

Tesina de Licenciatura en Economía  
Facultad de Ciencias Económicas - Universidad Nacional del Litoral

**Educación y Trabajo.**  
**Diferenciales en el acceso al trabajo en los jefes de hogar  
santafesinos desde un análisis multinivel.**

Tesinista: Rocco Bertinetti Giumelli

L.U.: 24.584

Director: Dr. Roberto Delfor Meyer

*Santa Fe, noviembre de 2016*

## ÍNDICE

Índice .....	1
Resumen .....	4
Capítulo I - Introducción .....	5
1.1. Problema de Investigación y estructura del presente trabajo .....	9
1.2. Objetivos .....	12
1.2.1. Objetivo General .....	12
1.2.2. Objetivos Específicos.....	12
1.3. Metodología.....	12
1.4. Justificación.....	15
1.5. Antecedentes .....	19
Capítulo II – Marco Teórico.....	24
2.1. La relación educación y trabajo. Aproximaciones Teóricas .....	24
2.2. Educación en la Modernidad.....	26
2.3. Economía de la Educación .....	28
2.4. Teorías Alternativas .....	33
2.5. Educación y Neoliberalismo .....	38
2.6. Breve conclusión sobre las teorías presentadas.....	41
Capítulo III – Modelos Multinivel.....	43
3.1. Modelos Multinivel (o Modelos de Efectos Aleatorios o Modelos Lineales Jerárquicos Generalizados y otras denominaciones).....	43
3.2. ¿Porqué utilizar modelos multinivel?.....	44
3.3. Un acercamiento a los Modelos Multinivel.....	46
3.4. Componentes implícitos en los Modelos Multinivel.....	47
3.4.1. El fenómeno de la anidación y sus consecuencias.....	47
3.4.2. La existencia de grupos y la heterogeneidad inobservada .....	49
3.4.3. El Modelo de Efectos Fijos.....	50
3.4.4. El Modelo de Efectos Aleatorios .....	52
3.4.5. El Coeficiente de Correlación Intraclase .....	54
3.4.6. El estimador de Máxima Verosimilitud.....	55
3.4.7. Del Modelo de Efectos Aleatorios a los Modelos Multinivel.....	56
3.5. Especificación del Modelo Multinivel utilizado en este trabajo .....	57

3.5.1. Modelos de regresión multinivel para datos longitudinales.....	57
3.5.2. Modelos de respuesta discreta.....	59
3.5.3. Modelos de regresión logística de intercepto aleatorio.....	65
3.5.4. Modelo Logístico Multinivel de intercepto aleatorio para el acceso al trabajo ..	67
3.6. Una aclaración sobre las “chances u odds” y el “cociente de chances u odds ratio” ...	71
Capítulo IV – Diferenciales en el acceso al trabajo.....	74
4.1. Base de datos y análisis descriptivo de la información .....	74
4.1.1. Analizando la variable dependiente .....	74
4.1.2. Trabajando con un panel no balanceado .....	78
4.1.3. Re-analizando las variables involucradas en el modelo.....	80
4.2. Estimación de los diferenciales en el acceso al trabajo de los jefes de hogar .....	90
4.2.1. Estudiando el patrón de datos perdidos .....	91
4.2.2. Estadísticos descriptivos para variables de panel .....	94
4.2.3. El modelo vacío .....	97
4.2.4. Modelo logístico multinivel de intercepto aleatorio para la estimación de diferenciales en el acceso al trabajo de los jefes de hogar santafesinos .....	102
4.2.5. Bondad de ajuste del modelo .....	108
Capítulo V – Conclusiones .....	115
Bibliografía.....	122
Apéndice – Aclaraciones sobre el cálculo de Tasas .....	131
Anexo – Otros modelos estimados .....	132

### Índice de Tablas

1. Proporción de acceso al trabajo de los jefes de hogar en el año 2009.....	76
1. Indicadores Laborales Comparados: Panel de Hogares y Encuesta Permanente de Hogares.....	77
2. Acceso al trabajo de los Jefes de Hogar Económicamente Activos 2009-2011.....	80
3. Estadísticos descriptivos para la variable “Edad” .....	83
4. Composición del género de los jefes de hogar 2009-2011 .....	87
5. Composición del nivel educativo de los jefes de hogar 2009-2011 .....	87
6. Nivel educativo de los jefes de hogar. Total agrupado 2009-2011 .....	89
7. Patrón de datos perdidos.....	92
8. Estadísticos descriptivos para variables de panel.....	94

**Índice de Figuras**

2. Tasa Anual de Crecimiento del Producto Bruto Geográfico (PBG) de la Provincia de Santa Fe .....	82
3. Histograma de la variable “Edad” de los jefes de hogar santafesinos económicamente activos para el período 2009-2011 .....	84
4. Proporción de los Jefes de hogar santafesinos económicamente activos que trabajan según nivel educativo alcanzado para el período 2009-2011 .....	90
5. Modelo Vacio de intercepto aleatorio .....	99
6. Coeficiente de Correlación Intraclase Incondicional para el modelo logístico multinivel de intercepto aleatorio sin variables explicativas.....	101
7. Modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas.....	102
8. Coeficiente de Correlación Intraclase para el modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas.....	103
9. Modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas estadísticamente significativas .....	106
10. Coeficiente de Correlación Intraclase para el modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas estadísticamente significativas.....	107
11. Modelo de regresión logística con todas las variables explicativas .....	113

## Resumen

Como todo proceso y relación que se define y forma socialmente, el vínculo entre la educación y el mundo del trabajo cambia a lo largo de la historia y asume características o rasgos propios de la sociedad en la que se desarrolla. En la actualidad, existen profundos procesos de transformación en múltiples aspectos de la vida humana y la relación educación-trabajo no es ajena a ellos. Partiendo de dicha concepción, se pretende analizar algunos aspectos de la misma en la ciudad de Santa Fe. Específicamente, el objetivo general de esta tesis es describir y cuantificar la existencia de diferenciales en el acceso al trabajo de los jefes de hogar santafesinos en relación al nivel educativo alcanzado y a otras características de los mismos para el período 2009-2011. Partiendo desde un enfoque cuantitativo, se estimará un modelo logístico multinivel para datos de tipo panel, provenientes de fuente secundaria. Los modelos multinivel presentan amplias ventajas sobre los modelos de regresión tradicional cuando se trabaja con datos con estructura jerárquica, como en este caso. Los mismos permitirán trabajar con diferentes niveles de análisis y así tener en cuenta las particularidades de cada individuo y las diferencias entre los mismos al considerar a cada uno de ellos como un clúster específico. Los resultados obtenidos permiten ratificar la relación positiva entre el nivel educativo alcanzado y las mayores posibilidades de acceso al trabajo, como así también, un gran diferencial en el acceso en relación al género, concluyendo así, que las chances de trabajar de los jefes de hogar santafesinos aumentan considerablemente a medida que aumenta el nivel educativo alcanzado y más aún, si el jefe de hogar es de sexo masculino.

**Palabras clave:** educación – trabajo – modelos multinivel – datos de panel – jefes de hogar

## Capítulo I

### Introducción

Afrontar el estudio de la relación entre educación y trabajo no es una tarea sencilla debido a la diversidad de disciplinas, enfoques y posturas desde la que puede ser abordada, como así también, por los múltiples aspectos de esta interrelación sobre los cuáles se puede focalizar el objeto de estudio. Justamente, la interrelación entre ambos mundos ha despertado desde siempre la atención de sociólogos, economistas, historiadores, profesionales de la educación, etc. Ahora bien, la importancia que tiene abordar el estudio de esta relación se hace evidente al observar que, como miembros de esta sociedad, transcurrimos gran parte de nuestra vida entre estas dos esferas.

Los complejos procesos de transformación acaecidos en el mundo en las últimas décadas han generado múltiples cambios en muchas dimensiones de la vida humana. La crisis del capitalismo industrial y del Estado de Bienestar en la década del setenta, dio paso a procesos de reestructuración que se intensificaron en los años noventa mediante mecanismos de desregulación, privatizaciones y la ruptura del contrato social entre el capital y la mano de obra. En esta nueva era de la información, la revolución tecnológica, la introducción de la micro-robótica y micro-ingeniería, los nuevos sistemas de comunicación, las nuevas formas de gestión “flexible” de las empresas, etc., han transformado fuertemente el mundo productivo, reestructurando el modo de producción capitalista y definiendo un nuevo modo de desarrollo que Castells (2000) define como “informacionalismo”. Este nuevo modo de desarrollo está constituido por el surgimiento de un nuevo paradigma tecnológico basado en las tecnologías de la información, en donde la fuente de la productividad radica principalmente en la tecnología de la generación del conocimiento y el procesamiento de la información y comunicación. Si bien el conocimiento y la información son elementos necesarios en cualquier modo de desarrollo, aquí lo que distingue el uso del conocimiento es su aplicación sobre sí mismo, “{...} en un círculo de interacción de las fuentes del conocimiento de la tecnología y la aplicación de ésta para mejorar la generación de conocimiento y el procesamiento de la información {...}” (Castells, 2000, vol. 1, p. 42). El nuevo modo de desarrollo que se define por la reestructuración del modo de producción capitalista, conlleva el surgimiento de una nueva estructura social, que se manifestará de diversos modos según las diferentes sociedades y culturas y que genera cambios en múltiples aspectos de las mismas, es decir, en sus formas

de organización, en las relaciones sociales, en sus principales instituciones, en las condiciones de vida, etc. Por ello, la educación, el trabajo y la relación entre ellos no están exentos de cambios.

Justamente, en la nueva era de la información y la sociedad en red, la educación y más específicamente el conocimiento, es considerado un factor estratégico del desarrollo, tanto para el individuo como para la sociedad en su conjunto. El conocimiento es considerado el ingrediente principal en un mundo cada vez más globalizado e interconectado, en el que la información y la comunicación es instantánea.

El mundo del trabajo tampoco está exento de cambios, sino todo lo contrario. Tal como postuló Castel (1997) “La diversidad y discontinuidad de las formas de empleo están reemplazando el paradigma del empleo homogéneo y estable” (pp. 336-337). La reestructuración del mundo laboral que comenzó con la crisis de la sociedad salarial, ha dado paso a nuevos tipos de trabajo y relaciones laborales caracterizadas por la movilidad y flexibilidad. Según Carnoy (2002), hay dos elementos clave en la transformación del trabajo dentro de la nueva economía global, la flexibilidad del proceso de trabajo y el trabajo en red de las firmas y de los empleados dentro de las firmas. La nueva “producción flexible” como define el autor, requiere que las tareas y el tiempo de trabajo se adapten constantemente a los nuevos mercados, procesos y productos. Así mismo, el trabajo en red refiere a una lógica nueva de funcionamiento y formas de organización basadas en conexiones interactivas dentro y entre firmas, como así también dentro y entre mercados. Precisamente, las nuevas tecnologías de la información son las que permiten una mayor interconectividad y flexibilidad, logrando adaptarse a los cambios en forma constante en un mundo cada vez más globalizado.

Es que precisamente, un rasgo distintivo del nuevo modo de desarrollo es que gran parte de los conocimientos nuevos son producidos por empresas y firmas. Entre los nuevos modos de organización, por ejemplo, se destaca el surgimiento de áreas de redes de empresas productoras de conocimientos, las cuáles crean las condiciones necesarias para el trabajo cooperativo y el intercambio de ideas innovadoras. Concomitantemente, las nuevas organizaciones de trabajo flexible son necesariamente organizaciones de aprendizaje. Carnoy (2002) y Castells (2000) entienden que la tecnología y la información hacen su gran contribución a la productividad cuando forman un círculo virtuoso de aprendizaje y enseñanza que es inherente al proceso de trabajo.

En este marco, los nuevos mercados y empresas requieren de trabajadores no solamente bien educados y disciplinados, sino también ágiles, flexibles, innovadores, rápidos para moverse de tarea en tarea y de mercado en mercado y competentes en la resolución de problemas. Además, un aspecto fundamental es que estén dispuestos al aprendizaje ya que, en esta nueva lógica del trabajo flexible, el aprendizaje no se termina en la escuela o en la Universidad sino por el contrario, debe continuar a lo largo de toda la vida del trabajador para poder adaptarse a la velocidad del cambio tecnológico y a la constante introducción de nuevos conocimientos e información (Tedesco, 2000).

En el mismo sentido, Castel (2004) afirma que las nuevas formas de trabajo desestandarizado y la individualización de las tareas exige a los trabajadores mayor movilidad y adaptabilidad. De este modo, las trayectorias de los trabajadores se vuelven móviles y depende de cada uno afrontar su recorrido profesional. La mayor autonomía en la toma de decisiones viene acompañada de un aumento del riesgo y la inseguridad. Si a esto le sumamos la falta de marcos y redes de contención producto de los procesos de descolectivización, desregulación y la decadencia de otras instituciones socializadoras tradicionales, como la iglesia o la familia, lo que se tiene es un aumento de la vulnerabilidad, sobre todo de los grupos más desfavorecidos. Esto llevó a distintos autores a hablar del incremento de las desigualdades, la marginalidad, la exclusión, es decir, un mundo contemporáneo sumido en la incertidumbre. Es decir, que ante la falta de marcos objetivos que estructuran la vida de los sujetos, lo que cobra importancia es la individualidad de los esquemas de acción y de pensamiento. (Castel, 1997; Beck, 1998). En el mismo sentido, Tedesco (2000) afirma que el aumento de la desigualdad, la exclusión, la polarización social, etc. se deben a que no existe un sistema institucional que se haga responsable del destino de las personas, como lo eran antes la familia, la comunidad local, la religión o la tradición, y por lo tanto, no queda más que depositar la confianza en uno mismo.

La educación no es ajena a estos procesos de cambios, o al menos, no debería serlo. Los cambios del sistema productivo, de la organización del trabajo y de la estructura social repercuten en los sistemas educativos. El desarrollo y la transformación de los mismos en cuanto al perfil de los conocimientos impartidos, la modalidad y la duración de los estudios, los diferentes métodos, tienen que ver en parte con el destino laboral esperado de los alumnos, es decir, con enseñar las habilidades y técnicas que necesitarán los mismos

para desempeñarse en el mercado laboral de su época. Además, la educación tiene un papel central como mecanismo de reproducción de la estructura social y del reparto del trabajo que esta estructura tiene asociado (Gallart, 1997).

Los sentidos que se le atribuyen a la educación van cambiando a lo largo del tiempo, al ritmo en que las sociedades, sus instituciones, estructuras y relaciones van cambiando. A su vez, las esperanzas, anhelos u objetivos que le depositan los individuos y la sociedad, también cambian. Desde ser considerada el elemento emancipador y liberador del hombre de todo lo tradicional en los orígenes de la modernidad (Terrén, 1999), pasando por ser considerada el principal medio para alcanzar la movilidad social ascendente durante los años del Estado de Bienestar (Torrado, 2007; Aronson, 2007) y también como el principal factor de desarrollo de las naciones (Schultz, 1961; Becker, 1975/1983), hasta entenderla como un derecho inalienable al que todos los niños, niñas y jóvenes deberían acceder en igualdad de oportunidades (Organización de las Naciones Unidas, Art. 28, 1989). La Comisión Económica para América Latina y el Caribe y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (CEPAL - UNESCO, 1992), por su parte, le atribuyeron a la educación y a la generación y acceso al conocimiento, ser el factor principal en la incorporación y difusión del progreso técnico que permitiría alcanzar la transformación productiva con equidad, con un doble desafío: el de generar ciudadanía en el plano interno y competitividad en el plano externo.

En el marco entonces de una organización del trabajo que está mutando, en la que está desapareciendo el empleo tradicional fordista de largo plazo por un empleo “flexible”, en el que el puesto de trabajo debe adecuarse a la necesidades del trabajo en red y que por lo tanto exige al trabajador nuevas competencias y calificaciones, es necesario pensar otra educación. Precisamente, uno de los aspectos centrales de la nueva educación es que ya no se trata de aprender un determinado acervo de conocimientos, mediante el análisis lógico-deductivo, sino el aprender a aprender, aprender mecanismos, procedimientos, la capacidad de resolución de conflictos, etc. En palabras de Fernández Enguita (2010):

{...} una capacitación más polivalente, lo que supone una formación básica más sólida y una capacidad de aprender desarrollada que permitan su repetida adaptación y actualización; y una disposición más proactiva, que facilite la integración y la participación en equipos o la iniciativa individual. (pp. 31-32).

Esta nueva educación que se está configurando apunta justamente a las nuevas necesidades que se le plantean a los trabajadores para lograr trayectorias exitosas: versatilidad, iniciativa, capacidad de adaptación y trabajo en equipo, creatividad. El sistema educativo que fue construido y pensado bajo el modo de organización tradicional fordista dónde imperaban la rutina, los procesos técnicos y las decisiones autónomas, no forma entonces para la iniciativa, la creatividad y la innovación (Gallart, 1997; Fernández Enguita, 2010). Es necesaria entonces, otra educación, ya que son necesarias otras capacidades. Esto queda claro en palabras de Estanislao Bachrach, especialista en neurociencias:

“En nuestro sistema educativo, desde sus inicios y hasta el tiempo presente, ha primado la importancia en la enseñanza y el desarrollo del análisis lógico y el razonamiento deductivo – que dominó el siglo XX – y no de nuestra capacidad empática y creativa – necesaria para conquistar el siglo XXI -.” (Bachrach, 2012, p. 15).

Ahora bien, la profundidad, el impacto y las consecuencias de estos cambios no se plasman de la misma manera en todo el mundo. Si bien se ha destacado los nuevos tipos de organización en el mundo del trabajo y los nuevos requerimientos en las capacidades de los trabajadores, la realidad es que la mayoría de los trabajadores no se desempeña aún en las llamadas empresas flexibles o productoras de conocimientos, ni siquiera en los países más desarrollados dónde abundan dichas empresas. (Carnoy, 2002). En el plano educativo sucede algo similar ya que, si bien se reconoce la necesidad de los cambios en la educación y hay algunos logros al respecto, la realidad en América Latina y el Caribe parece ser bien distinta y apunta a que la escuela no prepara para el trabajo (Banco Interamericano de Desarrollo [BID], 2014).

Reconociendo, tal como lo hace Riquelme (2006, p. 68), “{...} que la relación educación y trabajo es una articulación interactiva, compleja, multidimensional, radicada y determinada por un espacio y un tiempo, es decir, histórica” y entendiendo que para saber hacia dónde queremos ir, primero debemos saber dónde nos encontramos parados, me propongo estudiar algunos aspectos de la relación entre educación y trabajo en la ciudad de Santa Fe desde una mirada novedosa como lo es el análisis multinivel.

### **1.1 Problema de investigación y estructura del presente trabajo**

Nos encontramos inmersos entonces en un mundo que está viviendo grandes transformaciones y la ciudad de Santa Fe no es ajena a ello. El impacto de dichos cambios

en nuestra ciudad puede ser relativo actualmente, pero no pueden ser dejados de lado, ya que el mundo se encuentra cada vez más “conectado”. Se puede decir que el mercado de trabajo santafesino no muestra aún las características de la nueva economía flexible, pero eso no es lo mismo que decir que nunca lo hará. Los cambios en la educación tampoco pueden ser dejados de lado, ya que la educación que reciban los niños hoy, los prepara para el mundo de mañana.

Esta tesis no pretende poner en discusión los beneficios que tiene la educación para el individuo y para las sociedades en su conjunto, sin embargo, considero que no existe una concepción única entre la educación y el bienestar individual y colectivo que esta genera, precisamente, porque entiendo que las diferentes definiciones, aspectos y/o sentidos que se le otorgan a la educación y su conexión con el mercado laboral está relacionado con las diferentes circunstancias históricas y los enfoques teórico-metodológicos desde la que se lo aborda.

La literatura económica ha abordado innumerables veces el estudio de la relación entre el nivel educativo alcanzado y los ingresos o el acceso a mejores niveles ocupacionales, estableciendo una relación positiva entre ambos. También es cierto que han existido críticas y matices, e incluso actualmente dicha relación es a veces puesta en duda, tal como se verá a lo largo de este trabajo. La idea es dejar planteado que el vínculo entre educación y trabajo está cambiando y por eso es importante abordarlo.

Por ello, la pregunta que dispara el estudio de esta tesis es: ¿De qué manera influye el nivel educativo alcanzado por los jefes de hogar santafesinos en sus chances de conseguir trabajo?

Este trabajo no aspira a encontrar resultados extraordinarios, más bien, la hipótesis detrás de esta tesis es que en la ciudad de Santa Fe, existen diferencias en el acceso al trabajo que están positivamente correlacionadas con el nivel educativo alcanzado de los jefes de hogar. Del mismo modo se espera también que, por ejemplo, el género del jefe de hogar presente diferencias en las posibilidades de obtener trabajo, siendo mayores las chances de los hombres frente a las de las mujeres.

Lo que caracteriza a este trabajo en particular, es el hecho de poder *cuantificar* dichas diferencias como un “cociente de chances” o posibilidades. Más importante aún, es el enfoque estadístico que se utilizará para ello, enmarcado dentro de los denominados

modelos multinivel. Estos modelos, presentan amplias ventajas sobre los modelos de regresión tradicional cuando la información con la que se trabaja posee algún tipo de estructura jerárquica, como es el caso de los datos utilizados en esta tesis.

A fin de responder la pregunta que dio origen a este trabajo, el mismo se encuentra estructurado de la siguiente manera:

El capítulo uno, además de la introducción general al tema de estudio y los objetivos de la investigación, presenta la justificación del presente trabajo como así también los antecedentes y los aspectos metodológicos del mismo.

El capítulo dos presenta el marco teórico que abordará las principales teorías que han intentado explicar algunos aspectos del vínculo entre educación y trabajo, como así también, las implicancias que tiene la educación para el individuo en particular y para la sociedad moderna en su conjunto.

El capítulo tres se centrará en los llamados modelos multinivel. El objetivo del capítulo es ir incorporando progresivamente los conceptos que estos modelos implican. Por ello, la primera parte pretende, mediante un acercamiento sencillo, introducir al lector en los conceptos, características, ventajas y desventajas de los mismos, para pasar luego, a la presentación del modelo multinivel utilizado en esta tesis, el cual cuenta con ciertas características propias que serán explicadas con mayor detalle y lo hacen específicamente interesante.

El capítulo 4 presenta los resultados del modelo aplicado y seleccionado. Este capítulo también se encuentra dividido en dos grandes partes. La primera se centra en un análisis descriptivo de las variables utilizadas y algunas especificaciones sobre la base de datos, incluyendo la descripción de algunas características muestrales del mercado de trabajo santafesino para el período. La segunda refiere a la presentación del modelo seleccionado y sus resultados, como así también, la justificación estadística de dicha elección por sobre otros modelos y otros tipos de modelos.

El capítulo 5 exhibe las principales conclusiones del presente trabajo, las limitaciones del análisis pertinente y algunas posibles líneas de acción a seguir para profundizar y mejorar los aportes de esta tesis. Por último se exponen algunas reflexiones finales del autor.

## 1.2 Objetivos

### 1.2.1 Objetivo General

Describir y cuantificar las diferentes posibilidades o chances de trabajar entre los jefes de hogar santafesinos según el nivel educativo alcanzado, mediante la estimación de un modelo logístico multinivel para datos de panel en el período 2009-2011.

### 1.2.2 Objetivos Específicos

1. Describir algunos aspectos de la relación entre la educación y el trabajo a lo largo del tiempo, desde la postura que toman distintas corrientes teóricas.
2. Describir y explicar el modelo logístico multinivel para datos de panel.
3. Estimar un modelo logístico multinivel para los jefes de hogares santafesinos según el nivel educativo alcanzado.
4. Mejorar el modelo estimado considerando diferentes características de los jefes de hogares santafesinos en base al criterio de disminución del Coeficiente de Correlación Intraclase.
5. Justificar la selección del modelo seleccionado por sobre otros modelos.

## 1.3 Metodología

El objetivo de esta tesis es describir la existencia de posibles diferenciales en el acceso al trabajo de los jefes y jefas de hogar santafesinos en relación al nivel educativo alcanzado y a otras características de los mismos para el período 2009-2011. Partiendo entonces desde un enfoque cuantitativo, se aplicará un modelo logístico multinivel para datos de tipo panel, lo que permite trabajar con diferentes niveles de análisis y así tener en cuenta las particularidades de cada individuo y las diferencias entre los mismos al considerar a cada uno de ellos como un clúster específico.

Específicamente, se utilizará un modelo logístico multinivel para datos de panel de intercepto aleatorio. En un modelo de este tipo, se tienen distintas observaciones en el tiempo en el nivel uno, anidadas en grupos o clústeres en el nivel dos, que en este caso en particular son individuos. La variable respuesta es una variable dicotómica que toma el valor uno si la respuesta es afirmativa y el valor cero si no lo es, a la pregunta “En la última semana, ¿Usted trabajó?”.

El intercepto aleatorio permite tener en cuentas los factores inobservados de los distintos individuos que causan correlación en las respuestas brindadas por ellos a lo largo del tiempo. De este modo, se obtendrán errores estándar más conservadores y coeficientes que no presenten sesgos.

Mediante el cálculo del Coeficiente de Correlación Intraclase Incondicional, se verá si un modelo que tenga en cuenta la existencia de datos jerárquicos es necesario y, de ser así, se introducirán variables contextuales para intentar explicar la variabilidad existente *entre* los individuos.

Más precisamente, lo que se intentará es identificar un modelo de predicción que permita encontrar y cuantificar posibles diferenciales en el acceso al trabajo en función de características del individuo, principalmente, el nivel de estudios alcanzados, el sexo y la edad.

Para ello se propone utilizar información secundaria, generada por el Observatorio Social de la Universidad Nacional del Litoral. Dicha institución produce anualmente un relevamiento de hogares santafesinos, denominado “Panel de Hogares”, que tiene por objetivo ser un sistema de información social confiable acerca de la percepción de las familias santafesinas sobre diferentes temas a lo largo del tiempo, posibilitando así la construcción de información histórica sobre los mismos sujetos-actores, lo que permite analizar la comparación de resultados y la rotación de sus respuestas.

Esta posibilidad de confrontar a la familia informante con la información histórica que la misma produjo en un momento determinado, tiene además la característica distintiva de indagar sobre las *percepciones* de los actores sociales acerca de la realidad social en general y la propia en particular, desde un enfoque metodológico que tiene sustento en los conceptos de “necesidades” y “satisfactores” de Max-Neef (1986/2010). Según este enfoque, existe una diferencia fundamental entre lo que son necesidades y satisfactores de dichas necesidades. Las primeras refieren al conjunto de deseos, aspiraciones, anhelos, voluntad o carencias que motivan a los actores sociales y son manifestadas a través del deseo, la opinión o la demanda de los mismos. Las necesidades surgen así de la propia percepción de los individuos y son entendidas en un sentido amplio, es decir, no sólo limitadas a la subsistencia, por lo que pueden ser expresadas tanto en términos de una carencia o como de una potencialidad, ya que la falta o necesidad de algo actúan sobre el

ser humano como un movilizante y son, por ende, un potencial. Los satisfactores en cambio, son el conjunto de acciones y/o bienes que permiten cubrir las necesidades percibidas y explicitadas por los actores sociales, los cuales cambian a lo largo del tiempo y las culturas. Desde este sentido, los satisfactores se definen culturalmente, no así las necesidades humanas fundamentales, las cuales se consideran “(...) finitas, pocas y clasificables (...) [y] son las mismas en todas las culturas y todos los períodos históricos.” (Max-Neef, 1986/2010, p. 17). En el Panel de Hogares, se conduce al entrevistado a través del cuestionario en el contexto de una estrategia de “tiene o posee” versus “desea”, con el objetivo de poder identificar dichos satisfactores (Observatorio Social, 2016).

De este modo, el “trabajo” es considerado en el Panel de Hogares como una “necesidad”, mientras que el nivel educativo alcanzado es considerado un “satisfactor”, misma conceptualización que se adoptará en este escrito. El trabajo como necesidad es expresado a través de la demanda del propio actor, es decir, lo que se admite es que el propio sujeto considere si su “necesidad de trabajo” ha sido satisfecha o no. Esto lo diferencia de otros enfoques positivistas enfocados en “medir el empleo”, el cual, bajo dicha lógica, el mismo es entendido como el “satisfactor” hegemónico bajo el modo de producción capitalista. De más está decir, que tal como se mencionó en la introducción, el mismo se encuentra precisamente en un período de transformación, poniendo en duda su rol de satisfactor en las sociedades actuales, o mejor dicho, lo que está en dudas es el “empleo” como un satisfactor homogéneo de la necesidad de trabajo, tal como se presentó en la sociedad salarial, el cual estaba vinculado no sólo a la subsistencia sino como un elemento primordialmente identitario. Por su parte, el nivel educativo alcanzado es considerado un “satisfactor” de la necesidad de entendimiento. El mismo no es considerado una necesidad porque refiere al nivel educativo alcanzado por cada uno de los jefes de hogar, y no el nivel que “desearía” alcanzar.

El Panel General, es representativo de la ciudad de Santa Fe y la información constituida cada año por el mismo, se denomina “Onda de Panel”. Los hogares que se mantienen en el tiempo y responden la encuesta año tras año pasan a formar parte del denominado “Panel Detallista”. La Onda 2009 se la consideró reconstitutiva del panel, ya que se incorporaron nuevos hogares según cuotas geográficas establecidas por muestreo en base a datos censales, logrando así representatividad para la ciudad de Santa Fe con un total de 1091 hogares con 3877 integrantes relevados. El panel detallista para el período 2009-2011 que

se utilizará aquí, está compuesto por 582 hogares. La reducción del tamaño muestral es producto de aquellos hogares que resultan imposibles de relevar, lo que comúnmente se denomina mortalidad de panel, siendo éste uno de los principales inconvenientes que posee la encuesta de tipo panel (Observatorio Social, 2016).

Específicamente entonces, la investigación se llevará a cabo analizando los datos de tipo panel, los cuales combinan características de datos de corte transversal con características de datos longitudinales, que corresponden a jefes y jefas de hogar de la Ciudad de Santa Fe para los años 2009, 2010 y 2011. Es decir que para cada individuo, existen tres observaciones en el tiempo. La muestra de estudio está compuesta de 582 casos (jefes y jefas de hogar), constituyendo así 1746 observaciones en total, las que surgen de multiplicar el número de casos “N” por la cantidad de momentos en el tiempo “t” ( $N*t = 582*3 = 1746$ ).

Considero pertinente una última aclaración acerca del marco teórico que se presentará en el capítulo 2. Como toda investigación que utiliza información secundaria, me veo limitado a utilizar los datos brindados tal como han sido recabados, lo que se traduce en que muchas veces no se puede contar con la información que uno desearía en miras a un determinado objetivo. Es por ello que el modelo presentado en esta tesis, no se basa en una corriente teórica en particular, ya que las variables con las que uno puede contar vienen definidas previamente. En el mismo sentido, no se pretende refutar o validar mediante la modelización, cual o tal teoría, sino que las mismas contribuyan a explicar, entender y poner en contradicción los resultados del modelo aplicado. Por lo tanto, el marco teórico presentado en esta tesis pretende hacer un repaso por algunas de las teorías económicas que se enfocan en el vínculo educación-trabajo y las implicancias del mismo sobre los individuos y las sociedades, con el objetivo de exponer la articulación compleja, multidimensional e históricamente definida desde la que se aborda e interpreta aquí dicha relación.

#### **1.4 Justificación**

El estudio de la relación entre la educación y el trabajo, y las implicancias de dicha relación para los individuos y la sociedad en su conjunto, es un tema por demás interesante por su relevancia y alcance actual. Precisamente, el presente trabajo se justifica

principalmente en el hecho de que vivimos en un presente complejo y cambiante que exige constantemente ser investigado.

Los cambios y transformaciones acontecidos en las últimas décadas han llevado a poner en suspenso la idea de que la educación por sí misma constituye una herramienta de desarrollo, entendido este último en un doble sentido. Por un lado, desde el plano individual, el acceso a la educación facilitaba la movilidad social ascendente mediante el acceso a mejores puestos de trabajos, ya sea por una mejor retribución o por mejores condiciones laborales o ambas. Desde el plano de la sociedad en su conjunto, una educación impartida a un número cada vez mayor de ciudadanos generaba, por un lado, un aumento de la cohesión social y la fundamentación del sistema democrático mediante la difusión de valores, ideales, etc., y por el otro, mediante el incremento y acumulación de capital humano, lo que aumentaba la productividad del sistema económico y por ende, su crecimiento.

En el último “Global Education & Skills Forum” (GESF) realizado en Dubai este año, una de las mesas de debate giró en torno a que las habilidades que los empleadores requieren de sus empleados en el mundo actual están cambiando y no son necesariamente impartidas y fomentadas por las Universidades actuales. Este foro educativo internacional es considerado uno de los más importantes del mundo y este año los ejes del debate fueron “Educación. Equidad. Trabajo.” Allí se puso sobre la mesa la idea creciente de que el hecho de obtener un título universitario se está volviendo “irrelevante” a la hora de conseguir trabajo. Se planteó, como una paradoja, el hecho de que a nivel mundial haya aumentado el número de graduados y de universidades pero, al mismo tiempo, que los empleadores no encuentren en ellos las habilidades que necesitan. Entre los datos presentados que reflejan esta tendencia mundial, se destaca por ejemplo que en Estados Unidos, el porcentaje de graduados que accederán a un trabajo para el que no requieren título universitario asciende al 44% y en China, mientras que el 30% de los graduados en los últimos 15 años no encuentra trabajo, tan sólo el 4% de los que no tienen un título universitario se encuentran desempleados (Vázquez, 18 de marzo de 2016).

En el mismo sentido, el informe “¿Cómo repensar el desarrollo productivo? Políticas e Instituciones sólidas para la transformación económica” del Banco Interamericano de Desarrollo (2014), afirma que la escuela en América Latina y el Caribe no prepara a los estudiantes para el trabajo, a pesar de los grandes esfuerzos en educación y capacitación

que se han realizado en la región en los últimos años. La razón se encuentra según el informe en que las grandes inversiones realizadas fueron destinadas principalmente a ampliar la cobertura y en lograr mayores años de educación por habitante, sin tener en la misma consideración a la calidad de la educación impartida y la pertinencia o relevancia de las habilidades enseñadas. Por ello afirman que, si bien hubo una rápida convergencia en cuanto a la cobertura educativa y en los años de estudio de la fuerza laboral de los países de América Latina con los países desarrollados, dicha convergencia no se tradujo en una convergencia similar en productividad, ingresos y crecimiento económico. La hipótesis clave es que “[...] los años de estudios y capacitación (o de educación en general) no han generado los aprendizajes ni habilidades relevantes para el mercado laboral, comparables a los que han alcanzado otros países en condiciones similares.” (BID, 2014, p. 170)<sup>1</sup>.

Sin embargo, aunque la escuela no prepare para el trabajo, hay datos que nos demuestran que el bajo nivel educativo agrava el desempleo y que existen grandes diferencias en los salarios según el nivel educativo. Según el último informe de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), “Panorama de la Educación 2015”, más del 80% de los adultos con estudio universitario tienen empleo. Dicho porcentaje se reduce a un 70% en aquellos que poseen educación secundaria o media superior y menos de un 60% si no poseen educación media. Lo mismo ocurre con el nivel de ingresos, a mayor grado de escolaridad, mayores salarios (Organisation for Economic Co-operation and Development [OECD], 2015).

Según datos del “Bureau of Labor Statistics” (Oficina de Estadísticas Laborales), del Departamento de Trabajo de los Estados Unidos, en 2015 la tasa de desempleo en dicho país para quienes no terminaron la escuela secundaria era del 8%, casi el doble que el promedio general (4.3%). Incluso, la tasa de desempleo seguía estando por encima del promedio para aquellos que habían completado la secundaria y para quienes tenían grado de universitario incompleto. En cuanto a los ingresos, recién superaban el promedio general aquellos que contaban con grado de Licenciatura (Bureau of Labor Statistics, 2016).

En nuestro país, según datos de la Encuesta Permanente de Hogares del INDEC para el 2do trimestre del 2016 y teniendo en cuenta el total de aglomerados urbanos, si sólo se

---

<sup>1</sup> La expresión entre paréntesis es reproducida del original.

considera a la población con primaria completa, se encuentra ocupada un 47,0% de la misma. Dicho porcentaje aumenta a un 63,2% si sólo se considera a la población con estudios secundarios completos y alcanza un 77,2% si se tiene en cuenta sólo a la población con estudios superiores completos. Ahora bien, lo llamativo es que del total de ocupados según nivel educativo, el porcentaje más alto se encuentra entre aquellos que tienen secundario completo (27,1%), más alto incluso que el porcentaje asociado a los ocupados con estudios superiores completos (22,4%). Además, los ocupados con estudios superiores incompletos son el 13,5% del total, más bajo que los porcentajes correspondientes al total de ocupados con primaria completa (16,4%) y secundaria incompleta (16,4%) (INDEC, 2016).

Hay que destacar además, en el plano nacional, que la graduación universitaria se encuentra entre las más bajas del mundo, en donde de cada diez alumnos que ingresan sólo se reciben tres. Pero además, asociado con la idea de que existe una disociación entre las necesidades del mundo productivo y lo que el mundo educativo proporciona, existen en la Argentina muy pocos graduados en carreras estratégicas para el desarrollo de la productividad y competitividad de un país, es decir, técnicos e ingenieros. El perfil de los egresados sigue siendo muy cuestionado ya que de 100 abogados, se gradúan tan sólo 31 ingenieros (Braginski, 23 de septiembre de 2016).

Resulta evidente del análisis de estos datos que la relación entre educación y trabajo es compleja y multidimensional y que como todo proceso social se determina y configura según el tiempo y el espacio. Es por esta razón, que el principal justificativo que encuentro para este trabajo es el poder contar con información precisa, fidedigna y específica del contexto santafesino.

Además, el método de regresión multinivel que se utilizará tiene pocos antecedentes en este tema y menos aún, aplicados a los habitantes de la ciudad de Santa Fe. El análisis que se realizará nos permitirá saber, no sólo si existen diferenciales en el acceso al trabajo según el nivel educativo, el sexo o la edad, entre otras variables, sino también, si existen diferencias a lo largo del tiempo y variables contextuales que influyan en los resultados obtenidos. Y lo más importante, es que el modelo estimado tendrá en cuenta la estructura jerárquica de los datos utilizados. Precisamente, una de las críticas que se le realizan a muchos trabajos econométricos es no utilizar los modelos multinivel cuando la anidación existente es consecuente y genera dependencia en las observaciones (Bickel, 2007).

Asimismo, como ya se comentó, la información obtenida por el Panel de Hogares Santafesinos responde a una metodología no tradicional que intenta indagar en las percepciones y opiniones de los propios actores sociales. Esto es distinto a la información que puede obtenerse aplicando el paradigma positivista, para el cual, la generación de información adquiere un sentido finalista y generalista. Por eso, es diferente a la información generada por ejemplo por la Encuesta Permanente de Hogares que lleva a cabo el INDEC. En el panel de hogares, lo que se pretende es reconstruir la identidad social a partir de lo que la sociedad piensa de sí misma (Observatorio Social, 2015).

Es sabida entonces la importancia que tiene la educación para el hombre y la sociedad. El mundo nos está demostrando que cada vez son más los individuos, sobre todo los jóvenes, con un alto nivel educativo alcanzado y que no poseen trabajo, o que poseen un trabajo que no es acorde a sus competencias adquiridas. Esto no es sólo un problema a nivel individual, sino que el desarrollo mismo de las naciones puede verse comprometido. La teoría del capital humano postula que la inversión en capital humano es uno de los factores determinantes del desarrollo, pero no puede haber desarrollo si la “maquinaria” en la que se invierte está parada. Tal como se explica en este pasaje del informe de la Organización Internacional del Trabajo (2013) sobre las tendencias mundiales del empleo juvenil:

“Los costos económicos y sociales del desempleo, del desempleo de larga duración, de la falta de ánimo y de los empleos de baja calidad generalizados para los jóvenes siguen aumentando, socavando el potencial de crecimiento de las economías.” (p. 1)

Es por todo lo enumerado que considero pertinente la realización de una tesis de este tipo. Espero que la misma sirva para dar una mirada a la problemática actual de la sociedad santafesina en relación a la educación y al trabajo, al menos desde uno de sus múltiples aspectos. Es que para generar las políticas necesarias para solucionar los problemas de desempleo, deserción escolar y el desfasaje entre las competencias de los trabajadores y los puestos de trabajo que ocupan, es primordial contar con un conocimiento en profundidad de las causas y factores que explican dichos fenómenos en el contexto actual de nuestra ciudad.

## **1.5 Antecedentes**

Si bien a partir del análisis exploratorio realizado no se han encontrado antecedentes de la aplicación de modelos multinivel con datos longitudinales para estudiar posibles

diferencias en el acceso al trabajo en la ciudad de Santa Fe, hay que destacar dos cuestiones. Primero, que si bien no existen antecedentes de la aplicación específica de modelos multinivel, el tema del acceso al trabajo y la existencia de posibles diferenciales entre hombres y mujeres, entre diferentes niveles educativos, etc. ha sido abarcado por diversos autores en diferentes oportunidades. Como ya se comentó, la característica distintiva de este trabajo es la aplicación de una metodología poco implementada pero con grandes ventajas por sobre un análisis de regresión común cuando se utilizan datos con algún tipo de estructura jerárquica como es el caso. Segundo, la presente tesis tiene su génesis en un trabajo de Meyer, Díaz y Bulgarella (2014), titulado “Trabajo, edad, género y nivel educativo. Estimación de probables diferenciales entre jefes y jefas de hogar de la Ciudad de Santa Fe”.

En dicho trabajo, se aplicó un modelo logístico multinivel para analizar la existencia de posibles diferenciales en el acceso al trabajo según diferentes características de los individuos. Las principales diferencias con este trabajo es que el período de aplicación abarca los años 2005 a 2010 (a diferencia del período aquí analizado, 2009-2011) y una menor cantidad de hogares (406 contra 582), lo que nos brinda una mayor representatividad de la población santafesina. Además, hay diferencias en la recategorización de los niveles educativos, el tratamiento del tiempo y sobretodo en el procedimiento que se sigue para la identificación del modelo que mejor se ajusta a los datos.

Por otro lado, parte del marco teórico fue adaptado de un trabajo realizado por mí anteriormente en coautoría con Beker, Eberhardt, Gutiérrez y Perotti (2009). El mismo fue revisado y ampliado en concordancia con los objetivos de ésta tesis y utilizado con previa autorización de los coautores.

En cuanto al análisis multinivel, si bien aún se la considera una herramienta novedosa, debe decirse que a partir de la década del 80 su uso se ha extendido ampliamente, por lo que puede encontrarse una gran cantidad de trabajos aplicados, sobre todo en áreas como la medicina y la educación. En el ámbito econométrico son mayormente conocidos con el nombre de modelos de regresión de efectos aleatorios. Precisamente, la diversidad de denominaciones con las que se identifica a los diferentes modelos se debe en parte a la aplicación en distintas áreas del conocimiento. Para una introducción a la clasificación y las diferentes denominaciones de los modelos multinivel, sus principales áreas de

aplicación y una lista de referencia a las publicaciones más importantes relacionadas con los mismos, puede verse el trabajo de Amador y López-González (2007).

A su vez, existe un gran número de publicaciones que estudian diferentes aspectos de la relación entre educación y empleo, utilizando diferentes metodologías y apoyándose en distintas corrientes teóricas. Los trabajos de Llamas y Rubio (2015), Ventoso Mariño (2014), Freire Soane y Teijeiro Álvarez (2010) o Chevalier, Harmon, Walker y Yu Zhu (2004), por sólo citar algunos, ponen en discusión diferentes tópicos de dicha relación empleando diferentes modelos econométricos, que a pesar de ser distintos, tiene una particularidad en común: todos utilizan ecuaciones de regresión por mínimos cuadrados ordinarios, sin tener en cuenta que los datos empleados para la estimación de sus modelos cuentan con algún tipo de estructura jerárquica.

En referencia a esto, Bickel (2007) argumenta que muchos de los trabajos basados en la Teoría del Capital Humano sólo han puesto énfasis en las características de los individuos, es decir, que las consecuencias de invertir en uno mismo han sido tratadas como independientes de la variabilidad de los contextos sociales, políticos y económicos en dónde las personas viven.

En cambio, existen trabajos que sí han tenido en cuenta la existencia de estructuras jerárquicas en sus datos, como es el caso del estudio realizado por Dueñas Fernández, Iglesias Fernández y Llorente Heras (2014) quienes analizan la influencia que tienen los aspectos pre-market en el mercado de trabajo español, es decir, cómo condicionan de manera distinta las decisiones laborales de hombres y mujeres los factores sociales y sociológicos que influyen en la personalidad, analizando dos cuestiones, la segregación laboral y las diferencias salariales por género. Para ello utilizan diferentes modelos probit multinivel mediante los cuales se pretende identificar el peso explicativo de distintos tipos de variables (personales, familiares, laborales, antecedentes familiares y actitudes de género) en los procesos de asignación de las mujeres a las distintas ocupaciones laborales, concluyendo que los aspectos pre-market no ejercen una influencia significativa en la segregación ocupacional femenina, la que se encuentra fundamentada más bien en decisiones personales y el tipo de ubicación laboral detentada. Para el análisis de las diferencias salariales, utilizan ecuaciones multinivel para estimar el logaritmo del salario mensual para hombres y mujeres sobre la base de distintos tipos de variables para luego aplicar la descomposición de Oaxaca-Blinder en dos factores. Así pueden identificar cuál

es el aporte de las actitudes y valores en la explicación de las diferencias salariales entre hombres y mujeres, para concluir finalmente que los aspectos pre-market no son el principal aporte a la explicación de dichas diferencias, siendo los factores determinantes los relacionados al puesto de trabajo.

En el trabajo de Amara y Jemmali (2015) también se utilizan modelos multinivel para analizar simultáneamente los factores de nivel micro (los hogares) y de nivel macro (el barrio o la región) que podrían afectar la naturaleza y el patrón social de la pobreza en Túnez. Tal como ellos lo plantean, un análisis de la pobreza no puede sólo tener en cuenta factores individuales o de contexto sino que debería considerar ambos, ya que la experiencia indica que tanto el comportamiento de ricos y pobres y el contexto de los vecindarios como así también factores humanos y socioeconómicos que influyen en los medios de vida, todos están involucrados en la determinación de los patrones de pobreza de un país. Para este análisis, aplican diferentes modelos multinivel entre los que destaco un modelo logit multinivel que especifica la probabilidad de que un hogar viva o no por debajo de la línea de pobreza. La evidencia indica que la probabilidad de que un hogar sea pobre está positivamente correlacionada con el tamaño del hogar, el número de hijos por familia y el nivel educativo del jefe del hogar. En el análisis macro, una mayor tasa de desempleo por barrio está asociada a probabilidades más altas de pobreza, mientras que una mayor aglomeración industrial está asociada a una disminución de las chances de caer en la pobreza.

Otro trabajo que utiliza el análisis multinivel es el de Poston (2002), el cual estudia en qué medida ciertas características de capital humano y cultural afectan el nivel de ingreso de los inmigrantes asiáticos en los Estados Unidos. Utiliza un modelo multinivel ya que por un lado analiza variables a nivel micro, el individuo, y otras a nivel macro, el capital cultural, es decir, variables al nivel del país de procedencia de los inmigrantes asiáticos como por ejemplo, el nivel de instrucción medio de cada país y el porcentaje de población de cada uno de los países asiáticos que vive en Estados Unidos y se naturalizó. Precisamente, dichas variables al nivel de grupo resultaron significativas tanto para hombres y mujeres, así como también lo son la cantidad de años de educación completa y el status profesional al nivel individual. Lo interesante del estudio es que el efecto de correlación intra-clase es pequeño, tanto para hombres como para mujeres. Esto quiere decir, que la mayor proporción de la varianza ocurre al nivel individual y que tan solo un

porcentaje pequeño de la varianza en los ingresos anuales corresponde a las diferencias entre grupos de inmigrantes asiáticos de diferentes países.

Un trabajo similar es la disertación de Qin (2005) acerca de la influencia de factores individuales y contextuales sobre los salarios obtenidos por los chinos en el mercado de trabajo de los Estados Unidos. En el trabajo se llevaron a cabo los modelos multinivel para tres tipos de chinos: los descendientes de chinos, los nacidos en el extranjero pero nacionalizados norteamericanos y los nacidos en el extranjero no nacionalizados. Algunas de las características que utiliza en el análisis son la educación, la experiencia laboral y la capacidad en inglés al nivel individual, mientras que al nivel contextual utiliza el porcentaje de estadounidenses de origen chino, el porcentaje de origen asiático, el porcentaje de chinos que manejan su propio negocio, la tasa de desempleo, entre otras. Lo interesante del trabajo es ver cómo las mismas variables, tanto individuales como contextuales, afectan de diversa manera a las diferentes tipologías de chinos que el autor propone. A nivel individual los resultados en general apoyan la teoría del capital humano, mientras que al nivel macro los resultados son más heterogéneos, lo que sugiere, según el autor, que los trabajadores chinos con antecedentes de inmigración diferente se enfrentan al mercado laboral de manera diferente.

## Capítulo II

### Marco Teórico

En este apartado se intentará describir aspectos del vínculo que existe entre educación y trabajo a partir de los aportes de diversas teorías que tratan la temática. Se intentará además, dar cuenta de cuáles son las implicancias que tiene la educación para el individuo en particular y para la sociedad moderna en su conjunto, ya que no existe una concepción única entre la educación y el bienestar individual y colectivo que esta genera.

Como ya se comentó, en esta tesis se entiende a la relación entre educación y trabajo como una articulación compleja y multidimensional que se define históricamente, por lo que no se asumirá una perspectiva teórica en particular. Las diferentes teorías presentadas se irán exponiendo cronológicamente según su desarrollo en el tiempo, tratando de poner en evidencia cómo fue cambiando el (o los) sentidos que se le otorgan a la educación, sus implicancias y su conexión con el mercado laboral.

#### **2.1. La relación educación y trabajo. Aproximaciones Teóricas.**

Considero interesante comenzar esta sección desde la concepción del discurso pedagógico como instrumento para la emancipación del hombre en la modernidad. El elemento base de la sociedad moderna lo constituiría, precisamente, la razón. Y en ese sentido, tal como fue planteada desde comienzos de la Ilustración, la educación puede considerarse como paradigma del proceso de racionalización de la modernidad, en la que conviven dos discursos, uno utópico y uno burocrático (Terrén, 1999).

Es así que la educación era considerada el elemento emancipador y liberador del hombre de todo lo tradicional y el medio por el cual la razón llevaría a la Utopía del progreso. Pero desde la otra dimensión, la racionalización excesiva de todas las esferas sociales llevó a la burocratización de la educación, reivindicando la idea de control racional como funcionarización de la acción y control del destino, haciendo que las principales características de la institución escolar moderna sean el control social, el control del tiempo y la condena moral del ocio (Calvés, 2000).

Se encuentra cierto paralelismo entre estas dos dimensiones de la educación con la simplificación metodológica que realiza J. C. Hidalgo (2001) para definir la función de la Universidad desde dos enfoques diferentes.

Por un lado, como una unidad de producción en la que se considera a la capacitación un insumo y en la que el egresado profesional es considerado el output o producto final. De este modo para los neoclásicos, la Universidad y la educación en general es considerada una empresa y el modo en que se forma al individuo se asemeja a un proceso productivo y por lo tanto para su estudio se utiliza el instrumental teórico neoclásico.

Por otra parte, considera a la Universidad como una institución creadora de conocimientos que le pertenecen al conjunto de la sociedad, con el objetivo principal de formar ciudadanos libres. Desde este punto de vista, la misión de la Universidad consistiría en educar y generar conocimientos para permitirle al hombre integrarse al sistema democrático estimulando la autonomía del pensamiento y contribuyendo al ejercicio de la misma democracia (Hidalgo, 2001).

Continuando con las ideas de este autor, se puede decir que el campo teórico ha sido dominado por la concepción neoclásica de la educación desde que apareció hace ya más de 50 años la denominada *economía de la educación*. La formulación de la Teoría del Capital Humano de la mano de T. W. Schultz trasladó a la educación el análisis económico antes aplicado a las unidades de producción, y lo convirtió en el principal creador de lo que hoy se conoce con el nombre de la Economía de la Educación.

“El concepto de capital humano tiene en Schultz al menos dos dimensiones: en primer lugar se trata de la suma de conocimientos que adquiere una persona, que le permite desarrollar sus capacidades intelectuales en los ámbitos científico, tecnológico y cultural y que le posibilita el ingreso diferenciado respecto de quien no lo posee (...). En segundo lugar trata de investigar la relación entre educación y el desarrollo económico” (Hidalgo, 2001, p. 28).

A fines de la década del 60 se empieza a evidenciar que las ideas planteadas por esta teoría no se correspondían con la realidad, que la educación no producía los efectos de desarrollo económico esperados y que la conexión entre formación y remuneraciones era exigua, explica Aronson (2007). De esta manera surgieron corrientes teóricas nuevas que intentaron explicar de manera diferente la relación entre educación y empleo.

El credencialismo, como explica Recio (1997) parte de los mismos supuestos del enfoque neoclásico aunque considera la existencia de información imperfecta en el mercado, lo que

lleva a los empleadores a basarse en señales “observables”, como lo es un “título universitario”, al momento de la contratación.

Otra teoría alternativa es la de la correspondencia, propuesta por Bowles y Gintis (1975/1983) quienes desde una perspectiva marxista de la empresa capitalista, consideran a la producción como un proceso social y técnico, con dimensiones socio-políticas, donde el trabajo no es una simple mercancía y en la que la estructura de salarios no es exógena a la empresa sino uno de los instrumentos usados para maximizar sus beneficios. “La asignación de los trabajadores a los puestos de trabajo, la estructura de los puestos y la definición de los atributos del trabajador productivo no se pueden derivar simplemente, como harían los teóricos del Capital Humano, de un encuentro, mediado por el mercado, de las cualificaciones definidas técnicamente y de las necesidades de la producción definidas técnicamente. Las cuestiones de poder y, en definitiva, de clase entran en un nivel bastante fundamental” (p. 119).

De este modo, la teoría plantea que la educación no constituye la mejor forma de conseguir una mayor igualdad y movilidad social sino que por el contrario lo que hace es reproducir las desigualdades de clase, socializando a la población en los valores de la estrato dominante para cumplir así, en forma disciplinada, sus funciones productivas subordinadas en el mercado de trabajo (Medina, 1983).

Se propone entonces en base a estas teorías, describir cómo fue variando desde la modernidad hasta el presente el concepto de educación y las implicaciones que de allí se derivan, tanto sobre los beneficios, individuales como colectivos, como así también de política económica y las distintas relaciones que plantean con el mundo del trabajo.

## **2.2. Educación en la Modernidad**

En términos generales la modernidad ha sido el resultado de un vasto transcurso histórico, que presentó tanto elementos de continuidad como de ruptura; esto quiere decir que su formación y consolidación se realizaron a través de un complejo proceso que duró siglos e implicó tanto acumulación de conocimientos, técnicas, riquezas, medios de acción, como la irrupción de elementos nuevos: surgimiento de clases, de ideologías e instituciones que se gestaron, desarrollaron y fueron fortaleciéndose en medio de luchas y confrontaciones en el seno de la sociedad feudal. (Revueltas, 1990, “La modernidad como proceso histórico”, párr. 1).

En palabras de José Alsina Calvés (2000), la modernidad es un programa de transformación social que se ejerce en cuatro frentes: el político, el social, el

epistemológico y el pedagógico, donde cada uno de ellos puede ser identificado con ciertos procesos. En el ámbito político encontramos las revoluciones francesas y americana, el frente social definido por la reforma protestante y la revolución industrial y el epistemológico por la adopción del método científico como proceso para la obtención de conocimientos.

El frente pedagógico de la modernidad se fundamentaría en el programa de la Ilustración, como un proceso educativo que conecta razón, educación y emancipación destinado a la autonomía moral. Es decir, la clave de la reforma social radica en la armonía entre conocimiento y moral para alcanzar el bienestar social. (Terren, 1999).

La educación adquiere en la modernidad una nueva concepción,

cuando los autores del proyecto de una sociedad racionalmente educada creyeron haber sistematizado para siempre la posibilidad de conjugar el progreso moral y material de la humanidad; cuando creyeron haber encontrado en la educación nacional de la masa el instrumento decisivo para controlar el presente y conquistar el futuro. (Calvés, 2000, "Modernidad y pedagogía", párr. 2).

Siguiendo las ideas de este autor, la idea de educación durante el transcurso del desarrollo de la modernidad adopta diferentes matices pero nunca pierde de vista la noción de plasticidad del ser humano, que puede ser moldeado en forma pedagógica para hacerlo funcional y parte del proyecto social, y en concordancia con esto reconoce que ante cambios en el sistema educativo, el orden social se ve también modificado. Sin embargo, esta idea también es pensada en el sentido inverso, es decir, que detrás de las deficiencias del sistema social se esconden deficiencias en el sistema educativo.

El autor destaca un aspecto esencial del paradigma pedagógico y es que, si bien bajo una campana que suena al son de la liberación y la socialización pregonando por una cada vez mayor extensión de dicho paradigma pedagógico, se disimula la práctica burocrática de control y normalización. Para Terrén (1999) la configuración del espacio escolar se da en base a dos grandes dimensiones, la Utopía y la Burocracia. La dimensión Utópica se configuró debido al papel crucial que le otorgó el discurso ilustrado de la época a la educación, tanto como un mecanismo emancipador de la tradición y las creencias medievales, como así también, el generador de una voluntad colectiva instruida. Sin embargo, la extensión del proceso de racionalización a todas las esferas configuró el costado Burocrático, que se constituyó como una nueva forma de dominación tecnológica

y se ejerció, específicamente en el ámbito escolar, mediante el disciplinamiento, la administración y el control del espacio y el tiempo.

Es este segundo aspecto el que diferencia a la institución escolar de la modernidad con su antepasado medieval, ya que la escuela medieval si bien tenía como objetivo el ideal cristiano de la salvación del alma, la misma no estaba vinculada a ningún proceso productivo. “A medida que el proyecto educativo de la modernidad se va desplegando los aspectos relativos a la homogeneización y al control burocrático van tomando más protagonismo” (Calvés, 2000, “Modernidad y pedagogía”, párr. 12). De esta manera, las formulaciones pedagógicas de la modernidad, desde la dimensión “Burocrática” de Terrén (1999), generan las bases para que la institución escolar eduque a la población en la disciplina del trabajo fabril y en sus valores.

A principios del siglo XX se evidencia la consolidación del ámbito económico, con una sociedad industrial extendida como eje dominante y una fuerte creencia en la ciencia y en la técnica y una fe en el progreso. La enseñanza obligatoria paulatinamente se hace extensiva y cambia la concepción de la educación desde una idea de excelencia o perfección, a una instrucción generalizada, brindada al mayor número de sujetos en condiciones de homogeneidad.

### **2.3. Economía de la Educación**

En los treinta años gloriosos, la época de oro del capitalismo, el aumento de la productividad mediante la distribución de métodos de producción que ahorran trabajo se constituye en el motor de desarrollo. Este movimiento expansivo hacia el progreso vía industrialización y fortalecimiento de mercados internos, con un patrón de acumulación basado en la demanda interna y un desarrollo hacia adentro en que el Estado juega un papel fundamental, se va a convertir en el paradigma que exportará el occidente desarrollado a los países pobres o subdesarrollados, antes y después de la posguerra. (Ramírez Parra, 2002, p. 50).

En los años cincuenta comenzó a notarse que los factores productivos tradicionales (trabajo y capital) ya no explicaban totalmente el crecimiento económico, lo que llevo a diferentes autores a determinar que la parte no explicada del crecimiento podría atribuirse a la inversión en Capital Humano.

La Teoría del Capital Humano, fue inicialmente postulada por T.W. Schultz (1961) quien realizó estudios sobre los efectos de la educación en el desarrollo económico. Éste acuñó como variable de análisis relevante, el concepto de Capital Humano, otorgándole una

visión diferente a la aceptada hasta ese entonces, dando origen a lo que hoy se conoce como Economía de la Educación. La diferencia estriba en que los primeros economistas, como Smith (1776/1958), abordaron los efectos de la educación desde una perspectiva individualista (¿cómo contribuye la formación a la productividad individual del trabajador?). En cambio en los sesenta se empieza a estudiar los efectos de la educación sobre el crecimiento económico.

Capital humano tiene en Schultz al menos dos dimensiones: por un lado a modo de suma de conocimientos que adquiere una persona, y que le permiten desarrollar sus capacidades intelectuales y le posibilitan el ingreso diferenciado al mercado laboral respecto de quien no lo posee y por otra parte trata de investigar la relación entre educación y desarrollo económico. Este autor sostenía que los beneficios de la educación trascendían el ámbito individual, ya que la sociedad en su conjunto también se veía favorecida por los efectos positivos de la educación (externalidades). Estas ideas propiciaron la participación del Estado para solventar el gasto en educación en todos sus niveles. En el ámbito individual consideraba que la capacitación generaba satisfacciones presentes y futuras, elevando los ingresos esperados (Hidalgo, 2001).

Hidalgo explica que los beneficios más importantes de la educación según Schultz son:

- El fomento de la investigación.
- Descubrir, seleccionar y cultivar los talentos potenciales que de otro modo pueden permanecer ocultos.
- La flexibilidad de la fuerza de trabajo, es decir, el incremento en la capacidad del individuo para adaptarse a los cambios en las oportunidades de empleo asociadas al desarrollo económico.
- El reclutar y preparar estudiantes para la enseñanza.
- Como fuente importante del desarrollo económico.

Otro exponente de la Teoría del Capital Humano fue Gary S. Becker (1975/1983), quien complementó las ideas de Schultz y definió las múltiples formas del capital humano;

Las actividades que influyen en la renta monetaria y psíquica futuras aumentando los recursos de la gente (...) se denominan inversión en capital humano. (...) Sus principales formas son la educación, la formación en el trabajo, el cuidado médico, la emigración y la búsqueda de información sobre los precios y las rentas (Becker, 1975/1983, p. 39).

Si bien aclara que todas estas formas de inversión difieren tanto en los efectos que producen sobre el consumo y los ingresos, como en las cantidades que se invierten y la magnitud de los rendimientos esperados, aclara que todas ellas mejoran las cualificaciones del trabajador, sus conocimientos o su salud y, por lo tanto, aumentan su renta monetaria o su consumo.

De esta manera con los aportes de Schultz y Becker queda configurada la Teoría del Capital Humano, la cual establece, a nivel microeconómico, una relación causal directa entre educación, productividad y salarios, y a nivel macroeconómico, la relación entre el nivel educativo de la población y el desarrollo económico, es decir, la educación como fuente del desarrollo económico.

Enmarcada dentro de la teoría neoclásica, la educación sería considerada entonces un tipo de inversión destinada a aumentar la capacidad laboral futura de las personas. El individuo racional hace básicamente un análisis de inversión, considerando costos (los del gasto educativo propiamente dichos –directos-, mas el costo de oportunidad por no ingresar al mercado laboral en forma inmediata -indirectos-) y los ingresos futuros esperados. Siempre que la misma sea redituable, el individuo decidiría alcanzar un nivel educativo superior. De este modo, a lo largo de la vida activa, las personas con estudios ganan menos o tienen “ingresos negativos” en un principio, pero luego, los ingresos obtenidos por ser titulados superiores son mayores que los que presentan las personas sin estudios. En este sentido, el gasto en educación no es considerado sólo como consumo que satisface una necesidad presente, sino también como una inversión capaz de generar rendimientos en el futuro.

Como explica Albert Recio (1997), la teoría “supone que cuanto mayor nivel educativo se posee, más productivo es uno” (p. 106). En este sentido, los salarios se corresponden con los aportes productivos de la persona. La teoría propone entonces “una explicación de las diferencias salariales: las personas con mayores niveles educativos se supone que reciben mayores salarios debido a que la culminación de un proceso educativo ha acrecentado su productividad individual” (p. 106).

Esta relación causal directa en términos de educación, productividad y salarios que resalta la Teoría del Capital Humano, trasladada al ámbito macro, revierte en un mayor desarrollo económico; ya que éste depende del crecimiento de la capacidad productiva de los agentes

económicos. Por lo tanto la educación en este sentido genera beneficios tanto privados como sociales.

En el contexto de los progresos económicos que caracterizaron la segunda mitad de siglo, y sobre la base de los aportes realizados por Schultz y Becker, particularmente los relacionados con los beneficios de la educación en el desarrollo económico, es que se circunscribe la concepción educativa de la época. Como detalla Aronson (2007) estas nuevas ideas se hicieron palpables a través de una mayor participación del Estado que mediante políticas pretendió masificar la escolarización, alcanzar condiciones igualitarias de ingreso y a su vez el desarrollo de diferentes vocaciones profesionales. En consonancia con las ideas de la época, el Estado interviene fuertemente en la economía, adopta el papel de un agente económico racional cumpliendo funciones propias de una unidad de producción, planificando, coordinando, regulando, evaluando sus patrones de gastos minuciosamente, realizando inversiones previamente analizadas y adecuándose a las necesidades de la demanda social. Reorganizó prioridades, otorgándole al gasto en educación un lugar privilegiado gracias a la idea del capital humano como nuevo motor de desarrollo. El gran gasto de esta inversión educativa lo hizo recaer sobre los sectores privilegiados para favorecer la igualdad desde la base, en concordancia con las ideas del que luego se denominó “Estado de Bienestar”.

Según Hidalgo (2001), una de las principales funciones que debe cumplir el sistema público es “garantizar el acceso universal a la educación de las personas que integran la sociedad” (p. 44). La idea fundamental es lograr la igualdad de oportunidades para contribuir a la democratización al incluir al individuo en la sociedad y en el conjunto de sus conocimientos.

Estudiosos de la economía y sociología de la educación trabajaron en concordancia a fin de reorientar las ideas educativas, brindándole un carácter “instrumental” basado en “contenidos”, es decir, dotar al trabajador de una serie de conocimientos o “certidumbres” que le permitan optimizar su desempeño laboral e incrementar su futuro nivel de ingresos. El acoplamiento entre la división social del trabajo y el sistema educativo encontró su máxima expresión en la diferenciación de las ocupaciones que demandaba la organización fordista y la producción en masa (Aronson, 2007). Se genera una educación más práctica, acorde al curriculum y que respondiera, de esta manera, a la diferenciación de vocaciones y especialidades, consideradas vitales para el desarrollo. Pero no se trataba sólo de educar en

un ambiente económico en creciente especialización sino también en sociedades de posguerra, por lo que, desde un punto de vista más sociológico, se pensó a la educación como el principal factor para reforzar la cohesión social.

Por otro lado, uno de los sentidos que se le atribuyó a la educación fue el de constituir un medio eficaz para facilitar la movilidad social ascendente. La noción de meritocracia permitió al hombre creer en la posibilidad de alcanzar un status superior al que detenta valiéndose de sus propias capacidades, en una sociedad cada vez más jerarquizada y estratificada. Pero no es sólo el hombre a través de su esfuerzo personal quien se beneficia sino que la sociedad también se ve favorecida al intercambiar con el individuo ese esfuerzo por prestigio y lograr así una estructura social organizada. Partiendo de una “igualdad” desde la base, igualdad que el Estado intentaba garantizar, el status que el individuo alcanzase sería mérito propio, producto del esfuerzo personal y los conocimientos adquiridos mediante la educación.

Es gracias a este beneficio en particular y a las facilidades que otorgó en cierta medida el Estado, que la demanda social de formación experimentó un gran aumento, extendiéndose a sectores antes excluidos (mujeres, minorías raciales, etc.). Estos hechos fueron claramente evidenciados en el caso argentino. Como postula Torrado (2007) en el período en que la industria se consolida como eje del desarrollo en nuestro país se observa la expansión de la matrícula educativa en todos sus niveles y se le otorga un papel fundamental a la educación como principal canal para ascender socialmente.

La expansión de la demanda se debió en parte a que en un sistema capitalista estratificado de organización vertical, que retribuía al individuo, no sólo en términos económicos sino también sociales de acuerdo al status alcanzado, la ilusión de la época es entonces ascender en la pirámide social. La movilidad social intra e intergeneracional fue la utopía, el objetivo a alcanzar que dependería en primera instancia, de las calificaciones obtenidas por medio de la educación, las cuales permitirían, en un segundo plano, alcanzar los puestos de trabajo mejor remunerados para poder cumplir con dicho anhelo (Aronson 2001).

Sin embargo para otros autores (Terrén, 1999; Revueltas, 1990) es en este período dónde el contexto económico, la dimensión burocrática, comienza a ganar espacio y empieza a extender la racionalidad planificadora de los economistas de la época a todos los espacios sociales, entre ellos, la escuela. El nuevo capitalismo se caracterizará por un gran

desarrollo de la ciencia y la técnica, dónde se dará prioridad a la organización, a la planeación y a la racionalidad técnica, logradas gracias a la fuerte intervención estatal que ampliará e intensificará el control burocrático sobre la sociedad. En el marco educativo, el problema es conducido también, a un análisis de medios y fines, a un problema de cálculo racional. Es interesante el paralelismo entre la educación como burocracia y la educación funcional ligada a ideas neoclásicas, es decir, una educación que sea funcional a la reproducción del sistema, una educación instrumental que le brinde al hombre las herramientas adecuadas para desempeñarse en un mercado laboral basado en la producción en serie, el trabajo fragmentado, la línea de montaje y los protocolos taylorianos, desatendiendo al mismo tiempo el carácter emancipatorio de la educación.

#### **2.4. Teorías Alternativas**

El Fordismo fue –hasta mediados de la década del setenta- el modelo de producción dominante de los principales centros capitalistas y junto al Estado de Bienestar habían signado la época precedente. Claus Offe (1990) plantea una visión completa del asunto al afirmar que:

...el Estado de Bienestar ha sido el resultado combinado de diversos factores (...) El reformismo socialdemócrata, el socialismo cristiano, elites políticas y económicas conservadoras ilustradas, y grandes sindicatos industriales fueron las fuerzas más importantes que abogaron en su favor y otorgaron esquemas más y más amplios de seguro obligatorio, leyes sobre protección del trabajo, salario mínimo, expansión de servicios sanitarios y educativos y alojamientos estatalmente subvencionados, así como el reconocimiento de los sindicatos como representantes económicos y políticos legítimos del trabajo (p. 136).

La caída del Fordismo, como agotamiento del régimen técnico preexistente, junto al Estado Benefactor coincidió con causas externas tales como el resurgimiento de economías del sudeste asiático y la crisis del Petróleo (1973) entre otras (Rionda Ramirez, 2006).

Es en este contexto que surgen corrientes nuevas para explicar la relación entre educación y empleo. La relación causal directa entre educación, productividad y salarios que plantea la Teoría del Capital Humano no se evidenciaba en el plano empírico, ni tampoco los efectos de desarrollo económico que dicha relación produciría. Es esta cuestión la que intentarán explicar las nuevas teorías, dentro de las cuales se pueden citar al credencialismo (o teoría de la selección o del filtro), la teoría de la competencia por los

puestos de trabajo o de las “colas” y también la teoría de la legitimación o correspondencia de inspiración marxista.

El credencialismo plantea, al igual que la Teoría del Capital Humano, “que las empresas pagan salarios mayores o menores en función de la productividad individual. La diferencia estriba en que en este caso no se supone que la educación contribuya a aumentar la productividad de las personas, esta depende básicamente de capacidades innatas de cada uno, de orden genético” (Recio, 1997, pp. 108-109).

Sucede entonces que la empresa, al momento de contratar personal no conoce la verdadera productividad de los trabajadores y debe basarse entonces en “señales” observables, como lo es por ejemplo un título universitario. La teoría lo plantea entonces como un problema de información imperfecta, tal como postula Stiglitz (1975). De este modo, por el lado del empleador, éste analizará entonces al momento de la contratación la credencial o “etiqueta” que posee el aspirante, los costos y beneficios de dicha etiqueta, la institución que lo provee, la calidad de la misma, etc. La cuestión es que la información que refiere a las cualidades de un factor productivo o mercancía, es decir la “etiqueta”, a veces puede esconder “defectos de fabricación”. Sin embargo, como explica Medina (1983), el título o “etiqueta” de un individuo cumple la función fundamental de brindar información en un contexto de incertidumbre y es más, aún cuando los empleadores saben que la correlación entre títulos de educación superior y productividad puede ser exigua, igualmente se basan en ellos porque sirve como un mecanismo para establecer diferencias salariales. A su vez, visto desde la perspectiva del aspirante al empleo, éste también encuentra aquí incentivos para invertir en educación, es decir, los beneficios de obtener un mejor empleo o mayor salario al haber gastado recursos y tiempo en obtener un título universitario que aporte la más alta y prestigiada información al mercado de trabajo.

Según el credencialismo, la educación funciona entonces como un “filtro” selectivo y no como un proceso que mejora la capacidad productiva del individuo. Rawlins y Ulman (1974) proponen la existencia de un credencialismo fuerte y uno débil. Para el primero de ellos, al aumentar la oferta de mano de obra educada se produce a la vez un aumento paulatino de los estándares de contratación, exigiendo para una misma ocupación un nivel de educación cada vez mayor. El fenómeno, también conocido como devaluación de las credenciales, plantea entonces que la expansión educativa es incapaz de generar ganancias diferenciales. Mientras que para la versión débil del credencialismo, las dificultades de los

empleadores para conocer el futuro rendimiento de los postulantes, los llevan a basarse para la selección en la capacidad de los mismos como estudiantes, concediendo importancia a características no cognitivas sino más bien visibles de las personas, es decir, el sexo, la edad, la vestimenta, rasgos de personalidad, etc. (Medina, 1983; Albert Recio, 1997).

En el caso de la Argentina, el efecto de la devaluación de las credenciales educativas de nivel medio comienza a darse durante los años sesenta, debido principalmente a la expansión paulatina que sufrió la matrícula educativa sobre el conjunto de la población en años anteriores. El proceso de devaluación se acentuaría precisamente en la década del 70 y mostraría además un acceso diferencial de los distintos estratos sociales a cada nivel educacional (Torrado, 2007).

Otro representante del pensamiento de la época fue Lester Thurow (1972) quién, desde otro punto de vista, tampoco encontró explicación a las diferencias salariales en el vínculo entre duración del proceso formativo y posesión de capacidades intelectuales, ya que a su entender, estas diferencias estarían relacionadas con el funcionamiento del mercado de trabajo. Según el autor, el mercado se regía más por la competencia por puestos de trabajo que por la competencia de salarios, en un contexto de aumento de demanda educativa donde comenzaban a abundar los titulados en situación de subempleo o bien desempleados. A partir de los estudios que realizó, Thurow arribó a la idea de que a pesar de la expansión educativa, la riqueza no revelaba una mayor equidad distributiva.

Un aspecto crucial de la formación de un trabajador era la formación en el puesto de trabajo, es decir, la experiencia laboral. Para el autor el modo de organización del mercado de trabajo daba lugar a "colas de empleo" o listas de espera. Los empresarios clasifican a sus futuros empleados de acuerdo al tiempo y los recursos que le demandarían a la empresa durante el proceso de formación. La idea es lograr la mayor productividad posible con el menor costo de adiestramiento. Construye así una lista de empleo donde las primeras posiciones, es decir, los mejores empleos, serán ocupadas por aquellos individuos que probablemente le demanden menores costes de capacitación, según sus antecedentes y características. Es decir, en un mercado de trabajo basado en la competencia por los puestos de empleo, la función de la educación no es conferir una mayor habilidad y, por tanto, una mayor productividad y salarios más altos en el trabajador; sino más bien para certificar su "capacidad de entrenamiento" y conferirle así un cierto status en virtud de esta

certificación (Thurow, 1972, p. 68). Es planteado también como un problema de información, dónde la educación adquiere importancia al ser un indicador aproximado de las capacidades del individuo que el empleador usará para formar su lista, pero no será el único.

Esta crítica al Capital Humano introdujo dos novedades: por un lado, la existencia del fenómeno de sobre-educación de los postulantes y por otro lado, la divergencia entre nivel educativo y salarios. Thurow concluye que la productividad del trabajo no era sólo una función de la inversión educativa sino que dependía además “de medidas ligadas a la progresividad de las políticas fiscales, los controles salariales, los incentivos a las empresas y la activa intervención del sector público a fin de reducir las diferencias en la escala de retribuciones (Taberner Guasp, 1999: 232).” (Aronson, 2007, pp. 12-13). Pone en claro entonces que el vínculo entre educación y productividad no podía demostrarse con la información disponible y que la diferenciación salarial podría deberse a factores extraeconómicos, como la raza o el sexo, y no sólo al nivel educativo alcanzado.

Es importante destacar que, si efectivamente, tal como lo postuló Thurow, prevalece más una competencia por los puestos de trabajo en lugar de la competencia salarial, se ponen en duda las creencias acerca de los beneficios económicos de la educación y la eficacia de la educación como política social que contribuya a una mayor igualdad.

Un punto de vista similar sobre la eficacia educativa mantenía Raymond Boudon (1978), quién comprobó que la reducción de las desigualdades de oportunidades educativas no se correspondían con una atenuación de la “herencia social”<sup>2</sup> o una igualdad en el plano de las rentas, lo que sirvió para fundamentar el carácter dual de la educación como “bien de consumo” y “bien de inversión”. Entendida como bien de consumo, la educación era un objetivo fundamental de política social ya que era considerada un derecho al cual todos los individuos debían acceder. Sin embargo, como “bien de inversión”, no había nada que garantizara que el acceso del total de la población al capital educativo, un acceso igualitario, contribuyera a subsanar las desigualdades sociales. Advirtió entonces la necesidad de revisar los principios de la teoría del Capital Humano y las implicaciones de política que de ella se derivaban, considerando el hecho de que pese a su carácter

---

<sup>2</sup> La “herencia social” hace referencia a la inmovilidad social y profesional que se evidenciaba entre una generación y la siguiente, también definida como “inmovilidad social intergeneracional” (Boudon, 1978, p. 57)

meritocrático, lo que mostraba la realidad era una débil correspondencia entre el nivel de instrucción y la posición social. Sus postulados, ponían bajo lupa la relación fundamental entre los niveles educativos, la movilidad social y las retribuciones salariales.

En relación a la herencia o estructura social pero desde un punto de vista más radical, la Teoría de la correspondencia o legitimación plantea que la educación surge como respuesta a las necesidades económicas del capitalismo, planteando una mimetización de las estructuras y de las relaciones existentes en el mundo de la producción y en el mundo de la Escuela (relaciones de explotación y de dominación) de manera que se reproduce y legitima un patrón preexistente para entrenar y estratificar a la fuerza de trabajo. La educación era considerada un bien económico más, controlado por la burguesía dominante y que se hizo extensiva entonces para solucionar los problemas de control social que habían generado la urbanización e industrialización. La idea es que la función educativa socializa al estudiante en el estatus socioeconómico de los padres, perpetuando la estructura de clases, inculcando en el mismo valores como obediencia, respeto, responsabilidad y disciplina, en vez de proporcionarle conocimientos que le permitan comprender el mundo, conocimientos emancipadores que le brinden una verdadera posibilidad de movilidad social. Medina (1983) explica que, según la evidencia empírica que autores de la época como L. Thomas, (1956), L. Drake, (1972) y M. Keyser, (1974) supieron relevar, los empleadores al momento de la contratación se basan más en estos rasgos de obediencia, disciplina y formalidad de los individuos que en sus características cognitivas. Esto llevo a S. Bowles y H. Gintis (1976) a concluir que el sistema educativo ayudaba a integrar a la juventud al sistema económico debido a la correspondencia estructural existente entre ambos. Los mismos valores, normas, procesos y dinámicas se encuentran tanto en las instituciones educativas como en las organizaciones económicas burocráticas (Medina, 1983).

Se reconoce la función que cumplen la credenciales educativas en una economía capitalista, pero sostienen que se está dejando de lado la importancia de la clase y del conflicto de clases para dar cuenta de lo que ocurre en el mercado de trabajo. "... la estructura de salarios, los atributos del individuo valorados en el mercado de trabajo y las relaciones sociales del propio proceso educativo solo se pueden explicar mediante un análisis de clases explícito" (Bowles, S. y Gintis, H., 1975/1983, pp. 116-117).

Dado que la posesión de los medios de producción se encuentra en unos pocos, la mayoría de los individuos se ven forzados a vender su fuerza trabajo, así la educación, la instrucción laboral, la salud y el cuidado de la niñez juegan un rol importante en el proceso productivo, agregando que “Estos procesos no se pueden comprender si no se ponen en relación con los requisitos sociales para la producción de la estructura capitalista, de los individuos que la componen y las instituciones que la regulan” (Bowles, S. y Gintis, H., 1975/1983, p. 117).

De manera que una comprensión de los recursos humanos debe contener, por un lado, una teoría de la producción, como así también, de su reproducción social y esa es su crítica fundamental a la Teoría del Capital Humano, el pasar por alto las relaciones sociales de producción.

Partiendo de considerar que la producción posee una dimensión social y técnica, que consiste en la transformación de materia prima y de trabajadores cuyas cualificaciones y esfuerzos se utilizan para obtener beneficios, en el intercambio salario-trabajo no debe dejarse de lado el poder del capitalista por sobre el del trabajador ejercido por medio de un contrato. “El carácter de clase que tiene el proceso de producción se ve inmediatamente si se observa que los mecanismos de incentivos y de control y los tipos de coaliciones potenciales de los trabajadores dependen de las estructuras y propiedad de las empresas” (Bowles, S. y Gintis, H., 1975/1983, p. 119). La valoración que realiza un empleador no se limita a las cualificaciones técnicas sino que valorizan ciertos atributos (sexo, raza, edad etc.) utilizados para dividir y controlar potenciales coaliciones dentro de la organización capitalista.

Por lo tanto, para la Teoría de la Correspondencia la desigualdad económica y los niveles educativos de desarrollo que alcancen los alumnos serán definidos por el mercado, la propiedad y las relaciones de poder que definen el sistema capitalista. Lo económico domina y subordina a la educación, determinando que valores, normas y conocimientos es necesario que los niños aprendan.

## **2.5. Educación y Neoliberalismo**

En la década de los 70 el sistema capitalista atraviesa por una aguda crisis, aumenta el precio del petróleo y, en consecuencia, hay inflación y desempleo, así como estancamiento de las actividades productivas tradicionales, todo esto hace que la ideología del crecimiento ilimitado se vea seriamente afectada. Ahora bien, esta

crisis que pareciera volverse permanente se acompaña de profundas y aceleradas transformaciones que podrían marcar el inicio de una nueva época, cuyos rasgos empiezan a precisarse en la década de los '80. Los nombres para designarla varían: sociedad posmoderna; de consumo; del productivismo y la tecnocracia; postindustrial; del neocapitalismo; informacional; cibernética; "nuevo orden mundial", etc. (Revueltas, 1990, "Modernidad y mundialidad", párr. 1-2).

A partir de la década del '80 surge, de la mano de las ideas neoliberales que "ganan" el campo teórico, una nueva formulación de la Teoría del Capital Humano.

Alsina Calvés (2000) explica que con la crisis actual de la modernidad también se ha puesto en duda la capacidad de la educación como herramienta del desarrollo. A su vez indica que, paradójicamente, la sociedad le ha dado un lugar excepcionalmente importante, cargándola de tareas y atribuciones.

Dentro del mundo globalizado, las ideas neoliberales abarcan con fuerza todos los ámbitos sociales. No escapa de ellas el contexto educativo. Se pretende hacer de la educación, un bien de libre elección en el mercado, regulado por las leyes de la oferta y la demanda. El individuo consumidor (el alumno y su familia) deberá elegir de acuerdo a sus expectativas, gustos e intereses la educación que desee, dentro de una oferta global educativa, que competirá por diferenciarse en servicios y "calidad". El mercado queda configurado entonces como el ámbito natural donde se resolverán los problemas educativos. En cuanto a la Universidad en sí, Hidalgo (2001) afirma que su función de acuerdo a esta perspectiva es producir profesionales que den respuesta a las demandas del mercado. El autor considera a esta una "concepción individualista y utilitaria de la Universidad, que en gran parte ignora su función social" (p. 46). En palabras de Terrén (1999), el neoliberalismo asume la abolición de la burocracia y la libertad del individuo ("libertad" dentro de la lógica del mercado) como ejes principales. Por otro lado, la oferta educativa, despojada de todo tipo de regulación, mejoraría debido a la competencia que se generaría para atraer nuevos "clientes".

La concepción educativa de la nueva Teoría del Capital Humano presenta, en este contexto, diferencias respecto de la primera versión. Aronson (2007) cree que el carácter instrumental de la educación empieza a cambiar, dándole importancia al aspecto "inmaterial". La pedagogía de la época anterior intentaba especializar al individuo, brindándole habilidades técnicas y específicas para determinados oficios, con cuerpos de conocimientos definidos y estables. En el nuevo contexto mundial de economías abiertas y

de alta competitividad global, caracterizadas por la “incertidumbre”, las viejas competencias laborales no alcanzan. La nueva sociedad del conocimiento, de alto cambio tecnológico y velocidad de la información, hicieron necesaria una nueva concepción educativa que mutó hacia la competencia, en brindar conocimientos que le den al hombre ductilidad para resolver situaciones complejas, que despierten en él la creatividad, la autonomía y la capacidad para ser líder. En esta sociedad de la información y la tecnología, los conocimientos quedan obsoletos en poco tiempo, por eso es necesario una educación permanente y además de carácter “expresivo”, es decir, socializador de la educación, con el objetivo de despertar en él sus habilidades comunicativas y de relación, ya que también es necesario salvarlo de la crisis de legitimidad actual. Así, concluye Aronson (2007), “el conjunto configura los contornos de una formación orientada hacia las “competencias de empleabilidad” (p. 15).

Es que además, el individuo ya no debe sólo formarse y capacitarse para crear su identidad laboral y alcanzar puestos mejor retribuidos, sino que tiene la obligación de generar él mismo la forma de insertarse socialmente (Tedesco, 2000). Las formas de cohesión social e integración anteriores están rotas y el mejor ejemplo de ello es el trabajo. El trabajador ya no puede confiar en trazar su trayectoria laboral a largo plazo debido a la variabilidad de las competencias profesionales y a las políticas de flexibilidad laboral neoliberales, por lo que es consciente de que el nuevo mercado laboral lo obligará a cambiar de trabajo repetidas veces. El proceso de globalización, las políticas desreguladoras, la terciarización de la economía, han ido formando un mercado laboral generador de puestos de trabajo inestables, precarios y escasos, en el cual, incorporarse al mismo, ya no es sinónimo de movilidad social ascendente ni el principal mecanismo de inclusión social (Minujin, 1999). La competencia de educación alude ahora a una movilidad cognitiva y personal indispensable para mantener la posición, que permita al individuo “aprender a aprender” y así encontrar en el mercado los “pocos” lugares redituables. La sociedad pasa a estar organizada en forma horizontal o de red, “se trata de un entramado de carácter móvil que redefine continuamente su estructura, los vínculos que se establecen tienden a ser frágiles y fugaces, a diferencia de los fuertes y permanentes nexos que suscitaban las asociaciones duraderas” (Aronson, 2007, p. 17).

En los años '90, la Argentina sufre altos niveles de desocupación que llevan a una devaluación extrema de las credenciales educativas. Muchísimos ciudadanos quedaron

desocupados y sus conocimientos obsoletos de acuerdo a las demandas del mercado. La situación se volvería aún más preocupante, ya que los gastos educativos necesarios para este tipo de formación debían ser realizados por los propios individuos.

Características como la igualdad de oportunidades y la posibilidad de ascender de un estrato a otro que eran otorgadas a la educación y la convertían en la herramienta ideal para superar todo tipo de trabas como el sexo, la raza y el origen social, contrastan con las actuales políticas dirigidas a la efectividad y calidad educativa. Desde este enfoque, se debe garantizar un acceso universal al conocimiento pero el Estado subsidiará solo a escuelas efectivas, con profesores efectivos, dispuestas a competir en el “mercado” del conocimiento. Como bien aclara Aronson (2007), “si antes la igualdad de oportunidades significaba acceso amplio a la educación para contar con calificaciones laborales adecuadas, hoy se hace depender de la naturaleza de la educación que se imparte” (p. 20).

El objetivo es educar a la población en su totalidad, lograr que los jóvenes completen su formación básica, reduciendo por lo tanto la deserción escolar, y a la vez ofrecer oportunidades de completar sus estudios a la población ocupada que no cuente con la educación obligatoria. Pero como dijimos, el Estado debe influir sobre la calidad de la educación que se imparte, ya que se cree que los bajos rendimientos y la deserción escolar son producto de la escasa efectividad de la docencia. Por ello, de acuerdo a los postulados neoclásicos, las inversiones públicas deberán ser destinadas a instituciones eficientes, emprendedoras, con buena calidad del producto educativo y con personal altamente calificado. A su vez se debe fomentar el gasto privado en educación, formando un verdadero mercado educativo dónde los profesores compitan entre sí por los puestos de trabajo y los centros educativos hagan lo mismo para atraer más matrícula.

## **2.6. Breve conclusión sobre las teorías presentadas**

El recorrido recientemente realizado sobre las distintas teorías no debe tomarse como un recorrido exhaustivo, ya que no es ese su objetivo, como así tampoco se deben considerar a las mismas como las más “importantes”. Lo que se pretendía, era dejar en claro que la concepción de la educación, sus beneficios y su relación con el mundo del trabajo, se definen y re-definen históricamente, tal como se lo definió al comienzo de este trabajo. Es por ello que no existe y, por ende, aquí no se asume, una concepción única entre la educación y el mundo del trabajo, y el bienestar individual y colectivo que de allí puede

derivarse. Adoptar una postura, implicaría aceptar una concepción de la educación, del gasto incurrido en la misma, de la participación y el rol del Estado en la financiación y su forma de instrumentación, del acceso irrestricto e igualitario o no, de sus beneficios individuales y sociales, etcétera. Precisamente, lo que marca la historia es la inexistencia de una unicidad de criterios, hecho que se refleja en la continua transformación de las ideas.

Sin embargo, en base a lo recientemente expuesto se podría decir que, a grandes rasgos, el vínculo entre la educación y el bienestar que la misma implica, puede ser analizado desde dos dimensiones: La educación en abstracto (como fuente de emancipación) y la educación vinculada a la reproducción -ampliada- de la vida (como inversión en capital humano del *homo economicus*, como sistema de señalizaciones, como mecanismo de reproducción de la estructura social y del reparto del trabajo que esta estructura tiene asociado).

La relación entre educación y sus beneficios está hoy en el foco de la discusión. Ya no resulta fácilmente evidente, como lo fue tiempo atrás, que la misma sea el principal instrumento para garantizar la movilidad social ascendente, es decir, para romper con la "herencia social". Hoy en día, es puesto en duda incluso, conseguir una salida laboral acorde a las cualificaciones obtenidas. Particularmente creo además, que existe una deuda pendiente con el ideal utópico de la misma, como generadora de conocimientos que enriquezcan a toda la sociedad y formadora de hombres libres y plenos de derecho.

Sin embargo, pensar en un futuro sin educación, es pensar prácticamente en un futuro sin oportunidades. Por eso, en el mundo actual, marcado por profundos procesos de transformación en el que se están articulando nuevos modos de producción y nuevas formas de trabajo, considero que es muy importante pensar qué educación queremos para nuestra sociedad, sin dejar de desatender los beneficios de la misma en ninguna de sus dos dimensiones.

## Capítulo III

### Modelos Multinivel

El objetivo de este apartado es introducir al lector en los llamados Modelos Multinivel. En esta tesis, el modelo multinivel a utilizar tiene ciertas características que hacen que el mismo no se encuentre entre los modelos básicos, a partir de los cuáles, puede pensarse en ir incorporando progresivamente los conceptos que estos modelos implican. Por ello, se realizará primero un acercamiento a los Modelos Multinivel que pretende introducir en dichos conceptos para comprender luego más fácilmente el modelo a desarrollar en este trabajo. En este sentido, se realizará un acercamiento sencillo que no incluirá desarrollos matemáticos complejos. El álgebra y la matemática se dejarán para el final del capítulo dónde se expondrán las características propias del modelo a utilizar, sin embargo, la misma no llevará grandes complicaciones en su comprensión ya que no es ése el objetivo de esta tesis sino su aplicación e interpretación de los resultados finales.

#### **3.1. Modelos Multinivel (o Modelos de Efectos Aleatorios o Modelos Lineales Jerárquicos Generalizados y otras denominaciones)**

En la vida real, las personas en su vida cotidiana interactúan con el contexto, influenciando pero a su vez, siendo influenciadas por los grupos sociales a los que pertenecen. Existen en la sociedad, estructuras del tipo jerárquicas, en la que individuos y grupos son identificados, categorizados, ordenados, etc. En las investigaciones sociales, por lo tanto, es muy común intentar analizar la relación existente entre los individuos y los contextos o grupos en los que se desenvuelven.

La existencia de estructuras jerárquicas en los estudios sociales, llevó en los años 80 a desarrollar y profundizar técnicas de estimación que reconozcan y tengan en cuenta la presencia de dichas estructuras en los datos. El problema era que los modelos de regresión tradicionales, se basan en el supuesto de que las observaciones y/o sus residuos correspondientes, son independientes e idénticamente distribuidos pero, precisamente, la existencia de datos jerárquicos viola este supuesto fundamental. Específicamente, el problema con la no independencia de las observaciones en los modelos de regresión múltiple es que invalida sus métodos de estimación, subestimando los errores estándar y conduciendo a resultados significativamente espurios. Los modelos multinivel, también llamados modelos jerárquicos, es el término general que reciben los modelos estadísticos

que surgieron específicamente para dar respuesta a este problema, reconociendo explícitamente dicha estructura existente en la información y siendo por tanto, los apropiados para el análisis de este tipo de datos (Hox, 1995; Goldstein, 1999; Snijders, 2003).

Los modelos multinivel surgieron concretamente en la investigación educativa sobre rendimiento escolar pero se hicieron extensivos a otras ciencias como la economía, sociología, biología, etc. Precisamente, existen ciertas confusiones relativas a la terminología con la que cada disciplina denominó a los modelos multinivel, a pesar de que todos representan prácticamente los mismos modelos. En sociología o ciencias políticas son conocidos como modelos lineales jerárquicos o modelos multinivel. En biología o bioestadística son conocidos como modelos mixtos, de efectos mixtos o modelos de componentes de varianza, mientras que en economía son llamados, generalmente, modelos de efectos aleatorios.

La confusión se origina además porque en realidad existen tres modos de pensar los Modelos Multinivel:

1. Como una generalización de los modelos de regresión tradicionales.
2. Como una derivación de los análisis ANOVA.
3. Como modelos de variables latentes.

Sin embargo, se puede resumir que los Modelos Multinivel son básicamente una de las muchas técnicas y procedimientos que se han ido desarrollando a lo largo del tiempo como respuesta a alguna de las violaciones de los supuestos de regresión tradicional por Mínimos Cuadrados Ordinarios. Tal como explica Bickel (2007), con el tiempo, el uso de estos desarrollos se vuelve rutina y las técnicas se van perfeccionando, logrando así mejores y más aplicables modelos de regresión. Este es el caso de los Modelos Multinivel, que pueden ser entendidos como una mejor manera de hacer análisis de regresión bajo ciertas circunstancias. Es decir, cuando se tienen datos con algún tipo de organización jerárquica.

### **3.2. ¿Porqué utilizar modelos multinivel?**

Cuando se tienen datos jerárquicos, la información está contenida en diferentes niveles. Por ejemplo, estudiantes agrupados en aulas, votantes agrupados en jurisdicciones, pacientes agrupados en hospitales, etc. La idea básica de todos estos ejemplos es que se tienen

individuos en un nivel inferior, agrupados en un contexto o nivel superior. El problema crucial en el análisis de estructuras jerárquicas de este tipo, es la dependencia de las observaciones en el nivel más bajo de análisis, debido a la existencia de diferentes “contextos” a los que pertenecen dichas observaciones.

Como decía anteriormente, los modelos multinivel surgieron en el ámbito educativo. Tomando el ejemplo de estudiantes agrupados en aulas, lo primero que se debe hacer al tratar de analizar datos que contienen variables de individuo (estudiantes) junto con variables de contexto (aulas) es determinar la unidad de análisis. Por un lado, se pueden tomar los datos de los estudiantes en forma individual, en los cuáles existirán además ciertas variables de contexto (como características del profesor, la cantidad de horas de cada materia, características del colegio, etc.). O bien, se pueden tomar como unidad de análisis el aula o colegio, en cuyo caso los datos de los estudiantes serán agrupados y promediados, obteniendo así valores promedio por aula o colegio.

Ambos acercamientos conducen a errores ya que se están analizando variables de diferentes niveles en un único nivel. Por un lado, si se toman los datos individuales (estudiantes) y se sacan conclusiones a un nivel superior (aulas o escuelas) se puede caer en la falacia atomista, mientras que si se analizan los datos a un nivel superior y se sacan conclusiones a un nivel inferior, se cae en la llamada falacia ecológica, atribuyendo incorrectamente características de los contextos a los sujetos. Otra forma de entender esto es que en realidad, lo que se está analizando son dos unidades de estudio diferentes, los estudiantes por un lado y las aulas o escuelas por el otro, es decir, son dos muestras distintas.

Además de las fallas conceptuales, hay problemas de índole estadístico. Si se utilizan las medias agregadas, se pierde mucha información y potencia en el análisis estadístico. Si en cambio se utiliza la información individual pero no se tiene en cuenta la existencia de los grupos, se consideran a las observaciones como independientes cuando en realidad no lo son y este es casi siempre el caso en los datos multinivel. De este modo, existirán factores inobservables que afectarán a todos los individuos de un mismo grupo y generarán correlación en los residuos. Los errores calculados de este modo, serán subestimados y eso hará que se obtengan resultados significativos cuando en realidad no lo son (error de tipo I) (Hox, 1995; Goldstein, 1999; Hox y Mass, 2005).

La solución son los Modelos Multinivel que modelan explícitamente la forma en que los datos se encuentran anidados. Esto permite, desde un punto de vista estadístico, obtener estimadores estadísticamente eficientes de los coeficientes de regresión, como así también, errores estándar, intervalos de confianza y test de significancia que tienen en cuenta la presencia de clústeres, brindando así resultados más conservadores que los obtenidos por los métodos tradicionales de regresión que ignoran la existencia de dichos grupos (Goldstein, 1999).

Pero lo más interesante de los Modelos Multinivel es que permiten analizar simultáneamente las relaciones que pueden existir entre los individuos y los contextos en los que se desarrollan. Más específicamente, si se tienen variables medidas en los diferentes niveles de análisis, permiten modelar estadísticamente la influencia de las distintas variables y sus niveles (dos, tres o más), generando simultáneamente un modelo propio para cada subnivel. De este modo, representan la estructura jerárquica de los datos, donde cada sub-modelo expresa la relación entre las variables correspondientes a dicho nivel, pero permitiendo además ver cómo las variables de un nivel pueden influir en las relaciones de otro nivel, mediante las interacciones de variables de distintos niveles. Lo más importante es que además, con los modelos multinivel, es posible dar cuenta y encontrar explicaciones a la variabilidad que pueda llegar a existir entre los diferentes contextos (Hox y Maas, 2005; Bickel 2007).

### **3.3. Un acercamiento a los Modelos Multinivel**

En este trabajo se entenderá a los modelos multinivel como una generalización de los modelos de regresión tradicionales, tal como lo dice Bickel (2007), *los modelos multinivel pueden ser vistos como una mejor manera de hacer análisis de regresión bajo ciertas circunstancias. Dichas circunstancias son aquellas en que las observaciones, como por ejemplo estudiantes, son anidadas o agrupadas en contextos identificables, tales como clases, escuelas o distritos* (p. 8; traducción propia, las negritas y cursivas son del autor).

Siguiendo a este autor, se verá brevemente cómo pasar de la regresión tradicional de Mínimos Cuadrados Ordinarios (MCO) a los Modelos de Regresión de Coeficientes Aleatorios para finalizar en los Modelos Multinivel.

La diferencia principal entre la regresión tradicional por MCO y Modelos de Regresión de Coeficientes Aleatorios o Modelos Multinivel es el hecho de que el intercepto y/o las

pendientes no tienen por qué ser fijos, sino que pueden variar de un grupo o contexto a otro. El término aleatorio refiere justamente al hecho de que los valores de los coeficientes no son fijos sino que son seleccionados de una distribución de probabilidad. Dicha distribución de probabilidad es referida a una variable aleatoria, es decir, el coeficiente.

### **3.4. Componentes implícitos en los Modelos Multinivel**

#### **3.4.1 El fenómeno de la anidación y sus consecuencias**

En los modelos de regresión por MCO en los que se intenta encontrar el vínculo entre una variable dependiente y una o más variables independientes, los coeficientes estimados para la ordenada al origen y los predictores son fijos. Es decir, se asume que no varían entre los diferentes grupos y contextos. Sin embargo, el término de error si es aleatorio, ya que su valor difiere de una observación a otra. Según los supuestos asumidos en las regresiones por MCO, los residuos son incorrelacionados y homogéneos, ya que se distribuyen en forma independiente y con varianza constante.

Ahora bien, ¿por qué se debería permitir que los coeficientes varíen? O mejor dicho, ¿por qué podrían variar los coeficientes? La respuesta está en la naturaleza de los datos con los que se trabaja. La razón de ser de los modelos multinivel es el hecho de que las observaciones están anidadas en grupos. Al existir grupos, existen contextos diferentes. Al existir contextos diferentes, pueden existir explicaciones diferentes para las mismas relaciones.

Se supone que se está ante la presencia de datos con algún tipo de estructura jerárquica, como por ejemplo, estudiantes agrupados en escuelas. En este caso, los estudiantes son el nivel uno, y las escuelas son el nivel dos, la variable de grupo o contexto. Para simplificar, se supone que se quiere determinar la relación entre las horas de estudio dedicadas por cada uno de los estudiantes y las notas obtenidas en algún examen puntual. En este caso, la variable dependiente es la nota obtenida y la variable independiente son las horas de estudio.

Mediante MCO, es muy sencillo obtener la ecuación de regresión de nivel individual para el ejemplo anterior, ¿pero qué sucede si se desea introducir una variable de contexto, como por ejemplo, si la escuela es pública o privada?

En este caso se tendría, no sólo una variable independiente al nivel del individuo (el número de horas de estudio de cada estudiante), sino también una variable independiente al nivel de grupo (el carácter público o privado de la escuela). Sin embargo, el análisis de regresión por MCO tratará ambas variables como si fuesen características del individuo, cuando en realidad no lo son. Esta es la forma de trabajar del análisis por MCO, combinando variables de diferentes niveles en un solo nivel, sin tener en cuenta la estructura jerárquica de los datos.

En el análisis multinivel las variables de diferentes niveles no son simplemente agregadas en la ecuación de un solo nivel, sino que las mismas están unidas de forma que se hace explícita la existencia simultánea de distintas ecuaciones de regresión, una para cada nivel de análisis. Luego, las distintas ecuaciones se combinan en una sola, dando como resultado una sola ecuación completa que es similar en apariencia e interpretación a las ecuaciones de regresión de MCO.

Resumiendo entonces, al existir datos con algún tipo de estructura jerárquica, existirán diferentes contextos que pueden influir de distinta manera en las distintas observaciones. Lo importante aquí es que entonces, además de variables de nivel individual (nivel uno), existen variables del nivel de grupo (nivel dos), llamadas variables contextuales. Este tipo de variables pueden tomar muchas formas pero lo importante es entender que, sin el fenómeno de la anidación, no existirían contextos que difieran entre sí y que pueden ser medidos de forma que permitan la creación de variables contextuales (Bickel, 2007).

Es necesario aclarar, sin embargo, que la existencia de datos agrupados y variables contextuales no es un requisito “obligatorio” para utilizar modelos multinivel, ni mucho menos. Históricamente, se ha utilizado la regresión por MCO para trabajar con datos de este tipo. Sin embargo, también se verá porqué, bajo ciertas circunstancias, existen enfoques mejores que la regresión por MCO.

Retomando el ejemplo anterior, se podría decir que la variable de contexto, que define si una escuela es pública o privada, es también una característica del individuo. Efectivamente, se la puede considerar de ese modo y realizar una regresión, al nivel del individuo, por MCO.

Pero también se podría analizar la relación entre la media de las notas de los alumnos de cada escuela y la media de las horas de estudio dedicadas por los estudiantes para cada

colegio. De este modo, se realizaría una regresión por MCO al nivel de la escuela y la variable que define si un colegio es público o privado, sería analizada al mismo nivel.

Ahora bien, en cualquiera de los dos casos, todas las variables son medidas y analizadas al mismo nivel, es decir, no se toma en cuenta la anidación de los datos. En el primer caso, todas las variables son consideradas características del individuo, mientras que en el segundo, todas son consideradas características del colegio. Por más de que la variable sea medida al nivel de grupo y caracterizada como una variable contextual, la naturaleza de nivel único propia del análisis de regresión múltiple por MCO, hace que analíticamente todas las variables ocupen el mismo nivel de análisis. Además, se estarían cometiendo los problemas mencionados en el apartado anterior, de índole conceptual (falacia atomista y ecológica, respectivamente) y estadístico (dependencia de las observaciones, residuos correlacionados, errores estándar subestimados, etc.).

En cambio, los modelos multinivel permiten trabajar con variables medidas y especificadas a diferentes niveles, reconociendo explícitamente dicha realidad. Pero además, permiten considerar a una variable contextual, como el tipo de escuela, de las dos formas, es decir, como una característica del individuo y como una característica de la escuela en el mismo modelo, ya que admiten trabajar con más de un nivel de análisis.

### **3.4.2 La existencia de grupos y la heterogeneidad inobservada**

En el ejemplo que se está desarrollando, se comenzó diciendo que existe un determinado número de niños agrupados en escuelas. Hasta el momento, no se ha dicho nada concreto respecto a los grupos en sí, o a la posibilidad de que la existencia de dichos grupos genere relaciones diferentes entre la variable dependiente y las independientes, es decir, que exista variabilidad entre escuelas. Se puede pensar ahora que el número de escuelas en que los niños están agrupados es 10. ¿Cómo se podría tener en cuenta esta información en una ecuación de regresión múltiple por MCO?

Al comienzo del capítulo se dijo que al analizar información de tipo jerárquica, existen en los datos características inobservables en cada grupo que influyen (o no) de alguna manera a las observaciones pertenecientes a cada uno de ellos, haciendo que las mismas no sean independientes. Esto viola uno de los supuestos fundamentales de cualquier análisis de regresión. Será necesario entonces, tener en cuenta la existencia de dichas características y encontrar alguna forma de hacerlas explícitas en el modelo, para evitar que parte de dicha

dependencia se traslade a los residuos, a la parte no explicada en el modelo, generando correlación en el mismo. Por lo tanto, el objetivo es que el modelo a estimar contemple, en la medida de lo posible, dichas características comunes entre las observaciones anidadas.

Al tener datos jerárquicos, las observaciones pertenecientes a un mismo grupo serán entonces influenciadas por características inobservables propias de su grupo, que harán que las mismas tiendan a ser más similares entre sí y que se diferencien a su vez de las observaciones pertenecientes a otro grupo. De este modo, es muy posible que ante la existencia de datos jerárquicos existan dos tipos de heterogeneidad inobservada, la variabilidad “within-subject” o intra-sujeto y la variabilidad “between-subject” o entre-sujetos. Esto quiere decir que la variabilidad total existente en los datos provenientes de una estructura jerárquica es la suma de la variabilidad del nivel uno (la variabilidad al nivel del individuo) más la variabilidad al nivel dos (o la variabilidad existente al nivel de grupo)<sup>3</sup>.

Se ahondará en esto al retomar la pregunta que quedó sin responder, ¿cómo se podría tener en cuenta en una ecuación de regresión, la existencia de 10 escuelas diferentes?

### 3.4.3 El Modelo de Efectos Fijos

Respondiendo a la pregunta anterior, se podría mediante un análisis de regresión por MCO introducir nueve variables dummy, una por cada escuela, dejando una escuela por fuera como referencia. Para un alumno en particular, una de dichas variables tomará el valor uno, la que corresponda con el colegio al que asiste, mientras que el resto tomará el valor cero. Al introducir una variable dummy por cada escuela, lo que se está haciendo es modelar explícitamente las características inobservables de la misma. Es decir, se capta la variabilidad perteneciente a cada grupo, en este caso, a cada escuela.

Este método es también conocido como modelos de “Efectos Fijos”. Cada coeficiente de la variable dummy es como un “efecto fijo” que intenta captar las características inobservables de cada grupo que causan la no independencia de las observaciones. Al hacerlo, se hacen explícitas en el modelo las diferencias entre los grupos y se eliminan, al mismo tiempo, las razones que causaban correlación en el término de error y sesgaban el

---

<sup>3</sup> Se recuerda que es posible la existencia de estructuras jerárquicas de más de dos niveles. En dicho caso, para cada nivel existente, existirá una porción de variabilidad propia de cada nivel que puede ser desagregada y que explicará la variabilidad total del modelo.

valor del resto de los coeficientes. Además, se elimina también del término de error, la variabilidad relativa al nivel de grupo, es decir, la variabilidad “between”, ya que todas las posibles diferencias entre grupos quedan cuantificadas por el coeficiente de cada variable dicotómica.

De este modo, este tipo de análisis considera que existe un intercepto o término constante diferente para cada grupo, asociado con cada una de las variables dummy con coeficiente específico que intenta captar justamente las características inobservables propias de cada clúster. Cada variable dummy representará cada uno de los grupos del segundo nivel, bajo el supuesto fundamental de que los efectos de grupo son *independientes* entre sí (Mayorga y Muñoz, 2000; Urbina e Ibarra, 2013).

Pueden existir varios problemas con este tipo de enfoque. Por un lado, cuando el número de grupos es grande, también lo son el número de variables dummy que hay que generar y los coeficientes a estimar de cada una de ellas, lo que puede generar problemas de multicolinealidad. Además, se seguirían considerando variables contextuales o del segundo nivel como características del individuo, lo que resultaría en una especificación inadecuada del modelo.

Pero lo más importante aquí es que un modelo de efectos fijos, también conocido como Modelo de Mínimos Cuadrados con Variables Dummy (LSDVM, por sus siglas en inglés), no permite que las relaciones entre la variable dependiente y las variables independientes, varíen entre las categorías de una u otras variables. Es decir, que se asume que todos los coeficientes son fijos, o sea, son los mismos para todos los grupos.

La cuestión aquí es que un modelo de efectos fijos intenta encontrar diferencias significativas entre las medias de cada grupo. Básicamente, se intenta comprobar si las características inobservables del grupo, que causan homogeneidad al interior, producen al mismo tiempo que el grupo difiera o no de la media general. Por ello, en un enfoque de este tipo, el centro de atención es el comportamiento específico de cada grupo o contexto y no si las relaciones entre las variables independientes con la dependiente, varían entre grupos. Al contrario, asume que dicha relación es idéntica para todos los grupos. Con un enfoque de este tipo, hallamos un intercepto diferente para cada contexto pero no una *relación* diferente (John Nezlek, 2015).

Si bien un análisis de MCO me permite introducir términos de interacción entre niveles para analizar diferentes relaciones, esto está limitado a un cierto número de términos, ya que sino el modelo se vuelve inviable. Además, para la regresión por MCO el hecho de que el intercepto y las pendientes no varían entre grupos es un supuesto que se asume. Si efectivamente en la naturaleza de los datos con los que se trabaja, el intercepto y las pendientes varían entre contextos, se cometería otro error de especificación. (Bickel, 2007; Urbina e Ibarra, 2013).

Si bien hay otros problemas que surgen con el enfoque de efectos fijos y que serán vistos más adelante, es importante saber que más allá de estas complicaciones, cuando lo que interesa es analizar la variabilidad entre los grupos, la existencia de diferentes relaciones entre el intercepto y las pendientes y dar cuenta y encontrar explicaciones al efecto asociado al contexto en que las observaciones se encuentran agrupadas, el enfoque de MCO con variables dummy para cada grupo no es el adecuado.

Entonces, ¿de qué otro modo se puede analizar la existencia de 10 escuelas diferentes?

#### **3.4.4 El Modelo de Efectos Aleatorios**

Cuando existen datos de estructura jerárquica que producen que el intercepto y/o las pendientes varíen entre grupos es esencial el uso de un modelo de efectos aleatorios o modelo multinivel (Bickel, 2007; John Nezlek, 2015).

A diferencia del enfoque de efectos fijos, en un modelo multinivel, los factores inobservados que causan homogeneidad intra-grupo y heterogeneidad entre-grupos son considerados una variable aleatoria. O dicho más sencillamente, en un modelo de efectos aleatorios, los diferentes grupos o contextos son considerados una variable aleatoria, es decir, provenientes de una población de contextos de los que fueron aleatoriamente seleccionados. Al provenir de una población de grupos o contextos, están distribuidos aleatoriamente alrededor de un valor, por lo tanto no son independientes entre sí y no toman la forma de un coeficiente fijo.

Como ya se indicó, el problema de trabajar con datos anidados es la existencia de características inobservables propias de los grupos que producen correlación al interior de los mismos. El objetivo es intentar modelar dichas características para que no produzcan estimadores sesgados. En el modelo de efectos fijos, todo lo característico de cada grupo se

resumía en el coeficiente fijo de la variable dummy asociada. Aquí, en cambio, se considera que dichos factores provienen de una distribución de probabilidad, con su propia media y varianza. La varianza estimada del efecto aleatorio, es decir, la varianza al nivel de grupo, va a resumir la cantidad de variabilidad existente al nivel de grupo que no es explicada por el resto de las variables dependientes del nivel inferior.

Un modo de comprender esto más sencillamente es entendiendo que en los datos de estructura jerárquica, existen tantas muestras como niveles en la estructura de los datos. En el ejemplo que se viene desarrollando, se tiene una muestra al azar de estudiantes por un lado, pertenecientes al nivel uno y una segunda muestra al azar de escuelas pertenecientes al segundo nivel. De este modo, tanto los estudiantes como las escuelas poseen su propia distribución de probabilidad ya que provienen de una población de estudiantes y escuelas, respectivamente.

Por lo tanto, el término de error asociado al efecto aleatorio es una medida de las diferencias entre-grupos asociada a los factores inobservados. Lo que es importante aquí es saber si las varianzas de los diferentes grupos son constantes o no, ya que lo que interesa es analizar la variabilidad entre los grupos. Si el efecto aleatorio no tiene impacto en la variable dependiente, entonces su varianza será cero y esto quiere decir que no hay variabilidad entre-grupos, que no hay efectos no observados que produzcan diferencias significativas entre los distintos grupos o contextos (Mayorga y Muñoz, 2000).

Al introducir el efecto aleatorio que da cuenta de la existencia de diferentes grupos, se ingresa en el ámbito de los modelos de regresión de coeficientes aleatorios o modelos de efectos aleatorios. Como ya expliqué, en estos modelos el valor de los coeficientes no es fijo sino que proviene de una distribución de probabilidad. Estos modelos permiten que tanto el intercepto como las pendientes de las variables de nivel inferior, varíen entre los distintos grupos existentes. A su vez, es posible analizar correlaciones entre el intercepto y las pendientes o entre dos pendientes aleatorias, permitiendo que varíen juntos entre los distintos contextos.

Antes de continuar es necesario aclarar que, como se verá más adelante, los modelos multinivel y/o de efectos aleatorios son llamados también modelos mixtos o modelos de efectos mixtos. Esto se debe a que los mismos permiten tener una parte fija común a todos los contextos, con coeficientes fijos que no varían entre grupos, y también una parte

aleatoria, con coeficientes aleatorios que de hecho varían entre los distintos contextos. (Murillo, 2008).

Pero además, es interesante aclarar que los coeficientes de regresión aleatorios que pueden variar entre contextos tienen dos componentes, uno fijo y uno aleatorio, mientras que los coeficientes a los que no se les permite variar entre grupos cuentan con un sólo componente. Estos últimos son los tradicionales coeficientes fijos de la regresión por MCO. En cambio, los coeficientes aleatorios cuentan con un componente fijo, que se interpreta al igual que los coeficientes fijos de la regresión por MCO y es un promedio ponderado para todos los grupos, y uno aleatorio, que es la varianza estimada del coeficiente de regresión aleatorio que captura la variabilidad entre grupos. Esta diferencia se comprenderá mejor cuando se desarrolle el modelo multinivel de dos niveles.

### **3.4.5 El Coeficiente de Correlación Intraclase**

Se ha logrado entonces tener en cuenta la existencia de distintas escuelas mediante un enfoque diferente al denominado regresión por mínimos cuadrados con variables dummies. Con un modelo de efectos aleatorios, es posible tener en cuenta la estructura jerárquica de los datos y el hecho de que existen características inobservables en los diferentes grupos que hacen que las observaciones de nivel inferior no sean independientes, y por lo tanto, se violen los supuestos de la regresión por MCO. Esto dispara una serie de preguntas.

Por un lado, es razonable pensar que la existencia de distintos grupos o contextos no necesariamente tiene que dar lugar a la existencia de diferencias significativas entre los mismos. ¿Cómo saber si la existencia de los mismos genera que la homogeneidad intra-grupo sea lo suficientemente consecuente como para generar diferencias significativas entre ellos y que, por lo tanto, se violen los supuestos de la regresión por MCO? O dicho más sencillamente, ¿cómo saber si un análisis de efectos aleatorios es necesario?

Es posible saberlo mediante el Coeficiente de Correlación Intraclase (CCI). Este estadístico es una medida del grado de dependencia de las observaciones, es decir, una magnitud del grado de correlación de las observaciones en los grupos. Como se explicó anteriormente, la varianza total se puede descomponer en la varianza “entre-grupos” (“between-subject”) e “intra-grupos” (“within-subject”). El valor del CCI dirá cuál es la proporción de la varianza en la variable dependiente que es originada por la variabilidad entre grupos, en relación a la varianza total. Si su valor es bajo, significa que las observaciones dentro de un grupo son

tan diferentes entre sí como lo son las de otros grupos y en este caso la anidación no tiene consecuencias, es decir, se puede considerar a las observaciones como independientes. Por el contrario, si su valor es significativo, quiere decir que las observaciones difieren entre grupos más que dentro de los grupos y, en este caso, la anidación de los datos tiene consecuencias. Dicho de otro modo, un valor alto del CCI significa que la homogeneidad intra-grupo es lo suficientemente grande como para subestimar los errores estándar de los coeficientes estimados por MCO y que, por lo tanto, es necesario usar otro estimador que permita utilizar coeficientes aleatorios para tener en cuenta la variabilidad entre contextos (Murillo, 2008; Bickel, 2007).

Esto automáticamente dispara una nueva pregunta, si la existencia de datos anidados invalida el uso de los estimadores por MCO, ¿qué otro estimador es factible?

#### **3.4.6 El estimador de Máxima Verosimilitud**

Cuando se utiliza un modelo de efectos aleatorios o un modelo multinivel, se deben utilizar estimadores más adecuados que permitan tener en cuenta la autocorrelación en los errores, producto del valor positivo del CCI. Estimadores más adecuados para este tipo de enfoques son el estimador de máxima verosimilitud (MV) y el de máxima verosimilitud restringida (MVR). Estos estimadores tienen en cuenta la existencia de la correlación intraclase al permitir utilizar coeficientes con ambos componentes, fijos y aleatorios. Los mismos permiten encontrar los parámetros de población hipotéticos más probables, según los datos de la muestra, es decir, intentan encontrar aquellos valores de los parámetros de población que maximizan los valores observados para las variables dependiente e independientes (Agresti, 2002).

La importancia de los estimadores de máxima verosimilitud es que permiten utilizar intercepto y pendientes aleatorias y de este modo abordar, de una manera más eficaz que el estimador de MCO, los problemas de observaciones dependientes y errores correlacionados que surgen de la existencia de datos con estructura jerárquica. Además, estos estimadores tienen la ventaja de que los componentes fijos del modelo son interpretados de la misma manera que los coeficientes de regresión por MCO. Se verá que la interpretación de los componentes aleatorios no presenta tampoco mayores problemas (Bickel, 2007).

### 3.4.7 Del Modelo de Efectos Aleatorios a los Modelos Multinivel

Repasando lo hecho hasta el momento, se sabe que ante la presencia de datos jerárquicos pueden existir características inobservables comunes a los grupos que causen homogeneidad al interior de los mismos. Mediante el CCI se determina el grado de correlación entre las observaciones. Si este valor es significativo, el estimador por MCO deja de ser el más útil ya que la anidación de los datos produce la violación de algunos de sus supuestos: observaciones independientes, residuos incorrelacionados, intercepto y pendientes fijas. Ante esta situación, se puede usar como alternativa el estimador de MV, que admite emplear coeficientes con componentes fijos y aleatorios. De este modo, al introducir un efecto aleatorio, se permite que los coeficientes varíen de un grupo a otro, dando lugar a posibles diferentes interceptos y/o pendientes para los distintos grupos, reconociendo así la homogeneidad generada por la anidación.

Al comenzar este apartado, hice referencia al hecho de que para entender los modelos multinivel, haríamos el recorrido que nos lleva de la regresión por MCO a los modelos de efectos aleatorios para terminar finalmente en los modelos multinivel. Al igual que Bickel (2007), para los propósitos de esta tesis, la única diferencia entre los modelos de efectos aleatorios y los modelos multinivel, es que el primero mide las varianzas y covarianzas de los componentes aleatorios, mientras que el segundo además, intenta explicarlas.

El modelo de efectos aleatorios sirve de base estadística para el modelo multinivel. Como se verá más adelante, si mediante la introducción del efecto aleatorio de grupo, el CCI otorga un valor significativo, eso quiere decir que las pendientes y/o el intercepto no son los mismos para todos los grupos y se tiene entonces un fundamento estadístico para que el análisis de regresión multinivel sea factible. De este modo, se pasa de un análisis de efectos aleatorios a un análisis multinivel mediante la introducción de variables contextuales y términos de interacción entre niveles que intentarán explicar la variabilidad en los componentes aleatorios.

La introducción de variables contextuales, ayudará a explicar porqué las relaciones difieren entre grupo, es decir, intentan dar cuenta de la varianza y covarianzas en los componentes aleatorios. De este modo, una variable contextual ayudará a explicar la variabilidad entre grupos si logra reducir la varianza asociada a los componentes aleatorios. Estas pueden ser introducidas para explicar las diferencias en los interceptos aleatorios pero también en las

pendientes aleatorias. Cuando se introduce una variable contextual para explicar la variabilidad entre grupos en una pendiente aleatoria, eso lleva implícito la creación de un término de interacción entre niveles, que a diferencia de los términos de interacción de nivel único, utiliza variables de diferentes niveles.

En el ejemplo de las escuelas, el hecho de que la escuela sea pública o privada puede ser tratado como una variable contextual al nivel de la escuela que ayude a entender las diferencias en el logro escolar de los alumnos y además para tratar de dar cuenta de por qué las relaciones difieren entre las escuelas, reduciendo las varianzas de los componentes aleatorios. En este caso, el modelo de efectos aleatorios se convierte en un modelo de regresión multinivel.

Se ha visto entonces porqué un análisis multinivel es necesario ante la existencia de datos con algún tipo de estructura jerárquica. Se han introducido además, algunos conceptos importantes y se ha tratado de demostrar que el mismo no es un análisis completamente nuevo sino un tipo de análisis de regresión con características propias. Se procederá ahora a caracterizar el modelo multinivel utilizado en esta tesis.

### **3.5. Especificación del Modelo Multinivel utilizado en este trabajo**

El modelo multinivel utilizado en esta tesis tendrá dos características importantes que se deben destacar y detallar, ya que se trata de casos particulares de la aplicación de los modelos multinivel. Por un lado, se trabajará con datos de tipo panel, los que combinan características de datos transversales y longitudinales. Por el otro, la variable respuesta no será una variable continua sino una variable dicotómica. Estas características harán que el modelo multinivel tenga ciertas especificaciones propias que deberán ser tenidas en cuenta.

#### **3.5.1 Modelos de regresión multinivel para datos longitudinales**

Los modelos multinivel pueden ser utilizados para estudiar cualquier tipo de datos con estructura jerárquica. La cuestión principal es que las observaciones están agrupadas (anidadas) de alguna manera. Por ejemplo:

- Votantes agrupados en jurisdicciones.
- Estudiantes agrupados en aulas.
- Observaciones en el tiempo anidadas dentro de sujetos / individuos / personas.

Este último ejemplo es el caso de los datos de corte longitudinal, dónde se tienen diferentes observaciones del mismo individuo, país, etc. a lo largo del tiempo. Es decir, observaciones de un nivel inferior (puntos en el tiempo), agrupadas en un nivel superior (individuo).

Como se mencionó anteriormente, al trabajar con datos agrupados, las observaciones ya no son independientes. En los datos de tipo panel, en los que se observa al mismo individuo a lo largo del tiempo, no se puede decir que dichas observaciones sean independientes, ya que corresponden al mismo sujeto y por lo tanto existirán ciertas características inobservables que van a influir en cada una de sus observaciones a lo largo del tiempo. Es por eso que no se puede tratarlas como independientes. En otras palabras, los residuos van a estar correlacionados dentro de cada grupo, que en este caso es el mismo individuo.

Hay que distinguir dos clases de modelos que utilizan medidas repetidas en el mismo sujeto. Por un lado, se tiene el caso en el que las mediciones anteriores se tratan como covariables en lugar de respuestas. Este caso es más apropiado cuando los datos están recolectados en un número pequeño de ocasiones discretas y que es generalmente conocido como investigación de panel. En este caso, el tiempo “ $t$ ”, es decir, la cantidad de medidas repetidas sobre el mismo individuo, no suele ser un número grande. De este modo, el tiempo se modela como un efecto fijo sobre la variable dependiente, utilizando variables ficticias (dummy) que indican las diferentes ocasiones. Es útil cuando por ejemplo, se quiere probar la influencia de cada ocasión sobre la variable respuesta o probar la hipótesis nula de que las medias son iguales para todas las ocasiones (Goldstein, 1999; Hox, 2000).

Por otro lado, están los llamados “modelos de crecimiento”, en los cuáles, se tienen un número mucho mayor de medidas en el tiempo y en dónde todas las mediciones se tratan como respuestas. El tiempo es tratado como una variable continua y lo que se trata de modelar aquí es el efecto en “ $y$ ” del tiempo, es decir, el tiempo es la pendiente de la función. Los modelos multinivel son muy útiles para este tipo de estudios debido a su flexibilidad. En los casos en los que se desea estudiar el desarrollo de un sujeto como función del tiempo o la edad en las diferentes ocasiones, los modelos multinivel permiten modelar la evolución en el tiempo con diferentes coeficientes de regresión para los diferentes individuos, al considerar al tiempo un efecto aleatorio (y no un efecto fijo). De este modo, cada individuo recibe su propia curva de crecimiento, especificado por los coeficientes de regresión individuales que pueden depender de los atributos individuales. A

su vez, no sólo es útil para modelar tendencias lineales sino que también permite utilizar funciones cuadráticas y superiores para modelar dependencias no lineales al tiempo. Modelos de este tipo son vistos en Goldstein, 1999; Snijders, 1996 y Hox, 2000, entre otros.

Si bien no es el caso de este trabajo, vale aclarar que los modelos multinivel permiten extender los agrupamientos a más niveles. Por ejemplo, observaciones a lo largo del tiempo, anidadas en individuos, que a su vez están agrupados en barrios. Además, en este tipo de análisis, los modelos multinivel también presentan ventajas sobre otros tipos de enfoque, sobre todo cuando en el panel existen observaciones perdidas, ya sea por la ausencia de personas en alguna ocasión específica o por la llamada deserción o mortalidad del panel. El Análisis de Varianza Multivariado (MANOVA) o los modelos de efectos fijos para datos de panel, manejan la ausencia de datos faltantes mediante la supresión de todos los sujetos con datos incompletos. En cambio, en los modelos multinivel, los datos faltantes no plantean mayores problemas ya que el modelo no asume el mismo número de ocasiones idénticas para todas las personas, lo que permite trabajar con paneles no balanceados sin pérdida de información. A su vez, también es útil cuando los momentos en el tiempo y/o su espaciamiento difieren entre los individuos, es decir, cuando no todos los individuos son relevados en el mismo momento ni cuando el espaciamiento entre mediciones es igual para todos los sujetos (Snijders, 1996; Hox, 2000).

Tal como se explicó, en esta tesis en particular, el diseño longitudinal de los datos se encuadra dentro de lo que se conoce como investigación de panel. Específicamente, se cuenta con información de 582 jefes de hogar ( $n=582$ ) en tres momentos en el tiempo ( $T=3$ ), es decir, 1746 observaciones ( $n*T=582*3=1746$ ). Al tener pocas observaciones de cada individuo en el tiempo se utilizarán variables ficticias (dummy) para cada uno de los años, con el objetivo de probar si existen diferencias significativas en el acceso al trabajo entre los mismos o, por el contrario, si las medias son iguales para todas las ocasiones.

### **3.5.2 Modelos de respuesta discreta**

Hasta aquí se ha hecho referencia al análisis multinivel como una variante del análisis de regresión, diseñado específicamente para trabajar con datos con estructura jerárquica. De este modo, cuando se tiene una variable dependiente continua y se asume una distribución normal de los errores, el análisis multinivel es comúnmente llamado Modelo Lineal

Jerárquico (o Hierarchical Linear Model, HLM, en inglés) ya que es una extensión del Modelo Lineal General (o Generalized Linear Models, GLM, en inglés).

Sin embargo, cuando la variable respuesta no es una variable continua sino discreta, el supuesto de que los errores se distribuyen normalmente no se cumple, como así tampoco lo hace el supuesto de homocedasticidad de los mismos. Del mismo modo en que los Modelos Lineales Generalizados (Generalized Linear Models, GLM, en inglés) son una extensión del Modelo Lineal General para este tipo de situaciones, los llamados Modelos Lineales Jerárquicos Generalizados o Modelos Lineales Mixtos Generalizados (Hierarchical Generalized Linear Models o Generalized Linear Mixed Models, GLMM, en inglés) son una extensión del Modelo Lineal Jerárquico para versiones multinivel de, por ejemplo, los modelos Logit o Probit. (Snijders, 2003; Hox y Maas, 2005).

En esta tesis, la variable respuesta es una variable dicotómica, por lo que el modelo a especificar se enmarca dentro de los Modelos Lineales Mixtos Generalizados. Las variables dicotómicas admiten sólo dos respuestas, generalmente codificadas como 1 o 0, dónde 1 es interpretado como la respuesta “Si” y 0 como la respuesta “No” a alguna pregunta específica.

En este caso, el objetivo es encontrar un modelo de predicción que permita identificar la existencia o no de diferencias en las chances de obtener trabajo según ciertas características de los individuos. Por ello, la variable respuesta es una variable dicotómica que admite las respuestas “Si” y “No” a la pregunta “*En la última semana, ¿Usted trabajó?*”.

Desde el enfoque de los Modelos Lineales Generalizados, se solucionan los problemas que produce el hecho de que la variable dependiente no sea una variable continua mediante la introducción de una transformación de la función objetivo para lograr así normalidad y la elección de una distribución apropiada para los errores en el modelo, para reducir la heteroscedasticidad. De este modo, como explica Hox (1995), los Modelos Lineales Generalizados están definidos por tres componentes: la ecuación de regresión lineal (o componente sistemático), una distribución específica de los errores (componente aleatorio) y la función de enlace, que es la transformación introducida para lograr normalidad. El método de estimación de los Modelos Lineales Generalizados es el de máxima

verosimilitud, que permite interpretar los resultados como si fuesen los de una regresión lineal tradicional.

Siguiendo a Rabe-Hesketh y Skrondal (2012), se introducirá primero un modelo logit de nivel único para variable dicotómica, es decir, para datos que no tienen estructura jerárquica. Luego se demostrará que en el caso de los modelos multinivel, el procedimiento es similar.

En un modelo de regresión, lo que se intenta encontrar es la esperanza o la media de la variable respuesta como una función de las covariables. La esperanza de una variable respuesta binaria es simplemente la probabilidad de que la respuesta sea 1:

$$E(y_i|x_i) = \Pr(y_i = 1|x_i)$$

A diferencia de una regresión lineal, el problema aquí es que, cualquiera sea el valor que tome  $x_i$ , la probabilidad debe caer entre 0 y 1. Por ello, una función no-lineal puede ser especificada en alguna de las dos siguientes formas:

$$\Pr(y_i = 1|x_i) = h(\beta_1 + \beta_2 x_i)$$

$$g\{\Pr(y_i = 1|x_i)\} = \beta_1 + \beta_2 x_i = v_i$$

Dónde  $v_i$  es el predictor lineal que especifica las variables explicativas utilizadas. Este predictor lineal es el componente sistemático que hace que los modelos lineales generalizados se asemejen a las regresiones lineales y nos permitan utilizar cualquier tipo de variable explicativa. Para ello, necesitamos de la función de enlace que nos especifique la función de  $E(y)$  que haga equivalente el modelo al componente sistemático, o dicho más sencillamente, que conecte el componente aleatorio (la distribución de probabilidad de  $y$ ) con el componente sistemático (la combinación lineal de variables explicativas) (Agresti, 2002).

En este caso  $g(\cdot)$  es la función de enlace y si  $h(\cdot)$  es su inversa, entonces las expresiones son equivalentes y esta última es llamada función de enlace inversa. Para una variable de respuesta binaria, las funciones de enlace que generalmente se utilizan son la logit o la probit. Aquí me referiré sólo a la primera ya que es la que utilizaré en el modelo.

Si  $g(\cdot)$  es la función de enlace logit, entonces:

$$\text{logit}\{\Pr(y_i = 1|x_i)\} \equiv \ln\{\Pr(y_i = 1|x_i) / 1 - \Pr(y_i = 1|x_i)\} \equiv \ln\{\text{Odds}(y_i = 1|x_i)\} = \beta_1 + \beta_2 x_i \quad (1)$$

En la ecuación anterior,  $\text{Odds}(y_i=1|x_i)$  representa el número esperado de respuestas que tomen el valor 1 en relación al valor 0. Es decir, representa las posibilidades o chances (odds en inglés) de que  $y_i$  sea 1 dadas las variables explicativas. O dicho de otro modo, el número esperado de éxitos por fracasos.

Hasta el momento se ha introducido el componente sistemático o predictor lineal y la función de enlace, pero no se ha dicho nada del componente aleatorio o mejor dicho, de la distribución de la variable  $y$  dada las covariables. Si se asume  $\Pr(y_i = 1|x_i) \equiv \pi_i$  la distribución es especificada como Bernoulli, donde las respuestas 1 y 0 son el resultado de una prueba de Bernoulli con media condicional  $\pi_i$  y varianza condicional  $\text{Var}(y_i|x_i) = \pi_i(1-\pi_i)$ .

De este modo, podemos reescribir (1):

$$\text{logit}\{\Pr(y_i=1|x_i)\} \equiv \text{logit}(\pi_i) \equiv \ln\{\pi_i/(1-\pi_i)\} \equiv \ln\{\text{Odds}(y_i=1|x_i)\} = \beta_1 + \beta_2 x_i \quad (2)$$

Se puede ver como la función de enlace logit, linealiza la relación entre el componente aleatorio (la distribución Binomial  $(n_i ; \pi_i)$  con  $n_i=1$ , es decir, Bernoulli) con el componente sistemático  $(\beta_1 + \beta_2 x_i)$  convirtiendo los valores esperados a valores entre 0 y 1. Mientras que  $\pi_i$  posee una distribución binomial,  $\text{logit}(\pi_i)$  posee una distribución aproximadamente normal.

Es necesario tener en cuenta que la interpretación de los coeficientes de regresión en (2), es decir  $\beta_1$  y  $\beta_2$ , no es en términos de la proporción de respuesta  $\pi_i / (1-\pi_i)$ , que es lo que interesa encontrar, sino en términos de la variable subyacente definida por la transformación logit, es decir,  $\text{logit}(\pi_i) \equiv \ln\{\pi_i/(1-\pi_i)\}$ . Justamente, la introducción de la función de enlace logit se realiza con el objetivo de transformar las proporciones, las cuales por definición sólo pueden tomar valores entre 0 y 1, para que puedan tomar valores entre  $\{-\infty; +\infty\}$ . Para comprender más fácilmente cuales son las implicaciones de los coeficientes de regresión sobre las proporciones, se pueden transformar sus valores mediante la inversa de la función logit,  $h(\cdot)$ , definida anteriormente (Hox, 1995).

Entonces, a partir de la ecuación (1) o (2) se deduce que otra forma de especificar el modelo logit es como una función exponencial de los odds:

$$\text{Odds}(y_i = 1|x_i) = \exp(\beta_1 + \beta_2 x_i)$$

Además, siendo  $h(\cdot)$  la inversa de la función logit definida como:  $\Pr(y_i = 1|x_i) = h(\beta_1 + \beta_2 x_i)$  y teniendo en cuenta que la relación entre posibilidades (odds) y probabilidades es:

$$\text{Odds} = \text{Pr} / (1 - \text{Pr}) \quad \text{y} \quad \text{Pr} = \text{Odds} / (1 + \text{Odds})$$

Entonces:

$$\Pr(y_i = 1|x_i) = h(\beta_1 + \beta_2 x_i) = \text{logit}^{-1}(\beta_1 + \beta_2 x_i) \equiv \exp(\beta_1 + \beta_2 x_i) / \{1 + \exp(\beta_1 + \beta_2 x_i)\} \quad (3)$$

Entendida como la inversa de la función de enlace logit del predictor lineal. La inversa de la función logit es llamada a veces, función logística (logistic en inglés).

Lo interesante del enlace logit es que produce un modelo lineal para el logaritmo de las posibilidades como se puede ver en la ecuación (2), pero además, que el mismo es fácilmente transformable para hacer más fácil la interpretación de sus resultados en términos del cociente de chances u “odds ratio”. Veamos por ejemplo un cambio en una unidad de  $x_i$ , pasando de “ $a$ ” a “ $a + 1$ ”. El cambio en el logaritmo de las posibilidades sería:

$$\ln\{\text{Odds}(y_i = 1|x_i = a + 1)\} - \ln\{\text{Odds}(y_i = 1|x_i = a)\} = \{\beta_1 + \beta_2(a + 1)\} - (\beta_1 + \beta_2 a) = \beta_2$$

Los odds ratio se obtienen aplicando la exponencial a ambos lados de la ecuación:

$$\begin{aligned} \exp[\ln\{\text{Odds}(y_i=1|x_i=a+1)\} - \ln\{\text{Odds}(y_i=1|x_i=a)\}] &= \text{Odds}(y_i=1|x_i=a+1) / \text{Odds}(y_i=1|x_i=a) \\ &= \{\Pr(y_i=1|x_i=a+1) / \Pr(y_i=0|x_i=a+1)\} / \{\Pr(y_i=1|x_i=a) / \Pr(y_i=0|x_i=a)\} = \exp(\beta_2) \end{aligned}$$

La diferencia entonces estriba en que el coeficiente  $\beta_2$ , es interpretado como el cambio en una unidad de  $x_i$  sobre el logaritmo de las chances,  $\ln(\text{Odds})$ , y no de las chances en sí mismo. Por ello, al elevarlo exponencialmente su interpretación es directa como un cociente de chances (odds ratio).

Para finalizar, hay que destacar dos cuestiones. Por un lado, el hecho de que en la ecuación (2) no existe término de error porque en la distribución binomial la varianza de la proporción observada depende sólo de la proporción poblacional  $\pi$ . Como consecuencia, la varianza (al nivel más bajo) es completamente determinada cuando la media (que en este caso resulta ser la proporción) es conocida, por ello se la definió como *varianza condicional*  $\text{Var}(y_i|x_i) = \pi_i(1-\pi_i)$ . Por ello, desde el enfoque de los Modelos Lineales Generalizados, el error individual es determinado completamente por el valor predicho y

no aparece como un término separado en el modelo. Esto es importante porque en los modelos lineales multinivel, el término de error es un error compuesto por la suma del error individual (o de nivel uno) más el error del nivel de grupo (o error de nivel dos). Como veremos a continuación, desde su formulación de dos niveles, en los modelos logit multinivel tendremos sólo el error al nivel de grupo. Sin embargo, los modelos de regresión logística también pueden ser vistos como modelos de respuesta latente. La idea es que, subyacentemente a la respuesta dicotómica observada  $y_i$  (ya sea que el individuo trabaje o no), existe una respuesta continua no observada o latente  $y_i^*$  que representa la propensión al trabajo o la excesiva utilidad de trabajar en comparación con no trabajar. De este modo, si esta respuesta latente es mayor que 0, entonces la respuesta observada es 1, de lo contrario, la respuesta observada es 0. Al utilizar esta formulación, se puede especificar para dicha respuesta latente un modelo de regresión lineal  $y_i^* = \beta_1 + \beta_2 x_i + e_i$  donde  $e_i$  es un término de error residual con media condicional cero  $E(e_i|x_i) = 0$  y con términos de error independientes para distintos individuos. Esta formulación tiene una relación con la formulación que se venía desarrollando para modelos lineales generalizados<sup>4</sup> y lo importante de ello en el caso de la regresión logística, es que se asume que  $e_i$  tiene una función de densidad acumulativa logística estándar dada las variables explicativas:

$$\Pr(e_i < \tau | x_i) = \exp(\tau) / \{1 + \exp(\tau)\}, \text{ donde } \tau = \beta_1 + \beta_2 x_i$$

Por lo tanto, para este tipo de distribución,  $e_i$  tiene media cero y varianza  $\pi^2/3 \approx 3.29$  (A modo de aclaración,  $\pi$  es el número “pi” y no una probabilidad). Lo que es importante para este trabajo, es que dicho valor de varianza, es el valor correspondiente a la variabilidad *intra-grupo*, la cuál es importante en un modelo multinivel para determinar el valor de la varianza *entre-grupos* en relación a la varianza total, como se verá en el capítulo siguiente (Hox y Maas, 2005; Agresti, 2002; Rabe-Hesketh y Skrondal, 2012).

La otra cuestión a destacar es que debido a la naturaleza de la variable dependiente, debemos utilizar otro estimador de los parámetros, ya que el método de mínimos cuadrados ordinarios supone una distribución lineal y aquí estamos asumiendo una distribución binomial de la misma. En los modelos logit, al igual que en los modelos multinivel, se

<sup>4</sup>Una explicación más profunda de la relación entre los dos tipos de formulaciones para una variable dicotómica excede los propósitos de esta tesis, para ello puede referirse a Agresti (2002) o Rabe-Hesketh y Skrondal (2012).

utiliza el método de máxima verosimilitud para la estimación de los parámetros, que consiste en escoger los valores estimados de los parámetros que maximizan la función de verosimilitud, es decir, el valor de los parámetros bajo los cuáles, los datos observados tienen la mayor probabilidad de ocurrencia. Los estimadores proporcionados por el método de máxima verosimilitud son consistentes, eficientes y asintóticamente normales. Además, los valores de los parámetros que maximizan la función de MV también maximizan el logaritmo de dicha función, que es el que más se utiliza en la práctica (Agresti, 2002).

Ahora bien, hasta el momento no se ha tenido en cuenta el hecho de la existencia de datos jerárquicos. El modelo logit presentado recién asume que no existen factores inobservables que causen correlación entre las observaciones, ya que es un modelo de regresión logística de nivel único. Si bien se puede utilizar, la alternativa de los modelos multinivel es el enfoque adecuado.

### 3.5.3 Modelos de regresión logística de intercepto aleatorio

Teniendo en cuenta entonces que se está analizando individuos a lo largo del tiempo, existirá correlación entre las distintas observaciones de un mismo individuo. Por lo tanto, se permitirá que el intercepto varíe entre los “grupos” del segundo nivel, es decir, entre los individuos, para tener en cuenta la existencia de la homogeneidad intra-sujeto. Introduciendo un intercepto aleatorio al nivel del individuo, lo que se permite es que cada individuo tenga su propia ordenada al origen. Para hacerlo, utilizaré la formulación en dos etapas, como las que utilizan Bickel (2007), Hox (1995) o Hox y Maas (2005).

Me basaré en el modelo presentado por Rabe-Hesketh y Skrondal (2012) pero para simplificar, el modelo contará de un intercepto y una sola variable explicativa fija, la pendiente, que no variará entre individuos. A continuación se muestra la distribución de la variable dependiente, la función de enlace y el modelo estructural, que sería la ecuación de nivel uno:

$$y_{ij} \sim \text{Bernoulli}(\varphi_{ij})$$

$$\text{logit}(\varphi_{ij}) = \eta_{ij}$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} \tag{4}$$

Dónde  $i$  es la medida en el tiempo y  $j$  es el individuo. La ecuación de nivel dos para el intercepto  $\beta_{0j}$  viene dada por:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Se evidencia aquí lo descrito anteriormente, que un coeficiente aleatorio  $\beta_{0j}$  posee dos componentes, uno fijo  $\gamma_{00}$  y uno aleatorio  $\mu_{0j}$ . El coeficiente asociado a la variable  $x_{1ij}$  es un coeficiente fijo,  $\beta_1$ , que no varía para los distintos individuos.

Sin embargo, también se puede permitir que el mismo varíe entre individuos, por lo que se tendría una pendiente específica para cada uno de los sujetos. De este modo, el coeficiente de la pendiente sería  $\beta_{1j}$ , y por lo tanto, la ecuación de nivel dos para la pendiente aleatorio vendría dado por:  $\beta_{1j} = \gamma_{10} + \mu_{1j}$ . En este caso, el término de error se volvería más complejo, ya que estaría compuesto por dos componentes aleatorios, una para el intercepto y otro para la pendiente. Sin embargo, se mantendrá el modelo simple de intercepto aleatorio.

Introduciendo la ecuación de segundo nivel en la de nivel uno, obtenemos el modelo completo:

$$\text{logit}(\varphi_{ij}) = \gamma_{00} + \beta_1 x_{1ij} + \mu_{0j} \quad (5)$$

Como se puede observar, este modelo posee un efecto fijo  $\beta_1$  y también un efecto aleatorio  $\mu_{0j}$ , por lo que se enmarca dentro de los Modelos Lineales Mixtos Generalizados o Modelos Lineales Jerárquicos Generalizados.

Los supuestos sobre los interceptos aleatorios  $\mu_{0j} \sim N(0, \psi)$  es que son independientes e idénticamente distribuidos entre individuos  $j$  y además independientes de las covariables  $x_{ij}$ , es decir, de las variables explicativas de nivel uno que cambian en el tiempo. El intercepto aleatorio intenta tener en cuenta el efecto combinado de las características inobservables específicas de cada individuo que son *constantes* en el tiempo y causan la correlación entre sus observaciones. La idea es que la heterogeneidad inobservada producto de dichas características es modelada mediante la simple introducción del intercepto aleatorio, al igual que la heterogeneidad observada es modelada mediante la introducción de las variables independientes de coeficiente fijo (Rabe-Hesketh y Skrondal, 2012).

### 3.5.4 Modelo Logístico Multinivel de intercepto aleatorio para el acceso al trabajo

Finalmente, el modelo utilizado en esta tesis será un modelo logístico multinivel para datos de panel de intercepto aleatorio. En un modelo de este tipo, las observaciones en diferentes ocasiones en el tiempo son el nivel uno, las que se encuentran anidadas en individuos en el nivel dos. La variable respuesta es una variable dicotómica que toma el valor uno si la respuesta es afirmativa y el valor 0 si no lo es. El intercepto aleatorio permite tener en cuenta los factores inobservados de los distintos individuos que causan correlación en las respuestas brindadas por ellos a lo largo del tiempo. De este modo, se obtendrán errores estándar más conservadores y coeficientes que no presenten sesgos. Mediante el cálculo del Coeficiente de Correlación Intraclase Incondicional, se comprobará si un modelo que tenga en cuenta la existencia de datos jerárquicos es necesario y, de ser así, se introducirán variables contextuales para intentar explicar la variabilidad existente entre los individuos.

Más precisamente, lo que se intentará es identificar un modelo de predicción que permita encontrar posibles diferenciales en el acceso al trabajo en función de características del individuo, principalmente, el nivel de estudios alcanzados, el sexo y la edad.

En este trabajo, el diseño de los datos se encuadra dentro de lo que se conoce como investigación de panel, es decir, información que cuenta con características tanto de corte transversal, como longitudinal. Específicamente, la investigación se llevará a cabo analizando las respuestas sobre el acceso al trabajo de los jefes y jefas de hogares santafesinos de la Ciudad de Santa Fe para los años 2009, 2010 y 2011. Es decir que para cada individuo, existen tres observaciones en el tiempo,  $t=3$ . Por lo tanto, no se tratara al tiempo como un efecto aleatorio, sino que se modelara como un efecto fijo, utilizando variables dummy para cada ocasión, lo que sirve para saber si hay diferencias significativas en la respuesta al trabajo entre los diferentes años.

La muestra de estudio está compuesta de 582 casos (jefes y jefas de hogar), constituyendo así 1746 observaciones. En el modelo a especificar, estas 1746 observaciones son el nivel inferior o nivel 1, que estarán agrupadas en un nivel superior o nivel 2, es decir en los 582 individuos. De este modo, el panel cumple con dos características importantes que se suelen requerir para el manejo de este tipo de base de datos que son, por un lado, que el número de individuos “N” sea grande y el número de períodos “t” sea pequeño y, por el otro, que el panel se encuentre “balanceado”, es decir, que para cada una de las unidades

de corte transversal “N” se cuente con la misma cantidad de períodos “t”, o dicho más sencillamente, que se tenga información de todos los individuos en todos los períodos de tiempo (Wooldridge, 2010). Sin embargo, ya se mencionó que una de las ventajas de los modelos multinivel es que permite trabajar tanto con paneles balanceados como no balanceados, lo cual se verá en el próximo capítulo, será importante para este trabajo.

La variable dependiente, “Trabaja”, es una variable dicotómica que toma el valor uno si la respuesta es afirmativa y el valor cero si no lo es, a la pregunta “En la última semana, ¿Usted trabajó?”. Se recuerda que el Panel de Hogares utiliza como base metodológica los conceptos de “necesidades” y “satisfactores” de Max-Neef (1986/2010), indagando en las percepciones y opiniones de los propios actores sociales. De este modo, el trabajo es considerado una “necesidad” que se expresa a través de la demanda, el deseo u la opinión del propio actor, es decir, lo que se admite es que el propio sujeto considere si su “necesidad de trabajo” ha sido satisfecha o no. Es así que se puede definir la respuesta a dicha pregunta como una necesidad cubierta si el individuo responde afirmativamente, dando a entender que tiene acceso al trabajo, y como una necesidad no cubierta si responde negativamente, es decir, quedando excluido del mercado de trabajo.

De este modo, la proporción a estimar en el modelo logístico, se puede definir como la proporción de los jefes de hogar santafesinos con acceso al trabajo. Las siguientes variables explicativas son introducidas con el objetivo de encontrar posibles diferenciales en el acceso al mismo, es decir, con el objetivo de explicar la existencia (o no) de diferencias en las chances o posibilidades de cubrir la *necesidad de trabajo*, tal como se la definió anteriormente.

Se resumen a continuación las variables explicativas que se utilizarán en el modelo:

- “Año”: la variable temporal que indica el año de la observación y puede tomar tres valores, 2009, 2010 o 2011. La misma será recodificada, tomando los valores 1=2009, 2=2010 y 3=2011. Será considerada una variable ficticia que indicará los tres puntos en el tiempo. Se tomará el año 2009 como base y lo que se intentará es probar la influencia de cada ocasión sobre la variable respuesta, es decir, si existen diferencias significativas en la respuesta al trabajo entre los diferentes años.
- “Escolaridad”: variable cualitativa que mide el nivel educativo alcanzado según las siguientes categorías:

- Valor = 0 = Analfabeto
- Valor = 1 = Primario incompleto
- Valor = 2 = Primario completo
- Valor = 3 = Secundario incompleto
- Valor = 4 = Secundario completo
- Valor = 5 = Terciario incompleto
- Valor = 6 = Terciario completo
- Valor = 7 = Universitario incompleto
- Valor = 8 = Universitario completo
- Valor = 9 = Posgrado
- Valor = 10 = Escuela especial

Las categorías “Analfabeto” y “Escuela especial” son descartadas por no presentar valores. Como se indicó, esta variable será recodificada en la variable “Educ\_categor”, que incluye las siguientes categorías:

- Nivel educativo bajo {Primario incompleto; Primario completo}. La cual será identificada como la categoría de referencia en el modelo.
- Nivel educativo medio {Secundario incompleto; Secundario completo}
- Nivel educativo alto {Terciario incompleto; Terciario completo; Universitario incompleto; Universitario completo; Posgrado}
- “Sexo”: variable dicotómica que no cambia en el tiempo y toma los valores 1 = masculino, 2 = femenino. La categoría de referencia será el sexo femenino.
- “Edad”: variable cuantitativa continua que indica la edad en años cumplidos del individuo en el momento “t”, es decir, es una variable que cambia con el tiempo.

El modelo base que se intentará estimar entonces, es un modelo logístico multinivel de intercepto aleatorio. Como se explicó, la estructura jerárquica de los datos posee dos niveles: en el nivel uno tenemos las distintas medidas repetidas u observaciones del individuo y en el nivel dos, es decir, el nivel de grupo, está formado por cada uno de los individuos, lo que quiere decir que se tendrán tantos clúster o grupos como individuos haya en los datos. A continuación se presenta el mismo a partir de su formulación en dos niveles:

Ecuación de nivel 1:

$$\text{Logistic}(\text{Trabaja}_{ti}) = \beta_{0i} + \beta_{1i}\text{Educ\_categor}_{ti} + \beta_{2i}\text{Edad}_{ti} \quad (6)$$

Ecuación de nivel 2:

$$\beta_{0i} = \gamma_{00} + \beta_{01}\text{Sexo}_i + \beta_{02i}\text{Año2010}_i + \beta_{03i}\text{Año2011}_i + \mu_{0i} \quad (7)$$

En dónde  $i$  es el individuo (clúster o grupo) y  $t$  las distintas observaciones en el tiempo para el sujeto  $i$ .

En la ecuación de nivel uno,  $\text{Trabaja}_{ti}$  es la variable dependiente para el individuo  $i$  en el momento  $t$  y las variables Escolaridad (“Educ\_categor”) y Edad nos indican el nivel educativo alcanzado y la edad del sujeto  $i$  en el momento  $t$ , es decir, son covariables que cambian en el tiempo. Si bien no quedan dudas de que la Edad es una variable que cambia en el tiempo, no pareciera ser la misma situación para el nivel de escolaridad alcanzado, sobre todo si estamos analizando jefes y jefas de hogar. Sin embargo, como se demostrará más adelante, existen casos en la base de datos que *cambiaron* el nivel educativo alcanzado en los años analizados y por ello, a priori, se la considera una variable que cambia en el tiempo.

Por último, se recuerda que los modelos logísticos no poseen un término de error al nivel uno,  $e_{it}$ . Dicho término se define en los modelos multinivel como el error o residuos al nivel intra-grupo, es decir, el término de error que capta la variabilidad no explicada en la variable independiente y que se atribuye a factores no observados de los individuos que *cambian en el tiempo*. Como ya se definió, en un modelo logístico dicho término es determinado completamente por el valor predicho de la media o proporción y no aparece como un término separado en el modelo. Sin embargo, se verá en el próximo capítulo que el mismo tiene relación con la varianza total, y por ende, con la varianza o el error al nivel de grupo.

Por su parte, la ecuación de nivel dos es la ecuación al nivel del individuo, en la cual se incluyen las llamadas variables contextuales o de grupo. En este caso en particular, como el individuo es el grupo en sí mismo, las variables incluidas refieren también a características propias de los sujetos pero la diferencia estriba en que aquí se incluyen todas las variables *que no varían en el tiempo*. Por ello se incluye además el término de

error al nivel de grupo  $\mu_{0i}$  que intenta captar todas las características del individuo que no son incluidas en el modelo y se mantienen fijas en el tiempo. La introducción del efecto aleatorio en el intercepto nos permite tener una medida de la variabilidad *entre-sujetos*, o dicho de otro modo, una medida de la variabilidad en la respuesta de los individuos que se debe a factores *constantes en el tiempo* y que no están incluidos en el modelo, siendo esto precisamente lo que lo diferencia del término de error de nivel uno o intra-grupo.

De este modo, la ecuación (7) para el intercepto  $\beta_{0i}$  consiste en: el intercepto de la media general  $\gamma_{00}$ , las covariables que no cambian en el tiempo y un intercepto aleatorio específico para cada individuo  $\mu_{0i}$ . Introduciendo la ecuación de nivel dos (7) en la ecuación de nivel uno (6) y reagrupando términos obtenemos la ecuación de nivel único:

$$\text{Logistic}(\text{Trabaja}_{ti}) = \gamma_{00} + \beta_{01}\text{Sexo}_{ti} + \beta_{02i}\text{Año}2010_i + \beta_{03i}\text{Año}2011_i + \beta_{1i}\text{Educ\_categor}_{ti} + \beta_{2i}\text{Edad}_{ti} + \mu_{0i} \quad (8)$$

La expresión (8), es la ecuación de regresión logística multinivel de intercepto aleatorio del acceso al trabajo de los jefes de hogar santafesinos.

### 3.6. Una aclaración sobre las “chances u odds” y el “cociente de chances u odds ratio”

Para terminar este capítulo, me pareció importante realizar este apartado final aclaratorio acerca de la interpretación de los coeficientes fijos de una regresión logit (o logística), ya que lo considero de vital importancia para una mejor comprensión de los resultados que presentará el modelo. Se recuerda que en un modelo multinivel, posee coeficientes fijos y aleatorios. Los coeficientes fijos de los modelos multinivel se interpretan al igual que los coeficientes fijos de los modelos de regresión tradicionales, sin embargo, como se advirtió en el apartado 3.5.2, los coeficientes de los modelos logit tienen una interpretación particular en términos de chances y cociente de chances, que en inglés se definen como “odds” y “odds ratio”, respectivamente.

Se definió al odds como el número esperado de respuestas que tomen el valor 1 en relación al valor 0, es decir, es la posibilidad o chance de éxito en relación al número de fracasos. Es decir:

$$\text{Odds} = \text{Pr} / (1 - \text{Pr})$$

Dónde Pr es la probabilidad de éxito y (1-Pr) la probabilidad de fracaso. Por lo que, la relación entre chances (posibilidades) y probabilidades puede definirse también como:

$$\text{Pr} = \text{Odds} / (1 + \text{Odds})$$

Por ejemplo, si un odds es mayor a uno, quiere decir que las chances de ganar son más probables que las de perder. Los odds son siempre no negativos, por ello, cuando el valor del odds se encuentra entre 0 y 1, quiere decir que las chances o posibilidades de éxito son menores que las de ganar. Por ejemplo, si la probabilidad de “ganar” es 0.75, las posibilidades o chances de “ganar” son:  $\text{Odds} = \text{Pr} / (1 - \text{Pr}) = 0.75 / (1 - 0.75) = 3$ , es decir, se esperan a razón de tres éxitos por cada fracaso. En cambio, cuando la probabilidad de ganar es 0.25, las chances de perder son tres veces más que las de ganar,  $\text{Odds} = \text{Pr} / (1 - \text{Pr}) = 0.25 / (1 - 0.25) = 0.33$ , o dicho de otro modo, se espera a razón de un triunfo cada tres fracasos.

En cambio, un cociente o razón de chances, “odds ratio”, viene definido precisamente por la división de dos “odds”, dónde cada uno refiere a dos poblaciones diferentes. Por ejemplo, suponiendo que el éxito se asocia a la respuesta positiva del jefe de hogar hacia el trabajo, mientras que el fracaso se asocia a la respuesta negativa, para los jefes de hogar de sexo masculino, su chances de trabajar vienen dadas por:  $\text{Odds}_m = \text{Pr}_m / (1 - \text{Pr}_m)$ . Del mismo modo, se definen las posibilidades de trabajar del sexo femenino como:  $\text{Odds}_f = \text{Pr}_f / (1 - \text{Pr}_f)$ . La división de las chances de trabajar de los hombres en relación a las mujeres, es decir, entre  $\text{Odds}_m$  y  $\text{Odds}_f$ , viene dada por:

$$\text{Odds Ratio}_{m/f} = \text{Odds}_m / \text{Odds}_f = \{ \text{Pr}_m / (1 - \text{Pr}_m) \} / \{ \text{Pr}_f / (1 - \text{Pr}_f) \}$$

Del mismo modo que los odds, los odds ratio pueden tomar sólo valores no negativos. Cuando el odds ratio es mayor a uno, los individuos en la fila superior (en este caso, hombres) tienen más chances de ganar que los de la fila inferior. Cuando el mismo se encuentra entre 0 y 1, entonces los de la fila inferior tienen más chances que los de la fila superior, o lo que es lo mismo, las probabilidades de “ganar” asociadas son mayores para los de la fila dos (mujeres) que para los de la fila uno. Valores más lejanos a uno, en cualquiera de las dos direcciones, representa una asociación más fuerte, es decir, mayor diferencia en las chances de uno y otro. Tomando el ejemplo, el odds ratio se interpreta como las chances de trabajar de los hombres, en relación a la de las mujeres. Si el odds ratio es 2, entonces los hombres tienen el doble de chances que las mujeres de “trabajar”.

---

Como se explicó, los coeficientes del modelo logit no se interpretan directamente en términos de chances, posibilidades o proporción de respuesta, es decir,  $Pr / (1 - Pr)$ . Si no, en términos de la variable subyacente definida por la transformación logit, es decir,  $\text{logit}(Pr) \equiv \ln\{Pr/(1-Pr)\}$ . Por ello, sus resultados deben ser exponenciados o, lo que es lo mismo, utilizar la inversa de la función logit para que los mismos representen el cociente de chances. El software que se utilizará tiene la facilidad de presentar los coeficientes en este segundo sentido, lo que facilita su interpretación en forma directa, por eso los coeficientes de la función logit serán presentados como cocientes o razón de chances o posibilidades (odds ratio). Algunos autores diferencian el modelo “logit”, que reporta los coeficientes en términos de “odds”, del modelo “logistic” que reporta los coeficientes en términos de “odds ratio”. En esta tesis, ambos términos serán utilizados indistintamente, como así también su homónimo en español “logístico”.

## CAPITULO IV

### **Diferenciales en el acceso al trabajo en los jefes de hogar santafesinos**

En este capítulo se desarrollará el modelo logístico multinivel para datos de panel de intercepto aleatorio. En primer lugar, se realizará un análisis descriptivo de las variables que se utilizarán, a efectos de identificar el modelo de variación implícito para cada variable interviniente, como así también de algunas características muestrales del mercado de trabajo santafesino para el período. A su vez, se indicarán algunas especificaciones sobre la base de datos utilizada. Luego, se expondrán los resultados del modelo vacío, es decir, un modelo sin variables predictoras que tiene la importante función de indicar si un análisis multinivel es necesario, teniendo en cuenta que en este caso la variable de nivel dos (o variable de grupo) es el individuo. Dicha decisión se basará en el valor del Coeficiente de Correlación Intraclase Incondicional que nos brinda el modelo vacío. Si su valor es significativo, se tiene entonces el fundamento estadístico para llevar a cabo un análisis multinivel. Se procederá entonces a la inclusión de variables explicativas en el modelo para intentar dar cuenta de la variabilidad existente entre los individuos, es decir, variables estadísticamente significativas que reduzcan el valor de la varianza asociada a los efectos aleatorios. Por último, se explicarán y detallarán los criterios de selección del modelo bajo diferentes pruebas de bondad de ajuste.

Para la primer parte de este capítulo se utilizó el software estadístico SPSS 15.0, mientras que para la estimación del modelo y los métodos de post-estimación se utilizó el software Stata/SE 13.0.

#### **4.1. Base de datos y análisis descriptivo de la información**

##### **4.1.1 Analizando la variable dependiente**

Como se mencionó con anterioridad, para la estimación del modelo logístico multinivel de intercepto aleatorio se utilizará una fuente de información secundaria proporcionada por el Observatorio Social de la Universidad Nacional del Litoral.

Para este trabajo en particular, se considerará la información brindada por el jefe o jefa de hogar acerca de su situación laboral y algunas características personales, sin tener en consideración la información existente al nivel del hogar ni tampoco sobre otros miembros del hogar. A pesar de que la introducción de variables al nivel del hogar o la inclusión de

otros miembros de las familias hubiese sumado al análisis de los diferenciales en el acceso al trabajo, e incluso se proponen en las conclusiones algunas líneas de extensión del presente trabajo en dicho sentido, el recorte del objeto de estudio se debe en parte a los recursos limitados de cualquier investigación, como el tiempo, pero principalmente a una disponibilidad y organización de la información en miras al modelo que se aplicará aquí, el que hace necesario la ordenación de la misma en un formato particular que se explicará a continuación.

Para trabajar con datos de panel y poder confrontar al mismo individuo con la información brindada anteriormente, se debe reorganizar la misma de manera que las medidas repetidas de cada sujeto estén dispuestas como variables, lo que en la base de datos se ubica en columnas. Es decir, cada fila refiere a la información brindada por un individuo particular en un momento en el tiempo, mientras que en columnas aparecen cada una de las variables de análisis. Se debe generar entonces una variable que indique el año o período de observación y otra que identifique al individuo. De este modo, y en este caso en particular en las que hay tres observaciones en el tiempo para cada individuo, se tendrá por lo tanto, tres filas diferentes para cada uno de ellos. La variable que indica el tiempo es denominada "Año", mientras la que identifica al sujeto es llamada "NroCuest", la cual identifica a cada uno de los individuos con un número diferente, permitiendo de este modo reconocer a cada sujeto en un punto específico en el tiempo. Se recuerda que se cuenta con información de 582 jefes de hogar ( $N=582$ ) en tres momentos en el tiempo ( $t=3$ ), es decir, 1746 observaciones ( $N*t=582*3=1746$ ).

Se comenzará describiendo la variable dependiente "Trabaja". Seleccionando como base el año 2009, en la Tabla 1 se analiza la respuesta de los jefes de hogar a la pregunta "*En la última semana, ¿Usted trabajó?*".

Se observa que el 61% de los jefes de hogar respondió afirmativamente la pregunta, es decir, que considera haber trabajado en la última semana. Como esta metodología permite que sea el propio actor quien defina si ha trabajado o no la última semana, ello admite que existan diferencias con otro tipo de métodos de generación de información ya que, por ejemplo, puede darse el caso de alguien que haya realizado algún tipo de actividad, trabajo o "changa" y que ello no represente para él un "trabajo". Estudios previos del propio Observatorio Social han identificado que en la categoría "no trabajo" se suelen encontrar los casos considerados en la categoría "subocupado o trabajo demandante" cuyos

guarismos son provistos por la Encuesta Permanente de Hogares (INDEC). Es por ello que, generalmente, el porcentaje de los que no trabajan suele ser mayor según la metodología utilizada en el Panel de Hogares, lo que se ejemplificará más adelante. Sin embargo, que un 39% respondan que “no” trabajaron la última semana parece ser un número elevado, dada la condición de jefe de hogar, la cual está asociada a ser el sostén económico del mismo y a que la coyuntura económica del momento no mostraba indicadores tan negativos como para justificar dicho valor porcentual.

**Tabla 1:**  
*Proporción de acceso al trabajo de los jefes de hogar en el año 2009*

¿En la última semana, usted trabajó?	Frecuencia	Porcentaje
No	227	39%
Si	355	61%
Total	582	100%

Nota: Análisis del acceso al trabajo para todos los jefes de hogar santafesinos tomando como base el año 2009. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

La explicación se encuentra en el hecho de que dicha pregunta se le realiza a todos los jefes de hogar sin distinción, lo que puede incluir, por ejemplo, personas mayores que no trabajan por ser económicamente inactivos. De este modo, para poder medir si la necesidad de trabajo fue cubierta o no, se debe considerar sólo a los económicamente activos, esto es, a las personas que trabajan o no trabajan pero buscan trabajo, descartando así a los denominados inactivos (niños, personas mayores o cualquier persona que no trabaje ni busque trabajo).

Antes de pasar a considerar la respuesta al trabajo de los jefes de hogar económicamente activos, pretendo dejar en claro mediante un ejemplo concreto la diferencia entre la metodología utilizada aquí para medir la *necesidad* de trabajo con otro tipo de indicadores como, por ejemplo, la tasa de desocupación para el Gran Santa Fe elaborada por el INDEC. Dicha tasa para el segundo trimestre de 2009 (mismo período en que se realizó el relevamiento del Panel de Hogares) se ubicó en el 11,1%, mientras que la Tasa de No Trabajo<sup>5</sup> obtenida a partir del Panel de Hogares trepó al 20,5%. La diferencia estriba en

<sup>5</sup> Se obtiene de considerar a todos los integrantes de la Onda 2009 que sean económicamente activos, es decir, independientemente de si son o no jefes de hogar, alcanzando dicha condición a un total de 1885 personas.

que para el INDEC si una persona trabajo al menos una hora en forma remunerada la semana previa al relevamiento es considerada un Ocupado (o bien, Subocupado). Precisamente, la EPH considera al trabajo un *satisfactor*, por lo que termina incluyendo a individuos que cuando son consultados acerca de la *necesidad* de trabajo, responden afirmativamente. Mientras que para el Panel de Hogares, la condición del individuo quedará establecida en base a lo que cada persona responda, es decir, si el individuo no reconoce que la actividad que realizó la semana pasada haya sido “trabajo”, esa persona será considerada dentro de la Tasa de No Trabajo, expresando en efecto una carencia.

**Tabla 2:**

*Indicadores Laborales Comparados:  
Panel de Hogares y Encuesta Permanente de Hogares*

Panel de Hogares – Obs. Soc.	Tasa de “Actividad”	Tasa de "Trabajo"	Tasa de "No Trabajo"	Tasa de "Trabajo Demandante"
Junio 2009	48,60%	38,70%	20,50%	21,10%
Diciembre 2010	47,00%	38,80%	18,90%	19,60%
Febrero 2012 <sup>6</sup>	44,70%	37,50%	16,10%	17,80%
Encuesta Permanente de Hogares - INDEC	Tasa de Actividad	Tasa de Empleo	Tasa de Desocupación	Tasa de Subocupación Horaria
2 <sup>do</sup> trimestre 2009	42,80%	37,90%	11,30%	14,30%
4 <sup>to</sup> trimestre 2010	43,50%	39,70%	8,80%	12,60%
1 <sup>er</sup> trimestre 2012	44,90%	41,90%	6,60%	8,70%

Nota: Comparación entre los indicadores del mercado laboral para el Gran Santa Fe obtenidos mediante el Panel de Hogares, elaborados por el Observatorio Social, y los obtenidos mediante la Encuesta Permanente de Hogares, elaborados por el INDEC. Para una explicación más detallada sobre el cálculo de las tasas, ver anexo. Elaboración propia. Fuentes: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral e Instituto Nacional de Estadística y Censos – INDEC: E.P.H., procesado con Redatam+SP.

A modo comparativo, se presenta en la Tabla 2, los indicadores sobre el tema “trabajo” que se obtienen a partir del Panel de Hogares que realiza el Observatorio Social y los que realiza el INDEC a partir de la Encuesta Permanente de Hogares (EPH). Se puede observar que las tasas presentan valores similares pero no iguales, debido a las diferencias en la forma de medición de las variables. Dichas diferencias son intencionales, ya que como se mencionó anteriormente, lo que se pretende por medio del Panel de Hogares es cuantificar

<sup>6</sup> Es necesario aclarar que la fecha “Febrero 2012” corresponde a la fecha de finalización de la Onda que aquí se denominó 2011, debido a que la misma comenzó en Noviembre de dicho año. Debido a que la tabla utiliza información proporcionada por el Observatorio Social, se decidió utilizar en este caso la denominación allí presentada. Para el resto del documento, la misma continuará con la denominación que se le ha dado en el presente trabajo.

la “*necesidad*” de trabajo y no sólo el “*satisfactor*” que determina la situación laboral de la persona.<sup>7</sup>

A modo general, se puede decir que la Tasa de Actividad correspondiente al Panel de Hogares, presenta valores superiores a los elaborados por el INDEC, a excepción del año 2012. Las diferencias entre la tasa de “Trabajo” y su equivalente, la de “Empleo”, son incluso aún menores que las encontradas en las tasas de actividad. Las mayores diferencias se encuentran entre la “Tasa de No Trabajo” y la de “Desocupación”, sobre las cuales ya se hizo referencia, y en la comparación entre la “Tasa de Trabajo Demandante” y su análogo, la “Tasa de Subocupación Horaria”. Se puede observar como estas últimas dos tasas elaboradas con la información generada por el Panel de Hogares presentan valores mayores que las elaboradas por el INDEC, al medir la “*necesidad de trabajo*” desde la propia visión del actor social. La diferencia en las tasas de “no trabajo” o “trabajo demandante”, puede interpretarse como un alto grado de insatisfacción en la población, ya que la distancia entre las necesidades laborales y sus satisfactores es elaborada a partir de la propia visión que tienen los individuos acerca de las características percibidas sobre su trabajo (Observatorio Social, 2015).

#### **4.1.2 Trabajando con un panel no balanceado**

Tomando en consideración que la pregunta “En la última semana, ¿usted trabajó?” se realiza a todos los jefes de hogar sin distinción y el objetivo de esta tesis es describir la posible existencia de diferenciales en el acceso al trabajo, es necesario considerar solamente a los jefes de hogar *económicamente activos*, es decir, aquellos que trabajan o no trabajan pero buscan trabajo.

Se filtrarán entonces los casos en la base de datos según este criterio, lo que implica tener en consideración algunas cuestiones. Al filtrar la base, se reducirá el número de casos, descartando aquellos individuos económicamente inactivos durante todo el período. Sin embargo, hay que tener en cuenta que se está trabajando con información que incluye observaciones a lo largo del tiempo, por lo que si un individuo se considera inactivo en el año 2009, eso puede no ser cierto en el año 2010 y 2011, o viceversa. Esto se traduce en que no sólo se pierden los casos que declaran ser económicamente inactivos en los tres

---

<sup>7</sup> Para una explicación más detallada acerca de las diferencias en los cálculos de cada tasa, ver Apéndice: Aclaraciones sobre el cálculo de tasas.

períodos sino que además, pueden existir casos con sólo algunas observaciones perdidas, generando así un *panel no balanceado*.

De aquí se desprenden algunos tópicos interesantes. Lo primero es que gracias a la ventaja de los modelos de efectos aleatorios o multinivel que permite trabajar con paneles no balanceados, el hecho de que se pierdan una o algunas de las observaciones de un caso, no significa que el caso se descarte. Los datos faltantes no plantean mayores problemas ya que el modelo no asume el mismo número de ocasiones idénticas para todas las personas, lo que permite trabajar con paneles no balanceados sin pérdida de información. Esta es una ventaja que tiene el modelo por sobre los modelos de análisis de varianza o los modelos de efectos fijos.

Desde un punto de vista más analítico de la información, el hecho de que se encuentren datos faltantes con ciertos “patrones”, podría estar escondiendo efectos interesantes. Por ejemplo, es sabido que ante situaciones de desempleo elevado y persistente en el tiempo, puede darse la situación conocida en la literatura como “trabajadores desalentados”. Se denomina así a las personas desocupadas que luego de un tiempo de búsqueda de empleo pasan a formar parte del colectivo de los inactivos debido a la escasez del mismo, formando parte del llamado desempleo oculto (Neffa, Panigo y Pérez, 2000). Esto podría verse reflejado, a priori, en una distribución de la información que sigue el siguiente patrón: primero un aumento del número de individuos que buscan trabajo y luego un aumento en el número de inactivos (traducido en una disminución de las observaciones). Lo contrario también puede ser cierto, es decir, si en la información se encuentra una gran cantidad de observaciones perdidas en los primeros períodos debido a la existencia de un gran número de inactivos, seguida de un aumento en la cantidad de sujetos que trabajan o buscan trabajo, podría deberse a situaciones en las que se da un crecimiento de la economía, dónde los trabajadores desalentados “salen” de la inactividad para volcarse al mercado de trabajo.

Para poder confirmar o rechazar estas hipótesis, no basta con analizar el patrón de información de los datos, sino que también hay que realizar un análisis del contexto económico en el que los individuos se desenvuelven. Es decir, si no existen en el contexto económico coyunturas de crecimiento o depresión económica agudas que validen la hipótesis de los trabajadores desalentados, la existencia de patrones de datos faltantes puede encontrar otras explicaciones. Por ejemplo, si en la base de datos existe un número

elevado de personas mayores o cerca de la edad de jubilación, esto podría traducirse en un aumento de los inactivos en las últimas observaciones en el tiempo.

Por el momento se dejarán estas hipótesis planteadas que se irán discutiendo a lo largo de este apartado y en el próximo, cuando se analice el patrón de datos antes de comenzar a especificar el modelo y se dará lugar ahora al análisis de las variables involucradas en el modelo, teniendo en cuenta solamente a los jefes de hogar económicamente activos.

#### 4.1.3 Re-analizando las variables involucradas en el modelo

La Tabla 3, ilustra cómo cambia la respuesta a la pregunta “*En la última semana, ¿Usted trabajó?*”, considerando sólo a los jefes de hogar económicamente activos durante los tres años.

**Tabla 3:**

*Acceso al trabajo de los Jefes de Hogar Económicamente Activos 2009-2011*

Año	En la última semana, ¿usted trabajó?					
	No		Sí		Total	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
2009	34	8,7%	355	91,3%	389	100,0%
2010	33	8,9%	338	91,1%	371	100,0%
2011	23	6,9%	310	93,1%	333	100,0%

Nota: Análisis del acceso al trabajo para los jefes de hogar santafesinos económicamente activos para el período 2009-2011. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

Se puede observar cómo se redujo en el año 2009, el número de personas que respondieron que no trabajaron en la última semana si consideramos solamente a los jefes de hogar económicamente activos, pasando de 227 a tan sólo 34 casos. Obviamente, el número de personas que respondió afirmativamente no se ha modificado<sup>8</sup>. El 8,7% es la Tasa de No Trabajo para los jefes de hogar, un número considerablemente menor al 20,5% calculado para el total de los individuos de la Onda 2009 económicamente activos. Este resultado es más acorde a la condición de jefe de hogar que se asocia comúnmente con la figura del sostén económico de los hogares y también a la coyuntura económica de dicho momento. Se destaca nuevamente, la diferencia con la tasa de desocupación que realiza el INDEC, la cual vuelve a estar por debajo de la Tasa de No Trabajo, tomando un valor de 5,1%

<sup>8</sup> Vale aclarar que entre aquellos que respondieron afirmativamente, se incluyen dos categorías: “trabaja” y “trabaja y busca”. Dicha distinción, si bien guarda implicaciones de análisis que pueden ser relevantes, no será considerada en este trabajo ya que el modelo que se utilizará a continuación admite sólo dos respuestas “sí” y “no”, debiendo reducirse el espectro de respuestas a sólo estas dos categorías.

considerando sólo a los jefes de hogar económicamente activos (aplicada al mismo período). En cuanto a los años 2010 y 2011, las tasas de desocupación tomaron los valores de 3.8% y 2.7% respectivamente, evidenciando nuevamente, las diferencias en los indicadores que miden “necesidad” y los que miden “satisfactores”.

Otros resultados interesantes se desprenden de la Tabla 3. Primero y principal, que el número de observaciones por año se redujo de 582 a 389, 371 y 333 para los años 2009, 2010 y 2011, respectivamente. Es importante comprender qué representan dichos valores, ya que podría interpretarse *erróneamente* cómo el número de casos o individuos que quedaron en el panel luego del proceso de filtrado. Lo que esta información dice es que 389 jefes de hogares en el año 2009 son económicamente activos. 34 de ellos no trabajan y 355 sí lo hacen. Para el año 2010, son ahora 371 los jefes de hogar que se consideran económicamente activos, *independientemente* de su condición en el año 2009 o 2011. Dicho de otro modo, se está haciendo aquí una interpretación de los datos como si proviniesen de un análisis de combinaciones de cortes transversales independientes en el tiempo, es decir, no se está tomando en consideración que sean los *mismos* individuos a lo largo del tiempo, que es la característica principal de un panel de datos y lo que se intentará modelar en el capítulo siguiente. Para que la diferencia quede clara, con esta información se puede decir, por ejemplo, que el número de personas económicamente activas se redujo de 389 a 371 del año 2009 al 2010, pero eso no quiere decir que sean las *mismas* personas. Puede haber casos de individuos que se declararon inactivos en 2009 y se sumaron a la fila de los económicamente activos en 2010, o viceversa.

Teniendo en consideración entonces dichas limitaciones, se pueden sacar conclusiones importantes sobre los resultados presentados en la Tabla 3. La misma muestra que la Tasa de No Trabajo para los jefes de hogar, mostró un leve aumento del año 2009 al 2010. Sin embargo, si se consideran los valores absolutos, se puede observar en realidad que el número de jefes de hogar que declaro no trabajar, ha disminuido. Por lo tanto, el aumento en la Tasa se explica por una disminución de las personas económicamente activas, que va acompañado además de una disminución en el número de personas que trabajan.

Lo mismo ocurre a la inversa con la Tasa de Trabajo si se consideran los años 2010 y 2011. La misma registra un aumento porcentual pero una disminución en valores absolutos del número de personas que trabajan. Se observa que lo que disminuyó fuertemente fue el

número de personas económicamente activas, lo que se tradujo en parte en una disminución de los que trabajan, pero también de los que no trabajan.

En líneas generales puede decirse entonces que existe una tendencia a la baja en el número de jefes de hogar económicamente activos para el período 2009-2011, que se traduce en una disminución en valores absolutos tanto del número de desocupados como de ocupados<sup>9</sup>. Si bien en el año 2009 la economía Argentina registró un descenso pronunciado en el Producto Bruto Interno (PBI), sobretodo en el primer trimestre de dicho año, la misma se recuperó al año siguiente y mantuvo dicha tendencia positiva en el año 2011 (INDEC, 2016). Lo mismo puede observarse si se analiza la Tasa Anual de Crecimiento realizada por el IPEC (Instituto Provincial de Estadísticas y Censos, 2016) para la Provincia de Santa Fe, reproducida en la Figura 1. La misma está elaborada en base al Producto Bruto Geográfico, valuado a precios de productor<sup>10</sup>, y permite apreciar el pronunciado descenso en 2009 y la recuperación en 2010 y 2011 a niveles similares a los presentados en años anteriores.

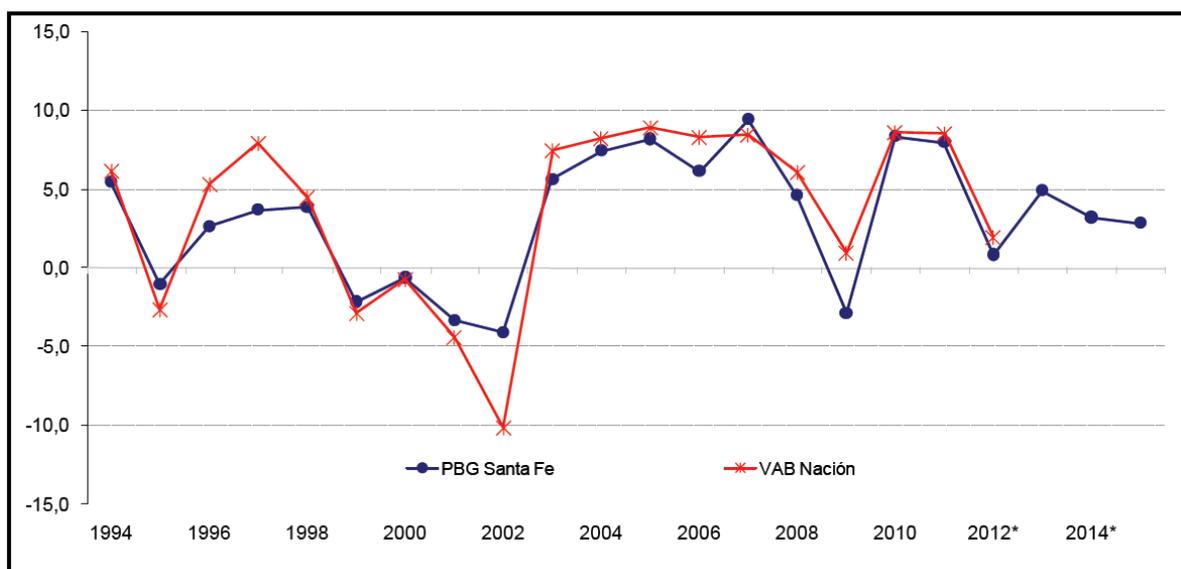


Figura 1. Tasa Anual de Crecimiento del Producto Bruto Geográfico (PBG) de la Provincia de Santa Fe. A precios constantes de 1993 (año base) en porcentaje. VAB Nación: Valor Agregado Bruto a precios de productor, mismo año base, a modo de comparación. Reproducido de "PBG. Producto Bruto Geográfico. 1993-2015. Resultados Preliminares", p. 7, por IPEC – Provincia de Santa Fe, Mayo 2016.

Teniendo en cuenta esta información, la hipótesis de los trabajadores desalentados pareciera no explicar estas tendencias, al menos no en su totalidad. Para que la misma sea

<sup>9</sup> Vale aclarar que no se puede afirmar que dicha disminución sea estadísticamente significativa, puesto que sólo se está realizando un análisis descriptivo de la muestra.

<sup>10</sup> Es decir, "...incluye los impuestos a la producción (sobre nómina salarial, inmobiliario, patente automotor) netos de subsidios, más los impuestos sobre los productos (impuesto a los ingresos brutos y derechos de exportación), sin incorporar el Impuesto al Valor Agregado (IVA)." (Santa Fe [provincia], 2016, p.5)

válida, la caída en la actividad económica debería haber sido, además de fuertemente pronunciada, constante a la baja por un período de tiempo prolongado, ya que la misma debería explicar, no sólo la disminución del número de personas sin acceso al trabajo que se vuelcan a la inactividad, sino también la disminución del número de los que si tenían acceso al trabajo, los cuales pasarían directamente a la inactividad, sin transitar por un período de búsqueda de trabajo. De este modo, la explicación a este fenómeno tal vez podría encontrarse y/o complementarse, estudiando la composición etaria de los jefes de hogar y el patrón de observaciones faltantes. Se analizará aquí la composición etaria de los jefes de hogar y se terminará de completar dicho análisis en la sección siguiente cuando se analice el patrón de observaciones.

**Tabla 4:**  
*Estadísticos descriptivos para la variable "Edad"*

Estadísticos		Valores
Número de observaciones		1093
Media		48,72
Error típico de la media		,338
Mediana		49,00
Moda		53
Desvío típico		11,176
Varianza		124,897
Asimetría		-,011
Error típico de asimetría		,074
Curtosis		-,467
Error típico de curtosis		,148
Rango		59
Mínimo		22
Máximo		81
Suma		53254
Percentiles	10	33,00
	25	41,00
	50	49,00
	75	57,00
	90	63,00

Nota: Estadísticos para la variable "Edad" de los jefes de hogar santafesinos económicamente activos para el período 2009-2011. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

Para el análisis de la edad, se tomará el total de las observaciones para los tres años luego de filtrar la base manteniendo sólo a los jefes de hogar económicamente activos<sup>11</sup>, lo que da un total de 1093 observaciones. La edad promedio es de 49 años aproximadamente, un valor que es prácticamente igual tanto para la media como para la mediana, lo que significa que la variable edad posee una distribución lo bastante simétrica respecto de su media. Esto puede observarse gráficamente en la Figura 2 como así también en los demás valores de dispersión de la Tabla 4. Se puede ver que hay una muy leve asimetría a la izquierda, casi imperceptible, y lo mismo ocurre con el valor de curtosis, que es pequeño, lo que indica una baja frecuencia de datos alejados y por ende, colas más finas.

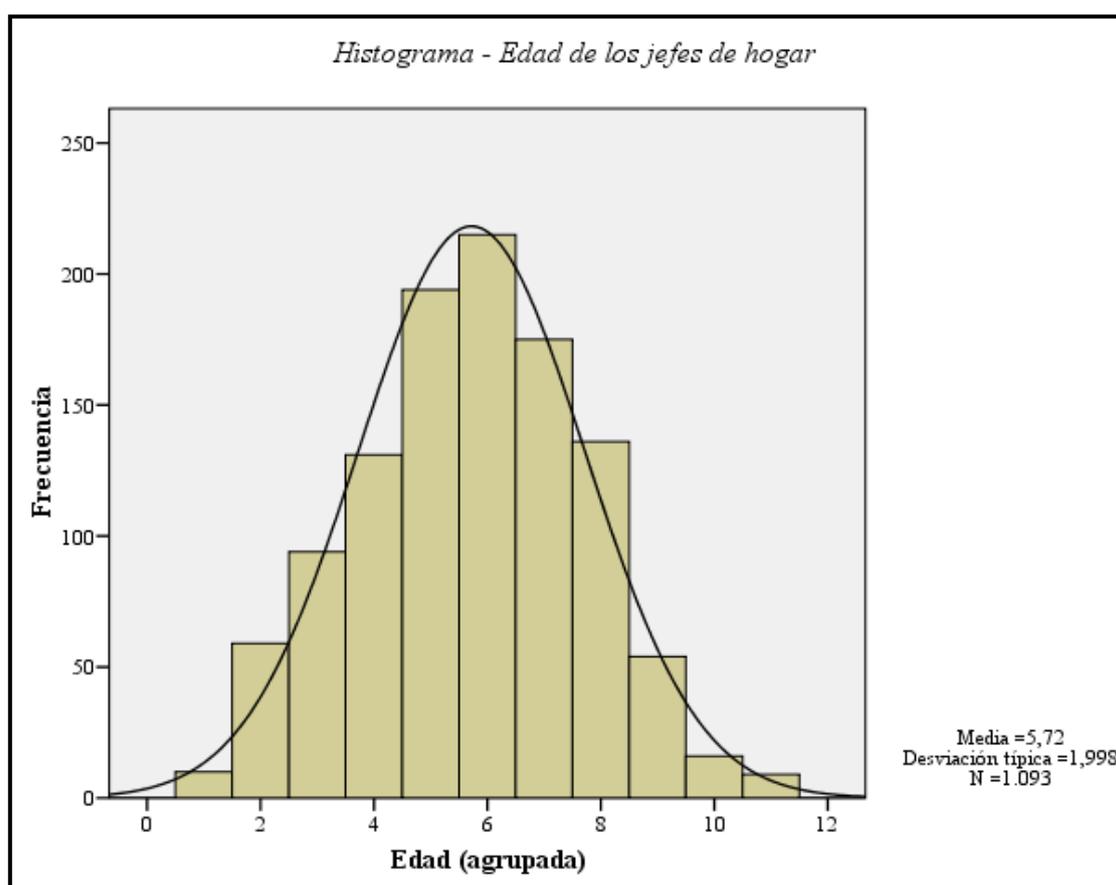


Figura 2. Histograma de la variable “Edad” de los jefes de hogar santafesinos económicamente activos para el período 2009-2011. La variable “Edad” fue recategorizada en 11 intervalos bajo la Regla de Sturges. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

Lo que se destaca también es el rango, es decir, la distancia entre la edad mínima y máxima observadas, el cuál es casi de 60 años. Como se puede observar en la Tabla 4, la edad mínima es de 22 años y la máxima de 81. La presencia de jefes de hogar con una edad

<sup>11</sup> De aquí en más, excepto que se aclare lo contrario, se trabajará sobre la base de datos de jefes de hogar económicamente activos para el período 2009-2011, que cuenta con 1093 observaciones en total.

mínima tan baja puede estar indicando por ejemplo, la presencia en los datos de jóvenes estudiantes que vienen a la ciudad de Santa Fe, provenientes de diferentes localidades, a cursar estudios superiores, como así también, de jóvenes de bajos recursos que suelen formar sus propias familias a una edad más temprana. Precisamente, el percentil 10 indica que un 10% de los jefes de hogar tienen 33 años o menos. Por otro lado, jefes de hogares con una edad tan avanzada podría deberse a la existencia de personas mayores que viven solas o que continúan siendo cabeza de familia aunque no aporten al ingreso familiar. Justamente, el percentil 90 cae prácticamente sobre la edad jubilatoria de los hombres, quedando menos del 10% por fuera de dicho límite.

Estos resultados son interesantes, ya que podrían explicar a priori porqué se produce la variación, o mejor dicho, la disminución en el número de jefes de hogar económicamente activos. Por un lado, el tener un número importante de jefes de hogar menores de 30 años puede explicar en parte dicha disminución, ya que se sabe que las dificultades en el ingreso de los jóvenes al mercado laboral aumentaron en las últimas décadas, y sobre todo en los últimos años. Un informe de la Organización Internacional del Trabajo para el año 2013, refiere precisamente a que la lenta recuperación de la economía mundial en el año 2012 había agravado la crisis del empleo juvenil que se venía manifestando desde años anteriores, dificultando el acceso para muchos de ellos, “hasta tal punto, que muchos están renunciado a seguir buscando” (OIT, 2013, p. 1), aportando así a la hipótesis de los trabajadores desalentados, al menos, entre los más jóvenes.

Más precisamente, lo que se acrecentó fue la dificultad de ingresar al mercado de trabajo con un puesto laboral formal. La inserción laboral de los jóvenes, que antes era entendido como el momento en que los mismos accedían por vez primera al empleo luego de pasar por su período escolar, es ahora visto más bien como un proceso, en el que se combinarán y alternarán períodos de ocupación, desocupación, inactividad, de vuelta a los estudios, etc. Es decir, que lo que está en discusión es cierta “linealidad” en las trayectorias, ya no se trata simplemente del pasaje de la educación al trabajo, sino que se conforma de múltiples entradas y salidas entre ambas esferas. Como explica Otero (2009), el pasaje no es sólo lineal, educación para el empleo, sino que también se observan situaciones simultáneas, educación y empleo, pero también reversibles, de la educación al empleo y de vuelta a la educación. (Otero, 2009, pag 71).

Si bien la tendencia general estaría llevando a una extensión de la condición juvenil, en la que los jóvenes postergan su independencia y la constitución de una familia propia, las nuevas formas “no lineales” de las trayectorias juveniles y las dificultades para acceder a un puesto de trabajo formal, lleva a la existencia de historias de vida diversas entre las cuales se encuentran casos de jóvenes que, por ejemplo, hoy tienen un trabajo y se van a vivir solos pero mañana lo pierden y deben regresar a casa de sus padres (Dávila, Ghiardo y Medrano, 2008).

Si por un lado entonces, la volatilidad de las trayectorias y las dificultades en el ingreso al trabajo formal podría explicar en parte la disminución del número de jefes de hogar económicamente activos, al menos entre los más jóvenes, por el otro lado, el tener un número importante de jefes de hogar justo por encima de la edad jubilatoria también puede explicar dicha disminución, ya que los mismos irían abandonando dicha condición a medida que se jubilen y pasen a ser considerados “inactivos”.

De este modo, pareciera ser que la explicación a la disminución en el número de jefes de hogar económicamente activos vendría dada por una combinación de factores y no por alguno de ellos en particular. Como se dijo anteriormente, se completará este análisis en la próxima sección, al estudiar el patrón de frecuencias de las observaciones de los datos del panel.

En cuanto al género de los jefes de hogar, la Tabla 5 muestra que más del 75% de los hogares en el panel tienen un jefe de hogar de sexo masculino en cualquiera de los 3 años. Al igual que cuando se analizó anteriormente la respuesta al trabajo en la Tabla 3, hay que destacar dos cuestiones. Primero, que aquí se interpretan los datos correspondientes a cada año como si fuesen independientes, es decir, como una combinación de cortes transversales independientes en el tiempo, y segundo, que el efecto de una gran disminución en el número de jefes de hogar económicamente activos lleva a interpretaciones diferentes de los resultados, ya sea que se analicen los valores absolutos o los porcentajes.

En este último sentido, si se considera a los jefes de hogar masculinos, se puede observar que si bien el porcentaje correspondiente a los mismos aumentó a lo largo de los tres años, en valores absolutos su número disminuyó. En cuanto a las jefas de hogar, en valores absolutos su tendencia es decreciente, mientras que si se examinan los valores porcentuales, hay un mínimo aumento del año 2009 al 2010.

**Tabla 5:***Composición del género de los jefes de hogar 2009-2011*

Año	Sexo					
	Masculino		Femenino		Total	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
2009	296	76,1%	93	23,9%	389	100,0%
2010	282	76,0%	89	24,0%	371	100,0%
2011	262	78,7%	71	21,3%	333	100,0%

Nota: Análisis del género de los jefes de hogar santafesinos económicamente activos para el período 2009-2011. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

La conclusión más importante a la que se puede arribar aquí es que, en términos porcentuales, el promedio de hogares con jefe de hogar masculinos para los tres años es aproximadamente el 77%, siendo el 23% restante, el promedio de hogares para el período que cuentan con un jefe de hogar de sexo femenino.

Se describirá ahora la principal variable explicativa en relación al objetivo general de esta tesis, el nivel de escolaridad de los jefes de hogar. Es importante aclarar que la misma es considerada aquí un “satisfactor”, ya que refiere al nivel educativo alcanzado por cada uno de los jefes de hogar, y no el nivel que desearía alcanzar. Además, si bien se expondrá a continuación un cuadro de frecuencias según las categorías originales, la misma será recategorizada con los fines de lograr resultados estadísticamente significativos y una categoría de referencia que permita una mejor interpretación de los resultados.

**Tabla 6:***Composición del nivel educativo de los jefes de hogar 2009-2011*

Nivel de Escolaridad Alcanzado	Año de la encuesta					
	Año 2009		Año 2010		Año 2011	
	Recuento	Porcentaje	Recuento	Porcentaje	Recuento	Porcentaje
Primario incompleto	37	9,5%	25	6,7%	21	6,3%
Primario completo	89	22,9%	93	25,1%	81	24,3%
Secundario incompleto	51	13,1%	55	14,8%	60	18,0%
Secundario completo	95	24,4%	95	25,6%	75	22,5%
Terciario incompleto	16	4,1%	10	2,7%	11	3,3%
Terciario completo	39	10,0%	28	7,5%	30	9,0%
Universitario incompleto	24	6,2%	26	7,0%	18	5,4%
Universitario completo	31	8,0%	34	9,2%	35	10,5%
Posgrado	7	1,8%	5	1,3%	2	,6%
Total	389	100,0%	371	100,0%	333	100,0%

Nota: Nivel educativo alcanzado de los jefes de hogar santafesinos económicamente activos para el período 2009-2011. Las categorías “Analfabeto” y “Escuela especial” no se muestran por no presentar valores. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

En la Tabla 6, se puede examinar la composición del Nivel Educativo alcanzado por los jefes de hogar en cada uno de los tres años. En los dos primeros años, la categoría con la mayor cantidad de observaciones es la de “Secundario completo”. Sin embargo, en el año 2011 se nota una disminución en dicha categoría, pasando a ser la categoría “Primario completo” la de la mayor cantidad de observaciones en dicho año. Se podría decir que en cada año, aproximadamente el 50% de los jefes de hogar se encontrará en una de estas dos categorías. Se destaca también el número de jefes de hogar con “Secundario incompleto”, siendo además la única categoría que, junto a los que obtuvieron título universitario, mantiene una tendencia creciente en el tiempo. Con una tendencia inversa, se encuentra el número de individuos que poseen “Primario incompleto”, que pasan de un 9,5% en 2009 a un 6.3% en 2011, por debajo incluso del número de personas con “Universitario completo”.

Es difícil sacar conclusiones y buscar factores que expliquen las tendencias y los cambios en la composición de las categorías educacionales para los jefes de hogar del período, por diversas razones. Primero y principal, porque hay un marcado descenso en el número de jefes de hogar económicamente activos, lo que modifica la base sobre la que se relacionan las distintas categorías. Segundo, porque como se explicó anteriormente, como la base fue filtrada según la condición de jefes de hogar económicamente activos, esto produjo que el panel deje de estar balanceado, es decir, ya no se cuenta con todas las observaciones para todos los individuos. Esto provoca que no se pueda definir que los cambios en los niveles educativos, año tras año, se deba única y exclusivamente a cambios en el nivel educativo de los mismos individuos, sino que, por el contrario, es más probable que se deba a la incorporación (o supresión) de distintos jefes de hogar en los distintos momentos en el tiempo, sobre todo al ser un estudio en tres años consecutivos, lo que reduce las chances de encontrar cambios significativos en los niveles de estudio para un individuo dado.

Sin embargo, a priori se podría decir que, probablemente, en las tendencias y los cambios en la composición de las categorías puedan esconderse ambos efectos, los producidos por la incorporación o eliminación de distintos jefes de hogar y los producidos por cambios al *interior* del individuo, es decir, jefes de hogar que completan alguno de los niveles de estudio pasando de una a otra categoría. Se verá en la próxima sección una medida de los cambios al *interior* del individuo, lo que permitirá saber si existen dichos casos y, lo que es

más importante aún, que se deba entonces considerar al nivel educativo alcanzado como una variable que *cambia* en el tiempo.

En la Tabla 7 se muestran los totales de cada categoría agrupados para los tres años. Queda en evidencia que más del 30% posee “Nivel Educativo Bajo”, lo que incluye las categorías “Primario incompleto” y “Primario completo”, tal como se la definirá en esta Tesis. Los jefes de hogar que posean “Secundario incompleto” o “Secundario completo”, se incluirán en la categoría “Nivel Educativo Medio”, los que suman un 39.4% para el total de los tres años. Si se tiene en cuenta las primeras cuatro categorías de la Tabla 7, más del 70% de los jefes de hogar posee educación media o básica, según la nueva recodificación. Por lo tanto, el casi 30% restante (28,9%) recae en la categoría definida “Nivel Educativo Alto”, que incluye desde la categoría “Terciario incompleto” hasta “Posgrado”, inclusive.

**Tabla 7:**

*Nivel educativo de los jefes de hogar. Total agrupado 2009-2011*

Nivel de Escolaridad Alcanzado	Total agrupado 2009-2011	Porcentaje	Porcentaje acumulado
Primario incompleto	83	7,6%	7,6%
Primario completo	263	24,1%	31,7%
Secundario incompleto	166	15,2%	46,9%
Secundario completo	265	24,2%	71,1%
Terciario incompleto	37	3,4%	74,5%
Terciario completo	97	8,9%	83,4%
Universitario incompleto	68	6,2%	89,6%
Universitario completo	100	9,1%	98,7%
Posgrado	14	1,3%	100,0%
Total	1093	100,0%	-

Nota: Nivel educativo alcanzado de los jefes de hogar santafesinos económicamente activos. Total agrupado para el período 2009-2011. Las categorías “Analfabeto” y “Escuela especial” no se muestran por no presentar valores. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

Para finalizar este apartado, se presenta a continuación en la Figura 3, la proporción de Jefes de Hogar que trabajan en relación al nivel educativo alcanzado, para cada uno de los tres años. El primer grupo de barras refiere a la proporción de los jefes de hogar con nivel educativo bajo que trabajan. Del mismo modo, el segundo grupo refiere a aquellos que trabajan y poseen nivel educativo medio y, por último, el grupo de los que trabajan y poseen nivel educativo alto. A simple vista, si bien la proporción de los jefes de hogar que trabajan nunca es menor al 0.8 para cualquiera de los niveles educativos, es evidente que a medida que aumenta el nivel educativo alcanzado por los jefes de hogar, aumenta la

proporción de los que trabajan en relación a los que no trabajan. Si se comparan los grupos, se puede observar que la proporción de los jefes de hogar que trabajan con nivel educativo bajo es menor a la proporción que tienen los jefes de hogar con un nivel educativo más alto, en cualquiera de los tres años. Si bien las diferencias se achican, de manera similar se puede decir que aquellos que poseen nivel educativo alto tienen una proporción mayor de jefes de hogar empleados que los poseen nivel educativo medio, particularmente en los últimos dos años.

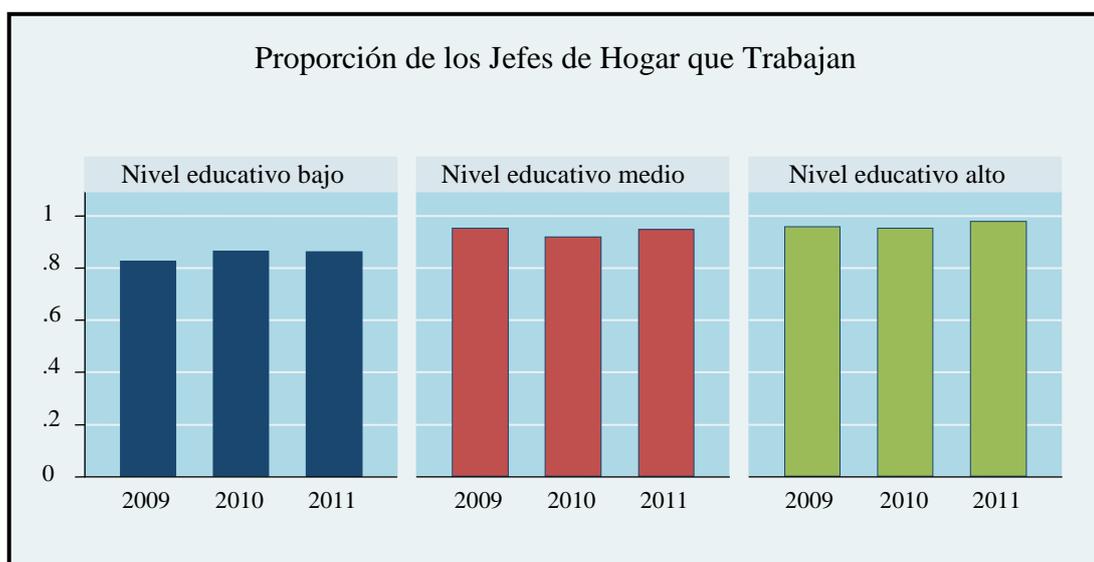


Figura 3. Proporción de los Jefes de hogar santafesinos económicamente activos que trabajan según nivel educativo alcanzado para el período 2009-2011. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral

#### 4.2. Estimación de los diferenciales en el acceso al trabajo de los jefes de hogar

Para comenzar, se reproduce el modelo que se estimará tal cuál fue definido en el apartado 3.5.4 del capítulo anterior:

$$\text{Logistic}(\text{Trabaja}_{i_t}) = \gamma_{00} + \beta_{01}\text{Sexo}_{i_t} + \beta_{02i}\text{Año}2010_{i_t} + \beta_{03i}\text{Año}2011_{i_t} + \beta_{1i}\text{Educ\_categor}_{i_t} + \beta_{2i}\text{Edad}_{i_t} + \mu_{0i} \quad (8)$$

Si bien la base original de los jefes de hogar de la ciudad de Santa Fe para el período 2009-2011 estaba compuesta por 582 individuos, como se explicó en el apartado anterior, en miras de cumplir con el objetivo de esta tesis que es estimar la existencia de posibles diferenciales en el acceso al trabajo de los jefes de hogar, se debió descartar aquellos sujetos que respondieron “no” trabajar pero debido a su condición de económicamente

inactivos. Si dichos individuos fuesen incluidos, se estaría sobreestimando el verdadero valor de la respuesta negativa al acceso al trabajo, representando un nivel de insatisfacción de la población santafesina respecto de si está cubierta la necesidad de trabajo, mayor al que realmente es.

Una vez descartados dichos casos, se mostró que el número de los jefes de hogar económicamente activos no era el mismo en los tres cortes temporales, reduciéndose de 389 en el año 2009 a 371 en el año 2010, para terminar en 2011 con 333 casos. Sin embargo, también se aclaró que esto no significaba que dichas observaciones correspondiesen a *exactamente* los mismos individuos a lo largo del tiempo, sino que podían existir diversas situaciones en donde algunos jefes de hogares se hayan declarado inactivos en un año determinado pero en otro no. Eso no impidió que no se realice un análisis de los datos considerándolos como si proviniesen de una combinación de cortes transversales independientes en el tiempo, tratando de plantear algunas hipótesis en base a la información brindada en cada período y teniendo en consideración otras variables como, por ejemplo, la edad. Para completar dicho análisis, se pretende ahora sumarle la información que puede brindar el estudio del patrón de datos perdidos.

#### **4.2.1 Estudiando el patrón de datos perdidos**

Cuando un panel cuenta con todas las observaciones en el tiempo para todos sus individuos, se denomina panel balanceado. En cambio, cuando no se cuenta con todas las observaciones para todos los casos, se dice que el panel es no balanceado. Es importante destacar que en realidad aquí no son observaciones específicamente “perdidas”, sino más bien “descartadas” por no cumplir las condiciones que se necesitan.

Por lo tanto, la variable de panel se definirá como no balanceada, mientras que la variable temporal puede ser descripta como balanceada pero con “brechas” o “vacíos”. La variable de tiempo se define como balanceada o equilibrada porque los valores son idénticos entre los individuos, es decir, todos fueron relevados en los mismos años, más allá de las brechas o espacios causados por las observaciones descartadas. Una variable tiempo no balanceada, es aquella en la que existen diferentes momentos de respuesta entre los distintos sujetos y es generalmente utilizada para estudios entre grupos, en los cuales se analizan las respuestas de cada grupo en diferentes momentos en el tiempo para analizar los efectos de, por ejemplo, un determinado fármaco.

El panel quedó compuesto por  $n = 417$  casos, tal como se puede apreciar en la Tabla 8, dónde se describe el patrón de datos perdidos. Evidentemente, este número es menor a los 582 individuos originales de la base de datos, pero es mayor a cualquier de los tres valores de jefes de hogar económicamente activos que se encontró en el análisis de corte transversal para cada uno de los tres años. Es que como se explicó, no se podía confirmar que fuesen exactamente los mismos jefes de hogar, ya que estos pueden cambiar su condición año tras año. Sin embargo, ahora sí se puede decir, que 417 jefes de hogar se han declarado económicamente activos en alguno de los tres años, de los cuales, 307 lo han sido en los tres momentos. En la primera fila de la Tabla 8 se puede observar que 307 jefes de hogar tienen datos completos, lo que se confirma con el hecho de que en ningún momento de los tres años el número de jefes de hogar económicamente activos superó dicho número. El resto del patrón es fácilmente describible, por ejemplo, a 46 jefes de hogar le falta la última observación, a 27 las dos últimas, a 10 las dos primeras, etc.

**Tabla 8:***Patrón de datos perdidos.*

Casos	Porcentaje	Porcentaje Acumulado	Patrón
307	73.62 %	73.62 %	111
46	11.03 %	84.65 %	11.
27	6.47 %	91.13 %	1..
11	2.64 %	93.76 %	.1.
10	2.40 %	96.16 %	..1
9	2.16 %	98.32 %	1.1
7	1.68 %	100.00 %	.11
417	100.00		XXX

Nota: Patrón de observaciones perdidas después de filtrar la base según jefes de hogar santafesinos económicamente activos. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

El patrón de datos perdidos parece ser bastante consistente con las hipótesis planteadas anteriormente. Por un lado, se observa que casi a un 18% (73 casos) le faltan sólo las dos últimas observaciones o la última, explicando así el gran descenso en el número de jefes de hogar económicamente activos. Esta información parecería sustentar la hipótesis de los jefes de hogar con una edad cercana a la edad jubilatoria, los cuales estaban por debajo del 10% del total de las observaciones, es decir, menos de 100 casos. Sería consistente entonces con la idea de que a medida que los jefes de hogar alcanzan la edad jubilatoria,

pasan a ser individuos económicamente inactivos. Por el otro lado, unos 37 casos, casi el 9%, siguen un patrón más irregular, faltándoles entre una o dos observaciones en forma más intermitente, lo que podría asociarse a la hipótesis de que el percentil 10 de la distribución de la variable edad corresponde a personas jóvenes, las cuáles, como ya se explicó, suelen experimentar problemas para encontrar un puesto laboral estable en sus primeros años de inserción en el mercado laboral. Precisamente, si antes la inserción laboral de los jóvenes era definida como el momento en que un joven accedía a un puesto de trabajo “formal”, ahora se piensa más bien en un “proceso de inserción”, que combinará períodos de desocupación con pasajes por diversidad de empleos y diferentes niveles de precarización, antes de lograr cierta estabilización (Dávila, Ghiardo y Medrano, 2008; Otero, 2009).

El patrón de datos puede definirse como intermitente o no monótono. Un patrón monótono de pérdida de datos es aquel en el que la no observación de una variable, implica la no observación de las siguientes. Si bien aquí, más del 90% de las observaciones se ajustan a dicha situación, estrictamente hablando, el patrón completo no lo hace y por lo tanto se dice que no sigue un patrón particular. En este sentido, es importante recordar que una de las ventajas de los modelos de efectos aleatorios que utilizan la estimación de máxima verosimilitud es que, en paneles de datos incompletos como este o que poseen variables temporales no equilibradas, permiten utilizar toda la información. Es decir, no sólo se considerará la información de los jefes de hogar económicamente activos en los tres años sino también la de aquellos con alguna observación perdida (es decir, inactivos en uno o dos años). Esta es una ventaja por sobre los modelos de efectos fijos, los cuales eliminan todos los casos que no cuenten con información completa. Si se utilizase dicho enfoque, el número de casos disponibles descendería, a priori, a 307 casos. Digo a priori porque además, el enfoque de efectos fijos sólo admite variables que cambien en el tiempo, tal como se explicó en el apartado anterior. De este modo, todos los individuos que no cambien su condición de ocupado en los tres años, serían también descartados. La poca variabilidad en las respuestas al *interior* de cada caso es lo esperable en estudios de este tipo y por ello el enfoque de efectos aleatorios es más atractivo para su modelización.

Además, otra de las ventajas de los estimadores de máxima verosimilitud es que, si el modelo es correctamente especificado, los mismos serán consistentes siempre que las respuestas sean perdidas aleatoriamente (MAR por sus siglas en inglés, missing at

random), que es el caso en el que la falta de información puede estar correlacionada con una de las variables observadas (Rabe-Hesketh y Skrondal, 2012). Tal como se analizó anteriormente, la condición de inactividad podría asociarse a la edad de los individuos, la cual es una variable observada en el modelo. Además, la existencia de datos faltantes se debe a una decisión de investigación con miras a no sobreestimar la respuesta negativa de los jefes de hogar sobre el acceso al trabajo, por lo que se puede concluir que la presencia de ellos no debería producir problemas en la estimación del modelo.

#### 4.2.2 Estadísticos descriptivos para variables de panel

Continuando con un análisis descriptivo, en la Tabla 9 se resumen algunos valores de interés para las variables del modelo, pero ahora teniendo en cuenta la estructura de panel existente en los datos.

**Tabla 9:**

*Estadísticos descriptivos para variables de panel*

Variable	Variabilidad	Desvío Estándar	Media	Min	Max	Observaciones
Año	overall	.8115077	2009.949	2009	2011	N = 1093
	between	.3363233		2009	2011	n = 417
	within	.7765449		2008.949	2.010.949	T-bar = 2.6211
NroCuest	overall	551.9666	938.0695	12	1755	N = 1093
	between	554.9101		12	1755	n = 417
	within	0		938.0695	938.0695	T-bar = 2.6211
Trabaja	overall	.2750112	.9176578	0	1	N = 1093
	between	.2432132		0	1	n = 417
	within	.1675799		.2509912	1,584324	T-bar = 2.6211
Sexo	overall	.421967	1.231473	1	2	N = 1093
	between	.4399202		1	2	n = 417
	within	0		1.231473	1.231473	T-bar = 2.6211
Edad	overall	11.17571	48.72278	22	81	N = 1093
	between	11.60692		23	80.33333	n = 417
	within	1.808355		28.38945	65.72278	T-bar = 2.6211
Educación en categorías	overall	.7781215	1.972553	1	3	N = 1093
	between	.7359731		1	3	n = 417
	within	.2803604		.6392193	2.972553	T-bar = 2.6211

Nota: Estadísticos descriptivos para variables de panel, aplicados a las variables involucradas en el modelo. Las palabras "Overall", "Between" y "Within" significan "Total", "Entre" e "Intra", respectivamente. Elaboración propia. Fuente: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral.

Lo que más me interesa resaltar aquí son los diferentes tipos de variabilidad que se pueden encontrar al trabajar con un panel de datos. La variación total o general (overall variation) de una variable es la variación *en el tiempo y los individuos*, es decir, la distancia de cada una de las observaciones en el tiempo respecto de la media general. Es la variabilidad total de una variable, comúnmente llamada desvío estándar. La variabilidad entre-individuos (between-subject), es la que mide la variabilidad entre los diferentes sujetos de la muestra, es decir, es una medida de la distancia entre la media de los individuos y la media general, comúnmente asociada al estimador de efectos aleatorios. Y por último la variabilidad intra-individuo (within-subject), es decir, la variabilidad en las respuestas de los individuos a lo largo del tiempo, medida como desviaciones en las observaciones de cada sujeto respecto de su propia media, la que va asociada al estimador de efectos fijos.

Lo más interesante para destacar de la Tabla 9, es que se puede visualizar el valor medio esperado para las variables sobre el total de las observaciones, junto con su desvío estándar (variabilidad total u overall). Algunos de estos valores tienen una interpretación válida, como por ejemplo, la edad media esperada es 48,72 (casi 49 años), o por ejemplo, la probabilidad media de la respuesta al trabajo, 0.91, la que se corresponde con los valores encontrados anteriormente (Tabla 3 y Figura 3).

Es interesante además, observar la composición de la variabilidad de cada una de las variables, lo que es importante para saber si dicha variabilidad se presenta más *entre* (between) individuos, sin considerar el tiempo, o al *interior* (within) de los individuos a lo largo del tiempo. Al observar la Tabla 9, es fácil deducir que las variables invariantes en el tiempo, como el sexo o la variable que identifica a los individuos, presenten variabilidad *within* igual a cero, ya que precisamente, no cambiarán de valor a lo largo del tiempo para cada individuo. Por ello, como se mencionó más arriba, variables de este tipo no podrían ser utilizadas con el estimador de efectos fijos. Toda la variabilidad en variables de este tipo se presenta *entre* sujetos, por ejemplo, dos personas pueden tener diferente sexo o religión. En el mismo sentido, variables que no presenten diferencias entre los distintos individuos, presentarán una variabilidad *between* igual a cero. No se observa ninguna variable que presente los mismos valores para todos los sujetos, sin embargo, ello se debe a que el panel ha dejado de estar balanceado después de filtrar a los jefes de hogar. Si el panel contase con las tres observaciones para todos los sujetos, la variable “Año” presentaría una variabilidad entre-individuos igual a cero, ya que todos los individuos

presentarían los mismos valores, es decir, 2009, 2010 y 2011, siendo el promedio individual de cada uno de ellos, igual al promedio general. En cambio, como ahora existen individuos con *algunas* observaciones, el promedio de cada uno de ellos será diferente y, por ende, también lo será el promedio general, tal como se observa en la Tabla 9. Sin embargo, puede observarse que aún la mayor parte de la variabilidad de esta variable se presenta al interior de cada sujeto.

El resto de las variables presentan también ambos tipos de variabilidad, y es interesante marcar algunas cuestiones. Por ejemplo, la variable que mide el nivel educativo alcanzado, suele ser una variable definida *teóricamente* como invariante en el tiempo. Sin embargo, se puede apreciar que este no es el caso y por ello hay que resaltar que siempre es importante realizar este tipo de análisis sobre las variables. Podría pensarse que al igual que con la variable “Año” la presencia de datos perdidos altera el valor de la variabilidad *within*, pero este no es el caso, ya que la misma lo que mide es la distancia entre las distintas observaciones de un individuo en particular respecto de su propia media, lo que no se ve alterado por las observaciones, perdidas o no, de otros individuos. Es más, este análisis descriptivo se realizó también con el panel de datos completo, dando por resultado un valor de variabilidad intra-individuo, prácticamente igual (.2983559 para el panel completo, 582 casos). La pequeña diferencia es producto, justamente, de la incorporación de jefes de hogar (o de observaciones filtradas de jefes de hogar económicamente inactivos) que *cambian* su nivel educativo en el período. Ahora bien, se puede observar que, tal como era de esperarse, la mayor parte de la variabilidad del nivel educativo se produce entre-sujetos. Sin embargo, como se mencionó al finalizar el apartado anterior, existen algunos casos en la base de datos que *cambiaron* el nivel educativo alcanzado en los años analizados y ésta es la información estadística que justifica que la consideramos una variable que cambia en el tiempo.

Por último se puede destacar que, tal como el sentido común lo predice, la variable “Edad” presenta ambos tipos de variabilidad, pero obviamente, la mayor parte se da entre-sujetos, siendo la variabilidad intra-sujetos mucho más pequeña, lo que concuerda con el hecho de que las tres observaciones son tomadas con una año de diferencia una de la otra.

Lo mismo sucede con la variable “Trabaja”, pero aquí hay que destacar algunas cuestiones. Al ser una variable dicotómica en un estudio de panel, de seguro presentará poca variabilidad, tanto intra como entre individuos, tal como lo confirma la información

presentada en la Tabla 8. En un estudio de panel, el problema con las observaciones es que suele existir correlación entre las respuestas de un individuo a lo largo del tiempo, por lo tanto, si alguien respondió “sí” en el tiempo  $t$ , es *más* probable que responda sí en  $t+1$ , o lo haya hecho en  $t-1$ . Por lo tanto, y sobre todo en variables dicotómicas, la variabilidad *within* suele ser poca y menor a la *between*. Justamente, la introducción de los efectos aleatorios que se realiza para corregir dicha correlación en las respuestas, suele ser preferido en el caso de que la variable respuesta presente poca variabilidad. En cuanto a la variabilidad entre-individuos para la variable “Trabaja”, se puede observar que esta es un poco mayor a la variabilidad *within*, sin embargo, ya cuando se analizó la información relativa a la respuesta al trabajo en la Tabla 3 se podía prever que la misma no sería tampoco tan elevada, debido a que aproximadamente el 91% respondía afirmativamente en cada una de las tres ocasiones. Este es un valor lo bastante alto, incluso teniendo en consideración el hecho de que no se podía confirmar que sean los mismos sujetos los que respondían.

Se puede aventurar por la información recogida en la Tabla 9 que, como casi todas las variables presentan una mayor variabilidad *entre-individuos*, el Coeficiente de Correlación Intraclase presentará un valor significativo que justifique la introducción de efectos aleatorios y, por ende, el modelado multinivel. Al mismo tiempo, la baja variabilidad intra-individuos nos informa que un análisis de efectos fijos puede no ser el adecuado.

#### **4.2.3 El modelo vacío**

En el capítulo anterior se expresó que el análisis multinivel es simplemente un forma de hacer regresión tradicional, bajo ciertas circunstancias. Es más, si se tienen dos regresiones, una bajo Mínimos Cuadrados Ordinarios y la otra Multinivel, que incluyan las mismas variables independientes y contextuales, es muy probable que ambas posean coeficientes similares. Sin embargo, se esperaría que las diferencias aparezcan en los errores estimados.

Esto es así porque como se explicó, al existir una estructura jerárquica en la información, es muy probable que el fenómeno de la anidación genere que las observaciones de un grupo tiendan a ser más similares entre sí (homogeneidad intra-grupo) y diferentes con las de otros grupos (heterogeneidad entre-grupos). Una manera más sencilla de decir eso es, simplemente, que las observaciones son dependientes al interior de cada grupo. Para saber

si la información con la que vamos a trabajar, presenta dependencia en sus observaciones, se calcula el Coeficiente de Correlación Intraclase (CCI). Como ya se explicó, este estadístico es una medida del grado de dependencia de las observaciones, es decir, una magnitud del grado de correlación de las observaciones en los grupos. Si su valor es cercano a cero, entonces, no hay evidencia de que la anidación genere correlación intra-grupo, en cambio, si su valor es positivo, la anidación tiene consecuencias, generando dependencia en las observaciones al interior de los grupos.

Cuando hay un gran número de casos u observaciones, anidadas en un número pequeño de grupos o clúster, se espera que el CCI sea positivo pero pequeño. Es decir, valores del CCI a partir del 0.10 ya justifican un análisis multinivel. A medida que aumenta el número de casos pero disminuye el número de observaciones al interior de cada uno, se requiere un valor del CCI mucho mayor para que la anidación genere consecuencias. Sin embargo, es bueno saber que incluso una correlación intraclase muy débil puede reducir sustancialmente los errores estándar de los coeficientes de regresión, aumentando la probabilidad de considerarlos estadísticamente significativos cuando en realidad no lo son (Bickel, 2007). Como aquí se está trabajando con datos de panel, se espera que el valor del CCI sea alto, debido a que se tienen pocas observaciones (tres como máximo) en un gran número de grupos, o mejor dicho, individuos ( $n=417$ ).

Se estimará ahora el modelo vacío con el objetivo de obtener el valor del Coeficiente de Correlación Intraclase Incondicional. El modelo vacío es un modelo en el cuál se regresa solamente la variable independiente, es decir, sin variables predictoras, y se tienen en cuenta la existencia de los individuos (es decir, los grupos) mediante la introducción de un efecto aleatorio en el intercepto. El CCI se denomina “Incondicional”, justamente, porque no está condicionado a ninguna variable explicativa. Su valor numérico indica entonces, la proporción de la varianza en la variable dependiente que se produce *entre* los individuos, en lugar de *dentro* de los individuos. Cuando se definió en el capítulo anterior la ecuación de nivel uno (6), es decir, al nivel de las observaciones, se explicó que los residuos de la misma  $e_{it}$  son definidos completamente por el valor predicho de  $y$ , por lo que no aparecía como un término separado en la ecuación. Sin embargo, se explicó que la regresión logística bajo su formulación de variable latente está relacionada mediante la función de densidad acumulativa con la formulación generalizada, y de allí se deriva que  $e_{it}$  tiene una función de densidad acumulativa logística estándar dada las variables explicativas, por lo

que se define su varianza como  $\sigma_e^2 = \pi^2/3$ . Dicho valor es una medida de la variabilidad intra-individuos en la variable dependiente, el cuál es importante para poder determinar el valor del CCI. La ecuación de nivel dos (7), definida también en el capítulo anterior, es la ecuación al nivel del individuo, siendo los residuos de este nivel,  $\mu_{0i}$ , una medida de la variabilidad entre-individuos.

El Coeficiente de Correlación Intraclase, también llamado “rho”, se puede definir entonces como:

$$CCI = rho = \text{Varianza Entre-Grupos} / \text{Varianza Total}$$

$$CCI = rho = \text{Varianza Entre-Grupos} / (\text{Varianza Entre-Grupos} + \text{Varianza Intra-Grupos})$$

$$CCI = rho = \text{Var}(\mu_{0i}) / \{\text{Var}(\mu_{0i}) + \text{Var}(e_{it})\} = \sigma_{\mu_0}^2 / (\sigma_{\mu_0}^2 + \sigma_e^2) \tag{9}$$

Dónde  $\sigma_e^2 = \pi^2/3$

En la Figura 4 se pueden observar los resultados obtenidos de la regresión del Modelo Vacio de intercepto aleatorio, también definido como regresión logística de efectos mixtos en dónde no existen variables predictoras.

Mixed-effects logistic regression		Number of obs	=	1093		
Group variable: NroCuest		Number of groups	=	417		
		Obs per group: min	=	1		
		avg	=	2.6		
		max	=	3		
Integration method: mvaghermite		Integration points	=	7		
Log likelihood = -282.49423		Wald chi2(0)	=	.		
		Prob > chi2	=	.		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	77.39671	45.59443	7.38	0.000	24.3938	245.5645
NroCuest var(_cons)	7.981349	3.78715			3.149061	20.22886
LR test vs. logistic regression: <u>chibar2(01) =</u>		56.83	Prob>=chibar2	=	0.0000	

Figura 4. Modelo Vacio de intercepto aleatorio. Regresión logística de efectos mixtos. Variable dependiente “Trabaja”. Odds Ratio, Cociente de chances; Std. Err., Error estándar; Conf. Interval, Intervalo de confianza; Likelihood-ratio test, Test de razón de probabilidad. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

Los coeficientes de la regresión están expresado en “Odds ratios” o cociente de chances o posibilidades. Al no existir variables predictoras, sólo se tiene el valor de la constante, la cual podría definirse como el cociente de chances medio, general para el total de las observaciones. Sin embargo, es importante diferenciar aquí una cuestión. Si bien el CCI es incondicional por no presentar ninguna variable explicativa (es decir, que intente dar cuenta de la variabilidad *entre* individuos), el modelo completo no lo es. Un modelo con la introducción de efectos aleatorios es siempre un modelo *condicional*. Esto es así, porque el modelo ajusta la probabilidad *sujeto-específica* o, dicho de otro modo, *condicional* al efecto aleatorio *individual*,  $\mu_{0i}$ . Es diferente a la regresión logística tradicional, que ajusta las probabilidades *marginales* o al *promedio de la población*, es decir, es un estimador agregado, ya que no tiene en cuenta la existencia de los grupos (individuos). Esto se ve claramente cuando se escribe el modelo vacío en su formulación de nivel único que se acaba de estimar:

$$\text{Logistic (Trabaja}_{it}) = \gamma_{00} + \mu_{0i}$$

Como se puede observar, no existe ninguna variable predictora que intente dar cuenta de la variabilidad existente en la variable independiente, pero si la variable sujeto-específica o efecto aleatorio al nivel del individuo, que lo que intenta es dar cuenta de *cuanta* variabilidad se produce entre-individuos y *cuanta* intra-individuos. Por ello, la constante estimada  $\gamma_{00}$  nos dice cuales son las chances promedio para un sujeto específico “medio”, es decir, con un valor medio de  $\mu_{0i}$  y no para el promedio de la población, tal como es el caso de la regresión tradicional.

Lo más importante aquí es la última parte de la Figura 4, la cual informa acerca de la parte aleatoria del modelo, es decir, del componente de varianza al nivel del panel. Lo que figura como “var(\_cons)” es la varianza de la constante, es decir, del efecto aleatorio al nivel del individuo introducido en el intercepto, para dar cuenta de la variabilidad producida a dicho nivel. Se puede observar entonces que  $\text{Var}(\mu_i) = \sigma_{\mu}^2 = 7.98$ , mediante la cual se puede calcular el CCI definido en (9):

$$\text{CCI Incondicional} = r(\rho) = \text{Var}(\mu_{0i}) / \{\text{Var}(\mu_{0i}) + \text{Var}(e_{it})\} = \sigma_{\mu_0}^2 / (\sigma_{\mu_0}^2 + \sigma_e^2)$$

$$\text{CCI Incondicional} = r(\rho) = \sigma_{\mu_0}^2 / (\sigma_{\mu_0}^2 + \pi^2/3) = 7.98 / (7.98+3.29) = 0.708$$

Se puede corroborar dicho valor con el proporcionado por el programa Stata, el cual puede apreciarse en la Figura 5.

Intraclass correlation				
Level	ICC	Std. Err.	[95% Conf. Interval]	
NroCuest	.7081177	.098073	.489066	.8601171

Figura 5. Coeficiente de Correlación Intraclase Incondicional para el modelo logístico multinivel de intercepto aleatorio sin variables explicativas (Figura 4). Variable dependiente “Trabaja”. Intraclass correlation, Correlación Intraclase; Level, Nivel; ICC, Coeficiente de Correlación Intraclase, Std. Err., Error estándar; Conf. Interval, Intervalo de confianza. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

Este valor significa que el 70.81% de la variabilidad en la variable dependiente (o sea, en la respuesta sobre el acceso al trabajo) se produce *entre* individuos, mientras que el 29.19% restante se produce al interior de cada individuo, es decir, en las distintas observaciones a lo largo del tiempo. Esto se traduce como un alto nivel de dependencia de las observaciones al interior de cada grupo (homogeneidad intra-grupo), lo que es esperable en un estudio de tipo panel en donde los grupos son cada uno de los individuos que integran el mismo. Al final de la Figura 4, se observa un test que compara la regresión obtenida de efectos mixtos con una regresión logística tradicional. Más precisamente, lo que el test realiza es probar si el valor de “rho” es o no diferente de cero. Si no existe evidencia estadísticamente significativa para rechazar la hipótesis nula, entonces el valor de rho no sería diferente de cero y, por lo tanto, la varianza al nivel del individuo (grupo) no es importante, de modo que las observaciones podrían considerarse como independientes y utilizar así un modelo logístico tradicional. El test precisamente compara el estimador agregado (logistic) con el estimador de panel o efectos aleatorios (logistic multinivel o de efectos aleatorios). En este caso, el valor resulta significativo, por lo que se rechaza la hipótesis nula y se obtiene evidencia estadísticamente significativa de que la anidación genera correlación entre las observaciones y, por lo tanto, será necesario el uso de un estimador más eficaz que el de Mínimos Cuadrados Ordinarios.

De este modo, se sabe ahora cuanta variabilidad en la variable dependiente se produce por diferencias entre los individuos y cuanto al interior de los grupos, y lo que es más importante aún, que un análisis multinivel es necesario. Se introducirán ahora las variables explicativas en el modelo, tal cual se definió en la ecuación (8).

### 4.2.4 Modelo logístico multinivel de intercepto aleatorio para la estimación de diferenciales en el acceso al trabajo de los jefes de hogar santafesinos

En la Figura 6, se presentan los resultados del modelo logístico multinivel de intercepto aleatorio que se definió en (8), tomando como base para la variable “Sexo” el género femenino, y para la variable “Educ\_categor”, el nivel educativo bajo. La variable “Edad” fue centrada respecto de la media general, debido a que es recomendable en los modelos multinivel centrar las variables continuas, sobre todo aquellas que se utilizan para realizar términos de interacción entre niveles (Bickel, 2007). En la primer parte de la figura se presenta la parte “fija” del modelo, mientras que en la segunda se muestran los componentes aleatorios. Una de las ventajas de los modelos multinivel, es que la interpretación de la parte “fija” del modelo se realiza del mismo modo que cualquier regresión logística.

Mixed-effects logistic regression		Number of obs	=	1093		
Group variable: NroCuest		Number of groups	=	417		
		Obs per group: min	=	1		
		avg	=	2.6		
		max	=	3		
Integration method: mvaghermite		Integration points	=	7		
		Wald chi2(6)	=	29.82		
Log likelihood = -263.55011		Prob > chi2	=	0.0000		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	5.418375	2.396431	3.82	0.000	2.277199	12.8925
Edad_centrada	1.019091	.0173633	1.11	0.267	.9856217	1.053697
Educ_Medio	4.33058	1.863497	3.41	0.001	1.863222	10.06532
Educ_Alto	7.535925	4.004358	3.80	0.000	2.659711	21.35201
Año2010	.9290511	.3061005	-0.22	0.823	.487067	1.772109
Año2011	1.360375	.4943651	0.85	0.397	.6673121	2.773245
_cons	5.723365	2.74272	3.64	0.000	2.237398	14.64063
NroCuest						
var(_cons)	4.701367	1.716802			2.298212	9.617411
LR test vs. logistic regression: <u>chibar2(01) =</u>		37.27		Prob>=chibar2 = 0.0000		

Figura 6. Modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas. Variable dependiente “Trabaja”. Odds Ratio, Cociente de chances; Std. Err., Error estándar; Conf. Interval, Intervalo de confianza; Likelihood-ratio test, Test de razón de probabilidad. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

Lo primero que se puede destacar es que se observa que existen variables estadísticamente significativas y variables que no lo son. La “Edad” pareciera ser una variable que, según

los datos analizados, no afectan la probabilidad de tener trabajo. Lo mismo se puede decir de las variables ficticias que intentan encontrar diferencias en las chances de obtener un empleo comparando los tres distintos momentos en el tiempo. Se puede decir, que no existen diferencias significativas en las chances de obtener en empleo en los tres años. En cambio, los hombres tienen 5.41 chances más de obtener empleo que una mujer, manteniendo todo lo demás constante, es decir, teniendo los mismos valores en el resto de las variables explicativas y el mismo valor del efecto aleatorio individual,  $\mu_{0i}$ . Igualmente significativos resultan los niveles educativos alcanzados. Una persona con nivel educativo medio tiene, aproximadamente, 4 chances más de trabajar que una persona en las mismas condiciones pero con nivel educativo bajo, lo que en términos de probabilidad equivale a un 81,23% de probabilidad de trabajar. El nivel educativo alto, ensancha las diferencias a 7 chances y media más de trabajar si se posee algún grado de estudio mayor al secundario completo en comparación con el nivel bajo, un 88,28% de probabilidad de trabajar.

Se observa además, el estadístico de Wald para probar la significación conjunta de las covariables, el cual es significativo por lo que se rechaza la hipótesis de que todos los coeficientes sean iguales a cero. También se rechaza la hipótesis de que la regresión logística tradicional sea conveniente. Se puede observar, comparando la Figura 6 con la Figura 4, la disminución en el valor de la varianza al segundo nivel, es decir, al nivel del individuo. La misma disminuyó de 7.98 en el modelo vacío a 4.70 en el modelo con todas las variables explicativas. Esto quiere decir, que la introducción de las variables contextuales ha podido explicar parte de la varianza entre-individuos, disminuyendo el valor del coeficiente de correlación intraclase, lo que puede observarse comparando la Figura 5 con la Figura 7. El CCI disminuyó del 0.7081 a 0.5883, una disminución del 16.92%.

Residual intraclass correlation				
Level	ICC	Std. Err.	[95% Conf. Interval]	
NroCuest	.5883154	.0884445	.4112704	.7451153

Figura 7. Coeficiente de Correlación Intraclase para el modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas (Figura 6). Variable dependiente “Trabaja”. Residual intraclass correlation, Correlación Intraclase Residual; Level, Nivel; ICC, Coeficiente de Correlación Intraclase, Std. Err., Error estándar; Conf. Interval, Intervalo de confianza. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

Sin embargo, a pesar de ser el modelo estadísticamente significativo, pareciera ser que el mismo no es el mejor ajuste de los datos que se puede encontrar, ya que tres variables resultaron no ser significativas estadísticamente, y si bien puede ser que no existan grandes diferencias en el ingreso al trabajo en un período de tres años consecutivos, es raro que la edad no presente ningún tipo de influencia ya que suele ser una variable tenida en cuenta por los empleadores al momento de contratar a una persona.

Diferentes modelos se llevaron a cabo para intentar encontrar aquél que mejor se ajuste a los datos<sup>12</sup>. Primero se intentó excluyendo las variables dicotómicas que correspondían a cada año. Luego, se agregó un efecto aleatorio en la pendiente de la variable “Edad\_centrada”, con el objetivo de permitir que la pendiente varíe para los distintos individuos, sin lograr tampoco resultados significativos, incluso con los términos de interacción implícitos que conlleva un modelo multinivel de intercepto y pendiente aleatorios. Se probó además con la introducción de otras variables, como por ejemplo:

- La existencia de hijos de 14 años o menos: Según la teoría, la presencia de hijos menores en los hogares reduce las posibilidades de tener trabajo, particularmente en las mujeres. No se hallaron resultados estadísticamente significativos.
- Estudiante: se introdujo con el objetivo de verificar si el hecho de ser estudiante, disminuía las chances de estar trabajando. En base al análisis realizado en la sección anterior, donde se verificó la existencia de jefes de hogar jóvenes, una de las hipótesis planteadas al respecto fue que posiblemente podía tratarse de jóvenes de distintas localidades que residen en la ciudad de Santa Fe con el objetivo de cursar estudios superiores. Esta hipótesis no puede verificarse estadísticamente, ya que los resultados obtenidos no son significativos.

Antes de continuar y describir los resultados encontrados del modelo que mejor ajuste presentó, quiero volver sobre la Figura 6 para explicar el coeficiente asociado a la Edad que se obtuvo en dicha regresión, ya que como se verá a continuación, finalmente la variable “Edad\_centrada” fue excluida por no presentar valores estadísticamente significativos en ninguno de los modelos. Sin embargo, considero que el valor del Odds Ratio encontrado para la variable Edad\_centrada, el cuál presentó valores similares en todos los modelos en los que se la incluyó, merece ser destacado. Cuando el cociente de

---

<sup>12</sup> Pueden verse los diferentes modelos estimados que no fueron considerados en la sección Anexo “Otros modelos estimados”.

chances toma el valor uno, está indicando que la probabilidad de éxito asociada es igual a la probabilidad de fracaso, es decir, que en función de la edad existen las mismas probabilidades de “Trabajar” que de “No trabajar”. En alguno de los modelos, el valor del Odds Ratio para la Edad se encontró unos puntos decimales por debajo de uno, lo que indicaría que las chances de trabajar disminuyen con la edad, tal como lo predice la teoría económica. Sin embargo, el hecho de que el valor del Odds Ratio gire en torno a la unidad en los diferentes modelos, estaría informando entonces que, para los datos de la muestra, no es un factor que modifique las chances de trabajar, lo que va de la mano con el hecho de no ser estadísticamente significativa. La explicación estadística posiblemente se encuentre en la distribución simétrica que presentó la variable al realizar el análisis exploratorio, lo que permite suponer que tal vez la distribución de las “no” respuestas en el acceso al trabajo, también presentó una distribución simétrica en relación a la edad. Dicho de otro modo, aquellos que respondieron “no” trabajar, se hallan distribuidos simétricamente a ambos lados de la media de la edad, cercana a los 49 años. De este modo, un cambio en una unidad de la variable “Edad\_centrada”, o sea, un año más respecto de la media de la edad, no cambiaría las posibilidades de trabajar, debido a que las mismas se encuentran distribuidas simétricamente en torno a dicha media.

La Figura 8 presenta el modelo final de regresión logística multinivel de intercepto aleatorio de acceso al trabajo. Todas las variables son estadísticamente significativas a nivel individual y también poseen significación conjunta, tal como lo indica el estadístico de Wald. Además, el modelo es preferido frente al modelo logístico tradicional, tal como se indica al final de la imagen. Esto significa que la anidación genera homogeneidad intra-clúster y por ello debe utilizarse un estimador de Máxima Verosimilitud que permita tener coeficientes aleatorios, en este caso, un intercepto.

En cuanto a la parte fija del modelo, se puede decir, que las chances medias de tener trabajo para un individuo de la muestra son 5.52, manteniendo constante, no sólo las variables explicativas, sino también el efecto aleatorio del individuo. Es decir, que la probabilidad media de tener trabajo manteniendo todo lo demás constante es del 84,66%, con un intervalo de confianza al 95% que va de 2.37 a 12.84. Esto es coincidente con la información presentada en la sección anterior, dónde la proporción de los que trabajan versus los que no trabajan era alta para cualquiera de los niveles educativos analizados (Tabla 3 y Figura 3).

Mixed-effects logistic regression		Number of obs	=	1093		
Group variable: NroCuest		Number of groups	=	417		
		Obs per group: min	=	1		
		avg	=	2.6		
		max	=	3		
Integration method: mvaghermite		Integration points	=	7		
Log likelihood = -264.83609		Wald chi2(3)	=	28.54		
		Prob > chi2	=	0.0000		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	5.260061	2.301697	3.79	0.000	2.231116	12.40108
Educ_Medio	4.191134	1.792608	3.35	0.001	1.812427	9.69176
Educ_Alto	7.365174	3.900279	3.77	0.000	2.608695	20.79422
_cons	5.519649	2.377686	3.97	0.000	2.372693	12.84049
NroCuest						
var(_cons)	4.708305	1.702767			2.317534	9.565399
LR test vs. logistic regression: <u>chibar2(01) =</u>		39.12	Prob>=chibar2 =	0.0000		

Figura 8. Modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas estadísticamente significativas. Variable dependiente “Trabaja”. Odds Ratio, Cociente de chances; Std. Err., Error estándar; Conf. Interval, Intervalo de confianza; Likelihood-ratio test, Test de razón de probabilidad. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

En cuanto a los Odds Ratio asociados a cada una de las variables explicativas, se puede decir que los mismos disminuyeron tan solo un poco respecto de los presentados en el modelo de la Figura 6, es decir que las chances prácticamente no se modificaron. Las chances de trabajar para una persona del sexo masculino son 5.26 veces más que las de una persona del sexo femenino que tenga las mismas características, es decir, el mismo nivel educativo y el mismo valor de intercepto aleatorio  $\mu_{0i}$ .

En cuanto al nivel educativo, tal como la teoría lo predice, las oportunidades de obtener trabajo se incrementan con el aumento en los años de escolaridad. Si bien no podemos distinguir si el aumento del nivel educativo se traduce en un aumento de la productividad del individuo o si sólo sirve como una “señal” para los empleadores, la verdad es que hay un alto grado de asociación positiva entre el nivel educativo alcanzado y las chances de trabajar. En comparación con un individuo que tiene nivel educativo bajo, dichas chances se multiplican por cuatro para aquellos que tienen un nivel educativo medio y por siete para aquellos que alcanzan un nivel educativo alto, si se mantienen todo lo demás constante. Si consideramos los intervalos de confianza, las chances para alguien que posee

nivel educativo medio, varían entre 1.81 y casi 10 veces más. Y si se considera a las personas con nivel educativo alto, se puede decir que sus chances son como mínimo el doble de una persona con nivel educativo bajo, llegando incluso a tener 20 veces más de posibilidades.

Residual intraclass correlation				
Level	ICC	Std. Err.	[95% Conf. Interval]	
NroCuest	.5886726	.0875693	.4132991	.744084

Figura 9. Coeficiente de Correlación Intraclase para el modelo logístico multinivel de intercepto aleatorio con todas las variables explicativas estadísticamente significativas (Figura 8). Variable dependiente “Trabaja”. Residual intraclass correlation, Correlación Intraclase Residual; Level, Nivel; ICC, Coeficiente de Correlación Intraclase, Std. Err., Error estándar; Conf. Interval, Intervalo de confianza. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

La última parte de la Figura 8 muestra el valor de la varianza para el intercepto aleatorio, que tiene como variable de clúster el número de identificación de la persona, permitiendo así distinguir entre los diferentes individuos. Comparada con el modelo vacío, la varianza registró una reducción importante, pasando de 7.98 a 4.71 con sólo una variable de contexto explicativa introducida, la variable “Sexo”. Es importante tener presente que en los estudios de panel, en los que el individuo es el nivel dos, todas las características de las personas invariantes en el tiempo, son consideradas variables contextuales. Las mismas son introducidas con el objetivo de dar cuenta de la varianza al segundo nivel, es decir, al nivel de grupo. Dicho de otro modo, el objetivo de introducir variables contextuales invariantes en el tiempo, es reducir la varianza en la variable independiente que se produce al nivel de grupo, en este caso, entre-individuos. En la Tabla 9 se pudo observar que, si bien la mayor parte de la varianza en la variable que mide el nivel educativo se produce entre-individuos, parte de la misma se produce también al nivel intra-individuo, lo que se traduce como que la misma no puede definirse como estrictamente invariante en el tiempo y por ello el “Sexo”, es la única variable contextual totalmente invariante en el tiempo introducida con el objetivo de reducir la varianza al nivel de grupo. La Figura 9, reproduce el valor del CCI asociado al modelo que se obtiene a partir del valor de varianza entre-individuos. Luego de la introducción de las variables, el mismo toma un valor de 0.5887, representando el 83.14% del valor total de la varianza entre-individuos obtenido mediante

el modelo vacío en la Figura 5 (CCI incondicional = 0.7081), lo que se traduce en una reducción de aproximadamente el 16.88%.

#### 4.2.5 Bondad de ajuste del modelo

Como se explicó anteriormente, la introducción de variables contextuales en los modelos multinivel, es decir, variables al nivel del grupo, tiene como objetivo reducir el valor de la varianza que se produce a dicho nivel, que en este caso específico refiere a la variabilidad entre-individuos. Por lo tanto, una buena medida de la bondad de ajuste de un modelo multinivel es la reducción en la varianza al nivel de grupo, en relación a la varianza total, o dicho más sencillamente, en la reducción del valor del CCI. Por lo tanto, el resultado recientemente encontrado es una medida de lo bien que ajusta el modelo especificado al conjunto de las observaciones, con una reducción del CCI de casi el 17%. Ahora bien, hay que destacar dos cuestiones interesantes que se desprenden de este resultado.

Por un lado, es que la gran disminución en el CCI se debe, no sólo a la introducción de la variable “Sexo”, sino también a la introducción del nivel educativo como variable explicativa, ya que si bien la misma mostró cierta variación intra-individuos, la mayor parte de la variabilidad de la misma se producía entre-individuos. Esto se confirma cuando la introducción de variables en el modelo vacío se realiza paso a paso, es decir, de a una variable por vez. Si al modelo vacío se le agrega la variable “Sexo” como variable explicativa, el CCI toma el valor de 0.6492, explicando un 8.33% del CCI. Si a dicho modelo se le agrega luego la variable que identifica los diferentes niveles educativos, el CCI alcanza el valor previamente mencionado de 0.5887. Se puede confirmar entonces que la introducción del nivel educativo actúa como una variable contextual, disminuyendo la varianza al nivel de grupo, es decir, explicando la variabilidad que se produce entre-individuos en un 8.55% en relación a la varianza total. Dicho de otro modo, casi el 9% de la variabilidad en la variable independiente que se produce *entre* individuos, es decir, en la variación de las diferentes respuestas de los distintos jefes de hogar, está explicado por el nivel educativo. Por su parte, el “Sexo” del jefe de hogar, explica un porcentaje similar. Otra forma de expresar dicho resultado es entendiendo que después de haber tenido en cuenta la variabilidad existente entre individuos producto de las diferencias en los niveles educativos y el sexo de los mismos, ahora la varianza restante atribuida a diferencias *entre* los individuos es del 58.87% en relación a la varianza total.

También es interesante entender que la disminución en los términos de la varianza entre-grupos se pueden utilizar en los modelos multinivel como un Pseudo  $R^2$ , el cual es entendido como la cantidad de la variabilidad en  $y$  que es explicada por las variables dependientes en el modelo. Este estadístico no es exactamente igual al  $R^2$  de las regresiones tradicionales, sin embargo, conceptualmente si lo es, ya que sirve como una medida para explicar la reducción proporcional en los errores de predicción cuando se compara el modelo estimado con el modelo nulo o incondicional (Bickel, 2007). En este caso en particular, refiere a la cantidad de variabilidad en  $y$  que se produce al nivel de grupo, es decir, entre-individuos y que es explicada por el modelo, sin tener en consideración la varianza al nivel uno, la que en los modelos logísticos, como ya se explicó, es constante. Precisamente, la diferencia con el CCI es que aquí se hace referencia sólo a la varianza entre-individuos y no en relación a la varianza total. El valor “total” de la varianza entre-individuos es el que se obtuvo a partir del modelo vacío, es decir, 7.98. Si al introducir las variables explicativas, la misma se redujo a un valor de 4.71, entonces:

$$\text{Pseudo } R^2 = (7.981349 - 4.708305) / 7.981349 = 0.410086$$

Lo que quiere decir que aproximadamente el 41.01% de la variabilidad en la respuesta sobre el acceso al trabajo al nivel de los jefes de hogar (*entre-sujetos*), es explicada por la introducción de las variables “Sexo” y “Nivel Educativo”.

La otra cuestión que puede desprenderse de dicho resultado, es la justificación del modelo seleccionado. Si se compara el valor del CCI obtenido con el modelo que incluye sólo variables explicativas estadísticamente significativas a nivel individual (Figura 9), con el valor obtenido por el modelo que incluía todas las variables que se consideraban, a priori, importantes para explicar el acceso de los jefes de hogar al mercado de trabajo (Figura 7), se puede observar que no existen entre ambos diferencias significativas. Más precisamente, los valores de los coeficientes de correlación intraclase obtenidos son prácticamente iguales, lo que significa que las variables que finalmente no fueron consideradas (Edad\_centrada, Año2010 y Año2011), no ayudaban a explicar la varianza en la respuesta al trabajo que se produce entre-individuos, lo que puede observarse también en términos análogos mediante los valores del pseudo  $R^2$ , el cual es de 0.410955 (41.09%) para el modelo de la Figura 6, un valor muy similar al obtenido por el modelo que contiene sólo variables significativas. Incluso, la introducción de dichas variables tampoco realiza un

gran aporte al ajuste total del modelo, ya que los valores del logaritmo de verosimilitud apenas difieren. Sin embargo, es interesante detenerse un minuto sobre esta cuestión.

Precisamente, una de las medidas de ajuste global para comparar la diferencia entre los valores predichos y los observados, es la Devianza. La misma se define multiplicando por menos dos (-2) el logaritmo de la verosimilitud, es decir,  $D = -2 * \text{loglikelihood}$ . El criterio establece que a mayor el valor de la devianza, peor el ajuste del modelo, es decir, es una medida de “falta de ajuste” del modelo. El programa Stata proporciona automáticamente el valor del logaritmo de la verosimilitud, es decir, *loglikelihood*, el cual es negativo, como puede observarse en las Figura 6 y 8. Esa es la razón por la que la Devianza se obtiene multiplicando por menos 2 el valor del logaritmo de la verosimilitud. Por lo tanto, podría decirse entonces que el modelo de la Figura 6, es decir, con todas las variables explicativas, tiene un mejor ajuste que el modelo de la Figura 8 con sólo variables significativas, ya que el valor de *loglikelihood* para el primero es menor (en valores absolutos) que el valor obtenido por el segundo, o lo que es lo mismo, el valor de Devianza es mayor para el modelo que tiene sólo variables significativas y por eso tiene un peor ajuste.

Llamando Modelo 1 al modelo con todas las variables explicativas y Modelo 2 al que tiene sólo las variables significativas, se pueden definir las devianzas de cada uno como:

$$D^1 = -2 * (-263.55011) = 527.100022 \quad < \quad D^2 = -2 * (-264.83609) = 529.67218$$

Tal como se advirtió, en base al criterio de la Devianza, el modelo 2 con sólo variables explicativas significativas tiene un peor ajuste, ya que posee un valor de devianza mayor.

Ahora bien, esto no es del todo cierto. El criterio de Devianza nos dice que un modelo ajusta mejor que otro pero sin tener en cuenta la cantidad de variables introducidas en el modelo. Generalmente, a medida que se introducen más variables se produce un “mejor” ajuste, pero se pierden grados de libertad. La mayoría de los criterios de selección o de bondad de ajuste, difieren porque cada uno utiliza diferentes fórmulas para “castigar” al analista por utilizar parámetros adicionales en la estimación de los coeficientes. El criterio de la Devianza, no incluye desincentivos por usar parámetros adicionales, por eso el Modelo 1 tiene un mejor ajuste que el Modelo 2. Sin embargo, los parámetros utilizados en el Modelo 1 son más de los utilizados en el Modelo 2, información que debería ser tenida

en cuenta para tomar una decisión acerca de qué modelo ajusta mejor. Sobre todo cuando la diferencia en los valores de las devianzas es tan pequeña.

Existe el criterio conocido como Diferencia de la Devianza el cual tiene en cuenta los parámetros utilizados en uno y otro modelo, castigando al modelo que tiene menos grados de libertad, es decir, que utiliza más parámetros. Los grados de libertad vienen dados por la cantidad de parámetros utilizados. En el Modelo 1 se tienen 9 grados de libertad (la constante, las seis variables explicativas y los dos términos aleatorios) mientras que en el Modelo 2 se tienen 6 (la constante, tres variables explicativas y los dos términos aleatorios). La diferencia de la devianza se calcula como:

$$\text{Diferencia de la Devianza} = 529.67218 - 527.100022 = 2.572$$

$$\text{Grados de Libertad} = 9 - 6 = 3$$

De este modo, se calcula la diferencia en el valor de la devianza como así también de los grados de libertad. Este estadístico tiene distribución aproximadamente Chi Cuadrado, con grados de libertad igual a la diferencia entre el número de parámetros utilizados en cada uno de los modelos. El valor crítico Chi Cuadrado con 3 grados de libertad y el nivel de significancia al 0.05 es 7.82, superior al valor encontrado. Esto indica, que el Modelo 2, con menos variables explicativas estadísticamente significativas, produce un mejor ajuste que el Modelo 1, justificando de este modo su elección (Bickel, 2007; Iglesias Cabo, 2013).

Dicha justificación del modelo seleccionado es en términos de modelos de efectos aleatorios con diferentes variables explicativas. Sin embargo, en el capítulo anterior se mencionó que otra forma de tener en consideración la existencia de características inobservables en cada individuo (o grupo) que causan la no independencia de las observaciones, es mediante el enfoque de “efectos fijos”. En este caso, se introducen tantas variables dummy como “grupos” existan en los datos, con la intención de hacer explícitas en el modelo las diferencias entre los mismos. De este modo, este tipo de análisis considera, al igual que el modelo de intercepto aleatorio, que existe un intercepto o término constante diferente para cada grupo, asociado con cada una de las variables dummy con coeficiente específico que intenta captar justamente las características inobservables propias de cada clúster, las cuales se suponen independientes entre sí bajo este enfoque. La diferencia estriba en que en el modelo de efectos aleatorios los factores inobservados son

considerados una variable aleatoria, es decir, que provienen de una población de grupos o contexto, por lo que están distribuidos aleatoriamente alrededor de un valor y por lo tanto, no son independientes entre sí y no toman la forma de un coeficiente fijo. La diferencia fundamental, como puede verse, está en los supuestos que se realizan sobre las características inobservables de los grupos o individuos.

Como se explicó anteriormente, el enfoque de efectos fijos tiene ciertas desventajas respecto del modelo de efectos aleatorios. Por un lado, se necesita que el panel esté balanceado, sino se pierden todos los grupos que no cuenten con observaciones completas. Cuando se mostro el patrón de observaciones perdidas una vez que se realizó el filtrado del panel respecto de los jefes de hogar económicamente activos, se dijo precisamente, que el número de casos disponibles para realizar un análisis de efectos fijos era, *a priori*, de 307 casos, que eran aquellos que contaban con observaciones completas. El término *a priori* se utilizó porque, justamente, otra de las desventajas de los modelos de efectos fijos, es que no admite la utilización de variables invariantes en el tiempo, es decir, que no presenten variabilidad “*within*” (el estimador de efectos fijos es comúnmente llamado estimador *within*). De este modo, no sólo no pueden utilizarse variables explicativas que no varían al interior de los grupos, como el “sexo” que resultó ser de gran importancia en el modelo presentado, sino que tampoco se pueden utilizar variables *dependientes* que presenten poca variabilidad. En modelos de panel con respuesta dicotómica como el presentado aquí, una variable de este tipo produce que el número de casos se reduzca aún más si la misma no cambia al interior de cada grupo, lo que suele suceder generalmente. De la Tabla 9 se sabe que la variabilidad al interior de los grupos de la variable dependiente “Trabaja” es poca, tan sólo, 0.1675, por lo que se deduce que efectivamente, el número de casos se reduciría a menos de 307.

Si bien existe una prueba, llamada Test de Hausman, para probar cual de los dos enfoques es el que mejor ajusta los datos, la misma no pudo ejecutarse ya que la disminución inicial en el número de casos producto del filtrado de la base y la poca variabilidad en la variable “Trabaja” al interior de los individuos, hizo imposible la estimación del modelo mediante el enfoque de efectos fijos, por lo que, se puede concluir que el modelo de efectos aleatorios es el indicado en este caso.

Habiendo justificado la elección del modelo de efectos aleatorios sobre el modelo de efectos fijos, como así también, la elección del “mejor” modelo de efectos mixtos, me

gustaría demostrar finalmente por qué, ante la existencia de datos con estructuras jerárquicas, modelos que tengan en cuenta dicha estructura son preferidos ante las regresiones tradicionales. Como ya se explicó, el programa Stata proporciona automáticamente un test de razón de probabilidad que compara el modelo con una regresión logística ordinaria y permite definir si la anidación es consecuente y genera correlación o si no lo hace. Si bien dicho valor resultó ser significativo y por ende, confirmó la utilización del modelo mixto por sobre la regresión tradicional, quisiera presentar a continuación los resultados del modelo (8) pero para una función logística tradicional, con el objetivo de visualizar cuales son los inconvenientes que se generan cuando la anidación es consecuente y no es tenida en cuenta en el modelo.

```

. logistic Trabaja Masculino Edad_centrada Educ_Medio Educ_Alto Año2010 Año2011
Logistic regression                               Number of obs   =       1093
                                                    LR chi2(6)      =       57.44
                                                    Prob > chi2     =       0.0000
Log likelihood = -282.1858                        Pseudo R2      =       0.0924
    
```

Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
Masculino	3.210117	.7506011	4.99	0.000	2.029971 5.076353
Edad_centrada	1.019503	.0104342	1.89	0.059	.9992561 1.04016
Educ_Medio	2.798014	.7211317	3.99	0.000	1.688378 4.636923
Educ_Alto	4.990825	1.680405	4.77	0.000	2.579727 9.655416
Año2010	.961285	.255007	-0.15	0.882	.5715415 1.616801
Año2011	1.230475	.3570468	0.71	0.475	.6967544 2.173031
_cons	2.774144	.7295684	3.88	0.000	1.65681 4.644996

Figura 10. Modelo de regresión logística con todas las variables explicativas. Comparable con el modelo mixto de la Figura 6. Variable dependiente “Trabaja”. Odds Ratio, Cociente de chances; Std. Err., Error estándar; Conf. Interval, Intervalo de confianza; Likelihood-ratio test, Test de razón de probabilidad. Elaboración propia. Fuente Datos: Observatorio Social - Secretaría de Planeamiento - Universidad Nacional del Litoral. Fuente Imagen: Stata/SE 13.0

En la Figura 10 se presenta el modelo de regresión logística con todas las variables explicativas. El mismo es comparable con el modelo multinivel o de efectos mixtos de regresión logística de la Figura 6, pero con la diferencia de que no existe un efecto aleatorio al nivel del individuo ya que el mismo no considera la existencia de una estructura jerárquica en los datos, sino que toma a los mismos como un “todo”. Se puede apreciar lo que se remarcó anteriormente: todos los errores estándar estimados de los coeficientes son subestimados. A modo de ejemplo, si tomamos el p-valor asociado del coeficiente de la variable “Edad\_centrada”, se puede ver que el mismo ahora es

significativo a un 10% y casi prácticamente significativo a un 5%. Es evidente con dicho ejemplo cual es el problema de no considerar la estructura jerárquica de los datos cuando la anidación es consecuente y genera correlación entre las observaciones: los errores estándar son subestimados, aumentando la probabilidad de obtener resultados significativos cuando en realidad no lo son (error de tipo I). Otros de los problemas que se pueden observar a simple vista son el sesgo en el valor de los coeficientes y un valor mucho mayor del logaritmo de la probabilidad, lo que se traduce en un peor ajuste del modelo.

Queda en evidencia entonces, que un modelo logístico tradicional que toma a los datos como un todo y ajusta los valores observados a una media general, sin tener en cuenta la existencia de observaciones dependientes producto de la homogeneidad intra-grupo, no es el adecuado. El modelo multinivel o de efectos mixtos con intercepto aleatorio, permite tener diferentes ordenadas al origen, medidas como desviaciones respecto de dicha media general por la introducción del efecto aleatorio, el cual representa las características inobservables de los individuos que son invariantes en el tiempo. La idea de la introducción del efecto aleatorio, es modelar dichas características que genera la dependencia entre las observaciones al interior de cada grupo, en este caso, individuo, ya que si no, la dependencia en las observaciones se traduce en residuos autocorrelacionados que generarán errores estándares de los coeficientes de regresión inexactos. La regresión de coeficientes aleatorios mediante el estimador de Máxima Verosimilitud, utiliza su capacidad para permitir que los coeficientes varíen entre los individuos para calcular un promedio ponderado, que toma la forma de la ecuación de regresión general. De este modo, el componente fijo es un promedio ponderado entre los individuos (es decir, el nivel de grupo), con “individuos” más grandes que contribuyen más al promedio total que los “individuos” más pequeños, y el componente aleatorio, por su parte, mide las desviaciones, respecto de dicho promedio, de las observaciones dependientes intra-grupo. Eso es lo que permite abordar eficazmente los problemas conjuntos de observaciones dependientes y residuos correlacionados (Bickel, 2007).

## CAPITULO V

### Conclusiones

Se comenzó esta tesis expresando la dificultad de abordar la relación entre educación y trabajo debido, por un lado, a los múltiples aspectos de dicha relación que se pueden tomar en consideración al momento de realizar una investigación, como así también, por las distintas posturas y enfoques desde la cual puede ser abordada. Tal es así, que se decidió poner en juego diversas teorías que intentan explicar dicha relación y no tomar una de ellas en particular, para intentar comprender la magnitud de la problemática del fenómeno bajo estudio y de la pluralidad de ideas que giran en torno al mismo.

Se asumió también a dicha relación como una articulación compleja y multidimensional definida históricamente, es decir, en un tiempo y espacio, lo que justificaba el estudio de algunos aspectos de la misma en el contexto santafesino. Más precisamente, el objetivo general de esta tesis pretendía describir las diferentes posibilidades o chances de trabajar de los jefes de hogar santafesinos según el nivel educativo alcanzado y otras características de los mismos, mediante la estimación de un modelo logístico multinivel para datos de panel en el período 2009-2011.

Se puede afirmar entonces, que los resultados encontrados en este trabajo no salen de la lógica común que relata que a mayores niveles educativos se corresponden mejores salarios o puestos de trabajo o, tal como se definió aquí, mayores posibilidades de trabajar. Sin embargo, no sólo se ha establecido una asociación positiva entre la proporción de los jefes de hogar que trabajan y el nivel educativo alcanzado de los mismos, sino que se ha podido cuantificar, mediante la estimación de un modelo logístico multinivel, las diferencias en las chances de trabajar que dependen del nivel de escolaridad.

Los resultados encontrados establecen que un jefe de hogar con nivel educativo medio, que en este caso comprende a aquellos que poseen secundario incompleto o completo, posee en promedio 4 chances más de trabajar en comparación con alguien que posee nivel educativo bajo (primario completo o incompleto), si los demás factores se mantienen constantes. Ahora bien, cuando el jefe de hogar alcanza un nivel educativo alto, aquí definido como terciario incompleto en adelante, las posibilidades medias de trabajar del mismo en comparación con un jefe de hogar con nivel educativo bajo, *ceteris paribus*, ascienden a más de 7.

Intervalos con 95% de confianza fueron construidos para cada uno de los coeficientes asociados a cada nivel educativo, demostrando un amplio margen de variabilidad para cada uno de ellos. En este sentido, las chances para alguien que posee nivel educativo medio en relación a un jefe de hogar con nivel educativo bajo, varían entre 1.81 y casi 10 posibilidades más de “trabajar”. Un jefe de hogar con nivel educativo alto posee como mínimo dos chances y medias más que un jefe de hogar de bajo nivel educativo. Lo llamativo es que dichas posibilidades pueden ser hasta 20 veces superiores. Los intervalos de confianza permiten afirmar, con un 95% de seguridad, que el verdadero valor del parámetro poblacional se encontrará entre dichos valores, o dicho de otro modo, que la chance media de trabajar de todos los jefes de hogar de la ciudad de Santa Fe con nivel educativo medio se encontrará entre {1.81; 9.69} y para aquellos con nivel educativo alto entre {2.60; 20.79}, siempre en comparación con un jefe de hogar que posee nivel educativo bajo y los demás factores iguales. Lo destacable de esto es que siempre las posibilidades de trabajar de aquellos que tienen un nivel educativo más alto son superiores a los que tienen nivel educativo bajo, debido a que ninguno de los intervalos incluye valores iguales o menores a uno, lo que equivaldría a que los primeros tengan en alguna oportunidad, las mismas o menos chances que los segundos, lo que se traduce en un altísimo grado de correlación positiva entre las variables analizadas.

Además, se encontró que los jefes de hogar de sexo masculino tienen ante sus pares femeninos más de 5 posibilidades de trabajar, cuando el resto de las características entre los mismos son iguales, lo que evidencia un mercado de trabajo santafesino fuertemente segregado en términos de género.

Resulta interesante destacar que no se encontraron diferencias significativas en las medias en el acceso al trabajo en los diferentes años bajo estudio, como así tampoco, un efecto significativo de la edad de los jefes de hogar en sus posibilidades de encontrarse trabajando.

Más allá de los resultados concretos, considero que uno de los aportes de ésta tesis es la aplicación de una metodología que, si bien no es del todo novedosa debido a que ya cuenta con más de tres décadas de aplicación, no se encuentra ampliamente difundida. Se han demostrado las amplias ventajas de la utilización de los modelos multinivel cuando se trabaja con datos con algún tipo de estructura jerárquica, es decir, con diferentes niveles de información. Los modelos multinivel permiten resolver las limitaciones existentes de los

modelos de regresión tradicional cuando se opera con datos con algún tipo de agrupamiento o anidación que invalida los supuestos fundamentales de la regresión por Mínimos Cuadrados Ordinarios. Se mostró que cuando la anidación es consecuente, las observaciones y/o sus residuos dejan de ser independientes e idénticamente distribuidos, debido a que existen características inobservables de los contextos que influyen en las observaciones de nivel inferior, generando homogeneidad intra-grupo y heterogeneidad entre-grupos. Los modelos multinivel son los apropiados para el análisis de este tipo de datos ya que tienen en cuenta la no independencia de las observaciones y sus residuos, permitiendo de este modo obtener estimadores estadísticamente eficientes de los coeficientes de regresión, como así también, errores estándar, intervalos de confianza y test de significancia que brindan resultados más conservadores que los obtenidos por los métodos tradicionales de regresión que ignoran la existencia de dichos grupos o contextos diferentes. Esto se pudo demostrar cuando se comparó el modelo logístico multinivel que incluía variables no significativas, como la edad, con el modelo tradicional, en el cual dicha variable pasaba a ser estadísticamente significativa al poseer un error estándar subestimado. Las ventajas de los modelos multinivel no son sólo en términos puramente estadísticos sino también conceptuales, ya que se pueden obtener modelos correctamente especificados al tener en consideración los diferentes niveles en que los datos se encuentran agrupados, haciendo explícitas las relaciones entre las variables que se desprenden de dicha estructura y evitando de este modo sacar conclusiones equivocadas, como las que pueden obtenerse al caer en las llamadas falacias ecológica o atomista.

El modelo multinivel aplicado permitió tener en consideración el hecho de que se estaba trabajando con datos de panel, es decir, una estructura de datos en los que se cuenta con diferentes observaciones del mismo individuo a lo largo del tiempo. Precisamente, en este tipo de datos, existe una fuerte correlación entre las distintas observaciones de un mismo individuo, producto de las características inobservables de los mismos que no cambian en el tiempo y no son tenidas en cuenta explícitamente en el modelo. El valor del coeficiente de correlación intraclase incondicional, demostró un alto grado de dependencia de las observaciones, al reflejar que un 70.81% de la variabilidad no explicada en la variable dependiente, es decir, en la respuesta en el acceso al trabajo, era originada por la variabilidad *entre* individuos. La introducción de variables contextuales, como el sexo o el nivel educativo, sirvieron para explicar casi un 17% de la varianza atribuida a diferencias

entre los individuos, destacando así la importancia de dichas variables para explicar el acceso al trabajo de los jefes de hogar en el mercado laboral santafesino. Y si bien restó una gran porción de varianza no explicada, la introducción del efecto aleatorio en el intercepto permitió, precisamente, tener en cuenta la existencia de factores no observados de los jefes de hogar que no cambian en el tiempo y pueden afectar, en este caso, su acceso al trabajo, evitando así que las mismas causen correlación entre las distintas observaciones de un mismo individuo. Por ello es importante recordar que en una regresión multinivel de intercepto aleatorio, cuando se analiza la influencia de una variable mediante el valor de su coeficiente y se mantiene todo lo demás constante, dicha expresión refiere también a mantener constante un mismo valor de efecto aleatorio sujeto-específico.

De aquí se desprenden cuestiones de análisis interesantes y, por qué no, líneas de acción a seguir para profundizar la investigación y mejorar la estimación del modelo mediante la reducción de la varianza no explicada que se produce al nivel del individuo. Si el efecto aleatorio sujeto-específico contiene la influencia que características inobservables de los jefes de hogar pueden tener sobre su condición de ocupado, se puede suponer que, por ejemplo, características como la capacidad innata de los jefes de hogar, la cual se define invariante en el tiempo, puede estar contenida en el efecto aleatorio individual y ser una fuente de explicación de la variabilidad en el acceso al trabajo entre los jefes de hogar. Más allá de las complicaciones de medición que una variable como la “capacidad innata” pueden generar, serían muy interesantes las conclusiones que podrían obtenerse de un análisis que permita reflejar si la educación produce realmente un aumento en la productividad de las personas, tal como postula la Teoría del Capital Humano, o si sólo sirve para “descubrir”, o mejor dicho, reflejar en un título las capacidades innatas de las mismas, tal como lo anuncia la teoría del filtro.

También resulta muy interesante introducir más niveles de análisis, como por ejemplo, un tercer nivel que tenga en consideración la vivienda de los jefes de hogar, lo que permitiría no sólo tener en cuenta características de los hogares a los que pertenecen los jefes y jefas, sino también, la introducción al nivel del individuo (nivel dos) de todos los miembros del hogar. Un modelo con estas características resulta muy atractivo para profundizar la investigación pertinente, dando lugar a nuevas preguntas y objetivos. Se podría, por dar un ejemplo, estudiar la influencia que diferentes características de los hogares pueden tener sobre las posibilidades de trabajar de los jefes de hogar, como ser por ejemplo, el barrio al

que pertenece, la cantidad de necesidades básicas insatisfechas, la mayor o menor disponibilidad de transporte público, etc. Sabiendo que las sociedades no construyen la desigualdad sólo desde las escuelas esto permitiría, tal vez, tener en consideración la influencia en el acceso al trabajo de otro tipo de desigualdades y no sólo las atribuidas al nivel educativo y poder poner en discusión por ejemplo, la teoría de la correspondencia o porque no, la influencia de la “herencia social” sobre la movilidad socio-profesional entre generaciones, al incluir además a los otros miembros del hogar. Esto último también podría dar lugar a analizar otros tipos de efectos que se dan en el mercado de trabajo, como el denominado “trabajador adicional” que puede evidenciarse en momentos en que el volumen y la tasa de desocupación son elevadas, la vulnerabilidad entendida como la probabilidad de quedar desocupados durante un cierto período tiempo, lo cual puede ser interesante para analizar la situación de los jóvenes en el hogar al concluir sus estudios, como así también, las diferentes trayectorias entre el empleo y el estudio que pueden darse en los jóvenes y su relación con la condición laboral de los padres, entre otros.

Al trabajar con datos secundarios, existen siempre limitaciones en las variables que uno puede incluir en el modelo debido a que la información disponible fue previamente relevada y no precisamente teniendo en cuenta los objetivos de investigación que uno se propone. Hubiese sido interesante por ejemplo, cuantificar las posibilidades de trabajar en relación a la experiencia de los jefes de hogar, la cual suele ser una variable positivamente relacionada con el acceso al mercado de trabajo. En el mismo sentido, también se podría haber extendido el estudio a más períodos en el tiempo. Como se mencionó, el recorte temporal se debió a fundamentalmente a que se disponía de dicha información en el formato necesario y que no se contaba con el seguimiento de dichas familias en las ondas posteriores que realizó el Observatorio Social, en 2014 y 2015, principalmente porque la Onda 2014 fue reconstitutiva del panel general. Más allá de la disminución en el número de casos, sería interesante poder encontrar las familias que se mantuvieron en la Onda en dichos años y analizar las mismas en un horizonte temporal más amplio. Considero particularmente que cobraría mucha importancia analizar la variabilidad “whitin” de la variable nivel educativo en los jefes de hogar, o de todos los integrantes si se dispone de dicha información, ya que como se pudo ver en este trabajo, la variable nivel educativo mostró no sólo variabilidad entre-individuos sino también al “interior” de los mismos, es decir, en las diferentes observaciones en el tiempo, dando lugar a casos en que los jefes de

hogar pasaron de un nivel educativo alcanzado a otro. Lo considero particularmente importante para el caso santafesino porque se podrían, por ejemplo, analizar los resultados del “Plan Vuelvo a Estudiar”, un acción de inclusión socioeducativa implementada desde 2013 por el Gobierno Provincial, la cual despliega diferentes estrategias con el objetivo de que jóvenes y adultos que no han podido terminar la educación secundaria, retomen, permanezcan y egresen de la misma (Santa Fe [provincia], Ministerio de Educación, 2016).

Más allá de los resultados concretos, las limitaciones del análisis y las posibles líneas de acción a seguir o profundizar, íntimamente relacionadas con las grandes posibilidades que brindan los modelos multinivel para el estudio de información jerárquica, considero importante destacar que, desde mi punto de vista, el estudio de los fenómenos sociales no debe circunscribirse únicamente a la teorización y posterior modelización de los mismos, enfocándose en una simple descripción y explicación de los hechos que son justificados por los resultados del modelo econométrico aplicado. A veces pareciera, que la disciplina económica se “olvida” de que es una ciencia social multidisciplinar que estudia fenómenos sociales, los cuales ya de por sí, son complejos.

Por estos motivos, más allá de haber aplicado un modelo econométrico concreto con el objetivo de definir y cuantificar una relación de causa-efecto entre variables, espero haber dejado en claro que la relación entre educación y trabajo fue entendida aquí como todo proceso y relación que se define y forma socialmente, es decir, un proceso complejo de múltiples dimensiones que asume características o rasgos propios de la sociedad en la que se desarrolla. Y no sólo ello, los “sentidos” que la sociedad le atribuye a la educación, como así también al trabajo, también mutan y se transforman y ni siquiera son los mismos para todos los individuos.

En el marco de un mundo cada vez más globalizado y con un complejo entramado político, económico y social, caracterizado por el cambio constante, uno de los grandes debates actuales es el rol de la educación. Más allá de los cambios tecnológicos, de la nueva era de la información, de los nuevos modos de desarrollo, de empleo, de producción flexible, del acceso a un número cada vez mayor de bienes y servicios y de todo lo “innovador”, también coexisten elevados índices de desempleo, sobre todo juveniles, mayores brechas de desarrollo, como así también, mayores diferencias en las rentas per cápita y la calidad de vida, la cada vez mayor acumulación de capital y conocimientos en menos manos, el aumento de los conflictos bélicos en oriente medio, como así también sus consecuencias, el

incontable número de refugiados, el debilitamiento de los derechos civiles y humanos y ese círculo vicioso que juegan de la mano, el fortalecimiento del terrorismo y el resurgir de nacionalismos de ultraderecha de discursos racistas y xenófobos. Ante tal escenario mundial, ¿cómo no jugar un rol importante la educación?

En un panorama de cambios profundos y pocas certezas, el rol de la misma es puesto en duda muchas veces, tanto en retóricas que indican que la escuela no prepara para el mundo productivo, como así también, en datos que indican que cada vez son más los egresados de estudios superiores que no encuentran un trabajo “formal y estable” o “acorde” a la “inversión” realizada, o peor aún, que directamente no encuentran un empleo. Sin embargo, la demanda educativa ha experimentado un aumento sin precedentes en América Latina en las últimas décadas (Sistema de Información de Tendencias Educativas en América Latina [Siteal], 2015), evidenciando la importancia que la sociedad le otorga a la misma. Es que si bien nada asegura la disminución de las desigualdades ni la movilidad social ascendente, pensar en tales logros sin educación pareciera imposible.

No me quedan dudas de que la educación tal como fue impartida en el siglo XX, debe cambiar. Se requieren nuevas habilidades, no sólo técnicas, sino también mentales, emocionales, actitudinales y, por ende, se requieren nuevas formas de enseñar. Considero substancial entonces saber qué pretendemos como sociedad de la educación, ya que si bien creo que es importante que la escuela prepare en las habilidades técnicas y socioemocionales que demandará el mercado de trabajo del futuro cercano, creo que la educación no debe nunca perder de vista su costado utópico, como elemento emancipador del hombre y garante del libre desarrollo del mismo. Al mismo tiempo, considero imprescindible una educación innovadora y de calidad, pero por sobre todo, más igualitaria, integradora y formadora de lazos sociales.

El contexto actual en la provincia de Santa Fe en el que se está debatiendo una nueva ley de educación, no puede presentar un mejor escenario para pensar qué educación queremos los santafesinos para nuestra sociedad.

---

### Bibliografía

- Agresti, A. (2002). *Categorical Data Analysis* (2a ed.). Florida: Wiley-Interscience.
- Amador, M., y López-Gonzalez, E. (2007). Una aproximación bibliométrica a los modelos multinivel. *RELIEVE*, 13(1), 67-82. Recuperado de [http://www.uv.es/RELIEVE/v13n1/RELIEVEv13n1\\_3.htm](http://www.uv.es/RELIEVE/v13n1/RELIEVEv13n1_3.htm)
- Amara, M., y Jemmali, H. (noviembre, 2015). Household and contextual indicators of poverty in Tunisia: A multilevel analysis [Indicadores del hogar y contextuales de la pobreza en Túnez: un análisis multinivel]. *The Economic Research Forum, Working Paper Series*, 968. Recuperado de <http://erf.org.eg/publications/household-and-contextual-indicators-of-poverty-in-tunisia-a-multilevel-analysis/>
- Argentina. Ministerio de Hacienda y Finanzas Públicas. Instituto Nacional de Estadísticas y Censos – INDEC. Encuesta Permanente de Hogares (EPH). Buenos Aires. Recuperado de <http://www.indec.gov.ar/>
- Aronson, P. P. (2007). El retorno de la teoría del capital humano. *Fundamentos en Humanidades*, 8(16) 9-26. Recuperado de <http://www.redalyc.org/articulo.oa?id=18481601>
- Bachrach, E. (2012). *ÁgilMente* (3ª ed.). Buenos Aires: Sudamericana.
- Beck, U. (1998). *La sociedad del riesgo. Hacia una nueva modernidad*. Barcelona: Paidós. (Trabajo original publicado en 1986)
- Becker, G. (1983). Inversión en Capital Humano e Ingresos. En L. Toharia (ed), *El mercado de trabajo: Teorías y aplicaciones. Lecturas Seleccionadas* (pp. 39-63). Madrid: Alianza (Trabajo original publicado 1975)
- Beker, D., Bertinetti Giumelli, R., Eberhardt, M., Gutiérrez, F. y Perotti, R. (2009). *Educación, trabajo y empleo*. Cátedra de Economía Laboral, Facultad de Ciencias Económicas, Universidad Nacional del Litoral.

- Bickel, R. (2007). *Multilevel analysis for applied research: it's just regression!* [Análisis multinivel para la investigación aplicada: ¡sólo es regresión!]. New York: The Guilford.
- Boudon, R. (1978). Educación e Igualdad. En *Política, Igualdad Social y Educación. Textos seleccionados de Sociología de la Educación*. (vol. I, pp. 55-75) Madrid: Servicio de Publicaciones del Ministerio de Educación y Ciencia. Recuperado de [https://sede.educacion.gob.es/publiventa/download.action?f\\_codigo\\_agc=756\\_19](https://sede.educacion.gob.es/publiventa/download.action?f_codigo_agc=756_19)
- Bowles, S., y Gintis, H. (1983). El problema de la Teoría del Capital Humano; una crítica marxista. En L. Toharia (ed), *El mercado de trabajo: Teorías y aplicaciones. Lecturas Seleccionadas* (pp. 115-127). Madrid: Alianza (Trabajo original publicado 1975).
- Braginski, R. (23 de septiembre de 2016). Universidad: terminan pocos y en carreras que no son estratégicas. *Clarín*. Recuperado de [http://www.clarin.com/sociedad/Universidad-terminan-pocos-carreras-estrategicas\\_0\\_1655834606.html](http://www.clarin.com/sociedad/Universidad-terminan-pocos-carreras-estrategicas_0_1655834606.html)
- Calvés, J. A. (verano, 2000). La educación y la crisis de la modernidad. *Revista Hespérides*, (8)20. Recuperado de <http://www.geocities.ws/paginatransversal/alsinaeducacion/index.html>
- Carnoy, M. (2002). *Sustaining the New Economy. Work, Family and Community in the Information Age* [El sustento de la nueva economía. Trabajo, familia y comunidad en la era de la información] (2a ed.). Cambridge, Estados Unidos: Harvard University.
- Castel, R. (1997). *La metamorfosis de la cuestión social. Una crónica del salariado*. Buenos Aires: Paidós. (Trabajo original publicado en 1995)
- Castel, R. (2004). *La Inseguridad Social. ¿Qué es estar protegido?*. Buenos Aires: Manantial. (Trabajo original publicado en 2003)
- Castells, M. (2000). *La era de la información: Economía, Sociedad y Cultura. Volumen 1: La Sociedad Red* (2a ed.). Madrid: Alianza.

- Chevalier, A., Harmon, C., Walker, I., y Zhu, Y. (noviembre, 2004). Does education raise productivity, or just reflect it? [¿La educación aumenta la productividad, o sólo la refleja?]. *The Economic Journal*, 114(499), F499-F517. Recuperado de <http://www.jstor.org/stable/3590169>
- Comisión Económica para América Latina y el Caribe; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. 1992. *Educación y Conocimiento: Eje de la transformación productiva con equidad*. Santiago de Chile: CEPAL – UNESCO. Recuperado de [http://repositorio.cepal.org/bitstream/handle/11362/2130/S9250755\\_es.pdf](http://repositorio.cepal.org/bitstream/handle/11362/2130/S9250755_es.pdf)
- Crespi, G., Fernández-Arias, E., y Stein, E. H. (eds.). (2014). *¿Cómo repensar el desarrollo productivo?: políticas e instituciones sólidas para la transformación económica*. (s.l.): Banco Interamericano de Desarrollo. Recuperado de <https://publications.iadb.org/handle/11319/6634>
- Dávila, O., Ghiardo, F., y Medrano, C. (2008). *Los Desheredados. Trayectorias de vida y nuevas condiciones juveniles*. Valparaíso: CIDPA. Recuperado de <http://www.cidpa.cl/wp-content/uploads/2013/05/Desheredados.pdf>
- Dueñas Fernández, D., Iglesias Fernández, C., y Llorente Heras, R. (2014). La incidencia de los aspectos pre-market: segregación laboral y gap salarial por género. *Documentos de Trabajo, 4*. IAES - Instituto Universitario de Análisis Económico y Social, Universidad de Alcalá. Recuperado de [http://www3.uah.es/iaes/publicaciones/DT\\_04\\_14.pdf](http://www3.uah.es/iaes/publicaciones/DT_04_14.pdf)
- Estados Unidos. Department of Labor. Bureau of Labor Statistics [Departamento de Trabajo. Oficina de Estadísticas Laborales]. (2016). Recuperado de [http://www.bls.gov/emp/ep\\_chart\\_001.htm](http://www.bls.gov/emp/ep_chart_001.htm)
- Fernández Enguita, M. (2010). El desafío de la educación en la sociedad del conocimiento. En Instituto Internacional de Planeamiento de la Educación IPE-Unesco (ed.), *Educación y Trabajo: Articulaciones y Políticas* (pp. 25-34). Buenos Aires: IPE-Unesco. Recuperado de <http://www.buenosaires.iipe.unesco.org/sites/default/files/Educacion%20y%20Trabajo%20web.pdf>

- Freire Soane, M. J., y Teijeiro Álvarez, M. (mayo/septiembre, 2010). La inversión en capital humano de los jóvenes gallegos: ¿sigue siendo rentable la educación?. *Cuadernos de Economía*, 33(92), 45-70. Recuperado de <http://hdl.handle.net/10486/10038>
- Gallart, M. A. (1997). La interacción entre la sociología de la educación y la sociología del trabajo. *Revista Latinoamericana de Estudios del Trabajo*, 3(5), 94-115. Recuperado de [http://relet.iesp.uerj.br/Relet\\_05/](http://relet.iesp.uerj.br/Relet_05/)
- Goldstein, H. (1999). *Multilevel Statistical Models* [Modelos estadísticos multinivel]. London: Wiley.
- Hidalgo, J.C. (2001) *Economía Política y Educación Superior, crítica al pensamiento neoliberal*. Centro de publicaciones Secretaria de Extensión UNL. 2001.
- Hox, J.J. (1995). *Applied Multilevel Analysis* [análisis de multinivel aplicado]. Amsterdam: TT-Publikaties. Recuperado de <http://joophox.net/publist/amaboek.pdf>
- Hox, J.J. (2000). Multilevel Analysis of Grouped and Longitudinal Data [Análisis multinivel de datos agrupados y longitudinales]. En T. D. Little, K. U. Schnabel, y J. Baumert (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples* [Modelado de datos longitudinales y grupos múltiples: Cuestiones prácticas, enfoques aplicados y ejemplos específicos.](pp. 1-15). Hillsdale, NJ: Erlbaum (en preparación). Recuperado de <http://dspace.library.uu.nl/handle/1874/23615>
- Hox, J.J., y Mass, C.J.M. (2005). *Multilevel Analysis* [Análisis Multinivel] En K. Kempf-Leonard (ed.), *Encyclopedia of Social Measurement* (vol.2, pp. 785-793). Netherland: Elsevier Inc. Recuperado de [http://joophox.net/publist/ESM\\_MLA05.pdf](http://joophox.net/publist/ESM_MLA05.pdf)
- Iglesias Cabo, T. (2013). *Métodos de Bondad de Ajuste en Regresión Logística*. (Tesis de Maestría). Universidad de Granada, Granada. Recuperado de [http://masteres.ugr.es/moea/pages/tfm-1213/tfm\\_iglesiascabo\\_tania/](http://masteres.ugr.es/moea/pages/tfm-1213/tfm_iglesiascabo_tania/)
- Max-Neef, M., Elizalde, A., y Hopenhayn, M. (2010). *Desarrollo a Escala Humana: una opción para el futuro*. Madrid: Ciudades para un Futuro más Sostenible (CF+S)

- (Trabajo original publicado 1986). Recuperado de <http://habitat.aq.upm.es/deh/adeh.pdf>
- Mayorga, M. y Muñoz, E. (2000). *La técnica de datos de panel un guía para su uso e interpretación*. Recuperado de [http://www.bccr.fi.cr/investigacioneseconomicas/metodoscuantitativos/Tecnica\\_datos\\_panel\\_una\\_guia\\_para\\_su\\_uso\\_e\\_interpretacion.pdf](http://www.bccr.fi.cr/investigacioneseconomicas/metodoscuantitativos/Tecnica_datos_panel_una_guia_para_su_uso_e_interpretacion.pdf)
- Medina, E. (1983). Educación, Universidad y Mercado de Trabajo. *Revista española de investigaciones sociológicas*, 24, 7-46. Recuperado de [http://www.reis.cis.es/REIS/PDF/REIS\\_024\\_03.pdf](http://www.reis.cis.es/REIS/PDF/REIS_024_03.pdf)
- Meyer, R., Díaz, P., y Bulgarella, E. (2014). *Trabajo, edad, género y nivel educativo. Estimación de probables diferenciales entre jefes y jefas de hogar de la Ciudad de Santa Fe*. Recuperado de <http://www.unl.edu.ar/observatoriosocial/indicadores/34.pdf>
- Minujin, A. (1999). ¿La Gran Exclusion? Vulnerabilidad y Exclusion en America Latina. En D. Filmus, (comp.) *Los noventa: política, sociedad y cultura en América Latina y Argentina de fin de siglo* (pp. 53-77) Buenos Aires: Eudeba.
- Murillo Torrecilla, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis, Revista Internacional de Investigación en Educación*, 1, 45-62. Recuperado de: [https://www.uam.es/personal\\_pdi/stmaria/jmurillo/curriculum/publicaciones.htm](https://www.uam.es/personal_pdi/stmaria/jmurillo/curriculum/publicaciones.htm)
- Nezlek, J. (24 de febrero de 2015). An Introduction to Multilevel Modeling - basic terms and research examples. [Una Introducción al Modelado Multinivel - términos básicos y ejemplos de investigación.] [Archivo de video]. Recuperado de: <https://www.youtube.com/watch?v=f817HdHJneo>
- Observatorio Social - Sistema de Monitoreo Social del Litoral (septiembre, 2015). *Informe Tema TRABAJO. Panel de Hogares*. Secretaría de Planeamiento, Universidad Nacional del Litoral. Santa Fe. Recuperado de <http://www.unl.edu.ar/observatoriosocial/>
- Observatorio Social - Sistema de Monitoreo Social del Litoral (octubre, 2016). *Documento Metodológico del Panel de Hogares*. Secretaría de Planeamiento, Universidad

- Nacional del Litoral. Santa Fe. Recuperado de <http://www.unl.edu.ar/observatoriosocial/>
- Offe, C. (1990). *Contradicciones en el Estado de Bienestar*. Madrid, Alianza.
- Organisation for Economic Co-operation and Development – OECD. (2015). *Education at a Glance 2015: OECD Indicators* [Panorama de la Educación 2015: indicadores OCDE]. OECD Publishing. Recuperado de <http://dx.doi.org/10.1787/eag-2015-en>
- Organización de las Naciones Unidas – ONU. (1989). Convención sobre los Derechos del Niño. Recuperado de <http://www.ohchr.org/SP/ProfessionalInterest/Pages/CRC.aspx>
- Organización Internacional del Trabajo – OIT. (2013). *Tendencias Mundiales del Empleo Juvenil: Una generación en peligro*. Ginebra: OIT. Recuperado de [http://www.ilo.org/global/research/global-reports/youth/2013/WCMS\\_222658/lang--es/index.htm](http://www.ilo.org/global/research/global-reports/youth/2013/WCMS_222658/lang--es/index.htm)
- Otero, A. (2009). *Procesos de transición a la vida adulta: un estudio cualitativo con jóvenes argentinos*. (Tesis de Doctorado). Facultad Latinoamericana de Ciencias Sociales – FLACSO, Buenos Aires. Recuperado de <http://repositorio.flacsoandes.edu.ec/handle/10469/2656#.WDQ79LJ9670>
- Poston, D. L. (junio, 2002). Human Capital, Cultural Capital and the Economic Attainment Patterns of Asian-born Immigrants to the United States: Multi-level Analyses [El capital humano, el capital cultural y los patrones de logros económicos de los inmigrantes asiáticos en los Estados Unidos: análisis multinivel]. *Asian and Pacific Migration Journal*, 11(2), 197-219. Recuperado de [http://www.smc.org.ph/administrator/uploads/apmj\\_pdf/APMJ2002N2ART2.pdf](http://www.smc.org.ph/administrator/uploads/apmj_pdf/APMJ2002N2ART2.pdf)
- Qin, B. (diciembre, 2005). Wage earnings of chinese in the United States: Individual and contextual determinants [Salarios obtenidos por los chinos en los Estados Unidos: determinantes individuales y contextuales]. (Tesis de Doctorado). Texas: A&M University. Recuperado de <http://hdl.handle.net/1969.1/3213>
- Ramírez Parra, P. (2002) *¿Todo vale? Latinoamérica ante la encrucijada de la postmodernidad*. Colombia: UNAB. Recuperado de:

[http://books.google.com.ar/books?id=VsYUPtW9IDc&dq=Latinoam%C3%A9rica+ante+la+encrucijada+de+la+postmodernidad&printsec=frontcover&source=bl&ots=tBfG\\_pzLHM&sig=6jAL89YuA4L6ZvMilf-G\\_IJ\\_m5o&hl=es&ei=TY0hS4neCoyIuAfyosHOCg&sa=X&oi=book\\_result&ct=result&resnum=1&ved=0CAgQ6AEwAA#v=onepage&q=&f=false](http://books.google.com.ar/books?id=VsYUPtW9IDc&dq=Latinoam%C3%A9rica+ante+la+encrucijada+de+la+postmodernidad&printsec=frontcover&source=bl&ots=tBfG_pzLHM&sig=6jAL89YuA4L6ZvMilf-G_IJ_m5o&hl=es&ei=TY0hS4neCoyIuAfyosHOCg&sa=X&oi=book_result&ct=result&resnum=1&ved=0CAgQ6AEwAA#v=onepage&q=&f=false)

Rabe-Hesketh, S. y Skrondal, A. (2012) Dichotomous or binary responses [Respuesta binaria o dicotómica]. En *Multilevel and Longitudinal Modeling Using Stata*. [Modelado Multinivel y Longitudinal con Stata] (3a ed.) (vol II, cap. 10, pp. 501-574). (s.l.): Stata Press. Recuperado de: [http://www.statapress.com/books/mlmus3\\_ch10.pdf](http://www.statapress.com/books/mlmus3_ch10.pdf)

Recio, A. (1997). *Trabajo, Personas, Mercados: manual de Economía Laboral*. Barcelona: Icaria.

Revueltas, A. (1990). Modernidad y Mundialidad. *Filosofía, Historia, Letras, Instituto Tecnológico Autónomo de México*, 23, pp. 119-134. Recuperado de: [http://biblioteca.itam.mx/estudios/estudio/letras23/notas/sec\\_1.html](http://biblioteca.itam.mx/estudios/estudio/letras23/notas/sec_1.html)

Rionda Ramírez, J.I. (2006) El Estado como objeto económico. [Edición electrónica]. Recuperado de: [www.eumed.net/libros/2006b/jirr-est/](http://www.eumed.net/libros/2006b/jirr-est/)

Riquelme, G. (2006). La relación entre educación y trabajo: continuidad, rupturas y desafíos. *Anales de la Educación Común*, 2(5), pp. 68-75. Recuperado de <http://servicios.abc.gov.ar/lainstitucion/revistacomponents/revista/archivos/anales/numero05/archivosparaimprimir/8.riquelme.pdf>

Santa Fe (provincia). Ministerio de Economía. Secretaría de Planificación y Política Económica. Instituto Provincial de Estadísticas y Censos – IPEC. (Mayo, 2016). *PBG. Producto Bruto Geográfico. 1993-2015. Resultados Preliminares*. Santa Fe: Autor. Recuperado de <https://santafe.gob.ar/index.php/web/content/download/227821/1193321/file/PBG%20Mayo%202016.pdf>

Santa Fe (provincia). Ministerio de Educación. Plan Vuelvo a Estudiar. (2016). Santa Fe: Autor. Recuperado de <https://www.santafe.gov.ar/index.php/web/content/view/full/190728>

- Sevilla, C. (2004). *El Capital Humano y su contribución al crecimiento económico*. España: Ediciones de la Universidad de Castilla- La Mancha. Recuperado de: [http://books.google.com.ar/books?id=i0oQda0p\\_oC&pg=PA28&lpg=PA28&dq=schultz+capital+humano&source=bl&ots=GQYfujtmX-&sig=vy5eDmTXas-Ri-gRpFeEmVk860Y&hl=es&ei=GvgCS627OdOWtgf3zaTzDQ&sa=X&oi=book\\_result&ct=result&resnum=9&ved=0CCEQ6AEwCA#v=onepage&q=schultz%20capital%20humano&f=false](http://books.google.com.ar/books?id=i0oQda0p_oC&pg=PA28&lpg=PA28&dq=schultz+capital+humano&source=bl&ots=GQYfujtmX-&sig=vy5eDmTXas-Ri-gRpFeEmVk860Y&hl=es&ei=GvgCS627OdOWtgf3zaTzDQ&sa=X&oi=book_result&ct=result&resnum=9&ved=0CCEQ6AEwCA#v=onepage&q=schultz%20capital%20humano&f=false)
- Schultz T. W. (1961). Investment in Human Capital [Inversion en Capital Humano]. *The American Economic Review*, 51(1). 1-17. Recuperado de: <http://la.utexas.edu/users/hcleaver/330T/350kPEESchultzInvestmentHumanCapital.pdf>
- Sistema de Información de Tendencias Educativas en América Latina - Siteal. (noviembre, 2015). La Educación Superior en América Latina ante una demanda sin precedentes. *Diálogos del Siteal*. Recuperado de [http://www.siteal.iipe-oei.org/sites/default/files/siteal\\_dialogo\\_pedro\\_henriquez\\_guajardo.pdf](http://www.siteal.iipe-oei.org/sites/default/files/siteal_dialogo_pedro_henriquez_guajardo.pdf)
- Smith, A. (1958) *Investigación sobre la naturaleza y causas de la riqueza de las naciones*. Fondo de Cultura Económica, México (Trabajo original publicado en 1776)
- Snijders, T. (1996). Analysis of longitudinal data using the hierarchical linear model [Análisis de datos longitudinales usando el modelo lineal jerárquico]. *Quality & Quantity*, 30, 405-426. Países Bajos: Kluwer Academics. Recuperado de [https://www.stats.ox.ac.uk/~snijders/Snijders\\_QQ\\_1996.pdf](https://www.stats.ox.ac.uk/~snijders/Snijders_QQ_1996.pdf)
- Snijders, T. (2003). Multilevel Analysis [Análisis Multinivel]. En M. Lewis-Beck, A.E. Bryman, T.F. Liao (eds.), *The SAGE Encyclopedia of Social Science Research Methods* (vol. II, p. 673-677). Thousand Oaks, CA: Sage. Recuperado de <https://www.stats.ox.ac.uk/~snijders/MultilevelAnalysis.pdf>
- Stiglitz, J. E. (junio, 1975). The Theory of "Screening," Education, and the Distribution of Income [La teoría del "filtro", la educación y la distribución del ingreso]. *The American Economic Review* 65(3), 283-300. Recuperado de: <http://sites-final.uclouvain.be/econ/DW/DOCTORALWS2004/bruno/Mafia/stiglitz.pdf>

- Tedesco, J. C. (2000). *Educación en la sociedad del conocimiento*. Buenos Aires: Fondo de Cultura Económica. Recuperado de <https://socioeducacion.files.wordpress.com/2011/05/tedesco-carlos-educar-en-la-sociedad-del-conocimiento.pdf>
- Terrén, E. (1999) *Educación y Modernidad: entre la utopía y la burocracia*. Barcelona: Ed. Anthropos.
- Torrado, S. (2007). Estrategias de desarrollo, estructura social y movilidad. En *Población y Bienestar en Argentina del Primero al Segundo Centenario. Una historia social del siglo XX*. (Tomo I, pp.31-67). Buenos Aires: EDHASA.
- Thurow, L. C. (verano, 1972). Education and economic equality [Educación e igualdad económica]. *The Public Interest*. 28, 66-81. Recuperado de: [http://www.nationalaffairs.com/public\\_interest/detail/education-and-economic-equality](http://www.nationalaffairs.com/public_interest/detail/education-and-economic-equality)
- Urbina, A. A. e Ibarra, C. J. (2013). Análisis multinivel en la investigación sociológica cuantitativa: implicancias y conveniencias. *Revista de Temas Sociológicos*, 17, 93 – 118. Recuperado de: <http://ediciones.ucsh.cl/ojs/index.php?journal=TSUCSH&page=article&op=view&path%5B%5D=195&path%5B%5D=152>
- Vázquez, L. (18 de marzo de 2016). Educación: ¿Vale la pena ir a la universidad?. *La Nación*. Recuperado de <http://www.lanacion.com.ar/1880841-educacion-vale-la-pena-estudiar-una-carrera-en-la-universidad>
- Ventoso Mariño, J. (2014). *Situación actual y perspectivas del mercado de trabajo de los universitarios españoles: inserción laboral y competencias*. (Tesis de Grado). Universidade da Coruña, Facultade de Economía e Empresa, A Coruña. Recuperado de <http://hdl.handle.net/2183/12447>

## Apéndice

### *Aclaraciones sobre el cálculo de Tasas*

Como se explicó, la forma de medición de las variables en el Panel de Actores Sociales del Observatorio Social difiere de la forma aplicada por la Encuesta Permanente de Hogares, debido a que la intención del primero es cuantificar la “necesidad” de trabajo y no sólo el “satisfactor” que determina la situación laboral de la persona, tal como lo realiza el segundo. Por ello, en la Tabla 2 se expusieron a modo de ejemplo y comparación una serie de tasas relativas al mercado de trabajo. Se explica a continuación, el cálculo de cada tasa según la metodología de cada Institución<sup>13</sup>:

#### Encuesta Permanente de Hogares – INDEC

**Tasa de Actividad:** Se calcula como porcentaje entre la población económicamente activa y la población total.

**Tasa de Empleo:** Se calcula como porcentaje entre la población ocupada y la población total.

**Tasa de Desocupación:** Se calcula como porcentaje entre la población desocupada y la población económicamente activa.

**Tasa de Subocupación:** Es la proporción de ocupados que trabajan menos de 35 horas semanales

#### Panel de Actores Sociales – Observatorio Social

**Tasa de “Actividad”:** Se calcula como porcentaje entre los económicamente activos (que trabajan y/o buscan trabajo) y el total de la muestra.

**Tasa de “Trabajo”:** Se calcula como porcentaje entre los que afirman trabajar y la muestra total (comparable con la Tasa de empleo de la EPH).

**Tasa de “No trabajo”:** Se calcula como porcentaje entre los que afirman que no trabajan y simultáneamente buscan trabajo, sobre la muestra económicamente activa (comparable con la Tasa de desocupación de la EPH).

**Tasa de “Trabajo demandante”:** Se calcula como porcentaje de personas que trabajan y que también buscan otro trabajo, sobre la muestra económicamente activa (comparable con la Tasa de Subocupación Horaria de la EPH).

---

<sup>13</sup> Esta información fue proporcionada por el Observatorio Social: “Informe Tema TRABAJO. Panel de Hogares.” (septiembre, 2015).

## ANEXO

### *Otros modelos estimados*

A continuación se presentan otros modelos que fueron estimados pero finalmente descartados por presentar variables estadísticamente no significativas o un peor ajuste que el modelo seleccionado.

#### 1. Modelo Logístico Multinivel de intercepto aleatorio, con la variable “Edad al cuadrado centrada”.

Mixed-effects logistic regression		Number of obs	=	1093		
Group variable: NroCuest		Number of groups	=	417		
		Obs per group: min	=	1		
		avg	=	2.6		
		max	=	3		
Integration method: mvaghermite		Integration points	=	7		
		Wald chi2(7)	=	31.19		
Log likelihood = -262.35499		Prob > chi2	=	0.0001		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	5.53622	2.451167	3.87	0.000	2.324572	13.18511
Edad_centrada	1.213945	.1389664	1.69	0.090	.9699684	1.519289
Edad_cuad_centrada	.9982108	.0011516	-1.55	0.121	.9959562	1.00047
Educ_Medio	4.128285	1.765136	3.32	0.001	1.785749	9.543749
Educ_Alto	7.49988	4.000904	3.78	0.000	2.636151	21.33725
Año2010	.923575	.304919	-0.24	0.810	.4835566	1.763994
Año2011	1.318131	.4808238	0.76	0.449	.6448514	2.69437
_cons	4.630091	2.247344	3.16	0.002	1.788278	11.98792
NroCuest						
var(_cons)	4.643851	1.662617			2.302112	9.367638
LR test vs. logistic regression: <u>chibar2(01) =</u>		37.63	Prob>=chibar2 =	0.0000		

2. Modelo Logístico Multinivel de intercepto aleatorio, sin las variables dicotómicas correspondientes a cada año.

Mixed-effects logistic regression		Number of obs =		1093		
Group variable: NroCuest		Number of groups =		417		
		Obs per group: min =		1		
		avg =		2.6		
		max =		3		
Integration method: mvaghermite		Integration points =		7		
Log likelihood = -262.8708		Wald chi2(5) =		30.50		
		Prob > chi2 =		0.0000		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	5.55965	2.458231	3.88	0.000	2.337135	13.22547
Edad_centrada	1.222288	.1395268	1.76	0.079	.9772532	1.528763
Edad_cuad_centrada	.9981475	.0011485	-1.61	0.107	.9958991	1.000401
Educ_Medio	4.125838	1.758257	3.33	0.001	1.789642	9.511698
Educ_Alto	7.505921	3.996877	3.79	0.000	2.643274	21.31404
_cons	4.783383	2.147028	3.49	0.000	1.984597	11.52917
NroCuest		var(_cons)		4.605098		1.650821
				2.280882		9.297687
LR test vs. logistic regression: <u>chibar2(01) =</u>				37.28 Prob>=chibar2 = 0.0000		

3. Modelo Logístico Multinivel de intercepto aleatorio, sin las variables dicotómicas correspondientes a cada año, ni la edad al cuadrado centrada.

Mixed-effects logistic regression		Number of obs =		1093		
Group variable: NroCuest		Number of groups =		417		
		Obs per group: min =		1		
		avg =		2.6		
		max =		3		
Integration method: mvaghermite		Integration points =		7		
Log likelihood = -264.16017		Wald chi2(4) =		29.01		
		Prob > chi2 =		0.0000		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	5.441113	2.402681	3.84	0.000	2.289893	12.92887
Edad_centrada	1.019856	.0173162	1.16	0.247	.9864757	1.054366
Educ_Medio	4.342976	1.863161	3.42	0.001	1.873343	10.06833
Educ_Alto	7.542812	3.999251	3.81	0.000	2.668212	21.3229
_cons	6.014274	2.654864	4.06	0.000	2.53186	14.28653
NroCuest		var(_cons)		4.659719		1.705766
				2.273859		9.548957
LR test vs. logistic regression: <u>chibar2(01) =</u>				36.86 Prob>=chibar2 = 0.0000		

4. Modelo Logístico Multinivel de intercepto aleatorio y pendiente aleatoria en la edad al cuadrado centrada, sin términos de interacción.

Mixed-effects logistic regression		Number of obs	=	1093		
Group variable: NroCuest		Number of groups	=	417		
		Obs per group: min	=	1		
		avg	=	2.6		
		max	=	3		
Integration method: mvaghermite		Integration points	=	7		
Log likelihood = -262.87067		Wald chi2(5)	=	30.51		
		Prob > chi2	=	0.0000		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	5.557751	2.456622	3.88	0.000	2.33697	13.21737
Edad_centrada	1.222197	.1394809	1.76	0.079	.9772359	1.528561
Edad_cuad_centrada	.9981483	.0011482	-1.61	0.107	.9959004	1.000401
Educ_Medio	4.124418	1.75718	3.33	0.001	1.789428	9.506294
Educ_Alto	7.502543	3.993981	3.79	0.000	2.642841	21.29835
_cons	4.78234	2.145882	3.49	0.000	1.984715	11.52346
NroCuest						
var(Edad_centrada)	1.32e-35	1.10e-19			.	.
var(_cons)	4.601457	1.648838			2.279737	9.287653
LR test vs. logistic regression: <u>chibar2(01) =</u>		37.28	Prob>=chibar2 =	0.0000		

5. Modelo Logístico Multinivel de intercepto aleatorio, con las variables que indican si el jefe de hogar es estudiante y si tiene niños de 14 años o menos (variable "A3C").

Mixed-effects logistic regression		Number of obs	=	659		
Group variable: NroCuest		Number of groups	=	385		
		Obs per group: min	=	1		
		avg	=	1.7		
		max	=	3		
Integration method: mvaghermite		Integration points	=	7		
Log likelihood = -148.37321		Wald chi2(5)	=	13.82		
		Prob > chi2	=	0.0168		
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Masculino	8.650317	6.055983	3.08	0.002	2.193391	34.11521
Educ_Medio	3.854027	2.600337	2.00	0.046	1.027078	14.46192
Educ_Alto	11.87125	11.26801	2.61	0.009	1.84735	76.28585
A3C	1.188439	.7149978	0.29	0.774	.3654861	3.864411
Estudiante	1.058505	.1747063	0.34	0.730	.7659531	1.462796
_cons	6.389267	8.43561	1.40	0.160	.4804279	84.97161
NroCuest						
var(_cons)	8.445041	6.057404			2.070415	34.44657
LR test vs. logistic regression: <u>chibar2(01) =</u>		18.43	Prob>=chibar2 =	0.0000		

6. Modelo Logístico Multinivel de intercepto aleatorio y pendiente aleatoria en la edad, con término de interacción entre niveles (variables sexo y edad). Estructura de la matriz de covarianza: identidad. El mismo modelo fue probado con otras estructuras de la matriz de covarianza sin lograr convergencia.

Mixed-effects logistic regression		Number of obs	=	1093			
Group variable: NroCuest		Number of groups	=	417			
		Obs per group: min	=	1			
		avg	=	2.6			
		max	=	3			
Integration method: mvaghermite		Integration points	=	7			
		Wald chi2(5)	=	36.84			
Log likelihood = -274.52701		Prob > chi2	=	0.0000			
( 1) [var(Edad_centrada[NroCuest])]_cons - [var(_cons[NroCuest])]_cons = 0							
Trabaja	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]		
Educ_Medio	2.622122	.8004188	3.16	0.002	1.441513	4.76966	
Educ_Alto	5.242283	2.138058	4.06	0.000	2.356997	11.65955	
Sexo							
Masculino	2.305328	.7435018	2.59	0.010	1.225202	4.337681	
Edad_centrada	1.025501	.0259765	0.99	0.320	.9758309	1.077699	
Sexo#c.Edad_centrada							
Masculino	.9398347	.027812	-2.10	0.036	.8868748	.9959571	
_cons	4.503759	1.346241	5.03	0.000	2.506908	8.091183	
NroCuest							
var(Edad_centrada)	.0090631	.0045588			.0033815	.0242906	
var(_cons)	.0090631	.0045588			.0033815	.0242906	
LR test vs. logistic regression: <u>chibar2(01) =</u>		11.09	Prob>=chibar2 =	0.0004			