



Facultad de Ciencias
Instituto de Estadística
Ingeniería en Estadística

Estudio de brechas salariales entre jefes y jefas de hogar según variables de la Encuesta CASEN en Pandemia 2020

Franco Ogalde Briceño

Bajo la supervisión de:

Profesor Guía

Carlos Henríquez Roldán. Ph. D.

Instituto de Estadística, Universidad de Valparaíso

Profesor Co-Guía

Alejandra Christen. Ph. D.

Instituto de Estadística, Universidad de Valparaíso

Para optar al:

grado académico de: *Licenciado en Estadística*

título profesional de: *Ingeniero en Estadística*

minor en: *Modelización Estadística*

22 de diciembre de 2022

Resumen

En los últimos años se mencionó a Chile como el quinto país con la mayor brecha salarial del mundo la cual había superado por bastante a sus países vecinos de latinoamérica, esto sumado al estado actual de los ingresos en el país, en donde una gran parte de la población lucha por mantenerse mes a mes debido a dificultades como el alza de los productos y la calidad de vida a causa de la inflación rampante dejando al país en un estado complejo en este aspecto. Se realizó un análisis al instrumento realizado por el Ministerio de Desarrollo Social y Familia en conjunto al Instituto de Estadísticas llamado Encuesta CASEN en Pandemia 2020, esto con el propósito de comparar los ingresos entre los distintos sexos y determinar el estado de la brecha salarial entre los jefes y las jefas de hogar en el país durante el periodo del año 2020. Además se buscó conocer los posibles factores que apoyaban o afectaban directamente a la creación de esta inequidad en estas diferencias. También, a lo largo del análisis se realizaron métodos y técnicas multivariadas tales como la descomposición Oaxaca-Blinder y el análisis de correspondencias múltiple para observar como actúan estos ingresos entre los sexos de los jefes de hogar y el cómo algunas variables observadas en la Encuesta CASEN se asocian entre sí. Entre los resultados obtenidos se encuentra que la brecha salarial por género entre los jefes de hogar es equivalente al 33.3% a favor de los hombres, la cual supera al promedio de la brecha entre las naciones de la OCDE en el año 2018, lo cual no demuestra una disminución relevante de la brecha en el país. Además, se logró caracterizar las diferencias entre los jefes de hogar por sexo con respecto a la inserción laboral. Con este trabajo se espera que se tomen en consideración de mayor manera la equidad entre ambos sexos en el ámbito de los ingresos, se espera que se tomen medidas en las leyes públicas para que el país mejore de una manera más igualitaria y este sea mucho más próspero.

Agradecimientos

Quienes me conocen sabrán que nunca he sido muy bueno hablando de temas personales de manera abierta, aún así, me gustaría que supiesen lo importante que ustedes son para mí y que merecen todo mi agradecimiento y que son demasiado relevantes en mi vida.

Quisiera iniciar estos agradecimiento partiendo con las personas más cercanas que tengo, y que sin ellas nada de esto sería posible, mi familia.

Agradezco a mis padres, Marcos Eduardo Ogalde Briceño y Patricia del Rosario Briceño López quienes me regalaron la vida, me criaron y me formaron para llegar a este hito relevante de mi vida, sin ellos dos no hubiese sucedido nada de lo que he vivido hasta aquí. Una de las cosas que he tenido en la mente desde pequeño ha sido el cómo devolverles algo de lo que han hecho ustedes por mi, puesto a que criar a un hijo no es fácil, y han sido 24 años en donde ellos han estado ahí para mí en todo momento, por lo que siento que al menos al realizar este trabajo puedo demostrarles a mis padres que todo lo que han realizado en mi tutela y crianza ha rendido frutos, no desmereciendo a mis logros anteriores pero sí considerando el hecho de que he llegado hasta un punto de educación en el cual me siento totalmente orgulloso de mis logros. Les estoy sumamente agradecido y me vuelve muy feliz el que estén como un apoyo constante a todos mis proyectos.

A mi hermano Eduardo Andrés Ogalde Briceño quisiera agradecerle todo lo que ha hecho por mí, si no fuese por él lo más probable es que yo no hubiese realizado este trabajo ni tampoco hubiese estado en algo similar, puesto a que gran parte de mis gustos, hobbies y actividades que hago o he intentado han sido gracias a que él desde pequeño me ha introducido en las actividades que realizaba y a su gran apoyo incondicional a lo largo de los años, tanto en las cosas que hemos desarrollado juntos como en las que he conocido por mi cuenta, en donde él ha sido un pilar sumamente importante en donde confiar. No considero que exista otra persona que haya sido un mayor aporte en lo que soy actualmente como lo es él, y de ello le estoy sumamente agradecido por todo lo que ha hecho.

A mi hermano Marcos Adrian Salvador Ogalde Gatica con quien he tenido una mejor relación a lo largo de estos últimos años, quisiera agradecerle por todo el apoyo y todo lo que ha hecho durante todo este tiempo. También quisiera mencionar a su hijo y mi sobrino Martín Ogalde, quien ha sido alguien muy importante en el círculo familiar en los últimos años.

A mis amigos Bruno Inostroza, Francisco Pozo, Samuel Díaz, Darko Castillo, Jorge Gallo, Rodolfo Calderón, Gabriel Zuleta, Felipe Mc Clinton, Cristobal Cabello, José Santa Ana, Lucas Galleguillos y a todos los que no he podido mencionar, quiero que sepan que los aprecio demasiado, son una gran parte del apoyo emocional que he tenido a lo largo de los años y espero que ese apoyo se siga manteniendo a través del tiempo ya que los momentos que he pasado con todos ustedes son irremplazables para mí, puesto a que fuese simplemente por juntas de amigos o por pasar noches

hablando a través de Discord, ustedes realmente me hacen muy feliz y al igual que como ustedes lo han sido conmigo yo estaré para ustedes siempre que lo necesiten.

Al Dr. Carlos Henríquez quisiera agradecerle por toda la formación que me ha dado a lo largo de la carrera, ha sido una persona que me ha enseñado demasiado y con la que me ha gustado trabajar a lo largo de mi tiempo en el área de la estadística desde el primer momento al que llegué a ella. Además, quisiera agradecerle por todo lo que ha realizado al ser mi profesor guía en este trabajo como el ser excesivamente cuidadoso en el escrito, el cual muestra los frutos de mi formación a lo largo de estos cinco años.

A la Dra. Alejandra Christen quisiera agradecerle también por la formación que me ha otorgado en los dos últimos años de la carrera, es una excelente persona la cual también es muy exigente y me ha gustado su tutela a lo largo de estos años. Quisiera agradecerle nuevamente por darme la oportunidad de realizar este trabajo con ella como profesora guía, en donde sin su exigencia semana a semana no creo que nada de esto haya sido posible.

Quisiera agradecer además al Dr. Harvey Rosas que, si bien no ha estado muy presente en los últimos años de mi formación en la carrera, él fue la persona que me introdujo a la estadística y al área de la minería de datos, si él no hubiese estado lo más seguro es que simplemente me hubiese mantenido en mi primera carrera y posiblemente hubiese abandonado mucho después, por eso le estoy agradecido.

A todos mis compañeros de la carrera Ingeniería en Estadística, tanto los que entraron a la carrera junto a mí en el año 2018 como a los que he conocido en medio de esta carrera, quienes estuvieron conmigo a lo largo de estos cinco años a través de varias victorias y derrotas, quisiera mencionar especialmente al grupo que llamamos "los espartanos" quienes han sido las personas más cercanas a mí a lo largo de todo este camino.

Quisiera agradecer de igual manera al Observatorio Social del Ministerio de Desarrollo Social y Familia por facilitar los datos que fueron analizados a través de este trabajo de titulación y además al grupo fundador del software estadístico R y a las personas de la compañía StataCorp LLC, de donde proviene el software Stata que es el que ha sido utilizado a lo largo de todo este trabajo.

Finalmente, quisiera agradecerles a todas las personas que de una u otra manera hicieran esto posible, ya que no esto no es sólo hecho por las personas mencionadas anteriormente sino que es la suma de todas las experiencias que he tenido hasta el momento las que hicieron posible que llegase aquí.

Índice general

Resumen	2
Agradecimientos	4
1. Introducción	6
1.1. Objetivos	8
1.2. Discusión bibliográfica	9
1.3. Recursos utilizados	11
2. Metodología	13
2.1. Diseño muestral	13
2.1.1. Marco muestral	14
2.1.2. Cuestionario	15
2.1.3. Factores de expansión	16
2.2. Descomposición Oaxaca-Blinder	17
2.3. Regresión lineal	20
2.4. Análisis de correspondencias	21
2.5. Análisis de correspondencias múltiple	23
2.6. Integración del survey data y factores de expansión	26
3. Resultados	27
3.1. Análisis descriptivo	27
3.2. Descomposición Oaxaca-Blinder	41
3.3. Regresión lineal	44
3.4. Análisis por correspondencia múltiple	45
4. Conclusiones y discusión	50
A. Códigos	52
Referencias	58

Capítulo 1

Introducción

Un tópico que suele ser bastante delicado en general es lo referente a los ingresos, esto es debido a que Chile suele tener una gran cantidad de factores que dejan en una extraña posición a los empleos de las personas dentro del país, ya que estos empleos no alcanzan generalmente a cubrir necesidades básicas debido a las bajas remuneraciones que tienen por lo general la población chilena (Considerando el sueldo promedio de \$681.039 y el sueldo medio de \$457.690 durante el año 2021 (Godoy, 2021)) entonces, al existir asuntos tales como el precio excesivo de los medicamentos (Romo, 18), las problemáticas con respecto al sistema de pensiones chileno (Muñoz y Aguilar, 2021) o la constante inflación que ha estado en un aumento constante con el paso de los últimos años al punto de que la tasa de variación del IPC (Índice de precios del consumidor) de septiembre del actual año ha aumentado al de septiembre del año 2017 por 10.6% (Datosmacro, 2022)z, lo cual crea una constante necesidad búsqueda de empleo o de la mantención de este, en especial considerando el aumento del desempleo durante los años de pandemia, en especial si no es sólo una persona la que se tiene que mantener, sino toda una familia.

En el Ministerio de Desarrollo Social y Familia de Chile (MDSF, anteriormente conocido como el Ministerio de Desarrollo Social, Ministerio de Planificación y, Ministerio de Planificación y Cooperación) se realizan medidas para promover mejores condiciones en la familia como tal, pero se mantienen casos en donde se tienen pobres medidas con respecto a las familias en los sectores más bajos de ingresos en la población.

Con todo lo anterior mencionado, al tener una situación económica en gran parte de la población en donde se tienen problemas con la obtención de los insumos básicos mes a mes, además hay que considerar la situación de que se disminuyan los salarios en relación a los pares sólo por el sexo o género, debido a la alta discriminación que ocurre en el país hacia el sexo femenino con el paso del tiempo en múltiples ámbitos tales como culturales, políticos, sociales y económicos ya que Chile fue de los cinco países con la mayor brecha salarial en el mundo según la Organización para la Cooperación y Desarrollo Económicos (OCDE) durante el año 2018 con diferencia mayor al 30% (durante aquel periodo, los sueldos de los hombres solían superar los \$600.000 mientras que los sueldos de las mujeres era cercano a los \$410.000 en promedio), en donde su puesto estaba detrás de países tales como Lituania, Japón, Estonia y Corea del Sur (CNN, 2020) superando con creces el promedio de los países que integraban la OCDE en aquel año el cual era de 14.3% (OCDE, 2018). Por lo tanto, se vuelve necesario el que estos temas constantemente en boca de la gente para poder causar cambios, debido a que si bien, servicios tales como el Servicio Nacional de la Mujer y la Equidad de Género (SernamEG) y órganos administrativos como el Ministerio de la Mujer y Equidad de Género (MinMujeryEG) han hecho avances y han servido un rol importante

en la disminución de la discriminación e igualdad hacia la mujer en distintas áreas tales como la integración de temas como la equidad de género en las políticas públicas o el velar para que estas sigan efectuándose, aún sigue existiendo discriminación hacia el sexo femenino en distintas áreas, mostrándose claramente en el mundo laboral a través de las diferencias de salarios en el país y hasta en otras aristas como las oportunidades que se encuentran para entrar al mundo laboral como tal o las dificultades que las mujeres tienen para avanzar en sus puestos dentro de las empresas en donde se encuentran, entre otras razones.

Pero al hablar de discriminación laboral, ¿a qué se refiere este término? ¿qué es la brecha salarial por género? La comisión Europea menciona que la brecha salarial mide un concepto mucho más amplio que sólo la discriminación laboral, y que esta se expande a un número de desigualdades a las que las mujeres se enfrentan constantemente en el mundo laboral siendo el acceso, la progresión y las recompensas que se encuentran dentro de esta área. Entre estos problemas se encuentran:

- Segregación por sector, debido a que cerca de un 24 % de la brecha salarial entre hombres y mujeres está causada por la sobrerrepresentación de las mujeres en sectores de paga menor a los que están mayormente representados por hombres. Los trabajos que suelen ser considerados más “femeninos” tienden a tener pagas inferiores a los que no suelen ser catalogados de esta manera.
- Reparto desigual del trabajo remunerado y no remunerado, con la mayor cantidad de trabajo que suelen tener las mujeres en comparación a los hombres considerando el trabajo no remunerado, lo que afecta a tanto a las opciones profesionales para obtener como mantener un empleo.
- Techo de cristal, término con el que se suele referir a las dificultades que tienen las mujeres para ascender dentro de las organizaciones en donde se encuentran. Al menos existe un 8 % de directores ejecutivos (CEO) o equivalentes dentro de las empresas suelen ser mujeres, además, las mujeres en altos cargos suelen tener un 23 % de ingresos menor que los hombres en las mismas condiciones o similares.
- Discriminación salarial, debido a que en algunas ocasiones, las mujeres suelen recibir un salario inferior al de los hombres netamente por discriminación salarial sin una razón aparente.

A pesar de que se conozcan algunas de las razones del por qué ocurre este tipo de discriminación, gran parte de la existencia de esta generalmente no puede ser explicada ni tampoco vincularse a las características del trabajador o del sitio de trabajo, por lo que el origen de esta discriminación debe de mantenerse como desconocida. Es por razones como estas que es de suma importancia el conocer las características que suelen causar las brechas que separan a ambos sexos en áreas tales como la laboral, para tener un ecosistema laboral mucho más próspero, sano e igualitario ([European Commission, 2021](#)).

En un contexto como este, es necesario contar con estimaciones pertinentes sobre la discriminación salarial por género que ocurre constantemente en el país, para así poder cambiar o mejorar las políticas públicas para que estas sean acordes y solventar las necesidades en estos ámbitos como corresponden.

Una de las medidas que otorga el Ministerio de Desarrollo Social y Familia para conocer las situación del país en relación a los ingresos, tales la situación, las carencias o las brechas de pobreza a través de su división del Observatorio Social es la Encuesta de Caracterización Socioeconómica

Nacional (llamada también Encuesta CASEN) la cual es un instrumento que dispone de información referente a los ingresos para conocer la situación en la que se encuentran los hogares y la población dentro del país en el ámbito económico, principalmente aquella en la situación de pobreza y de los grupos definidos como prioritarios a través de políticas sociales, evaluando el impacto de estas políticas. Este instrumento se lleva a cabo desde el año 1990, desde el cual se han realizado nuevas versiones en periodos de cada tres años, siendo la última versión la estudiada en esta investigación, denominada como Encuesta CASEN en Pandemia 2020 (MDSF, 2021b).

Una de las características que vuelve tan importante a esta versión de la CASEN es que debido a las restricciones sanitarias establecidas en el año de ejecución de la encuesta debido a la pandemia por Coronavirus, la Encuesta CASEN en Pandemia 2020 fue realizada a través de una modalidad mixta la que constaba de tres etapas, denominadas como Precontacto, Encuesta de manera telefónica y Recuperación presencial. También, a causa de las restricciones sanitarias y la nueva modalidad en la que se ejecuta el instrumento se optó por una reducción de los contenidos del instrumento con respecto a las versiones anteriores de este (MDSF, 2021b).

Al considerar lo anterior mencionado, esta investigación quiere responder a la pregunta de que durante el pasado año 2020, ¿Chile seguía manteniendo un alto nivel de discriminación salarial por género?

En el presente estudio se tiene el objetivo de determinar la brecha salarial entre los jefes y las jefas de hogar y conocer los factores que puedan ser determinantes de esta, para ello, se utilizará el método conocido como descomposición Oaxaca-Blinder utilizado habitualmente en estudios de género para la obtención de la magnitud de la brecha salarial explicada a través de la discriminación a lo largo de varios factores y la técnica de análisis de correspondencias múltiple (ACM, o por sus siglas en inglés MCA) para la obtención de patrones y características que puedan explicar parte del comportamiento de la brecha salarial.

En la siguiente sección, se mencionarán los objetivos considerados a lo largo de este proyecto.

1.1. Objetivos

Para esta investigación, se consideraron los objetivos pertinentes durante la ejecución de este:

Objetivo principal

- Determinar el estado en el año 2020 de la brecha salarial entre jefes y jefas de hogar y conocer posibles factores que sean determinantes en la explicación de esta brecha, a partir de la Encuesta CASEN en Pandemia 2020.

Objetivos específicos

Entre los objetivos específicos que se encuentran dentro de esta investigación son:

1. Estudiar el plan de muestreo utilizado en la Encuesta CASEN en Pandemia 2020 y el uso de factores de expansión en los análisis realizados.

2. Comparar los ingresos entre los distintos sexos estipulados dentro de la Encuesta CASEN en Pandemia 2020.
3. Encontrar variables que sean de utilidad dentro de la Encuesta CASEN en Pandemia 2020 para encontrar nuevos factores que puedan influir en la presencia de una brecha salarial entre sexos.
4. Aplicar métodos multivariados en la Encuesta CASEN en Pandemia 2020 con las variables encontradas que tengan la posibilidad de ser factores influyentes en la brecha salarial entre sexos.
5. Realizar un análisis global de la brecha salarial entre los jefes y las jefas de hogar durante la situación de pandemia 2020.

En la siguiente sección, se hablará de la discusión bibliográfica y de lo que se ha hablado en la literatura.

1.2. Discusión bibliográfica

En la literatura se ha mencionado recientemente la discriminación por género y la brecha salarial de una manera constante debido a que Chile, por lo general, ha tenido una clara discriminación hacia el sexo femenino en el ámbito laboral que se ha vuelto más comentado en los últimos años, a diferencia de otros países más desarrollados en donde el tema estaba en boga con anterioridad. Esto se puede notar con la disminuida cantidad de información sobre la brecha antes de la década de los 2000, pese a ello, existen investigaciones de la brecha tales como la de Paredes y Riveros (1994) en donde se estiman las diferencias en los salarios por género en el país por el periodo de 1958 hasta 1990 utilizando la técnica Oaxaca-Blinder (1973) para corregir el sesgo de selección, en donde se plantea que las diferencias de salarios son una respuesta al ciclo económico y si bien, los hombres en los primeros años estudiados superaban por cerca del 95 % a las mujeres, en 1990 se habría disminuido hasta un 47.4 % en donde la mayor participación de mujeres en el mundo laboral ocurría en casos donde se tenía mayor capital, pese a ello, al corregir el sesgo se obtenía que la brecha que debió de haber entre aquellos años sería de un 91.4 % en el año 1958 y un 74.4 % en el año 1990. Además, Szasz (1995) describe las desigualdades en la estructura poblacional y en los ingresos, en donde se menciona que la alta escolaridad de las mujeres contrasta con los cargos que obtienen y sus ingresos.

A partir de la década del 2000, se tomó con más consideración el tema de la brecha salarial teniendo muchas más investigaciones enfocadas en ella, tales como Fuentes (2005) quien describe como evoluciona la discriminación salarial por género en Chile durante el periodo de 1990 y 2003, los años en donde la Encuesta CASEN se había realizado hasta aquel momento utilizando los métodos Oaxaca-Blinder seguido de Oaxaca-Ransom. Sus resultados concluían que la discriminación salarial por género había disminuido durante aquel periodo pero que ésta se había situado en un 27,5 % durante el año 2003, lo que se mantendría como el periodo con menor brecha salarial del país según estudios posteriores.

A pesar del historial de discriminación por género en los salarios del país, según Castillo (2009) en su estudio referente a la brecha salarial teniendo en cuenta más la subjetividad e influencia y el cómo la percibe la población, en donde se considera que esta desigualdad es justa por parte de la población, lo que no incentivaba al cambio por parte de las autoridades.

Más tarde, Ferrada (2018) mide la magnitud de la discriminación salarial por género al utilizar los datos de la Encuesta CASEN 2015 a través del método de descomposición Oaxaca-Blinder (1973). En este estudio se obtiene una diferencia no significativa en la brecha salarial mencionando que se considera sobreestimada, y se considera una discriminación observable de un 11 %. Pese a ello, Castillo (2019) estudia las remuneraciones, las discriminaciones y las brechas por sexo en las entidades públicas de Chile. Los resultados de aquel artículo estipulaban que existen diferencias claramente notorias en el caso de las remuneraciones a favor de los hombres en un 42 % en comparación a las mujeres, teniendo mayores desigualdades en los casos de los empleos profesionales y técnicos.

Durante el año 2021 se tuvo una gran cantidad de investigaciones con respecto a la Encuesta CASEN en Pandemia 2020, entre estas se puede mencionar a Mandiola (2021) quien investiga sobre la disminución de la participación de las mujeres en el mundo laboral con respecto a los hombres, dándose a conocer que existe una brecha salarial de un 27,2 % en desmedro de las mujeres y que esta brecha va aumentando a mayor nivel educacional. En esta ocasión se menciona que el teletrabajo es mucho más complicado de realizarse en el caso de las mujeres puesto a que gran parte de los trabajos con mayor participación femenina son más difíciles de trasladar a un ambiente en ese estilo, en donde se mencionan casos tales como el área de la salud, los servicios de comercio y los trabajos domésticos.

Kremerman (2021) en su investigación identifica la cantidad de personas en una situación de pobreza al considerar sólo los ingresos laborales y de pensiones contributivas. La medición realizada fue más exigente a la formada en el Ministerio de Desarrollo Social y Familia la cual tiene como función evaluar la suficiencia de los ingresos obtenidos en el trabajo y las pensiones, para medir la cantidad de la población en estado de pobreza, partiendo por el supuesto de que los salarios y las pensiones en Chile son bajas por lo que es esperable que la pobreza calculada sólo con los ingresos laborales y de pensiones sea alta. En sus resultados se confirma la hipótesis, en donde se supera por una gran cantidad a las cifras oficiales de pobreza, pasando en el caso de las mujeres de un 11 % a un 42.2 % y en el caso de los hombres de un 10.8 % a un 39.9 %.

De una manera similar a Fuentes (2005), Salce (2021) realiza un análisis de la Encuesta CASEN en Pandemia 2020 durante el periodo de 1990 hasta el año 2017, esto con el fin de explorar la brecha y la discriminación salarial por género en el país. Para ello, también se realizan distintos métodos de la descomposición Oaxaca-Blinder (1973) para la medición de la brecha salarial y la descomposición de esta, además de la corrección por sesgo de selección. Entre sus resultados, menciona al igual que fuentes (2005) la reducción de la brecha durante el periodo de 1990 hasta el 2003, punto en donde la discriminación se vuelve estable hasta el año 2017. La discriminación salarial en el año 2017 llegó hasta un 49.7 % y que al calcular sin considerar el sesgo por selección subestima la discriminación que ocurre dentro del país. También se menciona que gran parte de la discriminación salarial ocurre en los extremos de menores y mayores ingresos en la distribución. También se menciona que por lo general las mujeres suelen tener mayor experiencia laboral en comparación a los hombres, aún así tienen sueldos menores por razones no aparentes.

Finalmente, Antognini, *et al* (2021) realizó un estudio en el cual se analizan las políticas de protección social realizadas en Chile frente a la crisis sanitaria a causa del Covid-19. En el estudio se plantea que las políticas sociales, especialmente las relacionadas con la protección social han agrandado las brechas de inequidades en la población en distintos ámbitos tales como el económico, por lo que deberían de ser estas políticas las que se tendrían que cambiar para tener un sistema más equitativo en el país.

En el ámbito internacional, Cepeda *et al* (2014) menciona que en los casos de universitarios recientemente titulados con un empleo formal en Colombia sufren una brecha salarial creciente de un 11 % el cual un 40 % es explicado por las diferencias de género como tal en las carreras escogidas y que se puede explicar hasta un 30 % en base al tipo de carrera universitaria a la cual se entra, demostrando la existencia de fenómenos tales como el techo de cristal dentro del país. Este fenómeno del techo de cristal también se observa en países tales como México, pero en una menor medida, ya que durante los periodos de 1990 hasta 2010 se disminuyó la cantidad de la brecha salarial hasta un 6 % durante el año 2010 y también hubieron patrones descendientes de techo de cristal, pese a ello, en el país se encontraron patrones estables del denominado “piso pegajoso”, los cuales debieran de tomarse a mayor consideración en futuras investigaciones (Arceo y Campos, 2014).

Años más tarde, en el caso de países como Argentina, Paz (2019) realizó una investigación de la brecha salarial por género en donde se enfatizaba en la disparidad salarial entre los cónyuges y las personas sin pareja, indagando en la distribución de los salarios y la existencia del techo de cristal y del piso pegajoso en el país, fenómenos que se mantienen observando en los países latinoamericanos. En los resultados se observa que existe una evidencia clara de la brecha como tal a favor de los hombres en comparación a las mujeres con un 32.5 % de ingresos más alto en el caso de los hombres pero con un buen manejo en las políticas públicas este podría bajar a un 13.2 %. También se implica que hasta 59 % de la brecha salarial total puede ser explicada por características observables.

Durante el periodo entre los años 2007 hasta el año 2019 la brecha salarial por género en Ecuador había tenido una disminución bastante notable, la cual era cercana a un 46 % teniendo un promedio de 52 dólares de diferencia, pese a ello se sigue percibiendo una inequidad en los ingresos dentro del mercado laboral, por lo que al formarse estas inequidades a través de factores sociales, económicos, culturales y políticos hay una cantidad de variables que no se pudieron considerarse, pero que debían de ser estudiadas para la disminución de la brecha existente (Sanchez, 2021). Otro estudio enfocado en Ecuador durante el periodo de 2011 hasta 2018 logró detectar una discriminación salarial para el género femenino, en donde algunos factores que apoyan a los menores ingresos y a parte de esta discriminación son la falta de experiencia debido a los patrones definidos tales como la conformación familiar y el cuidado de los menores de cinco años. También se considera una relación positiva entre las brechas salariales y el género de quien está trabajando, siendo en el caso de los hombres un crecimiento de los salarios mientras que con la inserción de las mujeres al mundo laboral estos ingresos decaen, atascando el acceso a un mercado laboral mucho más equitativo (Godoy, 2021).

En la siguiente sección, se mencionarán los recursos utilizados en los análisis de la brecha salarial de jefes de hogar utilizando los datos de la Encuesta CASEN en Pandemia 2020.

1.3. Recursos utilizados

Para todos los análisis realizados y las posteriores conclusiones de estos, se utilizaron las bases de datos de la Encuesta CASEN en Pandemia 2020 (última revisión de la base de datos en Septiembre, 2022), los cuales fueron obtenidas a través de la página del Observatorio Social del Ministerio de Desarrollo Social y Familia, de la cual se obtuvieron las respuestas de los nueve módulos en donde se realiza la caracterización socioeconómica de la población residente de las viviendas particulares de Chile, de la misma manera se obtuvieron los manuales referentes a la creación de la Encuesta CASEN en Pandemia 2020, tales como los manuales del investigador y del diseño muestral.

Los análisis que se realizarán tendrán en consideración la implementación de los factores de expansión calculados para la realización de la Encuesta CASEN en Pandemia 2020 y del plan de muestreo complejo realizado en el instrumento.

Para la realización de los análisis de este proyecto, al igual que los gráficos y materiales de elaboración propia se utilizó principalmente el software estadístico Stata en su versión número 17 de la compañía StataCorp (2021).

En el siguiente capítulo, se mencionará la metodología que se impartió en la realización de la Encuesta CASEN en Pandemia 2020, junto con los métodos y técnicas realizadas en la aplicación de los análisis del instrumento.

Capítulo 2

Metodología

En este capítulo se mencionará la metodología que se impartió en la realización de la Encuesta CASEN en Pandemia 2020, además de los distintos métodos y técnicas que se realizaron a lo largo de este trabajo, siendo la descomposición Oaxaca-Blinder, regresión lineal, análisis de correspondencias (AC) y la variante que denominada como análisis de correspondencias múltiple (ACM).

2.1. Diseño muestral

Toda la información impartida en esta sección fue obtenida de los documentos realizados por el Observatorio Social (2021) del Ministerio de Desarrollo Social y Familia, en donde se utilizaron principalmente los documentos “Diseño Muestral Casen en Pandemia 2020” y “Manual del Investigador Casen en Pandemia 2020”.

La Encuesta CASEN en Pandemia 2020 fue planificada por el Ministerio de Desarrollo Social y Familia a través de su división del Observatorio Social, de la misma manera que las versiones anteriores realizadas del instrumento pero con la diferencia de que en esta ocasión, el diseño muestral pasó a estar a cargo del Instituto Nacional de Estadísticas (INE) y pasándose a llamarse el diseño muestral como: Diseño Muestral Encuesta Socioeconómica Nacional CASEN 2019. Además, el Instituto Nacional de Estadísticas estuvo a cargo del cálculo de los factores de expansión utilizados en el diseño muestral y de las estimaciones de los errores de muestreo bajo un diseño de muestreo complejo.

La Encuesta CASEN en Pandemia 2020 fue diseñada con el fin de cumplir los siguientes objetivos:

- Conocer la situación de pobreza por ingresos de las personas y los hogares, así como la distribución del ingreso de los hogares.
- Identificar carencias de la población en las áreas de educación, salud, vivienda, trabajo e ingresos.
- Evaluar brechas de pobreza por ingresos y carencias entre distintos grupos de la población como niños, niñas y adolescentes; jóvenes; personas mayores; mujeres; pueblos indígenas; migrantes; entre otros.
- Evaluar brechas de pobreza por ingresos y carencias entre zonas urbanas y rurales, y entre territorios.

- Estimar cobertura, focalización y distribución del gasto fiscal de los principales subsidios monetarios de alcance nacional entre los hogares, según su nivel de ingreso, para evaluar el impacto de este gasto en el ingreso de los hogares y en la distribución del mismo.

Como fue mencionado con anterioridad, debido a la situación sanitaria en la que se encontraba el país al momento de la realización del instrumento este pasó de una modalidad totalmente presencial como se mantenía desde su primera realización hasta el año 2017 a una nueva modalidad mixta secuencial dividida en las siguientes etapas:

- Precontacto: contacto de los encuestadores a los hogares seleccionados para participar en la Encuesta CASEN en Pandemia 2020.
- Encuesta de manera telefónica: realización de la encuesta de modalidad telefónica.
- Recuperación presencial: se visitaron los hogares en los cuales no fue posible contactar, en esta ocasión de manera presencial.

Población objetivo

La población objetivo del instrumento corresponde a los residentes de viviendas particulares a lo largo del territorio chileno, en donde se excluyen las áreas de difícil acceso (ADA).

Unidad de información

La unidad de información de la Encuesta CASEN en Pandemia 2020 es el hogar y las personas dentro de cada hogar. Dentro de cada vivienda seleccionada se realizó una identificación de los hogares y las personas miembros de estos hogares. La entrevista es realizada con un informante por hogar con mayoría de edad y a través de él se recolectan los datos de los demás miembros del hogar que son residentes habituales de la vivienda. Este informante es denominado como informante idóneo el cual debe ser el jefe de hogar, en el caso de la ausencia del jefe de hogar debe de realizar la encuesta un miembro del hogar mayor a los 18 años.

2.1.1. Marco muestral

Para la creación del marco muestral, se utilizaron las actualizaciones realizadas a finales del año 2018 al marco de secciones 2002 como al marco de manzanas con foco CASEN 2015, esto realizado a partir de las viviendas totales obtenidas en el Precenso 2016. Estas actualizaciones fueron denominadas como Marco Muestral de Manzanas (MMM2016) para el área urbana y Marco Muestral de Secciones (MMS2016) en el caso de las áreas rurales y el resto del área urbana (RAU).

El marco está compuesto por los conglomerados, los cuales son correspondientes a los conjuntos de las viviendas agrupadas y que están contenidas en las áreas geográficas. Estos conglomerados son definidos como segmentos, y más específicamente en secciones y manzanas, siendo correspondiente de las unidades primarias de muestreo (UPM). A través de los conglomerados se accede a la vivienda y al hogar, conformados por las personas que son el objetivo de estudio.

Dominios de estudio

Los dominios de estudio son coincidentes con los dominios definidos en las tres versiones anteriores de la CASEN (correspondientes a los años 2011, 2013 y 2017), por lo que la muestra continúa teniendo presencia en las 324 comunas que en las versiones anteriores del instrumento.

Objetivos de precisión y tamaño de muestra

Los criterios de precisión considerados para la Encuesta CASEN en Pandemia 2020 se mantienen de la CASEN en su versión anterior del año 2017, en donde los errores absolutos no deben superar los 4 puntos porcentuales y los errores relativos no deben superar los 30 puntos porcentuales.

El tamaño de muestra objetivo en el caso de la Encuesta CASEN en Pandemia 2020 se estableció con un mínimo por comuna de 50 viviendas en el área urbana y 30 viviendas en el área rural.

Estrategia muestral

El diseño correspondiente a la Encuesta CASEN en Pandemia 2020 es definido como probabilístico, estratificado y bietápico, teniendo los estratos conformados por la dupla de Comuna-Área. En el caso de las áreas urbanas las manzanas fueron clasificadas según su tamaño, donde se crearon cinco estratos por tamaño, por lo que los estratos de áreas urbanas estuvieron conformados por la terna Comuna-Área-Grupo.

Selección de unidades primarias y secundarias de muestreo

En el marco de manzanas se realizó una selección sistemática de igual probabilidad, a nivel Comuna-Área-Grupo de tamaño. En cuanto al marco de secciones, se realizó una selección sistemática de probabilidad proporcional al tamaño en cuanto a número de viviendas, a nivel Comuna-Área. En el caso de las viviendas, se realizó una selección sistemática y con igual probabilidad al interior de cada uno de los conglomerados.

2.1.2. Cuestionario

A causa de la situación sanitaria del país debido a la pandemia del Covid-19 y del cambio de modalidad que surgió en la implementación de este instrumento, en el cuestionario de la Encuesta CASEN en Pandemia 2020 se considerará una reducción de la cantidad total de preguntas, priorizando las preguntas que tuviesen relación con los ingresos para realizar la medición de pobreza por ingresos.

Como fue mencionado anteriormente, el informante idóneo correspondiente al jefe de hogar es a quien se le solicita la información de todas las personas residentes del hogar en cuestión el cual debe de cumplir con la mayoría de edad.

Los módulos que componen la Encuesta CASEN en Pandemia 2020 son los siguientes:

1. Registro de Residentes, módulo en el cual se registra la información referente a la identificación de las distintas personas del hogar, al igual que la identificación de núcleos familiares de las viviendas.
2. Educación, módulo que incluye preguntas que sirven para la estimación de indicadores referentes a la escolaridad de la población y los casos en los cuales la población no fue incorporada al sistema educativo.
3. Trabajo, módulo que incluye preguntas que recogen información referente a la estimación de indicadores sobre la situación ocupacional de la población y caracterización de la situación laboral y previsional de la fuerza de trabajo ocupada.

4. Ingresos, módulo que incluye preguntas de información sobre las distintas formas de ingreso que reciben en las personas y hogares, ya sea como participación activa o como receptor de transferencias.
5. Identidades, módulo que incluye preguntas con respecto a la pertenencia hacia los pueblos originarios o indígenas y de migración internacional.
6. Vivienda, módulo que incluye preguntas para la estimación de estimadores sobre las características básicas en las viviendas del país al igual que condiciones de habitabilidad de los hogares.

En la siguiente subsección se mencionarán los factores de expansión que fueron calculados para la Encuesta CASEN en Pandemia 2020.

2.1.3. Factores de expansión

El análisis realizado en la Encuesta CASEN en Pandemia 2020 utiliza los ponderadores conocidos como los factores de expansión, los cuales son ponderadores útiles para la estimación de los parámetros asociados a las variables de interés del estudio, para que estas variables tengan validez sobre la población objetivo.

En la realización de cada CASEN se han utilizado dos factores de expansión por observación, un factor que expande a la población regional (expr) y otro que expande a la población comunal (expc). En esta versión se agrega un factor provincial (expp) que se asume válido para todas las provincias que no sean ADA (las provincias de Isla de Pascua, Palena y Antártica Chilena son definidas como ADA) y que todas las comunas estén presentes en el marco muestral del INE o cuyas comunas faltantes no representen más del 10% de la población total de la provincia.

Entre las 56 Provincias del país, en la Encuesta CASEN en Pandemia 2020 se utilizaron factores de expansión provinciales para tres casos, siendo las Provincias de Coyhaique, Nagakkabes y Última Esperanza, siendo las Provincias no consideradas dentro de los factores de expansión las catalogadas como ADA y las Provincias de Parinacota, Capitán Prat y Tierra del Fuego debido al incumplimiento de los requisitos de los factores de expansión o estar excluidas del marco del INE.

El cálculo de los factores de expansión consiste en la obtención secuencial de cuatro ponderadores:

- Ponderador de selección, refleja cuantas viviendas en la población representa la vivienda selecciona en la muestra.
- Ponderador corregido, corrige las fallas que llevan a la inclusión de las viviendas no elegibles en la muestra.
- Ponderador corregido por no respuesta, reduce el sesgo referente a la falla en conseguir la cooperación de los hogares seleccionados en la muestra.
- Ponderador de calibración, cuadra las estimaciones poblacionales mediante la corrección de fallas en la cobertura de subpoblaciones de interés.

Los ajustes de carácter “comunal” se establecen como la razón entre la totalidad de personas en la comuna según las estimaciones demográficas del INE y el número total de personas residentes en las viviendas particulares estimadas por comuna a partir de la encuesta.

Los ajustes de carácter “provincial” se establecen como la razón entre la totalidad de personas en el grupo poblacional según las estimaciones demográficas del INE y el número total de personas residentes en las viviendas particulares para cada provincia a partir de la encuesta.

Los ajustes de carácter “regional” se realizan de una manera distinta a los dos anteriores, realizando una separación entre las áreas urbanas y rurales a través de la razón entre la población de viviendas particulares de las estimaciones poblacionales realizadas por el Subdepartamento de Demografía y Estadísticas Vitales del Instituto de Estadísticas y la totalidad de las personas estimadas a partir de la encuesta.

La metodología detrás de estas estimaciones poblacionales utilizadas para los ajustes regionales, provinciales y comunales, fueron desarrolladas en base a la información del Censo de Población y Vivienda 2017.

En la siguiente sección, se presentará del método de descomposición Oaxaca-Blinder.

2.2. Descomposición Oaxaca-Blinder

Esta sección utiliza información obtenida principalmente del trabajo de Jann (2008) sobre la implementación de la descomposición Oaxaca-Blinder en el software Stata.

Generalmente la metodología utilizada para el estudio de los resultados del mercado laboral a través de grupos es la descomposición de las diferencias de las medias en los salarios logarítmicos basados en los modelos de regresión de una manera contrafactual. Este procedimiento es conocido como la descomposición Oaxaca-Blinder, método que es utilizado generalmente para los estudios relacionados al mercado laboral por grupos, siendo en el caso de este proyecto por sexo. Este método divide el diferencial de los salarios entre dos grupos en donde la primera parte es ‘explicada’ por las diferencias de los grupos a través de las características en la productividad, características tales como la experiencia o la educación y un error estimado que no puede ser explicado con las características anteriores. Esta parte no explicada o error es utilizada generalmente como una medida de discriminación, pero también incluye a los efectos de las diferencias de los grupos en los predictores no observados. Si bien, generalmente es utilizada esta técnica para el área del mercado laboral, puede ser utilizada para el estudio de diferencias de grupo en cualquier variable de resultado.

Dados dos grupos A y B, una variable de resultado Y correspondiente a los ingresos, y un conjunto de predictores correspondientes a las variables.

$$R = E(Y_A) - E(Y_B) \tag{2.1}$$

En donde $E(Y)$ denota el valor esperado de los ingresos, contabilizada por la diferencia de grupos en los predictores.

Basado en el modelo lineal,

$$Y_\gamma = \mathbf{X}_\gamma^\top \beta_\gamma + \epsilon_\gamma \quad , \quad E(\epsilon_\gamma) = 0 \quad , \quad \gamma \in A, B \tag{2.2}$$

donde \mathbf{X} es un vector que contiene a los predictores y una constante, β contiene los parámetros de la pendiente y el intercepto, y ϵ es el error, la diferencia de la media de salida puede ser expresada como la diferencia en una predicción lineal de las medias específicas de grupo de los regresores. Esto es:

$$R = E(Y_A) - E(Y_B) = E(\mathbf{X}_A)^\top \beta_A - E(\mathbf{X}_B)^\top \beta_B \quad , \quad (2.3)$$

Ya que,

$$E(Y_\gamma) = E(\mathbf{X}_\gamma^\top \beta_\gamma + \epsilon_\gamma) = E(\mathbf{X}_\gamma^\top \beta_\gamma) + E(\epsilon_\gamma) = E(\mathbf{X}_\gamma)^\top \beta_\gamma \quad ,$$

Con $E(\beta_\gamma) = \beta_\gamma$ y $E(\epsilon_\gamma) = 0$ por suposición.

Para identificar la contribución de las distintas diferencias de grupo en los predictores de la diferencia del resultado general, la ecuación 2.3 se puede escribir de la siguiente manera:

$$R = [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top \beta_B + E(\mathbf{X}_B)^\top (\beta_A - \beta_B) + [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top (\beta_A - \beta_B) \quad . \quad (2.4)$$

Esta es una descomposición “triple”, es decir, que la diferencia de salida es dividida en tres partes:

$$R = E + C + I$$

El primer sumando de la ecuación 2.5 equivale a la parte de los diferenciales que se debe a las diferencias de grupo en los predictores.

$$E = [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top \beta_B \quad . \quad (2.5)$$

El segundo componente de la ecuación 2.6 mide la contribución de las diferencias de las diferencias en los coeficientes.

$$C = E(\mathbf{X}_B)^\top (\beta_A - \beta_B) \quad . \quad (2.6)$$

El último sumando de la ecuación 2.7 es un término de interacción que explica el hecho de que las diferencias en dotaciones y coeficientes existen simultáneamente entre los dos grupos.

$$I = [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top (\beta_A - \beta_B) \quad . \quad (2.7)$$

La descomposición de la ecuación 2.4 es formulada del punto de vista del grupo B. Es decir, las diferencias de grupo en los predictores están ponderadas por los coeficientes del grupo B para determinar el efecto de dotaciones (E). En otras palabras, el componente E mide los cambios esperados en las medias del grupo B, si es que el grupo B tuviese los niveles de predicción del grupo A. De manera similar, para el segundo componente (C), las diferencias en los coeficientes están ponderadas por los niveles de predicción del grupo B. Entonces, el segundo componente mide el cambio esperado de la media en el grupo B, si es que el grupo B tiene los coeficientes del grupo A.

Los diferenciales pueden ser expresados de una manera análoga desde el punto de vista del grupo A, dando la descomposición triple inversa

$$R = [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top \beta_A + E(\mathbf{X}_A)^\top (\beta_A - \beta_B) - [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top (\beta_A - \beta_B) \quad . \quad (2.8)$$

Ahora el “efecto de dotaciones” de la ecuación 2.5 equivale al cambio esperado del resultado de la media del grupo A, si es que el grupo A hubiese tenido los niveles de predictores del grupo B. El “efecto de los coeficientes” cuantifica el cambio esperado en el resultado de la media del grupo A, si el grupo A tuviese los coeficientes del grupo B.

Una descomposición alternativa que es prominente en la literatura de la discriminación resultante del concepto que existen algunos vectores de coeficientes no discriminatorios que debiesen de ser utilizados para determinar la contribución de las diferencias en los predictores. Dejando β^* ser uno de esos vectores de coeficiente no discriminatorios. La diferencia resultante puede ser escrito como:

$$R = [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top \beta^* + [E(\mathbf{X}_A)^\top (\beta_A - \beta_B) + E(\mathbf{X}_B)^\top (\beta^* - \beta_B)] \quad . \quad (2.9)$$

Se considera una descomposición doble

$$R = Q + U \quad ;$$

En donde el primer componente de la ecuación 2.9 es la parte explicada por las diferencias de grupo en los predictores (El “efecto de cantidad”)

$$Q = [E(\mathbf{X}_A) - E(\mathbf{X}_B)]^\top \beta^* \quad . \quad (2.10)$$

El segundo componente es la parte no explicada, lo cual generalmente es atribuido como discriminación pero hay que considerar que también captura todos los efectos potenciales en las diferencias de las variables no observadas.

$$U = E(\mathbf{X}_A)^\top (\beta_A - \beta^*) + E(\mathbf{X}_B)^\top (\beta - \beta_B) \quad . \quad (2.11)$$

La parte no explicada (ecuación 2.11) a veces se suele descomponer más aún. Sea $\beta_A = \beta^* + \delta_A$ y $\beta_B = \beta^* + \delta_B$ con δ_A y δ_B como vectores de parámetros fijos de discriminación. Entonces, U puede ser expresado como

$$U = E(\mathbf{X}_A)^\top \delta_A - E(\mathbf{X}_B)^\top \delta_B \quad . \quad (2.12)$$

Luego, el segundo componente del diferencial puede ser subdividido en dos partes, siendo la primera

$$U_A = E(\mathbf{X}_A)^\top \delta_A \quad ,$$

que mide discriminación en favor al grupo A y una segunda parte

$$U_B = E(\mathbf{X}_B)^\top \delta_B \quad ,$$

la cual cuantifica la discriminación sobre el grupo B. Esto es, realizando la suposición de que no existen valores relevantes sin observar en los datos.

Sea $\hat{\beta}_A$ y $\hat{\beta}_B$ las estimaciones de mínimos cuadrados para β_A y β_B , obtenido por separado de las muestras específicas por grupo y utilizando \bar{X}_A y \bar{X}_B como estimaciones para $E(\mathbf{X}_A)$ y $E(\mathbf{X}_B)$.

Basado en las estimaciones de las descomposiciones, se puede mencionar a \hat{R} como:

$$\hat{R} = \bar{Y}_A - \bar{Y}_B = (\bar{X}_A - \bar{X}_B)^\top \hat{\beta}_B + \bar{X}_B^\top (\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}_A - \bar{X}_B)^\top (\hat{\beta}_A - \hat{\beta}_B) \quad , \quad (2.13)$$

y

$$\hat{R} = \bar{Y}_A - \bar{Y}_B = (\bar{X}_A - \bar{X}_B)^\top \hat{\beta}_A + \bar{X}_A^\top (\hat{\beta}_A - \hat{\beta}_B) - (\bar{X}_A - \bar{X}_B)^\top (\hat{\beta}_A - \hat{\beta}_B) \quad . \quad (2.14)$$

La determinación de componentes de la descomposición doble (ecuación 2.9) es más complicada, por la estimación del vector de coeficientes no discriminatorios desconocidos β^* . Según Reimers (1983) se propone que se utilice los coeficientes promedio para ambos grupos como un estimado de un vector de parámetros no discriminatorios, lo que sería:

$$\hat{\beta}^* = 0,5\hat{\beta}_A + 0,5\hat{\beta}_B \quad . \quad (2.15)$$

De manera similar, Cotton (1988) sugiere que para la ponderación de los coeficientes por tamaño de grupos n_A y n_B , siendo

$$\hat{\beta}^* = \frac{n_A}{n_A + n_B} \hat{\beta}_A + \frac{n_B}{n_A + n_B} \hat{\beta}_B \quad .$$

Además, Neumark (1988) menciona una preferencia por el uso de coeficientes de una regresión conjunta sobre ambos grupos como un estimado de β^* .

Para evadir la distorsión de los resultados de la descomposición debido a las diferencias de grupo desbordándose sobre los parámetros de la pendiente del modelo conjunto, es recomendable incluir en el modelo conjunto un indicador de grupo como una covariable adicional.

En la siguiente sección, se mencionará una breve explicación del método de regresión lineal.

2.3. Regresión lineal

La mayor parte de la información utilizada en esta sección fue obtenida del libro “*Linear regression*” de Gross y Groß (2003).

El método de regresión lineal es un enfoque lineal en la modelización de la relación entre una respuesta escalar y una o más variables explicativas.

Si se tiene un conjunto de datos $\{y_i, x_{i1}, \dots, x_{ip}\}$ de n unidades, un modelo de regresión lineal asume que la relación entre la variable dependiente y_i y el p -vector de regresores x_i es lineal. Esta relación es modelada a través de un término de perturbación o variable de error ϵ_i , la cual es una variable aleatoria no observada que añade ruido a la relación lineal entre la variable dependiente y los regresores. De esta manera, el modelo adopta la siguiente forma:

$$Y = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i = x_i^\top \beta + \epsilon_i \quad , \quad i = 1, \dots, n,$$

donde $^\top$ denota la transpuesta y $x_i^\top \beta$ es el producto interno entre los vectores x_i y β .

En la siguiente sección se hablará del análisis de correspondencias.

2.4. Análisis de correspondencias

La mayor parte de la información que se encuentra en esta sección fue obtenida del libro “Análisis de datos multivariantes” de Peña (2002), en donde se basó para obtener la información del análisis de correspondencias.

El análisis de correspondencias es una técnica descriptiva en la representación de tablas de contingencias (también conocidas como tablas cruzadas), en las cuales se recogen datos de frecuencias entre dos variables cualitativas en un conjunto de datos para encontrar las interacciones entre ellas. El análisis de correspondencias cumple una función equivalente al análisis de componentes principales (ACP) en el caso de las variables cualitativas. Esta técnica inicia con una matriz de dimensiones $I \times J$, las cuales representan frecuencias absolutas que están observadas de dos variables cualitativas en n elementos. Generalmente, con las tablas de contingencia se tiene un conjunto de números positivos dispuestos en una matriz, en donde cada casilla es representada como la frecuencia absoluta observada para la combinación de dos variables.

El análisis de correspondencias puede interpretarse de dos maneras, entre las cuales la primera, es una forma de representación de las variables en una dimensión menor. De manera análoga al análisis de componentes principales, pero también explicando la distancia entre los puntos de manera coherente con la interpretación de datos acorde a esta técnica y utilizando la distancia ji-cuadrado. La segunda manera está mucho más cercana al escalado multidimensional, en donde es un procedimiento con el objetivo de asignar valores numéricos a las variables cualitativas.

Tablas de contingencia

Una tabla de contingencia generalmente es construida de la manera mostrada en la figura 2.1. Para los individuos n , los valores están disponibles en dos categorías, siendo estas V_1 (con I categorías) y V_2 (con J categorías). Las tablas de contingencia llevan el término general x_{ij} como el número de individuos figurados en la categoría i de V_1 y j de V_2 .

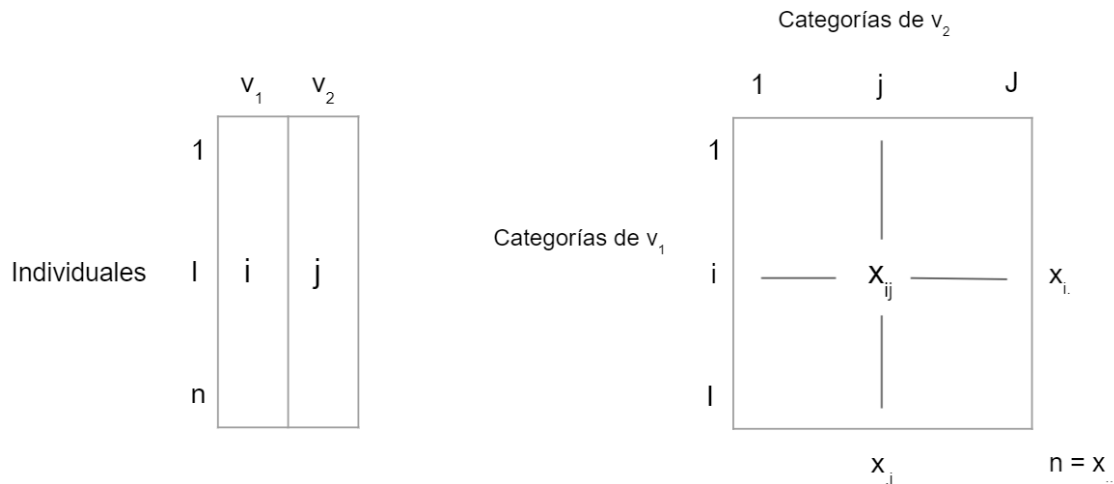


Figura 2.1: Tabla de contingencia.

Fuente: Elaboración propia.

Entonces, para una tabla de contingencia con dos variables categóricas V_1 y V_2 definidas por n individuos; individuo l lleva las categorías i y j consideradas x_{ij}

Las sumas de la tabla son denominadas como márgenes y son denotadas al sustituir el índice en x_{ij} por sobre el cálculo realizado por un punto. Además, en el análisis de correspondencias se consideran las tablas de probabilidad asociadas a las tablas de contingencia como el término general $f_{ij} = x_{ij}/n$, la probabilidad de llevar ambas categorías de i y j .

Uno de los principios fundamentales en la construcción de tablas de contingencia implica el estudio de las relaciones entre las respuestas de las dos preguntas involucradas en la construcción de la tabla, las cuales por lo general se supone o se tienen indicios que sí se tiene una relación entre ambas, ya que al no encontrar una relación entre los datos podría poner en duda la calidad de la información como tal.

El análisis de la relación que se encuentra entre las dos variables requiere que se sitúen los datos en función de un punto de partida determinado; en este caso, la ausencia de relación.

El estudio de la relación entre dos variables requiere que se sitúen los datos en función a un punto de partida determinado; en este caso, la ausencia de relación. El modelo de independencia especifica este criterio como el punto de partida, siendo la relación estándar de la independencia entre dos eventos ($P[A \cap B] = P[A]P[B]$) aplicable para ambas variables categóricas. Dos variables categóricas se consideran independientes si es que:

$$\forall i, j \quad f_{ij} = f_{i\bullet}f_{\bullet j} \quad (2.16)$$

Entonces, la probabilidad conjunta (f_{ij}) es dependiente de las probabilidades marginales.

Distancias

Para el análisis de una tabla de contingencia, las I filas pueden considerarse como I puntos del espacio \mathbb{R}^J , por lo que se busca una representación de estos puntos en un espacio de dimensiones menor para una mejor interpretación de las distancias relativas. De manera similar al análisis de componentes principales, pero considerando las características distintas de estos datos, los cuales son provenientes de diferencias entre las frecuencias relativas de las filas, lo que implica:

1. Las filas entre sí no tienen el mismo peso, debido a la diferencia de datos entre estas.
2. La distancia euclidiana entre los puntos no se considera una buena medida de las diferencias reales entre las filas, por lo que debe de modificarse utilizando la distancia ji-cuadrado, ya que esta tiene en consideración las distintas precisiones de las coordenadas.

En el caso de las columnas, se puede otorgar un análisis equivalente al análisis de las filas pero considerándolas puntos en un espacio \mathbb{R}^I .

En la siguiente sección, se hablará de la técnica de análisis de correspondencia múltiple que se empleará en el proceso de análisis de la Encuesta CASEN en Pandemia 2020.

2.5. Análisis de correspondencias múltiple

La información y los datos que se mencionan en esta sección son obtenidos del libro “*Exploratory Multivariate Analysis by Example Using R*” escrito por Husson *et al.* (2011).

El análisis de correspondencia múltiple es una forma más específica de aplicar el análisis de correspondencias a tablas de contingencia, en donde se permite el análisis de los patrones de relaciones en variables dependientes categóricas. Como es de esperar, si bien los datos obtenidos suelen ser similares a los del análisis de correspondencia habitual, estos deben de ser corregidos y adaptados a las nuevas interpretaciones de distancias de esta técnica.

Esta técnica es generalmente utilizada para el estudio de variables seleccionadas en encuestas, en donde cada pregunta resulta ser una variable y sus respuestas corresponden a las categorías de la variable.

Estudio de individuos, variables y categorías

Como los datos son estudiados en base a los individuos, las variables y las categorías hay que considerar ciertos aspectos:

- En el caso de los individuos se quiere conocer las similitudes entre estos en términos de las variables, es decir, cosas tales como conocer las similitudes entre los individuos o la consistencia entre estas similitudes. Por ello, los individuos se comparan en términos de presencia o ausencia de categorías que se han seleccionado, entonces, la distancia entre dos individuos depende solamente de sus características y no la de otros individuos. A pesar de ello, es importante el conocer las características de los otros individuos al calcular aquella distancia, siendo comparados de categoría por categoría.
- En el caso de las variables, de una manera similar al análisis de componentes principales en donde se resumen las relaciones entre estas variables, estas pueden ser estudiadas en pares o en su totalidad, en donde la información puede ser estudiada en base a sus categorías.
- Para el estudio de las distancias entre las categorías (considerando k y k' categorías), se debe de contar los individuos que se encuentran en ambas categorías y estas suelen estar mucho más alejadas mientras menos individuos se tengan en común. Es de suma importancia conocer que el tamaño de cada grupo de individuos al calcular la distancia de las categorías.

En la siguiente subsección se hablará de las distancias entre los individuos y las distancias entre categorías.

Distancias entre individuos y categorías

De una tabla de individuos y categorías se puede crear una matriz indicadora con individuos en las filas y categorías para cada variable en las columnas, en donde se tiene un valor de 1 si los individuos i llevan la categoría k y 0 si es que no, teniendo $I \times K$ dimensiones (siendo $K = \sum_{j=1}^J K_j$).

Utilizando la matriz indicadora mencionada anteriormente, las distancias entre los individuos pueden ser calculadas añadiendo las diferencias entre las categorías, $(x_{ik} - x_{i'k})^2$, contrarrestadas mediante una función inversamente proporcional a I_k . Esto es expresado como:

$$d_{i,i'}^2 = C \sum_{k=1}^K \frac{(x_{ik} - x_{i'k})^2}{I_k} \quad , \quad (2.17)$$

en donde C es una constante.

De manera similar, la distancia entre dos categorías k y k' es calculada contando los individuos los cuales llevan categoría k k' (como $I_{k \neq k'}$ y contrarrestando utilizando una función inversamente proporcional a I_k y $I_{k'}$. Esto es expresado como:

$$d_{k,k'}^2 = C^\top \frac{I_{k \neq k'}}{I_k I_{k'}} \quad . \quad (2.18)$$

en donde C es una constante. Sin embargo, según la codificación ($x_{ik} = 0 \cap 1$), el número de individuos que llevan una o más categorías es igual a $I_{k \neq k'} = \sum_{i=1}^I (x_{ik} - x_{ik'})^2$. Entonces,

$$d_{k,k'}^2 = C' \frac{1}{I_k I_{k'}} \sum_{i=1}^I (x_{ik} - x_{ik'})^2 \quad .$$

Desarrollando la ecuación y utilizando propiedades de codificación:

$$d_{k,k'}^2 = C^\top \left(\frac{1}{I_k} \frac{1}{I_{k'}} - 2 \frac{\sum_{i=1}^I x_{ik} x_{ik'}}{I_k I_{k'}} \right) \quad .$$

Sin embargo,

$$\frac{1}{I_k} = \frac{I_k}{I_k^2} = \frac{\sum_{i=1}^I x_{ik} x_{ik'}}{I_k^2} \quad .$$

La distancia al cuadrado entre las categorías pueden ser expresadas como:

$$d_{k,k'}^2 = C' \left(\frac{\sum_{i=1}^I x_{ik'}^2}{I_{k'}^2} + \frac{\sum_{i=1}^I x_{ik}^2}{I_k^2} - 2 \frac{\sum_{i=1}^I x_{ik} x_{ik'}}{I_k I_{k'}} \right) \quad ,$$

donde

$$d_{k,k'}^2 = C' \sum_{i=1}^I \left(\frac{x_{ik}}{I_k} - \frac{x_{ik'}}{I_{k'}} \right)^2 \quad . \quad (2.19)$$

En la siguiente subsección se mencionarán las nubes de individuos y las nubes de categorías.

Nubes de individuos y categorías

Al crear la nube de individuos, se utiliza el mismo enfoque que se utiliza para el análisis de componentes principales, en donde se maximiza la inercia de la nube de individuos proyectados en una serie de ejes ortogonales. De manera similar al análisis de componentes principales, las dimensiones pueden ser interpretadas usando los individuos.

Al calcular los coeficientes de relación entre las coordenadas de los individuos de los componentes y las variables categóricas, si la relación de la variable j y el componente s es cercano a 1, los individuos con la misma categoría tendrán coordenadas similares para el componente s .

Las categorías pueden ser representadas en el baricentro de los individuos en aquellas categorías, de manera similar a las variables suplementarias en el análisis de componentes principales. La representación es optimizada debido a que corresponde a la representación obtenida al maximizar la inercia de la nube de categorías en una secuencia de ejes ortogonales por transformaciones homotéticas.

En la siguiente subsección se mencionará algunos componentes que se tienen que tener en cuenta en la interpretación de los datos en los análisis de correspondencias múltiple.

Interpretación de los datos

En esta subsección se hablarán de algunos componentes relevantes en la interpretación de los datos tales como el porcentaje de inercia, la calidad de un individuo o categoría y los elementos suplementarios.

Porcentaje de inercia

Los porcentajes de inercia asociados a los componentes en este tipo de análisis son calculados de la misma manera que en el método de los componentes principales, y si bien el método de obtención es el mismo, los porcentajes de inercia asociados suelen ser mucho menores a los obtenidos en el análisis de componentes principales debido a que en este tipo de análisis se estudian relaciones mucho más generales y se requiere al menos de un mínimo de $(K_j, K_l) - 1$ dimensiones para poder representar la relación entre dos variables, cada una por la cual se tiene K_j y K_l categorías, respectivamente, a diferencia del ACP en donde sólo se estudian las relaciones lineales, por lo que un solo componente debería de ser suficiente para la representación de las variables si están correlacionadas. Es por esto que se suele estudiar una mayor cantidad de dimensiones en el análisis de correspondencias múltiple que por análisis de componentes principales.

Calidad de un individuo o categoría

Los cálculos e interpretaciones de las contribuciones y la calidad de representación de los individuos y categorías se mantiene igual al análisis de correspondencias. Pero, debido a las escalas de los conjuntos de datos, la calidad de representación suele ser algo más débil si es que esto se compara con la calidad de representación del análisis de correspondencia tradicional o del análisis de componentes principales.

La mayor escala de los conjuntos de datos no afecta a su contribución, debido a que esto se calcula para cada componente por separado. La contribución de una variable categórica dividida por $J\lambda_s$ es igual a la correlación entre los componentes principales y la variable categórica. Lo que en análisis de componentes principales se considera el componente principal se considera de la misma manera en el análisis de correspondencias múltiple.

Elementos suplementarios

Estos elementos pueden ser individuos o variables categóricas o cuantitativas de la misma manera que en análisis de componentes principales.

Para un individual suplementario i' y una categoría suplementaria k' , la fórmula de transición es expresada como:

$$F_s(i') = \frac{1}{\sqrt{\lambda_s}} \sum_{j=1}^J \sum_{k=1}^{K_j} \frac{x_{i'k}}{J} G_s(k) \quad , \quad (2.20)$$

$$G_s(k') = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^I \frac{x_{ik'}}{I'_k} F_s(i) \quad . \quad (2.21)$$

Estas fórmulas de transiciones son idénticas a los elementos activos.

Las variables cuantitativas suplementarias se representan de igual manera que en el análisis de componentes principales, esto es en un círculo de correlación entre la variable y los componentes principales.

En la siguiente sección se hablará de la obtención de datos para muestras complejas y de los factores de expansión utilizados en los análisis.

2.6. Integración del survey data y factores de expansión

Para esta sección, se utilizó principalmente información del documento “Manual del Investigador Casen en Pandemia 2020” otorgado por el Observatorio Social del Ministerio de Desarrollo Social y Familia (2021).

En los distintos análisis para obtener y representar con total fidelidad las estimaciones a partir de información recolectada a través de encuestas con un diseño muestral complejo como lo es en la Encuesta CASEN en Pandemia 2020, se deben considerar las características del diseño muestral lo que permite una estimación adecuada de la varianza de los estimadores.

El software estadístico Stata posee algoritmos que permiten realizar las estimaciones de la varianza bajo un muestreo complejo, en donde por análisis Stata asume el diseño muestral que se defina por el analista. Estas estimaciones bajo un muestreo complejo se realizan a través del comando *svyset*.

Debido a que la información de la Encuesta CASEN es de carácter muestral complejo, es necesario que se apliquen factores de expansión o de ponderación, los cuales asignan un peso distinto a los elementos según la probabilidad de que estos sean seleccionados. Al considerar estas ponderaciones se puede realizar inferencias estadísticas correctas sobre los indicadores que buscan representar las características del universo considerado en los análisis.

Para la realización de los análisis que se observarán en el capítulo de los resultados, estos se realizaron bajo la presencia de las características del diseño muestral complejo utilizado en la realización de la Encuesta CASEN 2020, además del uso de factores de expansión regionales puesto a que este ponderador es pertinente para la obtención de resultados de personas, hogares y núcleos expandidos que buscan ser representativos a nivel nacional, regional y por zona.

En el siguiente capítulo, se presentan los datos obtenidos en la aplicación de los métodos en la Encuesta CASEN en Pandemia 2020 además de los distintos análisis realizados.

Capítulo 3

Resultados

En este capítulo se presenta inicialmente los recursos necesarios para realizar el análisis descriptivo, esto seguido del análisis descriptivo de la Encuesta CASEN en Pandemia 2020 y la aplicación de las metodologías utilizadas tales como la descomposición Oaxaca-Blinder y la técnica de análisis de correspondencias múltiple en los datos de la CASEN en Pandemia 2020, considerando sus análisis y las conclusiones de estos.

3.1. Análisis descriptivo

Inicialmente se realizó una búsqueda de las variables que pudiesen ser factibles para el análisis de los resultados, para ello se observó en los distintos módulos las características que se creen factibles o relevantes en la realización de los análisis (sea Oaxaca-Blinder como análisis por correspondencia múltiple) y en algunos casos se modificaron estas variables para realizar más factible la observación de estas.

En el primer módulo considerado de la Encuesta CASEN en Pandemia 2020 (correspondiente al Módulo H: Registro de residentes) se pueden obtener varios datos relevantes con respecto a la caracterización de las personas, entre los que se pueden destacar:

- Sexo, sexo de las personas del hogar.
- Edad, edad de las personas del hogar.
- Numper, cantidad de personas que viven dentro del hogar.
- Relación con el jefe de hogar, ocupación de la persona dentro del núcleo familiar y su relación con el jefe de hogar.
- Estado civil, estado civil de las personas del hogar.

Entre las variables o características de los módulos que no sean correspondientes al primer módulo (Registro de Residentes) que se consideraron más relevantes para analizar se encuentran:

- Ingresos totales, variable que indica la suma total de ingresos per cápita de los hogares, el cual está dividido por el total de hogares.

- Nivel educativo, variable que indica el mayor nivel educativo adquirido por la persona, para mayor facilidad en la manipulación de los datos se utilizaron cuatro distintas categorías, siendo estas: sin educación, educación básica completa, educación media completa y educación superior completa.
- Tipo de empleo, variable que indica la clase de empleo en la cual se encuentra la persona, para mayor facilidad se disminuyeron las categorías de utilizadas en la Encuesta CASEN quedando en seis categorías distintas, las cuales son: patrón o empleador, trabajador independiente, empleado público, empleado privado, trabajador doméstico, FF. AA. y del Orden.
- Pertenencia o ascendencia indígena, variable que indica si la persona pertenece o es descendiente de alguno de los diez pueblos indígenas que son reconocidos por ley.

Estas variables se consideraron de vital importancia para la ejecución del análisis de correspondencias múltiple y la descomposición Oaxaca-Blinder que se utilizarán más adelante, en los cuales para algunos casos debieron de modificarse para poder obtener mejores resultados dentro de los métodos utilizados.

Chile es un país con una población relativamente pequeña y predominantemente femenina, esto se puede observar ya que de las 19.532.480 personas que conforman al país existen 10.526.345 casos del sexo femenino que corresponden a un 53.89% de la población, los hombres en comparación tienen una frecuencia de 9.006.135 casos equivalente a un 46.11%.

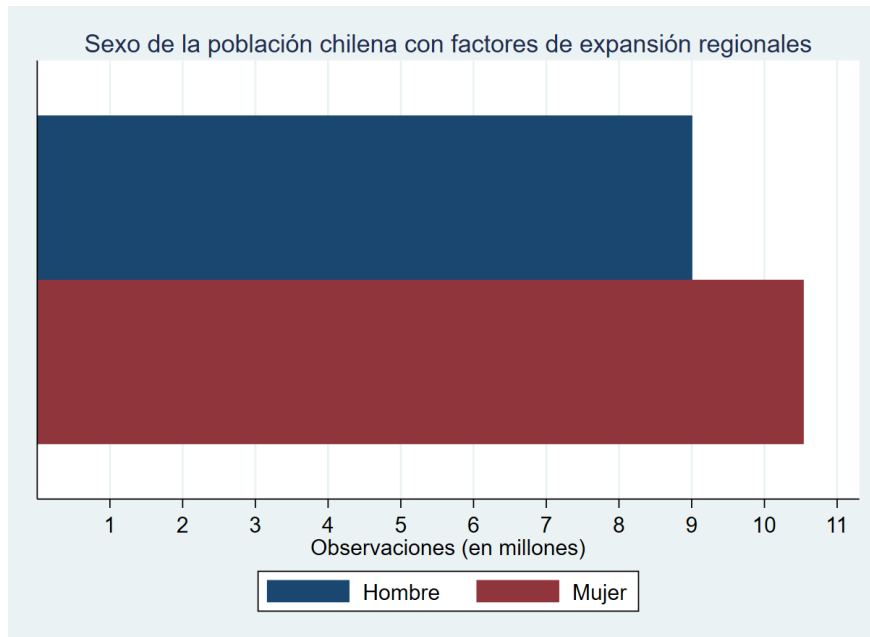


Figura 3.1: Gráfico de sexo de la población chilena con factores de expansión regionales.

Fuente: Elaboración propia.

Considerando a la población que es considerada como jefe de hogar quienes son las que se está interesado en analizar, se tiene una cantidad total de 8.050.009 casos equivalente a un 41.21% de la población, en donde se tiene una mayor cantidad de jefes de hogar de sexo femenino con una

cantidad de 4.215.119 observaciones equivalente a un 52.36% de la población de jefes de hogar en comparación a los 3.834.890 casos de los hombres (47.64%).

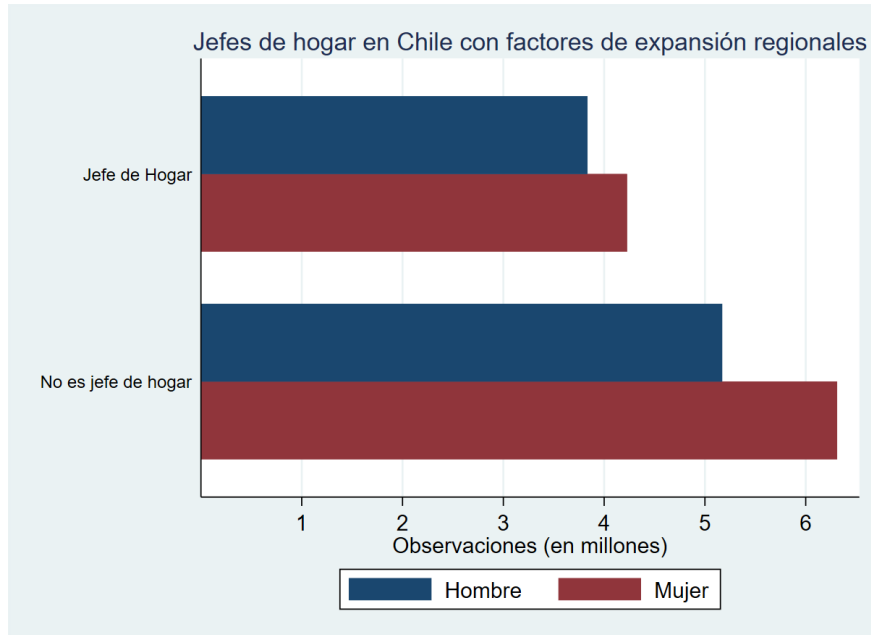


Figura 3.2: Gráfico de jefes de hogar.

Fuente: Elaboración propia.

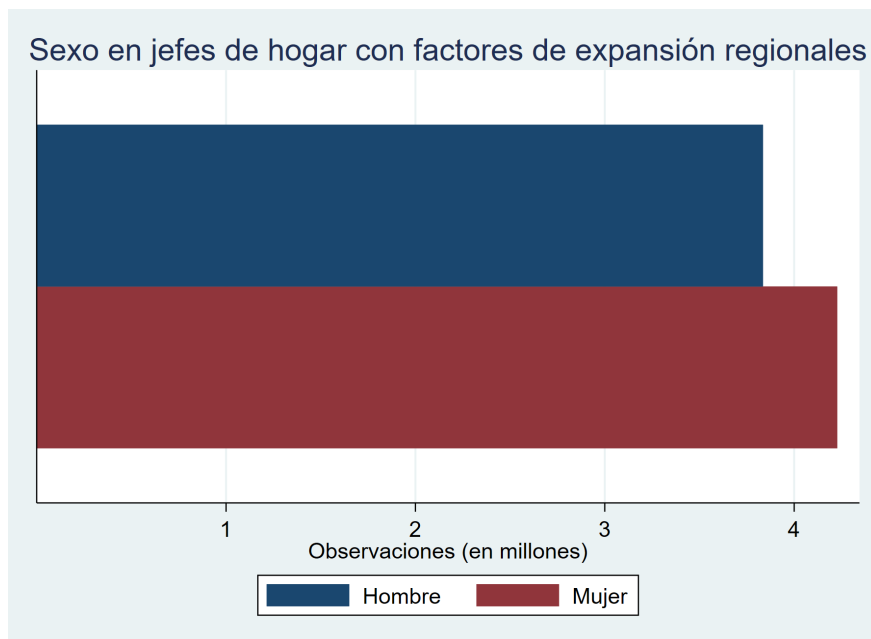


Figura 3.3: Gráfico de sexos en jefes de hogar con factores de expansión regionales.

Fuente: Elaboración propia.

Los ingresos de la población de jefes de hogar no suelen ser por lo general demasiado altos, ya que el ingreso promedio de la población es equivalente a un \$581.100 y se tiene una mediana de \$333.333 la cual es menor al sueldo mínimo a tiempo completo actual (equivalente a unos \$365.667), los ingresos suelen ser bastante dispares entre ambos sexos, teniendo en consideración que los hombres tienen un promedio de \$758.879 y una mediana de \$415.416 mientras que las mujeres tienen en comparación un sueldo de \$422.205 y una mediana de \$266.667, lo cual es muy por debajo de los ingresos que reciben los hombres, siendo una diferencia de por lo menos un 44.3 % entre los sueldos promedio y una diferencia de un 35.8 % en las medianas. Esto indica que no sólo los sueldos por lo general suelen ser bajos, sino que también suelen ser bastante inequitativos en relación a los sexos, en donde se observa que las mujeres suelen tener una cantidad de ingresos muy por debajo de los hombres.

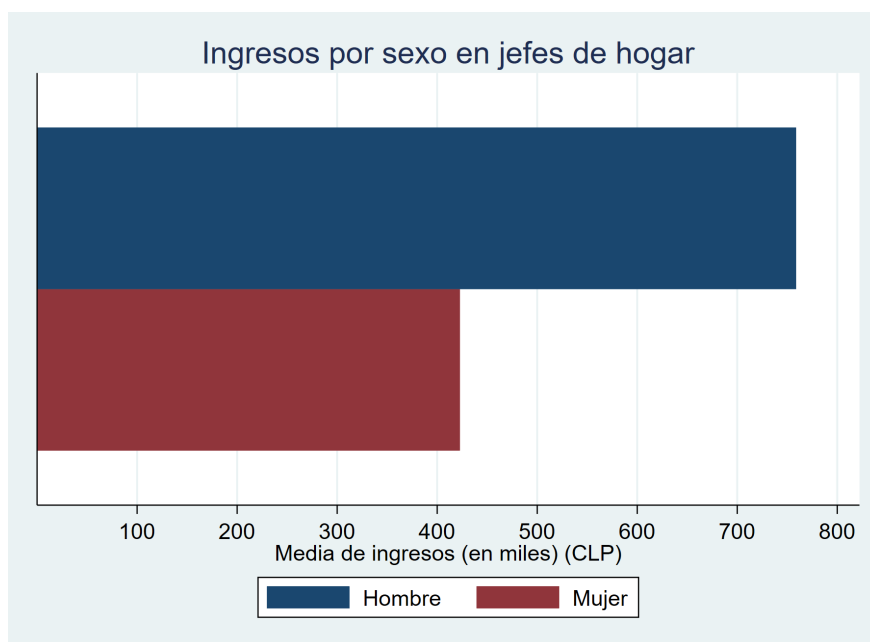


Figura 3.4: Gráfico de ingresos en jefes de hogar con factores de expansión regionales.

Fuente: Elaboración propia.

Las familias en las que los jefes de hogar del país están a cargo no suelen ser demasiado numerosas, esto es debido a que la mayor concentración en la cantidad de personas que se encuentran en los hogares no suele superar las 4 personas, siendo aproximadamente un 20 % de la población las que suelen tener una cantidad de 5 personas o más en su hogar. La cantidad de personas que es más común en los hogares de Chile suele ser de dos personas, en donde se obtuvieron los casos de 1.908.289 jefes de hogar con esa cantidad de personas en su hogar (siendo un 23.71 % de la población de jefes de hogar) del total de 8.050.009 al considerar los factores de expansión, seguido de tres personas por hogar quienes constituyen un 23.16 % con 1.864.245 casos.

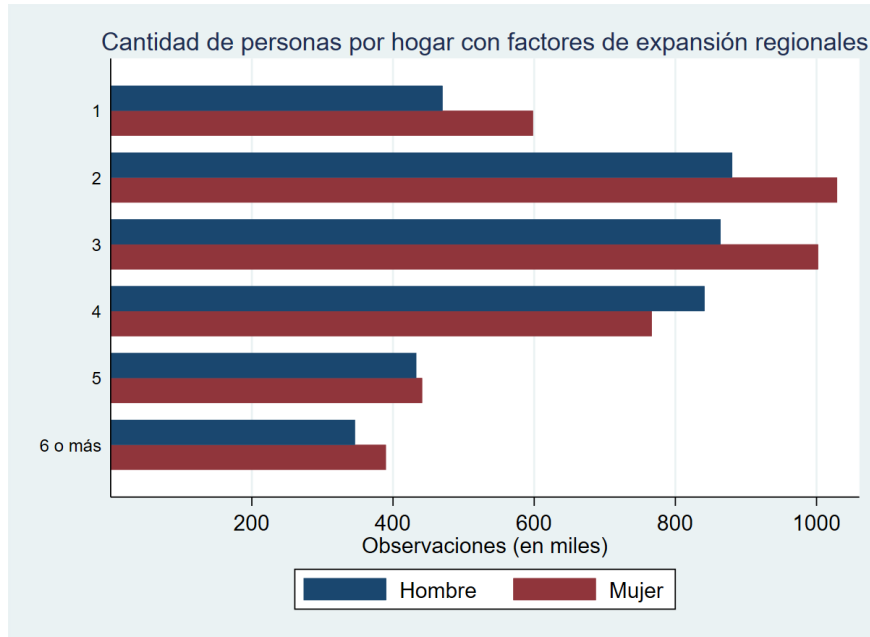


Figura 3.5: Gráfico de cantidad de personas por familia con factores de expansión regionales.

Fuente: Elaboración propia.

La población de los jefes de hogar es predominantemente mayor, esto es debido a que se tiene un promedio de 51.6 años en el caso de los hombres con una mediana de 54 años, mientras que en el caso de las mujeres se tiene una edad relativamente menor debido a que el promedio cae en 50.9 años y su mediana es de 51 años, por lo tanto se puede encontrar que las mujeres en el país que son jefas de hogar suelen ser menores que los hombres en la misma posición, aunque la diferencia no es excesiva. Como se mencionó anteriormente, la población de jefes de hogar es predominantemente mayor debido a que los jefes de hogar son principalmente personas en la tercera edad, siendo que 33.03% de los jefes de hogar tiene o supera los 60 años, superando con creces a las personas más jóvenes (con un rango etareo entre los 18 y los 29 años) ya que sólo un 11.07% de estas tienen la calidad de jefes de hogar.

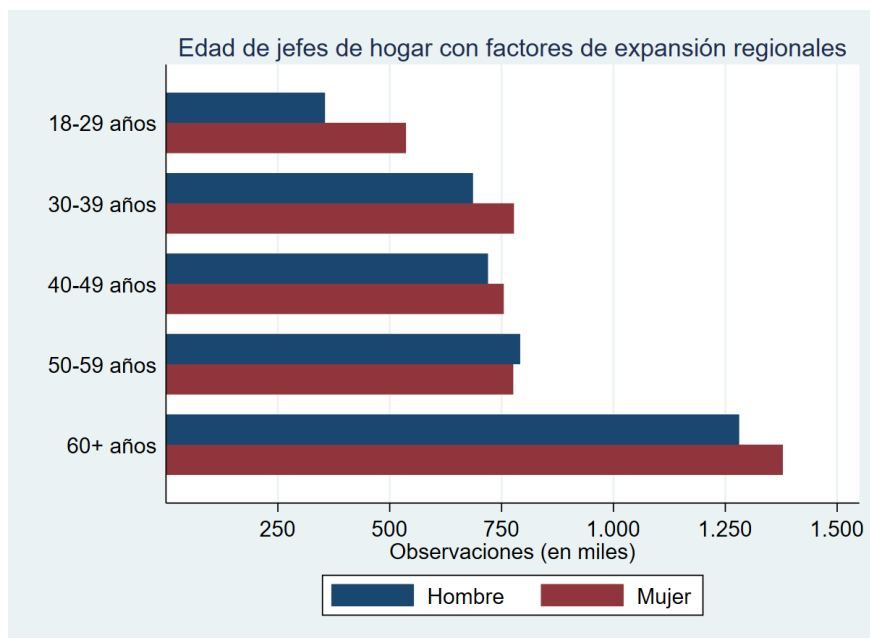


Figura 3.6: Gráfico de edad de jefes de hogar con factores de expansión regionales.

Fuente: Elaboración propia.

Se puede observar cómo se mantiene una cantidad estable entre los distintos rangos de edad que no sean los extremos, en donde se tiene un porcentaje entre 8.5 % y 9.9 % en los sexos por separado y en conjunto se tienen un porcentaje de 18.7 % entre los 30 y 39 años, un 18.3 % entre los 40 y 49 años y por último un 19.43 % entre los 50 y 59 años.

En cuanto a los ingresos de las personas, el rango con la mayor cantidad de ingresos promedio es el rango etareo entre los 40 y 49 años de \$784.735 con una media de %439.063 , lo cual significaría que los datos más externos a la media estarían elevando el valor del promedio total, considerando esto, el rango etareo con la mayor mediana sería entre los 30 y 39 años con una mediana de \$453.609 y que tiene un promedio de \$746.173 . En el caso inverso, se tiene que el rango etareo con menor promedio y menor mediana es el rango de personas con mayoría de edad (60 o más años) debido a que el promedio de sus ingresos es de \$414.410 y una mediana de \$225.000, lo cual explicaría la razón por la que las personas de mayor edad suelen mantenerse en sus empleos.

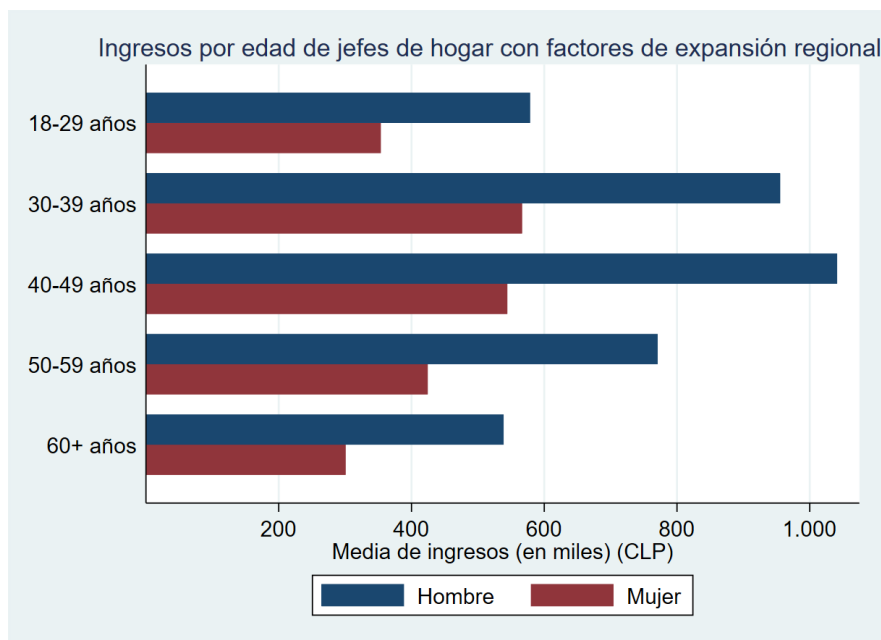


Figura 3.7: Gráfico de ingresos por edad con factores de expansión regionales.

Fuente: Elaboración propia.

Entrando a la educación de los jefes de hogar, hay que mencionar de que para el cálculo y análisis de los datos, se disminuyeron las categorías de las variables que corresponden al nivel educativo más alto de la persona, con esto se refiere a que en casos tales como la educación media como nivel más alto educativo también se considera la educación técnica comercial o el sistema de humanidades antiguo, puesto a que aquellos niveles educacionales también se encuentran luego del término de la educación básica y antes de la educación superior, considerando que la educación media científico-humanista es la opción más común entre la población en ese rango. De una manera similar, ocurre en el caso de la educación superior, ya que se asumen tanto los niveles técnico superior como los profesionales, y también dentro de este grupo se consideran las personas con postgrados tanto completos como incompletos.

La educación de los jefes de hogar suele estar bastante concentrada en la educación media y similares, en donde la mayor parte de la población de jefes de hogar con 2.928.624 casos equivalente a un 36.38% del total, lo que se concentra principalmente en el grupo etareo de personas entre los 30 y 39 años con 640.094 casos equivalente a un 21.87% aunque suelen ser bastante similares las cantidades a lo largo de todos los grupos etarios. El nivel educativo con la menor cantidad de personas es el que no tuvo ningún tipo de educación, correspondiente a la cantidad de 1.199.595 casos los cuales son un 14.9% del total de jefes de hogar, las personas en este nivel se encontraban principalmente en el rango etario de la tercera edad, 812.876 casos los cuales corresponden a un 10.09% de los jefes de hogar. También es destacable que en el único nivel educativo donde los hombres superan en cantidad a las mujeres suele ser en tener una educación superior completa, en donde los hombres cumplen un 13.77% (correspondiente a 1.108.106 casos) y las mujeres cumplen con un 13.24% (correspondiente a 1.066.103 casos).

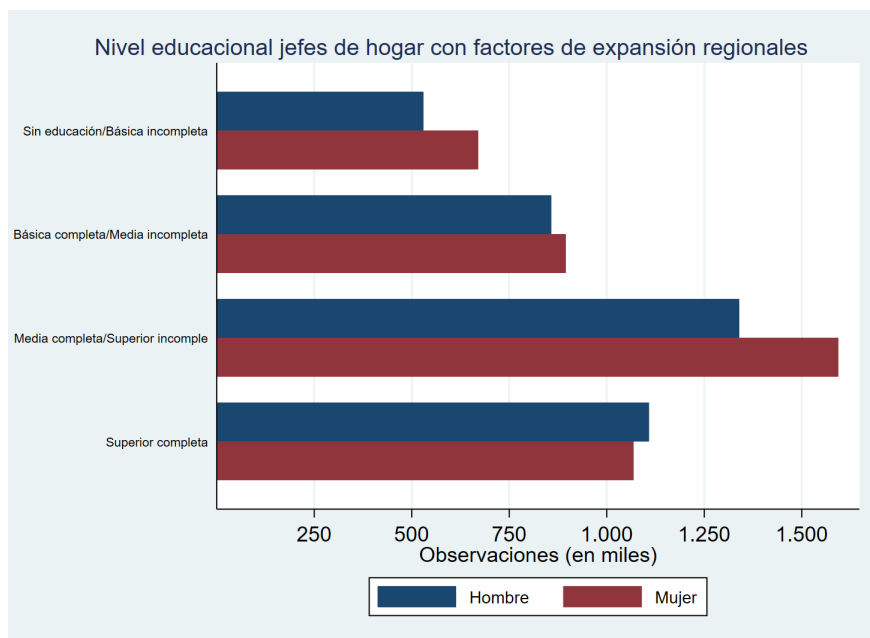


Figura 3.8: Gráfico de nivel educacional con factores de expansión regionales.

Fuente: Elaboración propia.

Con respecto a los ingresos dependiendo del nivel más alto educacional se puede mencionar que el grupo con mayor cantidad de ingresos es el grupo con una educación superior completa o con postgrados, puesto que este grupo alcanza un promedio de \$1.163.939 y con una mediana correspondiente a \$750.000 lo que comparado a los demás grupos es una cantidad bastante superior ya que el segundo grupo con mayores ingresos en promedio es de \$458.178 proveniente del grupo con la educación media completa. En el caso de menores ingresos se encuentran las personas sin educación con un ingreso promedio de \$233.719 y con una mediana de \$189.707, lo cual es bastante preocupante debido a que este grupo en especial en conjunto a las personas que sólo tienen la educación básica completa no suelen superar en promedio un ingreso total equivalente al sueldo mínimo actual (En el caso de las personas con sólo la educación básica completa, se tiene un promedio de \$312.360).

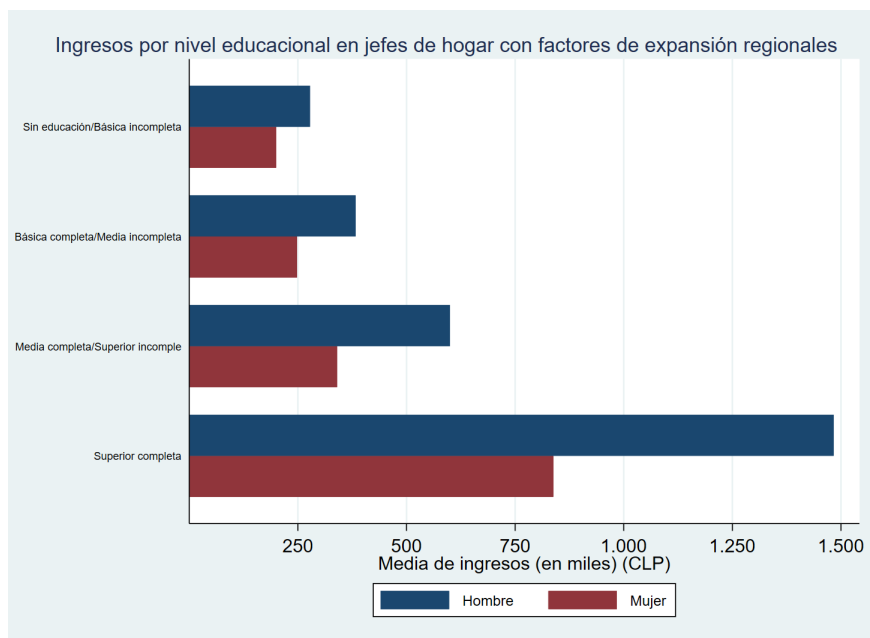


Figura 3.9: Gráfico de ingresos por nivel educacional con factores de expansión regionales.

Fuente: Elaboración propia.

En cuanto a los tipos de empleo, la población de jefes de hogar en Chile suele tener una preferencia hacia los empleos privados con respecto a los empleos públicos, esto es debido a que los jefes de hogar trabajando como empleados privados constituyen un 52.71% de los empleos totales en los jefes de hogar (teniendo una cantidad de 2.490.064 empleos en total) mientras que los empleos públicos tienen un total de 1.066.103 que equivalen a un 15.19% de los empleos totales. El tipo de empleo menos popular a lo largo de la población de jefes de hogar son las FF. AA. y del Orden, en donde se tiene una cantidad de 22.377 empleos correspondientes a un 0.47% del total de estos. En este ámbito no suelen ser tan equitativos ambos sexos, puesto a que existen casos tales como los empleos privados, el puesto de patrón o empleador y las FF. AA. y del Orden se tiene una cantidad mucho mayor de hombres que de mujeres mientras que las mujeres suelen tener predominancia en tipos de empleos tales como los de servicio doméstico o los empleos públicos.

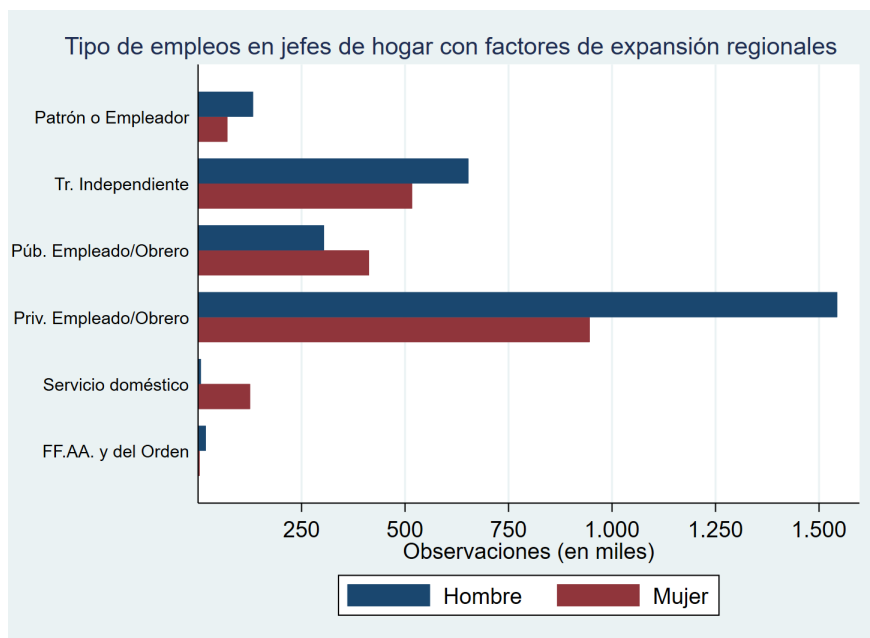


Figura 3.10: Gráfico de tipo de empleo con factores de expansión regionales.

Fuente: Elaboración propia.

En las ganancias por tipo de empleo, la clase que tiene mayores ganancias en promedio como es de esperar es la de patrón o empleador, en donde se tiene un promedio de \$1.387.906 y se tiene una mediana de \$578.000 , pese a que esta clase de empleos suele ser la con una menor cantidad de casos, también es la con una brecha mayor en comparación a los demás empleos, en donde se tiene una diferencia en los promedios de un 59.45% a favor de los hombres siendo los ingresos promedio de estos un equivalente a \$1.762.057 mientras que en el caso de las mujeres, su promedio de ingresos suele corresponder a sólo \$714.675 . Luego, a pesar de ser la clase de empleo menos popular, las FF. AA. y del Orden tienen el segundo lugar en cuanto a los ingresos, con un promedio de \$1.113.948 con una mediana de \$852.000 , siendo superior al promedio de los demás tipos de empleo excluyendo a quienes están en puesto de patrón o empleador y en empleos públicos. En el otro lado, el tipo de empleo con menores ganancias es el empleo doméstico con una media de \$355.335 y con una mediana de \$322.940 el cual era superior por poco al salario mínimo del año (el salario mínimo era equivalente a \$301.000 al inicio del año 2020, mientras que a partir de septiembre esta suma aumentó a \$326.500).

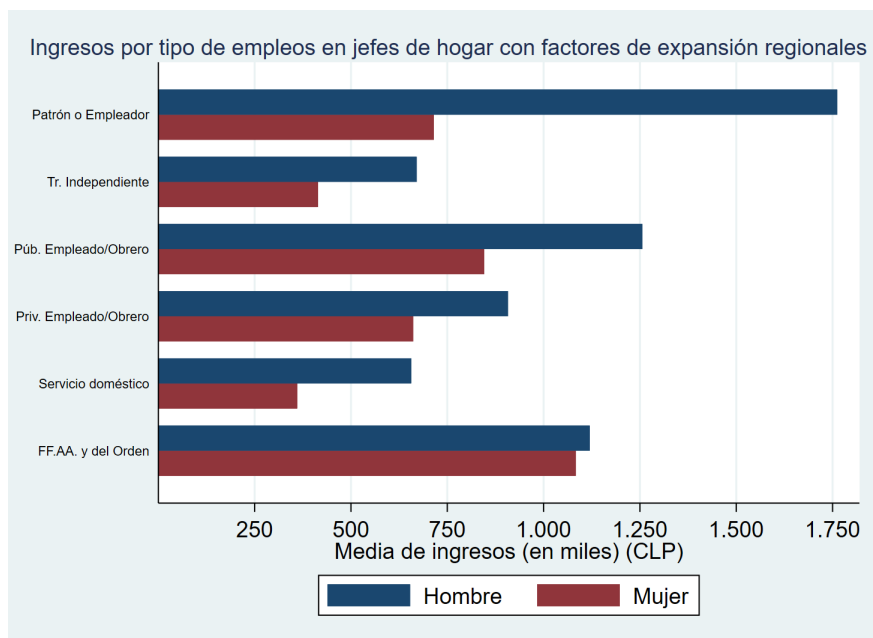


Figura 3.11: Gráfico de ingresos por tipo empleo con factores de expansión regionales.

Fuente: Elaboración propia.

Como se mencionó anteriormente y como se consideró en la realización de la Encuesta CASEN en Pandemia 2020, existen 10 tipos de pueblos indígenas los cuales son considerados por la ley 21.273 (publicada a la fecha del 17 de octubre del año 2020) los cuales son: Aimara, Rapa-Nui, Quechua, Mapuche, Atacameño, Colla, Kawashkar, Yagán, Diaguita y Changos. Si bien, en la pregunta como tal se consideraron como categoría cada pueblo por separado, por la cantidad de observaciones de cada pueblo en específico se modificó para sólo indicar si es que las personas se identificaban o tienen ascendencia de alguno de los pueblos indígenas reconocidos por la ley, de esa manera poder realizar observaciones de una mejor manera.

La cantidad de jefes de hogar que tiene ascendencia indígena en el país es bastante reducida en comparación a la cantidad de personas que no la tienen, siendo sólo un 9.04% de la población de jefes de hogar quien tiene algún tipo de ascendencia indígena, lo cual es un equivalente a 728.318 casos. Siendo un poco más específico en los datos, el pueblo indígena con la mayor población es el Mapuche, en donde se obtuvieron 577.184 observaciones o también un 7.17% de la cantidad total de jefes de hogar, mientras que en los casos de los demás pueblos indígenas se obtuvieron cantidades muy inferiores en comparación, en donde el siguiente pueblo originario fue el Diaguita con un 0.56%.

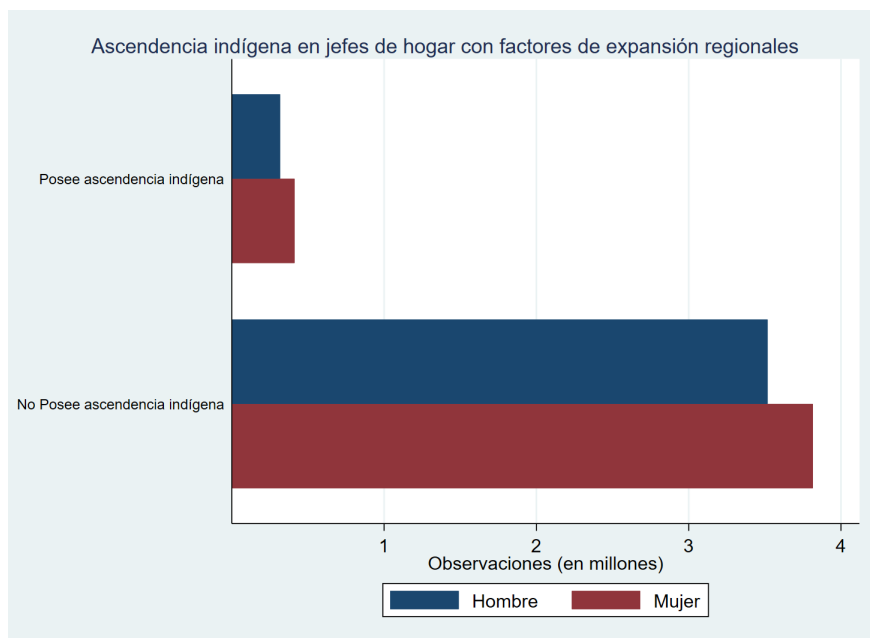


Figura 3.12: Gráfico de ascendencia indígena con factores de expansión regionales.

Fuente: Elaboración propia.

Las ganancias de las personas con ascendencia indígena suelen ser \$450.739 en promedio, con una mediana de \$309.743 mientras que en el caso de personas que no tienen ascendencia indígena se tiene un promedio de \$594.428 y una mediana de \$337.401 en ingresos, siendo una diferencia de un 24.18 %, la que si bien es una diferencia apreciable, no es tan grande en comparación a otras brechas tales como la que ocurre en torno a los géneros, debido a que las personas con ascendencia indígena suelen tener en promedio un ingreso promedio de \$569.807 mientras que las mujeres tienen un ingreso promedio de \$359.060 , lo cual corresponde a una diferencia de prácticamente un 37 % en los ingresos.

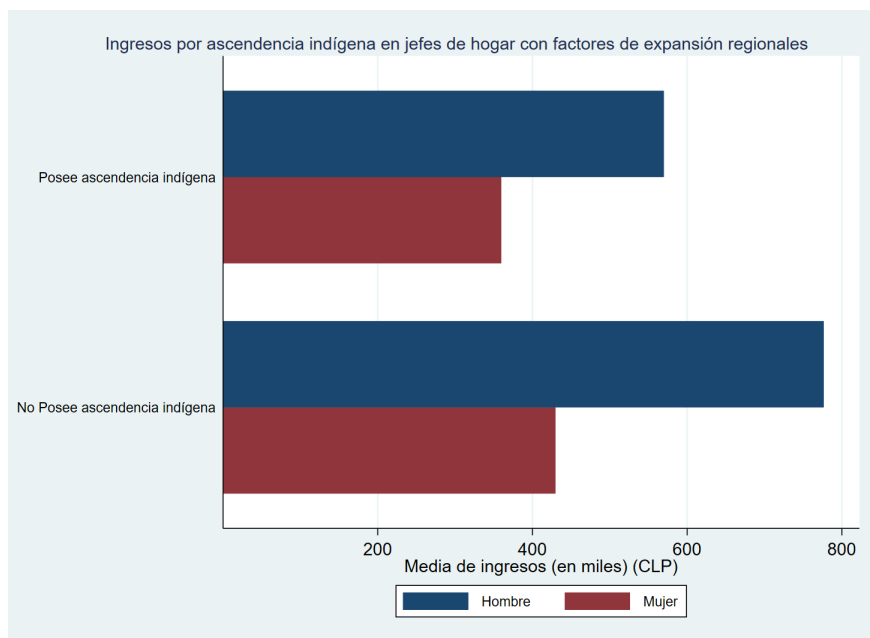


Figura 3.13: Gráfico de ingresos por ascendencia indígena con factores de expansión regionales.

Fuente: Elaboración propia.

En la última variable que se mencionará se encuentra si es que los jefes de hogar comparten o tienen la presencia de un cónyuge o pareja (sea con un acuerdo legal o no), en donde se tiene una cantidad bastante pareja en ambos casos ya que un 50.18 % de jefes de hogar indica que no tiene una pareja o cónyuge (una cantidad total de 4.039.358 observaciones) y un 49.82 % indica la presencia de una (una cantidad total de 4.010.651 observaciones). Una característica destacable es que los hombres jefes de hogar son quienes suelen tener la mayor cantidad de parejas, teniendo una cantidad total de 2.706.099 (equivalente a un 33.62 % del total de la población) jefes de hogar con pareja en comparación a las 1.304.552 mujeres (equivalente a un 16.21 % del total de la población) que tienen pareja.

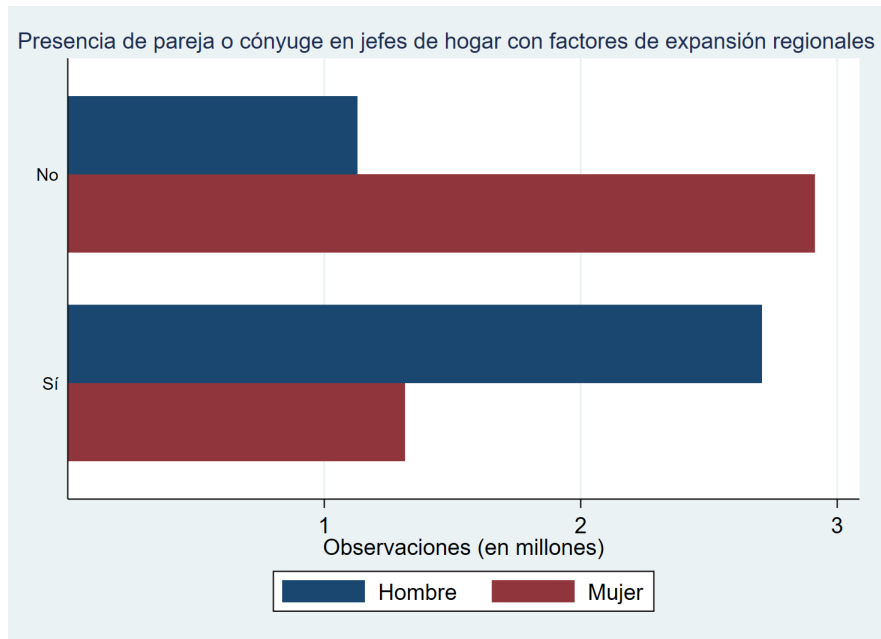


Figura 3.14: Gráfico de presencia de cónyuge con factores de expansión regionales.

Fuente: Elaboración propia.

Los ingresos de las personas que no suelen tener una pareja o cónyuge suelen ser inferiores a los ingresos de las personas con una pareja o cónyuge, esto se puede observar ya que los ingresos promedio de las personas sin pareja suele ser de \$482.515 con una mediana de \$300.000 mientras que las personas con pareja tienen un promedio de \$683.524 con una mediana de \$390.000 , lo cual podría indicar que las personas con pareja tal vez optasen por empleos de mayores sueldos debido a la vida en pareja que estos llevan.

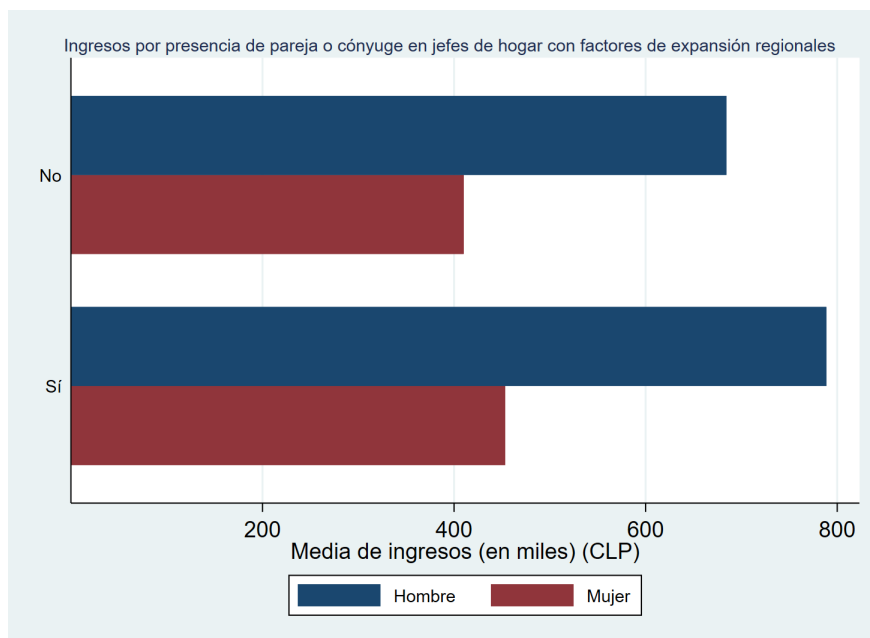


Figura 3.15: Gráfico de ingresos por presencia de cónyuge con factores de expansión regionales.

Fuente: Elaboración propia.

Si bien existen otras variables que se pudieron haber utilizado para el análisis de estos datos, varias fueron desechadas por las bajas cantidades de observaciones que estas tenían o la cantidad excesiva de categorías que estas incluían, lo cual hacía más complicado la tarea de observarlas para los análisis de datos posteriores.

En la siguiente subsección se mencionarán los resultados referentes a la descomposición Oaxaca-Blinder.

3.2. Descomposición Oaxaca-Blinder

En la ejecución de la descomposición Oaxaca-Blinder, la aplicación de esta técnica inicialmente divide la brecha salarial entre dos grupos distintos (los cuales en este caso son hombres y mujeres jefes de hogar) en una parte la cual es explicada a través de las diferencias entre los determinantes de los ingresos, los cuales son las variables que se han mencionado anteriormente y una diferencia no explicable por las diferencias entre los distintos grupos.

Cabe destacar que un problema que surgió en el análisis de los datos a través de la descomposición Oaxaca-Blinder es que esta no permite el uso de la función de *survey data* (comando `svy`) para el análisis de muestras complejas, por lo que se realizó la descomposición utilizando los factores de expansión regionales, por lo tanto, las estimaciones de varianza no son totalmente acertadas en este método.

Al realizar la descomposición a través del comando `oaxaca` (Jann, 2008) del software Stata en los datos de la Encuesta CASEN en Pandemia 2020 considerando los factores de expansión regionales, se obtuvieron los siguientes resultados:

Model for group 1

Source	SS	df	MS	Number of obs	=	2,396,122
Model	5.5029e+17	6	9.1714e+16	F(6, 2396115)	=	23389.89
Residual	9.3954e+18	2,396,115	3.9211e+12	Prob > F	=	0.0000
				R-squared	=	0.0553
				Adj R-squared	=	0.0553
Total	9.9457e+18	2,396,121	4.1508e+12	Root MSE	=	2.0e+06

ingresos_t~s	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
Edad_j	104786.3	1084.639	96.61	0.000	102660.4	106912.1
educacion_j	529839.1	1447.781	365.97	0.000	527001.5	532676.7
empleo_j	-61156.06	1301.593	-46.99	0.000	-63707.14	-58604.98
ascen_j	123275.8	4569.48	26.98	0.000	114319.7	132231.8
conyuge_j	52331.85	3281.647	15.95	0.000	45899.94	58763.77
numper_j	-12904.63	997.5047	-12.94	0.000	-14859.7	-10949.55
_cons	-1117589	12118.44	-92.22	0.000	-1141341	-1093837

Model for group 2

Source	SS	df	MS	Number of obs	=	1,939,505
Model	1.4457e+17	6	2.4095e+16	F(6, 1939498)	=	55399.72
Residual	8.4353e+17	1,939,498	4.3492e+11	Prob > F	=	0.0000
				R-squared	=	0.1463
				Adj R-squared	=	0.1463
Total	9.8810e+17	1,939,504	5.0946e+11	Root MSE	=	6.6e+05

ingresos_t~s	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
Edad_j	57649.43	410.5771	140.41	0.000	56844.71	58454.14
educacion_j	289992.5	564.4233	513.79	0.000	288886.3	291098.8
empleo_j	21985.33	470.1915	46.76	0.000	21063.77	22906.89
ascen_j	41542.08	1563.838	26.56	0.000	38477.01	44607.15
conyuge_j	112202.2	1087.872	103.14	0.000	110070	114334.4
numper_j	-54460.6	356.2356	-152.88	0.000	-55158.81	-53762.39
_cons	-513808.8	4391.804	-116.99	0.000	-522416.5	-505201

Figura 3.16: Tablas de modelos de los distintos grupos con factores de expansión regionales.

Fuente: Elaboración propia.

En estas tablas se puede observar cómo funciona el modelo en los distintos grupos, siendo el primer grupo de los jefes de hogar de sexo masculino mientras que el segundo grupo es para los jefes de hogar de sexo femenino. Inicialmente se realizan estimaciones de regresiones específicas para ambos grupos y luego se realiza la descomposición por tres partes (la cual fue realizada en esta ocasión).

La siguiente tabla indica los resultados que se obtuvieron a través de la descomposición Oaxaca-Blinder de tres partes:

1: sexo = 1

2: sexo = 2

ingresos_totals	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Differential						
Prediction_1	929540.9	1316.163	706.25	0.000	926961.3	932120.6
Prediction_2	619925.8	512.5196	1209.57	0.000	618921.3	620930.4
Difference	309615.1	1412.431	219.21	0.000	306846.8	312383.4
Decomposition						
Endowments	25063.13	552.7454	45.34	0.000	23979.77	26146.49
Coefficients	317578.7	1928.749	164.66	0.000	313798.4	321359
Interaction	-33026.73	1454.522	-22.71	0.000	-35877.54	-30175.92

Figura 3.17: Resultados de la descomposición Oaxaca-Blinder con factores de expansión regionales.

Fuente: Elaboración propia.

La primera parte de la tabla indica los resultados de los diferenciales, siendo estos las predicciones (Promedio de los logaritmos de los ingresos) de los distintos grupos (hombres y mujeres, respectivamente), además de los errores estándar, z-score, p-value, los intervalos de confianza (que en este caso ocurren con un 95% de confianza) y de la diferencia entre estas predicciones (la que se denomina como brecha salarial o R), por lo que se puede observar que existe una media de los logaritmos en los ingresos de aproximadamente \$929.540 en el caso de los hombres y de \$619.925 en el caso de las mujeres, lo que nos dejaría una diferencia aproximada entre los ingresos totales de ambos grupos de \$309.615 a favor de los hombres, lo que constituiría una diferencia entre los ingresos de un 33.3% y con un intervalo de confianza del 95% que se encuentra entre \$306.846 y \$312.383 lo que indica que los ingresos no varían demasiado de su valor estimado.

En la segunda parte de la tabla se indica los valores de la descomposición Oaxaca-Blinder de tres partes, siendo estos el efecto de las "dotaciones" (o endowments) lo que indicaría que si las mujeres tuviesen las mismas características de los hombres estas tendrían un aumento en los ingresos de aproximadamente \$25.063, lo que indicaría la diferencia de las dotaciones en la mitad de la brecha salarial, el efecto de los coeficientes que cuantificaría los ingresos en el caso de que las mujeres tuviesen aplicados los coeficientes de los hombres a sus características lo que significaría que tendrían un cambio de salario de aproximadamente \$317.579 y las interacciones que miden el efecto simultáneo de las diferencias en las dotaciones y los coeficientes que indicaría una baja de \$-33.027 aproximadamente.

Es notable destacar que los p-value de cada valor otorgado por la tabla de resultados es de prácticamente 0, por lo que los datos son altamente significativos.

En la siguiente sección se mostrarán los resultados que se obtuvieron al realizar una regresión lineal.

3.3. Regresión lineal

Se realizó además de la descomposición Oaxaca-Blinder una regresión lineal para confirmar la validez de los datos anteriormente obtenidos, debido a que la descomposición Oaxaca-Blinder es una variación de una regresión lineal realizada con una separación por grupos, por lo tanto, los resultados debiesen de ser similares a los obtenidos en una regresión lineal entre los ingresos y los distintos sexos de los jefes de hogar.

Los resultados obtenidos a través de la regresión lineal se realizaron utilizando la función de *survey data* de Stata y considerando los factores de expansión correspondientes, por lo que la estimación de las varianzas en este caso son totalmente correctas y son las indicadas para este tipo de análisis de muestreo complejo.

Survey: Linear regression

Number of strata =	293	Number of obs =	68,471
Number of PSUs =	1,467	Population size =	7,112,411
		Design df =	1,174
		F(1, 1174) =	492.29
		Prob > F =	0.0000
		R-squared =	0.0167

ingresos_t~s	Linearized				
	Coefficient	std. err.	t	P> t	[95% conf. interval]
sexo_j	-336184	15151.86	-22.19	0.000	-365911.7 -306456.2
_cons	1095058	30268.23	36.18	0.000	1035672 1154444

Figura 3.18: Resultados regresión lineal considerando survey data y factores de expansión.

Fuente: Elaboración propia.

Los resultados de la regresión lineal muestran que existe una diferencia entre ambos sexos de \$336.184, lo que es bastante similar a los resultados obtenidos por la descomposición Oaxaca-Blinder (\$309.615) teniendo una diferencia aproximada de \$26.569 entre ambos métodos. Además, se tiene un intervalo de confianza del 95 % que se encuentra entre \$306.456 y \$365.911, los resultados obtenidos en la descomposición Oaxaca-Blinder se encuentran dentro de este rango, por lo que se puede considerar que se aproximan por bastante a los resultados más óptimos de los datos analizados a través de un muestreo complejo.

Al encontrarse que los resultados de la descomposición Oaxaca-Blinder y sus intervalos de confianza relacionados a la brecha salarial se encuentran dentro del rango de los intervalos de confianza de la regresión lineal, método el cual está optimizado totalmente para un análisis de muestras complejas utilizando datos recolectados por encuestas, se podría considerar que los cálculos realizados resultan aproximarse bastante a los resultados específicos para este tipo de caso.

En la siguiente subsección se mostrarán los resultados del análisis por correspondencia múltiple.

3.4. Análisis por correspondencia múltiple

Antes de iniciar con el análisis, de una manera similar al caso anterior de la descomposición Oaxaca-Blinder las funciones para el análisis de correspondencia múltiple no soportan el *survey data* para muestras complejas, por lo que de la misma manera como se realizó anteriormente, se realizaron los análisis considerando los factores de expansión regionales, por lo que las estimaciones de varianza no son totalmente acertadas.

En la realización de la técnica de análisis por correspondencia múltiple se obtienen distintas tablas, en las que se observan varios resultados tales como las inercias principales, las cuales son los cuadrados de los valores singulares, estos valores son . En la figura 5.18 se puede observar cómo las tres primeras dimensiones explican al menos un 64.5 % de la inercia total que si bien explica la mayor cantidad de datos con una cantidad pequeña de dimensiones, en este caso se realizarán los análisis en dos dimensiones para realizar los análisis gráficos de manera más simple, por lo que se explicaría un 55,09 % de los datos con esta cantidad de dimensiones.

Dimension	Principal inertia	Percent	Cumul. percent
Dim 1	.011448	44.16	44.16
Dim 2	.002833	10.93	55.09
Dim 3	.0024385	9.41	64.50
Dim 4	.0006282	2.42	66.92
Dim 5	.0001062	0.41	67.33
Dim 6	.0000877	0.34	67.67
Dim 7	.0000361	0.14	67.81
Dim 8	7.80e-07	0.00	67.81
Dim 9	1.48e-07	0.00	67.81
Total	.0259231	100.00	

Tabla 3.18: Tabla de MCA de inercias principales con factores de expansión regionales.

Fuente: Elaboración propia.

Luego, se obtiene la tabla ?? con las distintas categorías, en donde se obtiene la masa (la suma de las filas y columnas en la tabla de proporciones), la medida de calidad (distancia entre el punto dentro del espacio y la proyección de los componentes) y el porcentaje de inercia (varianza total) en los datos generales, mientras que en el caso de las dimensiones en específico se obtiene las coordenadas de la categoría en la dimensión, las correlaciones cuadradas de las categorías y la contribución (influencia de los datos atípicos u *outliers* en la construcción de los componentes) de

cada una.

Categories	Overall			Dimension_1			Dimension_2		
	Mass	Quality	%inert	Coord	Sqcorr	Contrib	Coord	Sqcorr	Contrib
sexo_j									
Hombre	0.094	0.308	0.045	0.376	0.130	0.013	0.884	0.178	0.073
Mujer	0.073	0.308	0.057	-0.482	0.130	0.017	-1.132	0.178	0.094
educacion_j									
Sin educac~a	0.013	0.701	0.093	3.358	0.682	0.144	-1.137	0.019	0.017
Básica com~a	0.029	0.673	0.089	2.164	0.669	0.135	0.344	0.004	0.003
Media comp~a	0.065	0.478	0.029	-0.395	0.155	0.010	1.145	0.323	0.086
Superior c~a	0.060	0.574	0.097	-1.332	0.481	0.106	-1.181	0.093	0.083
Edad_j2									
18-29 años	0.020	0.339	0.037	-1.120	0.305	0.025	0.750	0.034	0.011
30-39 años	0.041	0.731	0.051	-1.419	0.727	0.083	-0.224	0.004	0.002
40-49 años	0.041	0.363	0.019	-0.494	0.230	0.010	0.758	0.134	0.023
50-59 años	0.039	0.636	0.022	0.904	0.636	0.031	0.008	0.000	0.000
60+ años	0.026	0.639	0.128	2.583	0.593	0.172	-1.442	0.046	0.054
empleo_j									
Patrón o Emr	0.007	0.307	0.007	0.843	0.306	0.005	-0.095	0.001	0.000
Tr Independe	0.041	0.790	0.062	1.642	0.789	0.111	0.089	0.001	0.000
Púb Empleo	0.025	0.706	0.072	-1.699	0.448	0.073	-2.587	0.257	0.169
Priv Empleo	0.088	0.579	0.025	-0.446	0.304	0.017	0.851	0.275	0.064
Servicio d~o	0.005	0.236	0.065	2.209	0.155	0.023	-3.203	0.081	0.048
FFAA y del~n	0.001	0.523	0.006	-2.512	0.361	0.005	3.380	0.162	0.009
ascen_j									
Posee ascen~a	0.015	0.093	0.014	0.334	0.054	0.002	0.562	0.038	0.005
No Posee ~a	0.151	0.093	0.001	-0.034	0.054	0.000	-0.057	0.038	0.000
numper_j									
1	0.019	0.392	0.010	0.260	0.057	0.001	-1.267	0.335	0.030
2	0.035	0.514	0.027	0.432	0.109	0.007	-1.669	0.404	0.099
3	0.040	0.451	0.003	-0.247	0.337	0.002	-0.287	0.113	0.003
4	0.037	0.421	0.019	-0.373	0.123	0.005	1.170	0.298	0.051
5	0.019	0.510	0.009	-0.033	0.001	0.000	1.475	0.509	0.042
6 o más	0.015	0.337	0.012	0.292	0.046	0.001	1.485	0.291	0.033

Tabla 3.19:Tabla de MCA por categoría con factores de expansión regionales.

Fuente: Elaboración propia.

Esto se puede observar de una mejor manera a través de una manera gráfica tal como en la figura 3.19 con las distintas categorías analizadas en dos dimensiones:

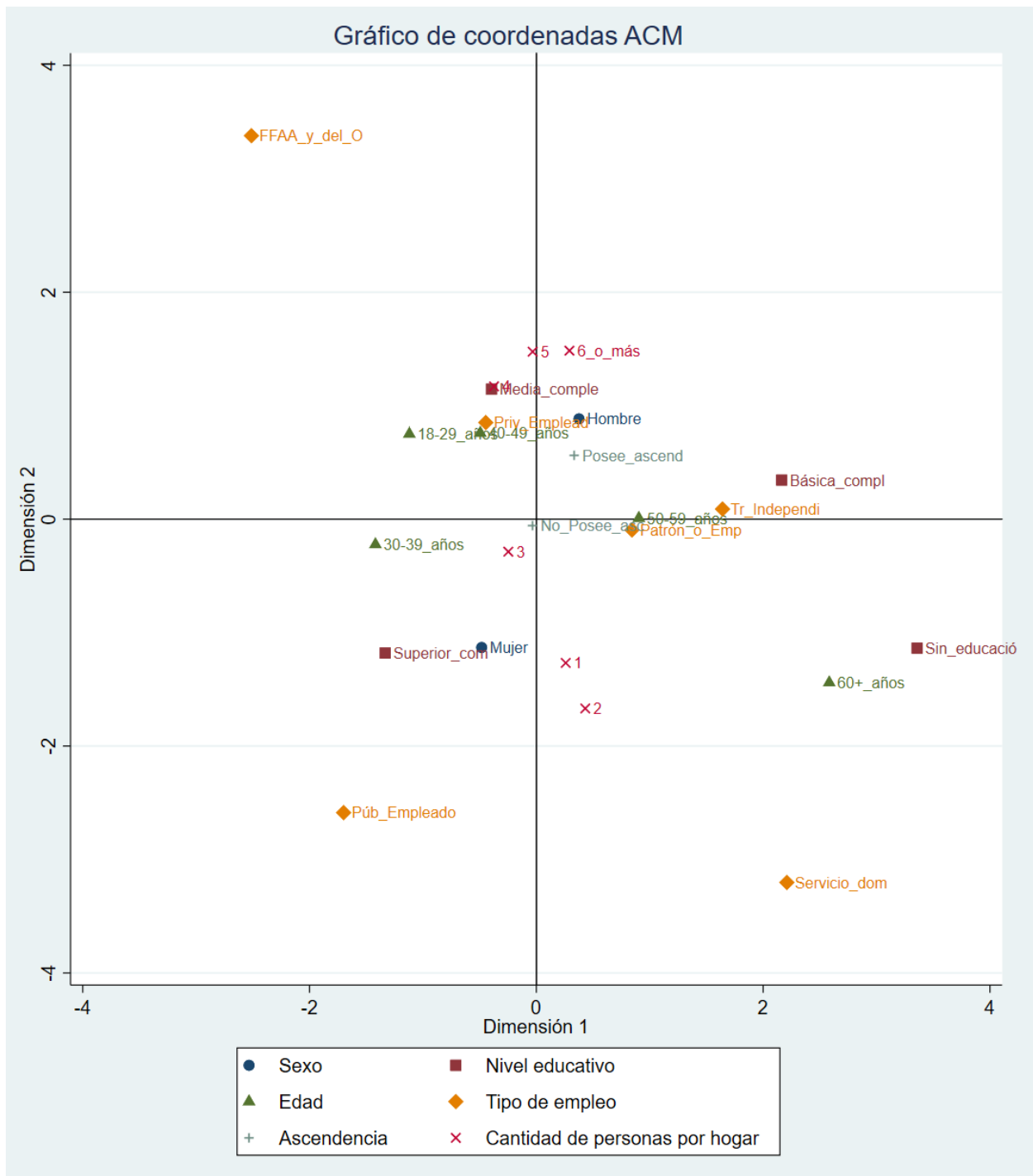


Figura 3.19: Gráfico de MCA.

Fuente: Elaboración propia.

El gráfico de análisis de correspondencias múltiple muestra el cómo se sitúan las distintas categorías a través de las coordenadas en dos dimensiones distintas, por lo que se pueden encontrar asociaciones entre las distintas categorías de las variables dependiendo de cómo están situadas sus coordenadas dentro del gráfico, en donde posteriormente puede realizarse una caracterización de

los sujetos de estudio.

De la figura 3.19 se pueden destacar distintas observaciones dependiendo de en qué dimensión se está observando, por lo que inicialmente se mostrarán observaciones a partir de la primera dimensión, seguida de la segunda dimensión y por último observaciones globales con respecto al análisis de correspondencias.

Al enfocarnos en la primera dimensión, las categorías que suelen estar más orientadas en esta dimensión son las categorías de las variables correspondientes a la edad y a la educación, en donde se puede observar patrones tales como que:

- Las personas a mayor edad suelen estar más ligadas a tener una menor educación que las personas más jóvenes o de mediana edad (especialmente las personas de tercera edad), quienes suelen tener una educación superior o por lo menos la educación media completa, patrón que se va alejando a medida que las personas suelen avanzar en edad.
- Las personas mayores de 50 años están principalmente caracterizadas por tener una menor educación y trabajar en empleos más independientes (sin superiores), entre los que destacan los servicios domésticos, trabajadores independientes y de patronos o empleadores, mientras que las personas menores a 50 años suelen caracterizarse por tener empleos más dependientes (considerando esto como empleos que se tengan superiores dentro del trabajo), tales como los empleos privados, públicos o de F.F. A.A.

En la segunda dimensión, las categorías que suelen estar más enfocadas en esta son las con variables correspondientes al sexo, el tipo de empleo, la ascendencia indígena y la cantidad de personas dentro del hogar, en donde se observan cosas tales como:

- Las mujeres que suelen ser jefas de hogar están principalmente caracterizadas a una cantidad de 3 o menos personas por hogar mientras que los jefes de hogar hombres suelen caracterizarse por tener mayores cantidades de personas dentro de sus hogares (4 o más personas).
- Las mujeres suelen caracterizarse por tener una educación más alta que los hombres, ya que estas tienden mucho más a tener una educación superior que en el caso de los hombres, mientras que estos tienden a tener una educación media completa.
- En el caso de los empleos, los hombres suelen estar más relacionados a los empleos privados y a las FF. AA. y del Orden que las mujeres, quienes tienden a tener más empleos públicos y empleos domésticos, en cuanto a los trabajos independientes estos suelen ser más cercanos a los hombres, pero sólo por una pequeña diferencia en comparación a las mujeres.
- Los hombres suelen tener una mayor representatividad de ascendencia indígena que las mujeres, las cuales no suelen tener tanta representatividad en este ámbito.

Si se considera el gráfico 3.19 de una manera global considerando ambas dimensiones o ejes, se pueden mencionar observaciones tales como:

- Se observa que los hombres jefes de hogar están en el primer cuadrante mientras que las mujeres jefas de hogar se encuentran en el tercer cuadrante, y si bien, estos están opuestos en el gráfico, no suelen estar demasiado alejados del origen y por lo tanto no suelen caracterizar en gran medida.

- En el primer cuadrante se observa una cercanía entre la educación básica completa y los trabajos que suelen ser independientes, si se considera además al segundo cuadrante se puede observar una cercanía entre los jefes de hogar hombres, una educación media completa, el considerarse con ascendencia indígena, empleos privados y una cantidad de 4 o más personas por hogar, por lo que se podría caracterizar a los hombres considerando estos datos.
- En el tercer cuadrante se puede observar que las jefas de hogar mujeres suelen estar cercanas a las coordenadas de la educación superior completa y a las coordenadas de los empleos públicos, por lo que las mujeres se estarían caracterizando en preferencia a este tipo de educación y de empleo, también si se considera al cuarto cuadrante se puede observar que suelen ser muy cercanas a una cantidad entre una y tres personas por hogar.
- En el cuarto cuadrante se observa una cercanía entre personas de tercera edad, sin educación y servicios domésticos, por lo que se puede indicar que las las personas de tercera edad podrían ser más propensas a tener este tipo de empleo y educación.
- Los tipos de trabajo más atípicos (F.F. A.A. y del Orden y Servicio doméstico) están completamente opuestos en los cuadrantes y bastante alejados del origen, esto también puede indicar que categorías tales como el sexo no suelen ser demasiado relevantes para la caracterización de estas categorías.
- Al observar el gráfico de una manera más general, se encuentra una cercanía entre los empleos de patrón o empleador y el rango de edad entre 50 y 59 años, o que las personas entre los rangos de 18-29 años y 40-49 años suelen caracterizarse por optar por empleos públicos y tener una educación media completa.

Por lo general, ambos sexos son bastante dispares en las variables y categorías estudiadas, en donde suelen haber pocos casos que no exista una tendencia hacia un sexo (tal como el caso de los trabajos independientes y de patrón o empleador), aún así, como las coordenadas por lo general no suelen estar demasiado alejadas del origen como en el caso específico de variables tales como la de los sexos o la ascendencia indígena, estas caracterizaciones no suelen mostrar demasiadas tendencias dicotómicas, por lo que funcionan para realizar suposiciones a través de ellas.

El siguiente capítulo incluye las conclusiones y la discusión que se consideraron en la realización de los distintos análisis de este trabajo.

Capítulo 4

Conclusiones y discusión

En este trabajo se realizó un análisis en la Encuesta CASEN en Pandemia 2020. Para ello se realizó una investigación de los estudios aplicados tanto a la Encuesta CASEN de años anteriores como para estudios de la brecha salarial en distintos países para obtener los métodos más eficientes o comunes para un análisis de la brecha salarial. Se estudió la metodología impartida en el diseño muestral de la Encuesta CASEN en Pandemia 2020, se realizaron diferentes métodos y técnicas entre los que se encuentran la descomposición Oaxaca-Blinder, la regresión lineal considerando datos de muestra complejos y análisis de correspondencias múltiple, para así cumplir los objetivos propuestos de determinar el estado de la brecha durante el año 2020 además de conocer los factores que estuviesen siendo determinantes para la explicación de esta brecha.

A partir de los resultados obtenidos a través de la descomposición Oaxaca-Blinder, se obtuvo una brecha salarial de género en Chile para las personas en condición de jefes de hogar equivalentes a un \$309.615, siendo una diferencia de 33.3 % en los ingresos lo que indica que la brecha se mantiene de manera muy similar a años anteriores, en donde esta no bajaba de un 30 % de discriminación, también se observó en el análisis descriptivo que esta brecha suele ocurrir con mayor frecuencia en sectores que suelen tener una mayor educación, un mayor poder adquisitivo o una mayor posición dentro del empleo, al igual que se tiene mayor ocurrencia entre las edades desde los 30 hasta los 59 años principalmente.

En cuanto a los resultados de la regresión lineal, se obtuvo a través de ella una diferencia entre los ingresos de ambos sexos equivalentes a \$336.184 con un intervalo de confianza de \pm \$29.727, en donde dentro de este intervalo de confianza se encuentran los valores de los resultados de la descomposición Oaxaca-Blinder, por lo que se puede mencionar que los resultados obtenidos, si bien no son totalmente óptimos en el caso de la descomposición, se acercan bastante a los resultados esperados por la regresión entre los ingresos y los sexos entre jefes de hogar.

Con respecto al análisis de correspondencias múltiple, se realizó una caracterización entre los distintos sexos de jefes de hogar, en donde se encontraron distintos patrones y características que suelen haber entre estos al considerar dos dimensiones del análisis de correspondencias además de una visión global de las variables y las categorías elegidas, entre las que se encuentran los distintos tipos de empleo, el nivel educativo de las personas, la presencia de ascendencia indígena y la cantidad de personas por hogar. Entre estas caracterizaciones se puede mencionar que las mujeres jefas de hogar suelen tener una mayor educación con respecto a los hombres jefes de hogar, además de optar mayoritariamente por empleos de tipo público, teniendo además una cantidad de tres personas o menos en el hogar, a diferencia de los hombres jefes de hogar que suelen caracterizarse más por

tener empleos privados, una cantidad de 4 o más personas en su hogar y tener una mayor cantidad de representación de ascendencia indígena que en el caso de las mujeres jefas de hogar.

En general, si consideramos los estudios anteriores y cómo se comporta la brecha en este momento, sí ha habido una disminución como tal de la brecha en relación a los años posteriores al 2003 en donde realizó la Encuesta CASEN, pero en el país no se logra cumplir con un porcentaje de discriminación tan bajo como lo hubo en el año 2003 que tuvo un valor de 27.5%, por lo que se puede observar que existe una mejora con respecto a la brecha y a su discriminación en el caso de jefes y jefas de hogar, por lo que se espera que esto pueda bajar a lo largo de los próximos años y así se puedan realizar cambios tan relevantes en la discriminación como se pudo lograr hace 20 años.

Al considerar lo anterior mencionado, se puede indicar que se logró cumplir con el objetivo principal de este trabajo que consistía en determinar el estado de la brecha salarial además de conocer algunos de los posibles factores que fuesen determinantes en la explicación de esta brecha, además se cumplieron los objetivos específicos debido a que se realizaron los distintos estudios del plan de muestreo, se realizó una comparación de los ingresos entre los sexos de los jefes de hogar, se encontraron distintas variables sirvieron de utilidad para encontrar nuevos factores de la presencia de la brecha, se aplicaron distintos métodos multivariados y se realizó un análisis global de la brecha salarial entre los jefes y jefas de hogar.

En cuanto a la pregunta realizada al inicio de la investigación, esta logró cumplirse debido a que Chile sigue manteniendo un alto nivel de discriminación con respecto a sus años anteriores ya que, a pesar de que hubo una baja en la discriminación desde el año 2017 de un 49.7% de discriminación a un 33.3% en el año 2020 según los resultados de la descomposición Oaxaca-Blinder, esta sigue superando al porcentaje de discriminación del año 2003 de un 27.5%, el cual a pesar de ser el año con menor discriminación salarial sigue superando con creces al promedio de discriminación otorgado por la OCDE dentro de los países que la integran, que en el año 2018 era de un 14.3%.

En cuanto a lo que se puede mejorar o se puede llevar a cabo en una próxima investigación, puede ser la realización de la descomposición Oaxaca-Blinder a través de otros métodos, tales como la descomposición por dos partes y también el análisis de correspondencias en una mayor cantidad de dimensiones, lo que podría dar paso a un análisis de mayor escala al producido en esta ocasión, además de estudiar con un mayor enfoque las variables correspondientes a los ingresos que no sean los totales, tales como las ganancias por pensiones o bonos.

Es necesario aplicar los distintos tipos de métodos y técnicas utilizadas considerando un tipo de muestreo complejo con datos recolectados a través de encuestas, ya que esto significaría que las varianzas del muestreo complejo tendrían estimaciones totalmente adecuadas para su análisis y su análisis sería más óptimo y recomendable.

Los datos analizados podrían analizarse de una manera más específica, al expandirse para las regiones o comunas con sus pesos muestrales necesarios (comunales o provinciales), de esa manera, se podrían realizar comparaciones de brechas entre distintos sectores, zonas o territorios del país.

Por último, se espera que en el futuro se puedan integrar más métodos y técnicas a los algoritmos de análisis de muestreo complejo en bases de datos de encuestas, ya que estos algoritmos actualmente son bastante limitados y no se pueden realizar acciones demasiado variadas al tenerlos en cuenta.

Apéndice A

Códigos

```
clear all
version 16

use "C:\(...)\Casen en Pandemia 2020 STATA.dta"

//variable jefe de hogar
generate jefe = 1 if pco2 == 1
replace jefe = 2 if pco2 >= 2
label define jefe 1 "Jefe de Hogar" 2 "No es jefe de hogar"
label values jefe jefe
//drop jefe

//rename ingresos
ren ytot ingresos_totales

//survey data para muestras complejas
svyset varunit [pw=expr], strata(varstrat) singleunit(certainty)

svy: tabulate sexo if nucleo !=0, level(95)
svy: tabulate jefe if nucleo !=0, level(95) count format(%12.0f)

svy: mean ingresos_totales if nucleo !=0 & jefe == 1, level(95)

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & sexo_j == 1, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & sexo_j == 2, level(95)
summarize ingresos_totales if nucleo !=0 & jefe == 1 & sexo_j == 2 [fweight=expr], detail

//Edad sin considerar jh
generate Edad = 1 if edad >= 0 & edad <= 17
replace Edad = 2 if edad >= 18 & edad <= 29
replace Edad = 3 if edad >= 30 & edad <= 39
replace Edad = 4 if edad >= 40 & edad <= 49
replace Edad = 5 if edad >= 50 & edad <= 59
replace Edad = 6 if edad >= 60
//drop Edad
label define Edad 1 "0-17 años" 2 "18-29 años" 3 "30-39 años" 4 "40-49 años" 5 "50-59 años"
6 "60+ años"
label values Edad Edad

//Edad jefes de hogar
generate Edad_j2 = Edad_j
replace Edad_j2 = . if Edad_j == 1
label values Edad_j2 Edad
```

```

//drop Edad_j2
tab1 Edad_j2

//svy
svy: tabulate Edad_j2 sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0, level(95) percent
svy: mean Edad_j2 if nucleo !=0, level(95) over(sexo_j)
summarize Edad_j2 [fweight=expr]
tab1 Edad_j2 [fweight=expr]

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & Edad_j2 == 2, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & Edad_j2 == 3, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & Edad_j2 == 4, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & Edad_j2 == 5, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & Edad_j2 == 6, level(95)

summarize ingresos_totales if nucleo !=0 & jefe == 1 & Edad_j2 == 3 [fweight=expr], detail

svy: mean edad if nucleo !=0 & jefe == 1, level(95) over(sexo_j)
summarize edad if nucleo !=0 & jefe == 1 & sexo_j == 2, detail
svy: mean edad if nucleo !=0, level(95) over(sexo)

//Educación no jh
generate educ2 = 1 if educ <= 2
replace educ2 = 1 if educ == 99
replace educ2 = 2 if educ >= 2 & educ <= 4
replace educ2 = 3 if educ >= 5 & educ <= 7
replace educ2 = 3 if educ == 9
replace educ2 = 4 if educ == 8
replace educ2 = 4 if educ >= 10
//drop educ2
label define educ2 1 "Sin educación/Básica incompleta" 2 "Básica completa/Media incompleta"
3 "Media completa/Superior incompleta" 4 "Superior completa"
label values educ2 educ2
label save using labels.do
ren educ2 educacion

//Educación jefe hogar
generate educacion_j = 1 if educacion == 1 & jefe == 1
replace educacion_j = 2 if educacion == 2 & jefe == 1
replace educacion_j = 3 if educacion == 3 & jefe == 1
replace educacion_j = 4 if educacion == 4 & jefe == 1
label values educacion_j educ2
//drop educacion_j

svy: tabulate educacion_j sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate educacion_j sexo_j if nucleo !=0, level(95) percent
svy: mean educacion_j if nucleo !=0, level(95)
tab1 educacion_j [fweight=expr]
tab1 e6a [fweight=expr]

svy: tabulate Edad_j2 sexo_j if nucleo !=0 & educacion_j == 1, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & educacion_j == 2, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & educacion_j == 3, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & educacion_j == 4, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & educacion_j == 4, level(95) percent

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & educacion_j == 1, level(95)

```

```

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & educacion_j == 2, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & educacion_j == 3, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & educacion_j == 4, level(95)

summarize ingresos_totales if nucleo !=0 & jefe == 1 & educacion_j == 1 [fweight=expr], detail

//Tipo de empleo
generate empleo = 1 if o15 == 1
replace empleo = 2 if o15 == 2
replace empleo = 3 if o15 >= 3 & o15 <= 4
replace empleo = 4 if o15 == 5
replace empleo = 5 if o15 >= 6 & o15 <= 7
replace empleo = 6 if o15 == 8
//drop empleo
label define empleo 1 "Patrón o Empleador" 2 "Tr. Independiente" 3 "Púb. Empleado/Obrero"
  4 "Priv. Empleado/Obrero" 5 "Servicio doméstico" 6 "FF.AA. y del Orden"
label values empleo empleo
//label drop empleo

//empleo jefes de hogar
generate empleo_j = 1 if empleo == 1 & jefe == 1
replace empleo_j = 2 if empleo == 2 & jefe == 1
replace empleo_j = 3 if empleo == 3 & jefe == 1
replace empleo_j = 4 if empleo == 4 & jefe == 1
replace empleo_j = 5 if empleo == 5 & jefe == 1
replace empleo_j = 6 if empleo == 6 & jefe == 1
label values empleo_j empleo
//drop empleo_j
//tab1 empleo_j [fweight=expr]

svy: tabulate empleo_j sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate empleo_j sexo_j if nucleo !=0, level(95) percent

svy: tabulate Edad_j2 sexo_j if nucleo !=0 & empleo_j == 1, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & empleo_j == 2, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & empleo_j == 3, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & empleo_j == 4, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & empleo_j == 5, level(95) count format(%12.0f)
svy: tabulate Edad_j2 sexo_j if nucleo !=0 & empleo_j == 6, level(95) count format(%12.0f)

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 1, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 2, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 3, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 4, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 5, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 6, level(95)

summarize ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 5 [fweight=expr], detail

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 1 & sexo_j == 1, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & empleo_j == 1 & sexo_j == 2, level(95)

//ascendencia
ren r3 ascendencia
generate ascen = 1 if ascendencia <= 10
replace ascen = 2 if ascendencia == 11
//drop ascen
label define ascen 1 "Posee ascendencia indígena" 2 "No Posee ascendencia indígena"
label values ascen ascen

```

```

//label drop ascen

//tab1 ascen_j [fweight=expr]
//generate ascen_j = 1 if ascen == 1 & jefe == 1
// replace ascen_j = 2 if ascen == 2 & jefe == 1
//label values ascen_j ascen
//drop ascen_j

svy: tabulate ascen_j sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate ascen_j sexo_j if nucleo !=0, level(95) percent

svy: tabulate ascendencia sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate ascendencia sexo_j if nucleo !=0, level(95) percent

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & ascen_j == 1 & sexo_j == 1, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & ascen_j == 1 & sexo_j == 2, level(95)

summarize ingresos_totales if nucleo !=0 & jefe == 1 & ascen_j == 1 [fweight=expr], detail
summarize ingresos_totales if nucleo !=0 & jefe == 1 & ascen_j == 2 [fweight=expr], detail

//sexo jefe hogar
tab1 sexo_j [fweight=expr]
generate sexo_j = 1 if sexo == 1 & jefe == 1
replace sexo_j = 2 if sexo == 2 & jefe == 1
label values sexo_j sexo
//drop sexo_j

//cantidad personas por hogar
generate numper_j = 1 if numper == 1 & jefe == 1
replace numper_j = 2 if numper == 2 & jefe == 1
replace numper_j = 3 if numper == 3 & jefe == 1
replace numper_j = 4 if numper == 4 & jefe == 1
replace numper_j = 5 if numper == 5 & jefe == 1
replace numper_j = 6 if numper >= 6 & jefe == 1
label define numper 6 "6 o más"
label values numper_j numper

svy: tabulate numper_j sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate numper_j sexo_j if nucleo !=0, level(95) percent

//conyuge en jefes de hogar
//tab1 conyuge_jh if jefe == 1 [fweight=expr]
//tab1 conyuge_j [fweight=expr]
//generate conyuge_j = 1 if conyuge_jh == 1 & jefe == 1
// replace conyuge_j = 0 if conyuge_jh == 0 & jefe == 1
//label values conyuge_j conyuge_jh
//drop conyuge_j
svy: tabulate conyuge_j sexo_j if nucleo !=0, level(95) count format(%12.0f)
svy: tabulate conyuge_j sexo_j if nucleo !=0, level(95) percent
browse conyuge_j
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & conyuge_j == 0, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & conyuge_j == 1, level(95)

summarize ingresos_totales if nucleo !=0 & jefe == 1 & conyuge_j == 0 [fweight=expr], detail
summarize ingresos_totales if nucleo !=0 & jefe == 1 & conyuge_j == 1 [fweight=expr], detail

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & numper_j == 1, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & numper_j == 2, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & numper_j == 3, level(95)

```

```

svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & numper_j == 4, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & numper_j == 5, level(95)
svy: mean ingresos_totales if nucleo !=0 & jefe == 1 & numper_j == 6, level(95)

//oaxaca con descomposición de tres partes, primera variable son los ingresos y las subsiguientes
//las variables explicativas
oaxaca ingresos_totales Edad_j educacion_j empleo_j ascen_j conyuge_j numper_j [fweight=expr],
    by(sexo) noisily

svy: regress ingresos_totales sexo_j

//oaxaca con descomposición de dos partes
oaxaca ingresos_totales educacion hijos empleo ascendencia [fweight=expr], by(sexo) pooled

//MCA con factores de expansión en dos dimensiones
mca sexo_j educacion_j Edad_j2 empleo_j ascen_j numper_j [fweight=expr]
//gráfico visible, todos los puntos en el mismo gráfico, no admite factores de expansión esta
//versión del gráfico
mcaplot sexo_j educacion_j Edad_j2 empleo_j ascen_j numper_j, overlay origin scale(0.45)
    aspectratio(1.1) title(Gráfico de coordenadas ACM) xtitle(Dimensión 1) ytitle(Dimensión 2)
    legend(label(1 "Sexo") label(2 "Nivel educación") label(3 "Edad") label(4 "Tipo de Empleo")
        label(5 "Ascendencia"))

//Gráficos
//Edad
graph hbar (count) [fweight=expr], over(sexo_j) over(Edad_j2) asyvars title(Edad de jefes de hogar
    con factores de expansión regionales, size(medium)) ytitle(Observaciones (en miles)) ylabel(250000
    "250" 500000 "500" 750000 "750" 1000000 "1.000" 1250000 "1.250" 1500000 "1.500",format(%12.0f))
//Edad salarios
graph hbar ingresos_totales [fweight=expr], over(sexo_j) over(Edad_j2) asyvars title(Ingresos por
    edad de jefes de hogar con factores de expansión regionales, size(medium)) ytitle(Media de ingresos
    (en miles) (CLP)) ylabel(200000 "200" 400000 "400" 600000 "600" 800000 "800" 1000000 "1.000",
    format(%12.0f))
//Educación
graph hbar (count) [fweight=expr], over(sexo_j) over(educacion_j, label(labsize(vsmall)))
    asyvars title(Nivel educacional jefes de hogar con factores de expansión regionales,
    size(medium)) ytitle(Observaciones (en miles)) ylabel(250000 "250" 500000 "500" 750000 "750"
    1000000 "1.000" 1250000 "1.250" 1500000 "1.500",format(%12.0f)) legend(size(small))
//Educación salarios
graph hbar ingresos_totales [fweight=expr], over(sexo_j) over(educacion_j, label(labsize(vsmall)))
    asyvars title(Ingresos por nivel educacional en jefes de hogar con factores de expansión regionales,
    size(medium)) ytitle(Media de ingresos (en miles) (CLP)) ylabel(250000 "250" 500000 "500" 750000
    "750" 1000000 "1.000" 1250000 "1.250" 1500000 "1.500",format(%12.0f)) legend(size(small))
//Empleo
graph hbar (count) [fweight=expr], over(sexo_j) over(empleo_j, label(labsize(small))) asyvars
    title(Tipo de empleos en jefes de hogar con factores de expansión regionales, size(medium))
    ytitle(Observaciones (en miles)) ylabel(250000 "250" 500000 "500" 750000 "750" 1000000 "1.000"
    1250000 "1.250" 1500000 "1.500",format(%12.0f)) legend(size())
//Empleo salarios
graph hbar ingresos_totales [fweight=expr], over(sexo_j) over(empleo_j, label(labsize(vsmall)))
    asyvars title(Ingresos por tipo de empleos en jefes de hogar con factores de expansión regionales,
    size(small)) ytitle(Media de ingresos (en miles) (CLP)) ylabel(250000 "250" 500000 "500" 750000
    "750" 1000000 "1.000" 1250000 "1.250" 1500000 "1.500" 1750000 "1.750",format(%12.0f)) legend(size())
//Ascendencia
graph hbar (count) [fweight=expr], over(sexo_j) over(ascen_j, label(labsize(small))) asyvars
    title(Ascendencia indígena en jefes de hogar con factores de expansión regionales, size(medium))
    ytitle(Observaciones (en millones)) ylabel(1000000 "1" 2000000 "2" 3000000 "3" 4000000 "4",
    format(%12.0f)) legend(size())
//Ascendencia salarios

```

```

graph hbar ingresos_totales [fweight=expr], over(sexo_j) over(ascen_j, label(labsize(small)))
  asyvars title(Ingresos por ascendencia indígena en jefes de hogar con factores de expansión regionales,
  size(medium)) ytitle(Media de ingresos (en miles) (CLP)) ylabel(200000 "200" 400000 "400" 600000 "600"
  800000 "800",format(%12.0f)) legend(size(small))
//sexo
graph hbar (count) [fweight=expr], over(sexo_j) asyvars title(Sexo en jefes de hogar con factores
  de expansión regionales) ytitle(Observaciones (en millones)) ylabel(1000000 "1" 2000000 "2"
  3000000 "3" 4000000 "4",format(%12.0f))
//sexo no jefes
graph hbar (count) [fweight=expr], over(sexo) asyvars title(Sexo de la población chilena con factores
  de expansión regionales, size(medium)) ytitle(Observaciones (en millones)) ylabel(1000000 "1" 2000000
  "2" 3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6" 7000000 "7" 8000000 "8" 9000000 "9" 10000000 "10"
  11000000 "11",format(%12.0f))
//ingresos
graph hbar ingresos_totales [fweight=expr], over(sexo_j) asyvars title(Ingresos por sexo en jefes de
  hogar con factores de expansión regionales, size(medium)) ytitle(Media de ingresos (en miles) (CLP))
  ylabel(100000 "100" 200000 "200" 300000 "300" 400000 "400" 500000 "500" 600000 "600" 700000 "700"
  800000 "800" ,format(%12.0f))
//Conyuge
graph hbar (count) [fweight=expr], over(sexo_j) over(conyuge_j, label(labsize(small))) asyvars
  title(Presencia de pareja o cónyuge en jefes de hogar con factores de expansión regionales,
  size(medium)) ytitle(Observaciones (en millones)) ylabel(1000000 "1" 2000000 "2" 3000000 "3",
  format(%12.0f)) legend(size())
//Conyuge salarios
graph hbar ingresos_totales [fweight=expr], over(sexo_j) over(conyuge_j, label(labsize(small)))
  asyvars title(Ingresos por presencia de pareja o cónyuge en jefes de hogar con factores de expansión
  regionales, size(medium)) ytitle(Media de ingresos (en miles) (CLP)) ylabel(200000 "200" 400000
  "400" 600000 "600" 800000 "800",format(%12.0f)) legend(size(small))
//jefe
graph hbar (count) [fweight=expr], over(sexo) over(jefe, label(labsize(small))) asyvars title(Jefes
  de hogar en Chile con factores de expansión regionales, size(medium)) ytitle(Observaciones
  (en millones)) ylabel(1000000 "1" 2000000 "2" 3000000 "3" 4000000 "4" 5000000 "5" 6000000 "6",
  format(%12.0f)) legend(size())
//numper
graph hbar (count) [fweight=expr], over(sexo_j) over(numper_j, label(labsize(small))) asyvars
  title(Cantidad de personas por hogar con factores de expansión regionales, size(medium))
  ytitle(Observaciones (en miles)) ylabel(200000 "200" 400000 "400" 600000 "600" 800000 "800"
  1000000 "1000",format(%12.0f)) legend(size())
//numper salarios
graph hbar ingresos_totales [fweight=expr], over(sexo_j) over(numper_j, label(labsize(small)))
  asyvars title(Ingresos por presencia de pareja o cónyuge en jefes de hogar, size(medium))
  ytitle(Media de ingresos (en miles) (CLP)) ylabel(200000 "200" 400000 "400" 600000 "600"
  800000 "800",format(%12.0f)) legend(size(small))

```

Referencias

- Abrahamsen, P. (1997). *A review of gaussian random fields and correlation functions*. Oslo, Noruega: Norwegian Computing Center.
- Antognini, A. F., y Trebilcock, M. P. (2021). Pandemia, inequidad y protección social neoliberal: Chile, un caso paradigmático. *Brazilian Journal of Latin American Studies*, 20(40), 189–209.
- Arceo-Gómez, E. O., y Campos-Vázquez, R. M. (2014). Evolución de la brecha salarial de género en México. *El trimestre económico*, 81(323), 619–653.
- Arellano-Valle, R. B., Gómez, H. W., y Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128(2), 427 - 443. Descargado de <http://www.sciencedirect.com/science/article/pii/S0378375804000254> doi: <https://doi.org/10.1016/j.jspi.2003.11.014>
- Castillo, D. O. V., y Loredó, M. A. F. (2019). Desigualdad salarial en los servicios públicos en Chile: Un estudio de caso servicio nacional de aduanas de Chile. *Neumann Business Review*, 5(2), 04–25.
- Castillo, J. C., y cols. (2009). ¿cuál es la brecha salarial justa? opinión pública y legitimación de la desigualdad en Chile. *Estudios Públicos*(113).
- Cepeda Emiliani, L., Cardona Sosa, L., y Barón, J. D. (2014). La brecha salarial de género entre los universitarios recién titulados ¿qué tanto influye el tipo de carrera universitaria? *El trimestre económico*, 81(322), 441–477.
- CNN. (2020). *De los países de la OCDE, Chile es el quinto con más brecha salarial entre hombres y mujeres*. Descargado de https://edition.cnn.com/pais/de-los-paises-de-la-ocde-chile-es-el-quinto-con-mas-brecha-salarial-entre-hombres_20180502/
- Datosmacro. (2022). *IPC de Chile*. <https://datosmacro.expansion.com/ipc-paises/chile>.
- European Commission. (2021). *The gender pay gap situation in the EU*. Descargado de https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en
- Ferrada, J. N. (2018). Discriminación salarial entre hombres y mujeres en Chile. *Revista Némesis*(14), 41–64.
- Fuentes, J., Palma, A., y Montero, R. (2005). Discriminación salarial por género en Chile, una mirada. *Estudios de economía*, 32(2), 133.
- Glynn, D. (2014). Correspondence analysis. *Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 443–486.
- Godoy, G. (2021). *El ingreso laboral promedio mensual en Chile fue de \$681.039 en 2021*. [https://www.ine.cl/prensa/2022/07/21/el-ingreso-laboral-promedio-mensual-en-chile-fue-de-\\$681.039-en-2021#:~:text=El%20ingreso%20laboral%20promedio%20mensual%20en%20Chile%20fue%20de%20\\$24681.039%20en%202021,-Jul%2021%2C%202022&text=El%2050%25%20de%20las%20personas,ingreso%20de%20hasta%20\\$24457.690%20mensuales](https://www.ine.cl/prensa/2022/07/21/el-ingreso-laboral-promedio-mensual-en-chile-fue-de-$681.039-en-2021#:~:text=El%20ingreso%20laboral%20promedio%20mensual%20en%20Chile%20fue%20de%20$24681.039%20en%202021,-Jul%2021%2C%202022&text=El%2050%25%20de%20las%20personas,ingreso%20de%20hasta%20$24457.690%20mensuales).

- Godoy Rodríguez, G. C. (2021). *Brechas salariales y género en el Ecuador del año 2011-2018* (B.S. thesis). Universidad Nacional de Chimborazo.
- Gross, J., y Groß, J. (2003). *Linear regression* (Vol. 175). Springer Science & Business Media.
- Husson, F., Lê, S., y Pagès, J. (2011). *Exploratory multivariate analysis by example using R* (Vol. 15). CRC press Boca Raton.
- Jann, B. (2008). The blinder–oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453–479.
- Kremerman, G. D.-M. (2021). La pobreza del 'modelo' chileno: la insuficiencia de los ingresos del trabajo y las pensiones.
- Mandiola, M. (2021). Desafíos para la igualdad de género en tiempos de pandemia. *Observatorio Económico*(152), 4–5.
- MDSF. (2021a). Diseño muestral casen en pandemia 2020 [Manual de software informático].
- MDSF. (2021b). *Encuesta CASEN en pandemia*. <http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-en-pandemia-2020#:~:text=Los%20objetivos%20de%20Casen%20en,%2C%20vivienda%2C%20trabajo%20e%20ingresos>.
- MDSF. (2021c). Manual del investigador casen en pandemia 2020 [Manual de software informático].
- Ministerio de Desarrollo Social y Familia. (2020). *Encuesta Casen en Pandemia*. Descargado de <http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-en-pandemia-2020#:~:text=Los%20objetivos%20de%20Casen%20en,%2C%20vivienda%2C%20trabajo%20e%20ingresos>.
- Morales H., C. (2017, 02). *Ocde: Chile es el quinto país con mayor brecha salarial por género*. Descargado de <https://www.latercera.com/noticia/ocde-chile-quinto-pais-mayor-brecha-salarial-genero/>
- Muñoz, C., y Aguilar, F. (2021). La sobrevivencia en la vejez en Chile: lo que muestra la pandemia sobre la protección social. *La pandemia social de COVID-19 en América Latina: reflexiones desde la Salud Colectiva. Ciudad Autónoma de Buenos Aires, Argentina: Teseo*, 251–272.
- OCDE. (2018). Estudios económicos de la OCDE: Chile 2018.
- Paredes Molina, R., y Riveros Cornejo, L. (1994). Gender wage gaps in Chile. a long term view: 1958-1990.
- Paz, J. A. (2019). La brecha salarial por género en Argentina: un análisis acerca de la segmentación laboral. *Sociedad e cultura*, 22(1), 157–178.
- Peña, D. (2002). *Análisis de datos multivariantes* (Vol. 24). McGraw-hill Madrid.
- Romo, G. (18, 18). *Medicamentos genéricos se imponen frente al alza de precios en Chile*. <https://www.eldinamo.cl/pais/Medicamentos-genericos-se-imponen-frente-al-alza-de-precios-en-Chile-20221118-0030.html>.
- Salce Díaz, F. (2021). Evolución y análisis de la discriminación salarial por género en Chile. *El trimestre económico*, 88(349), 39–75.
- Sánchez, C. P., Uriguen Aguirre, P., y Vega Jaramillo, F. (2021). Discriminación y desigualdad salarial. exploración de brechas por género en Ecuador.
- Sykes, A. O. (1993). An introduction to regression analysis.
- Szasz Pianta, I. (1995). Mujeres y migrantes: desigualdades en el mercado laboral de Santiago de Chile. *Revista de la CEPAL*.