10,38%	8,09%		10,37%	9,54%	9,97%	11,15%	10,18%	
Tasa de creación neta de plazas	11,13%	10,31%	8,29%	1,24%	1,18%		10,68%	15,44%	10,94%	5,44%	8,40%	
Plazas que permanecen	93,48%	94,09%	93,73%	89,62%	91,91%		89,63%	90,46%	90,03%	88,85%	89,82%	

Fuente: Laboratorio de Dinámica Laboral y Empresarial (LDLE)

4

Propuesta de Modelo Conceptual para
identificar patrones de
hechos estadísticos desde texto libre

Juan Carlos Delgado Loyola

RESUMEN

El presente documento es una propuesta de modelo conceptual para la extracción y reconocimiento de textos escritos en lenguaje natural, mediante herramientas de minería de textos automatizadas, cuyo objetivo es revelar información estructurada en términos aceptados internacionalmente, sobre algún aspecto de la realidad social y económica de un país, descrito en palabras como un hecho o fenómeno de relevancia estadística. El estudio se enfoca en realizar una descomposición sistemática del texto informativo de actividades económicas de empresas, a partir de narraciones de respuesta a la pregunta abierta de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), que indaga sobre esta temática, y cuyo contenido requiere ser comparado con la terminología del clasificador internacional. El objetivo es establecer los componentes del modelo conceptual, que permitan mejorar la comprensión del texto y evaluar así, su capacidad explicativa de la realidad. Los resultados muestran algún grado de validez de dichos componentes para explicar la correspondencia de hechos entre la respuesta del entrevistado y los patrones de texto codificados en base a la descripción formal de actividades económicas. Sin embargo, para generalizar el modelo a otros hechos estadísticos, hace falta realizar más experimentos que pongan a prueba su eficacia en la representación lingüística estructural de textos y la búsqueda de significado oculto en los datos extraídos de las encuestas.

Palabras Clave: reconocimiento de patrones de texto, minería de textos, procesamiento de lenguaje natural, modelos léxicos.

Descargo de responsabilidad: Las opiniones e interpretaciones expresadas en este documento pertenecen a los autores y no reflejan el punto de vista oficial del Instituto Nacional de Estadística y Censos (INEC). El INEC no garantiza la exactitud de los datos que figuran en el documento

1. Introducción

Toda investigación necesaria para realizar una innovación que pueda aplicarse al reconocimiento de patrones de texto aplicados a hechos estadísticos requiere de métodos de reconocimiento del lenguaje natural y sus técnicas (Pelayo, Moreno, Fraga, Moreno, & Corredor, 2012).

Es de conocimiento público que los hechos estadísticos, para ser difundidos, requieren de un proceso que interprete el texto del que están hechas las respuestas a preguntas abiertas de las encuestas y las conviertan a datos estructurados, esto es, a representaciones formales del lenguaje natural, llámese estos modelos conceptuales, plantillas o esquemas para extraer información y organizarla en una jerarquía de términos relacionados y universalmente aceptados al estar contenidos en las clasificaciones internacionales, tal es el caso de la Clasificación Internacional Industrial Uniforme (CIIU4), aquella que ha motivado este análisis. Es a través de los términos extraídos de encuestas como la ENEMDU de Ecuador, con los cuales se codifican muchas de las variables categóricas, como por ejemplo: la rama de actividad económica, los bienes y servicios producidos, el tipo de insumo utilizado o consumido, la técnica de producción empleada y la forma en que se realiza la producción (INEC, 2010).

De allí la importancia de innovar en el campo del reconocimiento y organización formal de conceptos extraídos a partir de hechos estadísticos, para los cuales, cabe indicar aquí, se ha tomado como unidad de estudio las estructurales morfológicas y sintácticas que conforman el texto descriptivo de las actividades económicas en la pregunta de la ENEMDU: ¿A qué se dedica principalmente la empresa o negocio donde trabaja(ba)?

El grado de precisión en la interpretación y categorización de dichos textos, presentes en todo tipo de encuestas es fundamental para asegurar la calidad de la información oficial que se publica en los medios oficiales. El proceso de mejoramiento continuo requiere de alternativas metodológicas innovadoras en el campo de la tecnología, en áreas tales como: la lingüística computacional, la lexicografía y la inteligencia artificial para el aprendizaje y reconocimiento de patrones que se repiten en el texto. La gran interrogante que surge aquí es: ¿con qué grado de precisión y certeza se puede, a través

de aplicaciones informáticas, reconocer, identificar, clasificar, categorizar y codificar de forma automática, aquellos hechos del dominio estadístico?. Más aún, si las descripciones textuales recogidas en campo, deben ser traducidas y enlazadas hacia entradas de una clasificación formal aceptada internacionalmente, como paso obligado para la instrumentalización de variables y la construcción de indicadores estadísticos.

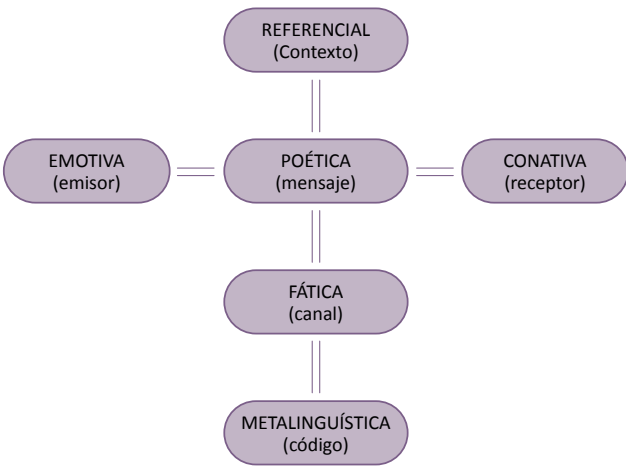
Para cumplir con este objetivo el presente documento se estructura así: la sección 2, describe el estado del arte de las técnicas de reconocimiento de texto enfocado a la construcción de diccionarios o léxicos para identificar patrones de hechos y explica la construcción del modelo conceptual extendiendo su utilidad al reconocimiento de patrones de texto de las actividades económicas. Los primeros resultados de aplicar el modelo se presentan en la sección 3, logrados luego de aplicar el modelo por medio de un algoritmo computacional de búsqueda de patrones, el cual es parte de la innovación de este estudio. La sección 4 presenta la evaluación del modelo conceptual y finalmente la sección 5 las conclusiones.

2. Marco conceptual

2.1. Los modelos de comunicación y la identificación de patrones de hechos

El punto de partida para la innovación en la identificación de patrones de hechos desde texto escrito en lenguaje natural, se centra en la función referencial del modelo de comunicación de Jakobson (Jakobson, 1988). Dicha función tiene como propósito el descomponer el mensaje en estructuras de “contexto”, es decir, en aspectos lingüísticos organizados en términos comunes, de modo que simplifiquen el proceso de codificación y emparejamiento con otras estructuras similares de texto (Figura 1).

Figura 1. El modelo de Jakobson y sus funciones en la comunicación.



Fuente: (Jakobson, 1988)

Además, las aportaciones de Lasswel, en una derivación del modelo de Jakobson, facilitan la extracción de información a través de preguntas declaradas por el emisor del mensaje, tales como: ¿quién?, ¿dice qué?, ¿en qué lugar?, ¿a quién? y ¿con qué efectos lo hace?, lo cual es útil para interpretar información incompleta, como la de las narraciones hechas por los encuestados al describir actividades económicas de las empresas. En este sentido, se ha considerado los avances del modelo de Shannon y Weaver, ya que califica a los significados inconexos del texto como fuente primaria de ruido entre la señal emitida y la señal recibida. Por último, la ejemplificación del modelo de Tuba de Schramm, descompone los actores del mensaje en: decodificador, intérprete y codificador, justificando así la necesidad de establecer, mediante análisis inductivo, un nexo lingüístico fuerte entre el emisor y el receptor (Alsina Rodrigo, 2003).

De todas las técnicas, la que se seleccionó aquí para al propósito experimental de la propuesta, fue la función referencial del modelo original de Jakobson, ya que ésta deja de lado los elementos de ruido relacionados con la información de contexto. Además, existen experiencias documentadas y patentadas sobre este tipo de extracción de conocimiento, como la de la patente de Wasson, la cual descompone un texto en piezas de información, que son luego utilizadas en la búsqueda de patrones de hechos (Wasson & Wiltshire JR, 2005). El objetivo de esta estructuración es reducir el mensaje a unos pocos elementos suficientes para extraerle significado a la comunicación sobre un determinado hecho.

2.2. Construcción del Modelo Conceptual

Para diseñar el modelo conceptual, en vista de que los contenidos de la información son expresiones del lenguaje natural, ha sido preciso incorporar diccionarios léxicos computacionales, cuyo contenido consiste en información adicional de contexto, escrita con el propósito de adquirir una comprensión profunda del mensaje. Hay que señalar aquí, que las palabras y frases escritas o habladas, al especializarse en un tema común, para una cierta lengua, en una localidad o tiempo particular, necesitan de estos diccionarios léxicos, los cuales, al ser utilizados como instrumentos computacionales para representación del conocimiento, explican de mejor manera los términos complejos del mensaje, esto es, las frases o sintagmas nominales, proposicionales o verbales presentes en sus estructuras sintácticas y semánticas. Se quiere comprobar por medio de este modelo, que a través de estos léxicos, se puede extraer significado de contexto en forma de patrones para categorizar descripciones de hechos estadísticos, como el de las actividades económicas de las empresas (Vazquez & Fernández, 2002).

Los pasos metodológicos para la construcción del modelo se indican en la Tabla 1. Para construir el modelo, se necesitó partir del análisis exploratorio del texto de los casos de estudio de las actividades económicas codificadas por expertos de la productora estadística, a partir del texto de la respuesta en la pregunta de encuesta de empleo relacionada con estas actividades, y recopilarlos en un “corpus lingüístico”.¹

Tabla 1. Pasos de Construcción del Modelo

Construcción del Modelo	
Paso	Descripción
1	Análisis exploratorio de texto
2	Especificación conceptual
3	Especificación formal
4	Estimación experimental
5	Ajuste y Modificación

Elaboración: Propia

1 (Gries, 2014), corpus lingüístico: representación esquemática de términos de mayor a menor frecuencia para un determinado contexto.

Este material debió ser ampliamente documentado en ejemplos para señalar palabras o frases más frecuentes, mediante etiquetas o marcas de texto, mismas que fueron descritas como metadatos. De esta manera, se pudo representar en patrones los significados de cada frase o palabra de contexto para dichas actividades económicas. Para reconocer estos patrones en nuevos casos de ejemplo, se utilizaron técnicas de procesamiento del lenguaje natural, debidamente automatizadas en un algoritmo computacional de búsqueda, adecuado al contexto del mensaje. Las marcas de texto en los metadatos que mejor explicaban el significado de parte o todo el texto analizado, eran aquellas relacionadas con el propósito de describir actividades, productos y lugares. Un detalle de estos resultados puede verse en las tablas 2 y 3, para un conjunto de casos con 283,993 ocurrencias de términos.

Tabla 2. Análisis Exploratorio- Entidades

Términos de entidad extraídos del corpus de texto	
Marca	Descripción de Entidad en el Texto
A	ACTIVIDAD
L	LUGAR
P	PRODUCTO
R	REFERENCIA: A QUIEN, PARA QUÉ, ETC
E	EMPRESA O ESTABLECIMIENTO

Elaboración: Propia

Tabla 3. Análisis Exploratorio – Actividades y productos

Términos de actividad y producto extraídos del corpus		Casos más frecuentes
1	ARTICULOS	92,254
2	COMIDA	43,917
3	MANTENIMIENTO	25,990
4	COMUNICACIÓN	18,695
5	MEDICO	14,957
6	PRENDAS DE VESTIR	10,523
7	BELLEZA	9,698
8	ENSEÑANZA	7,452
9	PRODUCTOS DE PANADERIA	7,172
10	MEDICINAS	5,659
11	ESPIRITUAL	5,031
12	BEBIDAS	3,493
13	ALIMENTOS	3,393

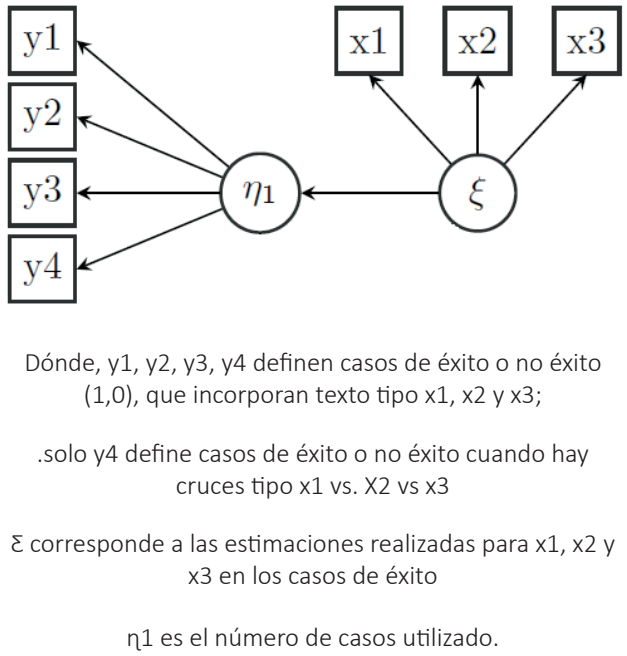
14	JURIDICO	3,360
15	CALZADO	3,174
16	VIVERES	2,345
17	MADERA	2,179
18	PUBLICICO	1,933
19	PRODUCTOS	1,876
20	VEHICULOS	1,816
21	ENTRETENIMIENTO	1,776
Subtotal		266,693
Otros términos		17,300
Total de casos		283,993

Elaboración: Propia

Para incorporar dichas consideraciones de marcado a los datos del modelo, fueron útiles los estudios de Ramchand y Folli, (Folli & Ramchand, 2001), y los conceptos lingüísticos tratados en Panagiotidis (Panagiotidis, 2001, págs. 161-170). Además de estas técnicas, una tarea imprescindible para construir el modelo fue la de utilizar lematizadores, esto es, descripciones abreviadas de raíces con las que se forman los términos del texto (Santana, Perez, Carreras, & Rodriguez, 1997), (Ashok & Kannathan, 2011).

Para generalizar la estructura de los componentes del modelo para la identificación de hechos estadísticos descriptivos sobre un determinado contexto, se establecieron las siguientes tres fases de descomposición del texto: a) La fase 1 consistió en la descomposición en elementos morfológicos y sintácticos del texto, tal como la recomendada en extracción de hechos de la patente de Wasson (Wasson & Wiltshire JR, 2005). b) La fase 2, es una innovación hecha en INEC y consistió en reconocer a estos elementos dentro de un conjunto de entidades de contexto, para el caso de la encuesta de empleo, estos elementos fueron: la actividad, el producto, el lugar y alguna que otra referencia sobre los anteriores . c) El tercer paso consistió en agrupar estos elementos en tres grandes componentes básicos, acordes con los diccionarios léxicos elaborados: 1) determinadores; 2) complementadores y 3) omisores. La Figura 2 muestra una primera aproximación a la especificación formal del modelo siguiendo estas fases.

Figura 2. Una aproximación a la especificación formal del modelo.



Elaboración: Propia

Los determinadores (X1).- Se han agrupado en la variable de número de casos X1.Son entidades del tipo verbal que expresan el significado de acción, esto es: actividad económica (A); seguido de los de tipo nominal que expresan el significado de producto (P); lugar de trabajo (L).

Los complementadores (X2).- Se han agrupado en la variable de número de casos X2. Son construcciones semánticas con entidades de tipo referencial (R) que contienen sintagmas verbales, nominales o proposicionales que expresan algún significado para complementar a la descripción de los determinadores o actividades económicas.

Los omisores (X3).- Se han agrupado en la variable de número de casos X3. Son la parte del texto descriptivo que no se escribió, es decir, los detalles que le faltó al emisor, en este caso al encuestador, para describir el hecho estadístico, es decir, la actividad económica. Consisten en información tipo (A), (L) o (P) no presente en el texto original, pero que se pueden decifrar con cierta probabilidad de éxito de acuerdo al significado oculto en las descripciones de los otros componentes. Para ilustrar la construcción del modelo conceptual se detalla en el Anexo 1 un caso experimental.

2.3. Trabajos Relacionados

El proceso de reconocimiento de hechos a partir de texto escrito en lenguaje natural, requiere de una evaluación de calidad que responda a la pregunta sobre qué tan válidas y efectivas pueden ser la categorías conceptuales (pre-definidas) en la interpretación a través de diccionarios léxicos, sin que estas categorias violen el significado implícito en la estructura lingüística original y el contexto en el cual ocurrió el hecho. Una correcta interpretación del texto es crucial en casi todas las aplicaciones de reconocimiento de patrones. Una información útil sobre las bases metodológicas relativas al aseguramiento de la calidad del proceso la muestra el proyecto de análisis de contenido de textos basado en corpus para artículos que cubren noticias (Kutter & Kantner, 2012, pp 6-12)

La explotación de las capacidades de análisis de significado en los textos, a través de estructuras de datos codificadas en patrones, en escenarios similares al de esta propuesta, se pueden encontrar en el Proyecto AKTeur, (Goldhammer, 2015, pp 2-4), para el cual se crearon dos escenarios similares: a) Codificación automática para categorizar medidas de calidad en la educación respecto al aprendizaje de la lectura y b) Codificación automática para diagnósticos psicológicos de los logros alcanzados en educación.

En otros escenarios de experimentación sobre contenido de los textos en el campo de la salud (Shu, 2005, pp 3-4), se ha propuesto un método de codificación automática para monitorear pacientes, el cual consiste en extraer información informal (no estructurada) de recetas médicas y preparar un vocabulario estandarizado que pueda ser entendido por el programa informático, para cotejarlo con información formal y normalizada (estructurada) de los léxicos médicos. Otros proyectos similares que manejan grandes cantidades de datos médicos, utilizan más de un sistema de codificación automático, con el de propósito de asignar con la máxima precision posibles los códigos correspondientes en la clasificación internacional de enfermedades. Un ejemplo clásico de esto es el Proyecto de Codificación Automática de la División de Oncología de la Universidad de Pennsylvania (Crammer, Dredze, Ganchev, & Ptratim Talukdar, 2009, pp 1-8).

El reconocimiento de patrones y la codificación automática, en la mayoría de los casos requieren de tareas complementarias de preprocesamiento

lingüístico, tales como el del análisis morfo-sintáctico (parsing) y la corrección ortográfica. Otros, a su vez, para reducir la complejidad del análisis semántico, requieren de tareas de categorización, reducción y simplificación de términos (lematización). Más aún, en áreas del conocimiento que aglutinan taxonomías extensas, se requiere de la elaboración de más de un thesaurus, o diccionario léxico, que incluya variantes, sinónimos, palabras y frases en varios contextos. Para casos especiales de ambigüedad máxima se incorporan estrategias informales del tipo prueba-error con la supervisión de sistemas de aprendizaje del corpus lingüístico.

3. Principales resultados

3.1. Muestras Experimentales de Datos

Para construir el modelo e identificar patrones de texto basados en hechos sobre las descripciones de la actividad económica de las empresas donde trabajaron los encuestados, se utilizó como universo a un conjunto de alrededor de 90,000 casos experimentales con descripciones codificadas a cuatro dígitos según la clasificación CIU4, para los últimos tres años anteriores a 2017. De este conjunto se tomaron al azar alrededor de 9,000 descripciones cuyo contenido incluía las frases y palabras más frecuentes para referirse a actividades económicas y a sus productos e insumos. Del análisis de estas descripciones se obtuvo los instrumentos para generalizar y construir el modelo. Su resultado se resume en la Tabla 4.

Tabla 4. Corpus de hechos con texto de respuestas a la pregunta sobre actividades de las empresas.

TIPO	QUE SE DESCRIBE EN TEXTO	CASOS	% MUESTRA	% POBLACIÓN
AA	Actividad solamente	313	3,46	0,34
EA	Empresa y Actividad	1877	20,73	2,04
EE	Empresa solamente	3901	43,08	4,25
SE	Servicio y Empresa	2916	32,20	3,17

SS	Servicio Solamente	48	0,53	0,05
	TOTAL MUESTRA:	9055	100,00	9,86
	TOTAL POBLACIÓN:	91861		100,00

Fuente: ENEMDU, INEC
Elaboración: Propia

Para identificar la validez de los parámetros del modelo, se seleccionó una nueva muestra nueva de 6485 casos de texto libre, escritos de la misma forma que en el corpus, para responder a la pregunta sobre actividades de empresas en el cuestionario de empleo para el mes de mayo de 2017. Los datos fueron seleccionados al azar en un inicio, verificando que la distribución de casos sea la misma que en el corpus, aunque luego se tomaron solo aquellos casos que incluían el servicio o actividad e información de ésta en cuanto al producto.

3.2. Algoritmo de identificación de patrones de hechos para evaluar el modelo.

Para evaluar la efectividad del modelo, se necesitó construir un algoritmo computacional que permitiese poner a prueba su efectividad para identificar hechos y reconocer patrones en ellos. El algoritmo de búsqueda experimental, fue diseñado e implementado en lenguaje ++, por la utilidad que presa este lenguaje de propósito general para el desarrollo técnicas de procesamiento de lenguaje natural. Además, se utilizó como insumo la Base de Datos Lexicográfica (BDDL), construida con los registros del corpus lingüístico de ejemplos de codificación manual y organizada en tablas de términos que se agrupan y clasifican en actividades, productos, lugares y referencias complementarias, obtenidas del texto de las respuestas a la pregunta de empleo sobre la actividad económica de la empresa en la que trabajaba el encuestado.

Los pasos del algoritmo se muestran en la Figura 3, y son los siguientes:

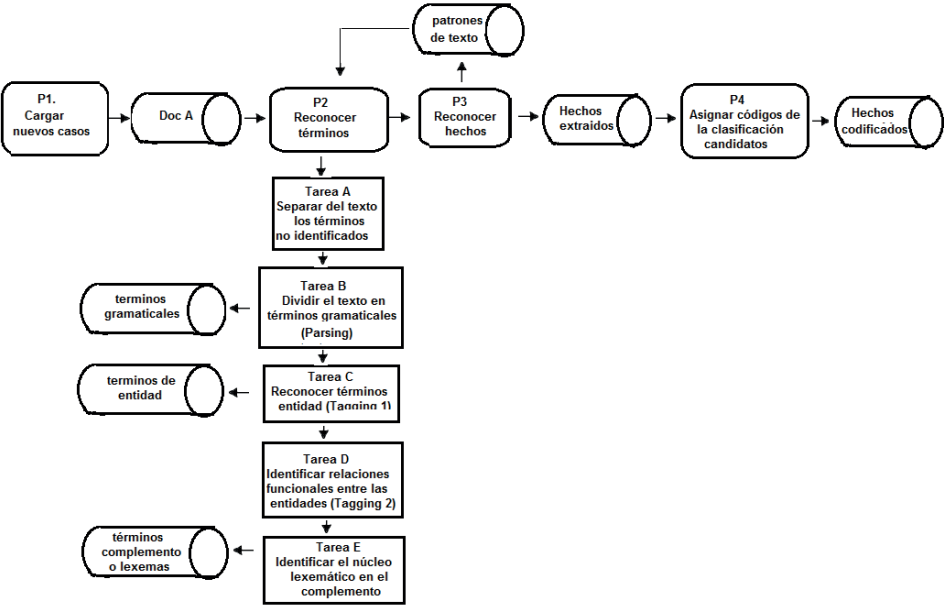
P1. Cargar nuevos casos.- Consistió en separar, corregir, depurar y reducir a su mínima expresión, lematizando los términos en el texto de los nuevos

casos de codificación de modo que al agruparlos se puedan comparar con los patrones léxicos de la base de datos lexicográfica.

Figura 3. Algoritmo para identificar patrones de hechos.

P2. Reconocer los términos.- Consistió en reconocer y catalogar los términos núcleo de los sintagmas presentes en los textos de nuevos casos de búsqueda en la BDDL. Para este paso se diseñaron y programaron tareas automáticas siguientes:

- a) Separar términos no identificados, es decir, no encontrados en la BDDL.
- b) Dividir el texto en términos de sintagmas verbales, nominales, proposicionales. A esta tarea se la nombró como “Parsing”, por su similitud en concepto con una de las tareas del procesamiento de lenguaje natural.
- c) Buscar e identificar términos entidad por la técnica de reconocimiento de entidades nombradas (RNE)², dichos términos se categorizaron en actividades (A), productos (P) y lugares de trabajo (L). A esta tarea se la dio el nombre de “Tagging 1”, por su naturaleza similar a una de las tareas de la técnica RNE.
- d) Buscar e identificar relaciones funcionales entre entidades. Sus resultados equivalen uno a uno con los patrones de texto de la BDDL. A esta tarea se la dio el nombre de “Tagging 2”
- e) Identificar el término principal o núcleo lexemático (argumento clave de búsqueda en patrones de la base datos BDDL).



Elaboración: Propia en base a diagramas de Wasson. (Wasson & Wiltshire JR, 2005)

P3. Reconocer hechos en los nuevos casos.- Este paso consiste en una única actividad, la cual realiza el emparejamiento aproximado entre los términos del nuevo caso, una vez leído, separado y esquematizado en entidades nombradas, versus los patrones léxicos de la base de datos BDDL. A esta actividad se la nombró como “Patting”, por su característica de reconocimiento y emparejamiento de patrones entre dos entidades nombradas.

P4. Asignar códigos candidatos a los nuevos casos.- En este paso, el algoritmo finaliza las tareas secuenciales de búsqueda, con la asignación automática de códigos candidatos en los patrones léxicos identificados en la BDDL para los textos de los casos de lexemas en las respuestas a las preguntas abiertas. Este paso es una tarea aparte del algoritmo pero no la menos importante, ya que permitirá clasificar al hecho estadístico según alguna entrada en la clasificación internacional, siendo para el caso de estudio, el código CIU4 a cuatro dígitos.

Un último paso, para adaptar el algoritmo a la encuesta de empleo, consistiría en una tarea asistida, en la cual debe seleccionarse el código CIU4 más adecuado de acuerdo al texto y a los patrones reconocidos en la BDDL en los pasos anteriores. Este paso fue necesario por cuanto hay casos en los cuales se da ambigüedad de términos para describir los códigos CIU4, por

² MUC-7 Named Entity Task Definition; también conocido en español como reconocimiento de entidades nombradas (RNE) consiste en tareas básicas para los sistemas de procesamiento de textos y técnicas de emparejamiento de patrones.

aparición de más de un patrón léxico reconocido por el sistema para el mismo caso. Los casos no resueltos con una valoración muy baja en similitud con algún patrón léxico de la BDDL, serían motivo de análisis posterior a este experimento.

Los resultados de la ejecución del algoritmo, corresponden a 2282 casos que pudieron ser identificados como actividades económicas presentes en la codificación CIU4. La distribución de estos casos se muestra como sigue a) 671 falsos negativos, correspondiente al 29,40 % para los cuales el algoritmo identificó como válidos los códigos de actividades económicas CIU4 asignados, sin embargo, éstos últimos no corresponden a los códigos manualmente asignados por los codificadores de la encuesta; b) 1211 casos de falsos positivos, para los cuales el algoritmo no pudo identificar patrones correspondientes a un código válido CIU4 y que representa el 53,03%; c) 401 casos en los cuales coincide la identificación de actividades

económicas del algoritmo basado en el modelo conceptual con las mismas actividades asignadas por el codificador y que representa el 17,57%.

En síntesis, si se considera como válida la identificación de actividades económicas realizada por los codificadores, se tendría apenas el 17,57% de casos válidos, sin embargo, si se adiciona el 29,40% de falsos negativos, se tendría un porcentaje del 46,97% de casos válidos de identificación según el modelo conceptual y su algoritmo de identificación de hechos estadísticos, donde los hechos son las actividades económicas. Más adelante se analizan los factores que hicieron que no exista una correspondencia entre los códigos CIU4 identificados como válidos por los encuestadores frente a los códigos que se identificaron únicamente mediante el algoritmo.

La precisión del algoritmo en la identificación del 46,97% de casos válidos se la obtuvo mediante la ecuación Factor F1 se obtuvo mediante la ecuación 1:

$$F1 = \frac{(2 * P * R)}{P + R}$$

, Dónde P= precisión y R = Relevancia

Ecuación 1

$$precisión = \frac{correctos}{correctos + falsos\ positivos}$$

$$relevancia = \frac{correctos}{correctos + falsos\ negativos}$$

La ecuación F1, es una de las métricas generalizadas que utilizan los autores de investigación y empresas innovadoras en varios campos de la información para evaluar el rendimiento de sus sistemas y que son publicadas por la Corporación Internacional de Aplicaciones Científicas, MUC-3.

Los falsos positivos (FP) corresponden a casos fallidos del algoritmo, y los falsos negativos (FN) son casos aparentemente fallidos del algoritmo, porque los codificadores a más del texto de la respuesta a la pregunta relacionada con la actividad económica, toman decisiones en base a otras variables no presentes aquí, siguiendo otras consideraciones de codificación CIU4 aplicadas únicamente a la encuesta ENEMDU. Un ejemplo de ejecución del algoritmo presentado se encuentra en la Figura 2 del Anexo 1.

4. Evaluación del Modelo Conceptual

Para evaluar el modelo conceptual propuesto, se utilizó una representación generalizada expuesta en la Ecuación 2. La hipótesis nula Ho y la hipótesis alternativa H1, pretenden establecer con un nivel de significación 0,05 < p <= 0,05, el papel preponderante de los componentes del modelo de predicción y categorización del hecho estadístico identificado por la variable independiente *Yi*, y que para este caso práctico se halla representado por la codificación a cuatro dígitos de la clasificación CIU4 de actividades económicas.

$$Yi = \frac{f(v1 * x1 + v2 * x2 + v3 * x3 + \dots .vn * xn)}{n}$$

Ecuación 2

Donde,

x1 = Número de determinadores del caso i

x2 = Número de complementadores del caso i

x3 = Número de omisores del caso i

v1 = Factor de aporte de x1 en el modelo

v2 = Factor de aporte de x2 en el modelo

v3 = Factor de aporte de x3 en el modelo

Mediante un análisis exploratorio de datos a partir de obtención de los valores distribuidos de las variables dependientes x1, x2 y x3 en el modelo en los datos de la muestra, de los 6485 nuevos casos de descripciones de actividades económicas en la encuesta ENEMDU de mayo 2017, se obtuvieron 2282 casos que pudieron ser reconocidos como actividades económicas, los

cuales representan el 35% frente a un 65% en los 4203 casos que no pudieron ser reconocidos como actividades económicas utilizando los patrones de texto con los cuales se estableció la comparación con casos de ejemplo.

Para la evaluación de los factores del modelo se establecieron tres hipótesis, a saber:

Ho: 0 < v1 <= 1 los determinadores x1 son necesarios

H1: 0 = v1 los determinadores x1 no son necesarios

Ecuación 3

Ecuación 4

Ho: 0 < v2 <= 1 los complementadores x2 son necesarios

H1: 0 = v2 los complementadores x2 no son necesarios

Ecuación 5

Ecuación 6

Ho: 0 < v3 <= 1 los omisores x3 son necesarios

H1: 0 = v3 los omisores x3 no son necesarios

Ecuación 7

Ecuación 8

$$v1 = x1 / (x1 + x2 + x3)$$

Ecuación 9

$$v2 = x2 / (x1 + x2 + x3)$$

Ecuación 10

$$v3 = x3 / (x1 + x2 + x3)$$

Ecuación 11

Para v1 se obtuvo: 0 <= 0,574 <= 1	Se acepta Ho
Para v2 se obtuvo: 0 <= 0,084 <= 1	Se acepta Ho
Para v3 se obtuvo: 0 <= 0,341 <= 1	Se acepta Ho

Para garantizar la generalización de los resultados de la investigación de modo que permita evaluar la interpretación de cualquier texto sobre hechos estadísticos, los reconocidos como actividades económicas, aplicando el modelo conceptual

propuesto, se han clasificado en tres tipos: determinadores (x1), complementadores (x2), y omisores (x3). Los primeros (x1) engloban contenidos implícitos en las actividades económicas, productos y lugares de trabajo. Los de tipo (x2) agrupan la parte del

texto que describe ciertas referencias narradas para complementar la idea de descripción de la pregunta, tales como: el por quien y para quién se elaboró el producto y el cómo se realizó la actividad económica. La última (x3) son los textos que no constan en la descripción y se refieren a la actividad, producto o lugar de trabajo omitidos, sobreentendidos o no

explicados en el texto.

La ecuación del modelo aplicado al reconocimiento de actividades económicas CIU4, consiste en tres componentes básicos: x1= determinadores; x2=complementadores; x3 = omisores, indicados en la Ecuación 12.

$$Y_i = \frac{f(0,574 * x_1 + 0,084 * x_2 + 0,341 * x_3)}{3}, \text{ siendo}$$

Ecuación 12

- .x1 = determinadores: Actividades (A), Lugares (L), Productos (P)
- .x2 = complementadores: referencias (R) al quién, cómo y para qué.
- .x3 = omisores: una o más actividades, lugares o productos omitidos en el texto.

Y_i, la variable independiente aquí, es de naturaleza dicotómica (1,0), cuyos valores no tienen máximos ni mínimos más que el de sus extremos 1 o 0. Representa el grado de capacidad del modelo para predecir un código según la conformación de los componentes determinadores, omisores y complementadores.

5. Conclusiones

El modelo conceptual de identificación de hechos estadísticos, tomando como caso de estudio a las actividades económicas codificadas a cuatro dígitos en la clasificación internacional CIU4, ha permitido dar una distribución ideal de parámetros de clasificación a los hechos, compuestos por tres tipos generalizados de relaciones entre entidades a) los determinadores, los cuales consisten en actividades (A), lugares de trabajo (L), y productos (P). b) los complementadores compuestos por referencias (R) a cualquier texto explicativo sobre el quién y a quién se dirige el producto, a más del cómo y para qué se realiza la actividad económica; y c) los omisores o casos de identificación de hechos del modelo, es decir, las actividades económicas que omiten uno o más de los determinadores A, L o P.

El análisis del factor de exactitud del algoritmo (F1) permite valorar, según los resultados obtenidos aquí, en un 30% la exactitud en la identificación de actividades económicas mediante reconocimiento de patrones de texto. Sin embargo, se debe tomar en cuenta que los patrones de texto de los casos prácticos que utiliza el algoritmo se remite únicamente a la identificación de los hechos estadísticos, únicamente

haciendo uso de texto, en este caso aquel que viene en las respuestas a la pregunta relacionada con actividad económica

La ecuación del modelo generaliza el problema en la Ecuación 12 para casos prácticos relacionados con la encuesta ENEMDU en casos que no requieren de un análisis más específico, basado en otras variables del cuestionario como las referentes a lugar de trabajo y la categoría de la empresa, información esta última que debe contrastarse con la codificación CIU4 utilizando únicamente el texto de la pregunta sobre actividades económicas.

6. Referencias

Alsina Rodrigo, M. (2003). Los modelos de comunicación. *Introducción a la teoría y la investigación en comunicación*. La Habana, Cuba: Editorial Félix Varela p. 40-113.

Ashok, K., & Kannathasan, N. (2011). A Survey on Data Mining and Patern Recognition Techniques for Soil Data Mining. *International Journal of Computer SCiences Issues, Vol 8, Issue 3. No. 1, May 2011*. Tamil Nadu, India: pag 422.

Folli, R., & Ramchand, G. (2001). Getting results: Motion verbs in italian and scottish gaelic. *Proceedings of WCCFL*. Roma, Italia: PP 192-205.

Gries, S. (2014). Corpus Linguistics and Linguistic Theory. Santa Bárbara, CA 93106-3100, Estados Unidos de Norteamérica.

INEC, V. A. (2010). Clasificación Uniforme de Actividades Económicas CIU-4. *Manual de Usuario*. Quito, Regional Norte: INEC.

Jakobson, R. (1988). Linguística y Poética. *Ensayos de Linguística General*. Madrid, España.

Llisterri, J. (2003). Linguística y Tecnologías del Lenguaje. *Panorámica de Estudios Lingüísticos*. Lynx, Francia: Vol 2,pp 9-71.

Panagiotidis, P. (2001). The categorial features of functional categories. *Essex Graduate Student Papers in Linguistics 3*. Essex, University of Essex, Estados Unidos: pp. 161-170.

Pelayo, V., Moreno, M., Fraga, A., Moreno, J., & Corredor, E. (2012). Método para generar patrones semánticos. *Method for generating semantic patterns*.

Madrid, Universidad Carlos III de Madrid, España.

Santana, O., Perez, J., Carreras, F., & Rodriguez, G. (1997). Flexionador y lematizador automático de formas verbales,. *Linguística Española Actual Vol 19,2*. Barcelona, España: citado por Llisteri, J en Linguística y Tecnologías del Lenguaje; pp 25.

Vazquez, G., & Fernández, A. (2002). Léxicos verbales computacionales. *Tratamiento del Lenguaje Natrual, citado por Llisterri, J (2002)*. Barcelona, España.

Wasson, M., & Wiltshire JR, J. (2005). Extraction of Facts from Text. *United States Patent Application Publication*. Nueva York, Estados Unidos de Norteamérica: Departamento de Patentes Federal.

Willenborg, L. (2013). Semantic networks for automatic coddng. La Haya, Holanda.

7. ANEXOS

Anexo 1. Caso experimental.

Figura 1-Anexo 1. Ejemplo de identificación de patrones de hechos estadísticos.

CLASIFICACIÓN INDUSTRIAL INTERNACIONAL UNIFORME		PATRONES LÉXICOS EXTRAÍDOS DEL CORPUS DE EJEMPLOS DE CODIFICACIÓN MANUAL REALIZADA POR LA PRODUCTORA ESTADÍSTICA EN INEC.									
CODIGO CIU4	CLASIFICACIÓN DE LA ACTIVIDAD ECONÓMICA	β	TIPO	MARCA ACTIV	ACTIVIDAD	MARCA PROD	PRODUCTO	MARCA LUGAR	LUGAR	MARCA REFER	REFERENCIA
A0111.31.01	Cultivo de granos y semillas de soya.	β	A-P	<A>	CULTIV	<P>	GRANOS	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> SOYA	-	-		
A0111.32.01	Cultivo de semillas de maní.		A-P	<A>	CULTIV	<P>	SEMIL <P> MANI	-	-		
A0111.39.01	Otros cultivos de semillas oleaginosas: semillas de ricino, semillas de linaza, semillas de mostaza, semillas de girasol, semillas de ajonjolí (sésamo), semillas de colza, semillas de cártamo, semillas de níger, etcétera.	β	A-P	<A>	CULTIV	<P>	SEMIL <P>	-	-		
			A-P	<A>	CULTIV	<P>	OLEAGIN	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> RICINO	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> LINAZA	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P>	-	-		
			A-P	<A>	CULTIV	<P>	MOSTAZA	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> GIRASOL	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P>	-	-		
			A-P	<A>	CULTIV	<P>	AJONJOLI	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> SÉSAMO	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> COLZA	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> CÁRTAM	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> NÍGER	-	-		
			A-P	<A>	CULTIV	<P>	SEMIL <P> ETC	-	-		
A0112.00.01	Cultivo de arroz (incluido el cultivo orgánico y el cultivo de arroz genéticamente modificado).	β	A-R	<A>	CULTIV			-	-	<R>	ORGÁNICO
			A-P	<A>	CULTIV	<P>	ARROZ	-	-	<R>	GENÉTIC MODIFIC
A0113.11.01	Cultivo de brócoli, col y coliflor.	β	A-P	<A>	CULTIV	<P>	BROCOLI	-	-		
			A-P	<A>	CULTIV	<P>	COL	-	-		
			A-P	<A>	CULTIV	<P>	COLIFLOR	-	-		

HECHO ESTADÍSTICO

PREGUNTA DEL EMISOR O ENCUESTADOR

TEXTO DEL MENSAJE

Pregunta 40

Encuesta de Empleo en Ecuador

RESPUESTA DEL RECEPTOR O ENCUESTADO

Al cultivo orgánico de semillas y arroz

PATRONES EXTRAÍDOS

<A> CULTIV <R> ORGANIC
<A> CULTIV <P> SEMIL
<A> CULTIV <P> OTROS
<A> CULTIV <P> ARROZ

INTERPRETACIÓN DEL HECHO Y SU CODIFICACIÓN AUTOMÁTICA

CODIGO CIU4	CLASIFICACIÓN DE LA ACTIVIDAD ECONÓMICA	SIMILITUD
A0111.39.01	Otros cultivos de semillas oleaginosas: semillas de ricino, semillas de linaza, semillas de mostaza, semillas de girasol, semillas de ajonjolí (sésamo), semillas de colza, semillas de cártamo, semillas de níger, etcétera.	75%
A0112.00.01	Cultivo de arroz (incluido el cultivo orgánico y el cultivo de arroz genéticamente modificado).	95%

Algoritmo para identificar Patrones de hechos

Figura 2-Anexo 1. Ejemplo de ejecución del algoritmo: P2: Reconocer términos.

ID1	Empresa/Establecimier	Lugar	Actividad	Servicio	Producto	A quien	Codigo
P011	PLANTACION		PLANTACION		FLORES DE EXPORTA		
P111			PLANTACION		FLORICOLA		
P011			PLANTACION Y EXP		FLORES		
P011			PRODUCCION		FLORES		
P011	BANANERA		SIEMBRA		PLANTAS ORNAMEN		
P111			CULTIVO		BANANO		
P011			ELABORACION		PLATOS DE PORCELA		
P011			ELABORACION		VAJILLAS DE LOSA		
P011			ELABORACION Y VEN		ADORNOS DECORAT		
P011			ELABORACION Y VEN		ARTICULOS CERAMI		
P011			VENTA AL POR MENI		ARTICULO DE BISUTE		
P011			VENTA AL POR MENI		ARTICULOS DE PAPE		

Anexo 2. Ejemplo de Descomposición de Términos de hechos desde texto...

Figura 1-Anexo 2.

ORDEN	C1	C2	C3	C4	C5 = SINTAGMA	NUCLEO
1	A	P	1	PREPAR	PREPAR PIZZA	1
2	A	P	1	VENTA	VENTA PIZZA	1
3	P	P	1	PIZZA	ACTIVIDAD PIZZA	2
4	R	F	0	AL	AL POR MENOR	3
5	F	F	0	POR	AL POR MENOR	3
6	F	F	1	MENOR	AL POR MENOR	3
7	F	F	0	DOMINOS	DOMINOS PIZZA	4
8	F	F	1	PIZZA	DOMINOS PIZZA	4

CAMPOS DE LA DESCOMPOSICIÓN	
C1	Tipo de entidad
C2	Tipo de estructura léxica: P=Palabra; F=Frase
C3	Clase: 1=Núcleo de Sintagma; 0=No núcleo
C4	Término lematizado o abreviado
C5	Sintagma extraído
C6	Numeración de Sintagmas

Elaboración: Propia

Elaboración: Propia

72

73



 **www.ecuadorencifras.gob.ec**

 Administración Central (Quito)

 Juan Larrea N15-36 y José Riofrío

 Teléfonos: (02) 2544 326 - 2544 561 Fax: (02) 2509 836

 Código postal: 17-15-135C

 Correo-e: inec@inec.gob.ec

ISSN 0000-0132

0 9 7 3 8



9 770000 013904