



Departamento de
Ingeniería Industrial
Universidad de Concepción

**BRECHA SALARIAL DE GÉNERO EN CHILE:
REGRESIÓN LINEAL VERSUS REGRESIÓN SIMBÓLICA
POR**

Josefina Jacqueline Barrera Labrín

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción
para optar al título profesional de Ingeniera Civil Industrial.

Profesores guía:

Marcela Parada Contzen
Carlos Contreras Bolton

Agosto, 2025
Concepción (Chile)

© 2025, Josefina Jacqueline Barrera Labrín Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo siempre la cita bibliográfica del presente documento y su autor.

Dedicado a mi familia, amigos y a toda persona que me entrego algún aprendizaje.

Son una inspiración cada día.

Agradecimientos

Quiero comenzar agradeciendo a todos los que han sido parte de este proceso. No podría haberlo realizado sola. Gracias por sus cariños, palabras, risas e interminables tazas de cafés.

Gracias a mi papá, Juan Carlos, quien siempre me ha motivado a soñar en grande. Me ha inculcado la valentía, compromiso y amor por las cosas que realizo. Gracias a mi mamá, Sara, por sus abrazos llenos de amor y sus palabras cargadas de sabiduría. Gracias a mi hermano, Martín, por ser la persona que me tiene más paciencia.

A mis amigos, Josué y Claudio, por siempre confiar en mí. A mi amiga Valentina, por tener siempre la palabra precisa y el calor de hogar. A Camila, por ser mi fiel compañera todos estos años en los cuales hemos crecido juntas. A Ónice y Daniela por siempre hacerme reír.

Quisiera agradecer a mi familia, por creer en mí y tener certeza de que lo podría cumplir. A mi tía Andrea, por ser mi fiel confidente, a mis primas y primos por siempre mostrarme que nada es tan serio. A mi tío Sergio, un saludo al cielo. No puedo olvidar a mis abuelos, por siempre estar, sean eternos. A mi familia y amigos en general, gracias por motivarme y acompañarme. Han sido una fuente de aprendizaje y risas.

No puedo olvidar a mis profesores guía, Marcela y Carlos. Mis más sinceros agradecimientos a ellos, siempre dispuestos a responder la infinidad de preguntas que tenía.

Finalmente, agradezco a todos los que han sido parte de este largo proceso, a los que han llegado y a aquellos que han permanecido muchas gracias por estar. Me agradezco a mí, porque lo que parecía imposible no lo fue.

Sumario

La brecha salarial de género se define como la diferencia de salario que recibe una persona con respecto a otra debido a su género, manteniendo constantes todas las variables. Es un tema ampliamente estudiado en la literatura y existe gran diversidad de resultados. En Chile, la literatura también presenta una gran diversidad de resultados dependiendo de los métodos empíricos utilizados.

La presente memoria de título estima la brecha salarial de género en Chile utilizando dos métodos de estimación. Por un lado, el método tradicional de la literatura de regresión lineal múltiple. Por otro lado, el método de regresión simbólica, un algoritmo que utiliza Machine Learning para estimar una ecuación en base a los datos.

Se utilizaron datos de la Encuesta de Caracterización Socioeconómica Nacional (CASEN) para 1992, 2000, 2011 y 2022. Para ambos métodos empíricos, se trabaja con una ecuación de salario a nivel individual donde la variable dependiente es el logaritmo natural del salario por hora del trabajador y el coeficiente clave es el parámetro que acompaña a la variable mujer (o sexo del trabajador).

Los resultados indican que el método que mejor ajusta los datos, considerando el error cuadrático medio (MSE), es la regresión lineal múltiple. La cual obtiene una disminución del 8,83% del MSE para 2022. La brecha salarial de género en Chile para 2011 y 2022 utilizando regresión lineal es de 16,29% y 11,76%, respectivamente. Mientras que utilizando regresión simbólica para 2011 es de 20,50% y para 2022 de 13,05%. Se aprecia que utilizando ambos métodos la brecha salarial de género disminuye en 2022 con respecto a las estimaciones para periodos anteriores. Dependiendo de la metodología, cambia la magnitud en la que disminuye la brecha salarial de género, obteniendo una mayor disminución cuando se utiliza regresión simbólica.

Se concluye finalmente que el método de regresión lineal permite mejorar la precisión de la estimación de la brecha salarial de género. Es importante destacar que los resultados respecto a esta brecha son distintos, pero cercanos en magnitud. Al utilizar regresión simbólica, la brecha salarial tiene mayores valores lo que puede tener impacto para el diseño de políticas.

Abstract

The gender wage gap is defined as the difference in salary that a person receives with respect to another due to their gender, keeping all variables constant. It is a widely studied topic in literature and there is a great diversity of results. In Chile, the literature also presents a great diversity of results depending on the empirical methods used.

This Thesis estimates the gender wage gap in Chile using two estimation methods. On one hand, the traditional method of multiple linear regression literature. On the other hand, the symbolic regression method, an algorithm that uses Machine Learning to estimate an equation based on data.

Data from the National Socioeconomic Characterization Survey (CASEN) were used for 1992, 2000, 2011 and 2022. For both empirical methods, a wage equation at the individual level is used, where the dependent variable is the natural logarithm of the worker's hourly salary, and the key coefficient is the parameter that accompanies the woman variable (or worker's sex).

The results indicate that the method that best fits the data, considering the Mean Square Error (MSE), is symbolic regression, obtaining an 8,83% reduction in MSE for 2022. The gender wage gap in Chile for 2011 and 2022 using multiple linear regression is 13,29% and 11,76%, respectively. Meanwhile, using symbolic regression, it is 20,50% for 2011 and 13,05% for 2022. Using both methods, the gender wage gap decreases in 2022 compared to estimates for previous periods. Depending on the methodology, the magnitude of the gender wage gap decrease varies, obtaining a greater decrease when using symbolic regression.

Finally, it is concluded that the multiple linear regression method improves the accuracy of estimation of the gender wage gap. It is important to note that results regarding this gap are different, but similar in magnitude. When using symbolic regression, the gender wage gap has larger values, which can impact policy design.

Índice de Contenido

| | |
|--|----|
| 1. Introducción..... | 1 |
| 1.1 Objetivos de la memoria..... | 2 |
| 1.1.1 Objetivo general | 2 |
| 1.1.2 Objetivo específico | 2 |
| 1.2 Alcances y limitaciones | 2 |
| 1.3 Organización del documento | 3 |
| 2. Revisión de literatura..... | 4 |
| 3. Metodología..... | 13 |
| 3.1 Método empírico | 13 |
| 3.2 Método de estimación..... | 13 |
| 3.2.1 Mínimos Cuadrados Ordinarios | 13 |
| 3.2.2 Metodología Regresión Simbólica | 14 |
| 3.3 Pruebas estadísticas | 15 |
| 3.3.1 R^2 y R^2 ajustado | 15 |
| 3.3.2 Error cuadrático medio | 16 |
| 3.3.3 Prueba F | 16 |
| 4. Datos..... | 18 |
| 4.1 Fuente de datos | 18 |
| 4.2 Creación de la muestra de estimación | 18 |
| 4.2.1 Modificación de las variables | 19 |
| 4.3 Descripción de la muestra de estimación | 20 |
| 4.4 Muestra para distintos períodos..... | 23 |
| 5. Resultados..... | 29 |
| 5.1 Regresión lineal múltiple..... | 29 |
| 5.1.1 Regresión lineal múltiple para 2022..... | 30 |
| 5.1.1 Regresión lineal múltiple para 1992, 2000 y 2011 | 32 |
| 5.2 Regresión simbólica | 36 |
| 5.2.1 Primeros experimentos de regresión simbólica | 37 |
| 5.2.2 Regresión simbólica para 2022..... | 37 |

| | |
|--|----|
| 5.2.3 Regresión simbólica para 1992, 2000 y 2011 | 38 |
| 5.3 Regresión lineal vs Regresión simbólica..... | 40 |
| 6. Conclusión..... | 42 |
| 7. Referencias | 45 |
| 8. Anexos | 50 |

Índice Figuras

| | |
|--|----|
| Figura 1: Resultados de la regresión lineal múltiple y regresión simbólica..... | 41 |
|--|----|

Índice Tablas

| | |
|--|----|
| Tabla 1: Revisión de la literatura | 8 |
| Tabla 2: Características de la encuesta CASEN 2022 | 18 |
| Tabla 3: Evolución del tamaño de muestra para 2022..... | 19 |
| Tabla 4: Variables de la investigación CASEN 2022..... | 20 |
| Tabla 5: Estadísticas descriptivas de las variables de CASEN 2022..... | 22 |
| Tabla 6 : Muestras CASEN para distintos períodos | 24 |
| Tabla 7: Estadísticas descriptivas de las variables CASEN 1992 | 24 |
| Tabla 8: Estadísticas descriptivas de las variables CASEN 2000 | 26 |
| Tabla 9: Estadísticas descriptivas de las variables CASEN 2011..... | 27 |
| Tabla 10: Variable eliminada en cada categoría en base a la mayor frecuencia para cada muestra. | 30 |
| Tabla 11: Resultados regresión lineal múltiple para el año 2022 | 31 |
| Tabla 12: Resultados regresión lineal múltiple para el año 1992 | 32 |
| Tabla 13: Resultados regresión lineal múltiple para el año 2000 | 33 |
| Tabla 14: Resultados regresión lineal múltiple para el año 2011 | 34 |
| Tabla 15: Resultados brecha salarial de género en Chile para los últimos 30 años utilizando regresión lineal múltiple. | 35 |
| Tabla 16: Comparación información entrega por la regresión lineal y regresión simbólica | 41 |
| Tabla 17: Variables de la investigación CASEN 1992 | 50 |
| Tabla 18: Variables de la investigación CASEN 2000 | 51 |
| Tabla 19: Variables de la investigación CASEN 2011 | 53 |

1. Introducción

La brecha salarial de género se define como la diferencia salarial entre hombres y mujeres considerando constantes otras características como variables de capital humano y de características del puesto trabajo (Kristjanpoller et al., 2023). La brecha salarial de género ha sido ampliamente estudiada en diferentes periodos y países considerando distintas características socioeconómicas y demográficas relevantes (Si et al., 2020; Meriküll & Tverdostup, 2023).

Según la Organización Internacional del Trabajo (2022), la brecha salarial de género a nivel mundial es en promedio de un 20%. Para los países desarrollados, la evidencia muestra que la brecha ha disminuido (Kunze, 2017). Hay evidencia diversa para Latinoamérica. Por ejemplo, para el caso de México la brecha salarial de género ha disminuido, mientras que en Costa Rica ha aumentado. En 2018, las mujeres ganaban 14% y 8,5% de los ingresos medios de los hombres en México y Costa Rica, respectivamente. Mientras que, en 2023, México aumento a 15% y Costa Rica a 4,1% (OCDE, 2023). En Chile, se encuentra un aumento de la brecha en los últimos años (Kristjanpoller et al., 2023). Según el INE (2023) la brecha salarial de género en Chile en 2021 fue de 18,9%.

Resultados anteriores, indican que la brecha salarial es mayor para los cuartiles más pobres (Vaccaro et al., 2022) y ocupados informales (OCEC udp, 2022). En Chile, hay evidencia que señala que la brecha salarial es mayor en la parte superior de distribución de salarios (Didier, 2021). La evidencia señala que la brecha salarial de género aumenta con la edad y número de hijos (Goldin, 2014), además le afectan las normas culturales, actitudes de género y prejuicio, característica predominante en Chile en comparación a otros países de Latinoamérica (Meriküll & Tverdostup, 2023; Perticará & Tejada, 2022).

En la presente memoria de título se estima la brecha salarial de género en Chile para 1992, 2000, 2011 y 2022. La contribución de esta investigación está asociada a las metodologías usadas. En este contexto, se comparan y discuten dos metodologías de estimación de brecha salarial de género. El primer método, está basado en una ecuación lineal de Mincer (1974) estimado vía regresión lineal múltiple. El segundo método es la regresión simbólica, la cual proviene del Machine Learning o aprendizaje automático (Kronberger et al., 2025.). Esta estima la brecha salarial de género sin especificar una forma funcional predeterminada, sino

que ésta va a ser resultado del problema de minimización de error cuadrático medio. De esta manera, se comparan los resultados y estimaciones utilizando ambos métodos. Así, esta memoria de título contribuye a la literatura con evidencia proveniente de la aplicación métodos de estimación modernos basados en aprendizaje automático.

Se utilizan las Encuestas CASEN para los años 1992, 2000, 2011 y 2022 con el fin de ver el cambio que ha tenido la brecha salarial de género en Chile en las últimas tres décadas (Ministerio de desarrollo social y familia, 1992, 2000, 2011, 2022). Se consideran entre 42 a 52 variables, divididas entre 11 a 14 categorías dependiendo de la información de cada encuesta.

1.1 Objetivos de la memoria

A continuación, se menciona el objetivo general y los objetivos específicos que tiene esta memoria de título.

1.1.1 Objetivo general

Estimar la brecha salarial de género en Chile para el periodo comprendido entre los años 1992 a 2022 con datos de la Encuesta CASEN para 1992, 2000, 2011 y 2022 bajo métodos de regresión lineal múltiple y regresión simbólica.

1.1.2 Objetivo específico

1. Revisar y analizar el estado del arte respecto a la brecha salarial de género.
2. Estimar la brecha salarial de género por medio de métodos de regresión lineal múltiple y regresión simbólica.
3. Comparar los métodos utilizados, además de analizar y discutir los resultados.

1.2 Alcances y limitaciones

La memoria de título entrega información sobre la brecha salarial de género en Chile a través de tiempo, mostrando su evolución. Para ello, se utiliza la Encuesta CASEN para los años 1992, 2000, 2011 y 2022, la cual es una encuesta representativa de la población chilena.

Las muestras de estimación están conformadas por todos los individuos que están trabajando y que tienen información de las variables clave del modelo. Las variables varían dependiendo de la información disponible en la encuesta.

Este trabajo utiliza dos métodos. Un enfoque tradicional como la regresión lineal múltiple. Mientras que la segunda metodología es un enfoque moderno, utilizando técnicas de Machine Learning, como la regresión simbólica.

La principal limitación del estudio de la brecha salarial de género es la falta de información en las bases de datos de hace más de 20 años. Aquello provoca una falta de conocimiento con respecto a variables relevantes para el análisis. Lo que produce que no se pueda hacer un análisis completo de la brecha salarial de género en Chile a lo largo del tiempo. Otra limitación de esta memoria tiene relación con que no se incorpora una ecuación que tenga relación con la decisión de participar o no en el mercado, solo se trabaja con individuos con empleo. Por otra parte, una limitación es la complejidad de la ecuación estimada por la regresión simbólica, debido a que se debe interpretar la contribución de la variable mujer en ella, es por ello, que llega a una expresión lineal.

1.3 Organización del documento

Esta memoria de título se divide por secciones. En la sección 2, se presenta la revisión de la literatura. En la sección 3, se presenta la metodología, tanto el modelo empírico como el método de estimación utilizado. Luego en la sección 4, se presentan los datos, considerando su fuente, creación y descripción de muestra. A continuación, en la sección 5, se muestran los resultados obtenidos y su discusión. Finalmente, se presentan las conclusiones en la sección 6.

2. Revisión de literatura

La brecha salarial de género históricamente ha afectado a las mujeres, obteniendo un menor ingreso por su trabajo en comparación a los hombres (Salazar Aramayo et al., 2025). La evidencia señala que la brecha salarial por género en el mundo ha disminuido de forma progresiva (Blau & Kahn, 2017). En los últimos años, la brecha salarial por género ha sido en promedio de un 20%. Mientras en América Latina, el salario de las mujeres con relación al de los hombres es de 87,7% en promedio. (OIT, 2020; OIT, 2024). La disminución de la brecha salarial de género se debe entre algunas de las razones al mayor acceso de la educación por parte de las mujeres y aumento de las experiencias laborales (Blau & Kahn, 2017).

Según la OCDE (2023), Estados Unidos posee una brecha salarial de género en 2019 de 18,5% y de 16,4% para 2023, muy similar a la de Canadá 16,1% para ese mismo año. En los países desarrollados, se destacan Bélgica, Irlanda y Dinamarca por sus bajos valores de la brecha; recibiendo las mujeres el 90% del ingreso de los hombres, valor superior a países como Japón y Corea del Sur donde las cifras son del 73% y 63%, respectivamente (Olivetti & Petrongolo, 2016). En Europa, las mujeres tienen un salario promedio correspondiente al 73,6% de los hombres (Cabrera et al., 2024). Si vemos países de la Unión Europea, se tiene el caso de Estonia, que posee la mayor brecha salarial de género de este conjunto de países, la cual es de 25% aproximadamente (Meriküll & Tverdostup, 2023). En Latinoamérica, países como Perú tuvieron una brecha salarial de aproximadamente de 20% (Ortogorin, 2024), mientras que México y Uruguay del 7% y 8%, respectivamente (Perticará & Tejada, 2022).

La evidencia muestra que la brecha salarial de género en Chile ha aumentado en el último tiempo (Kristjanpoller et al., 2023). Hay diversas estimaciones y valores para la brecha de Chile donde la brecha salarial de género se encuentra entre un 9% y 33% aproximadamente (Díaz, 2021; INE, 2023). Según la OCDE (2023), Chile tuvo una brecha salarial de género de un 13,6%, mientras que el INE (2023) señala una brecha de 18,9%. A su vez, Aguayo (2023) determina una brecha salarial para 2022 de 16%, valor que aumenta en las estimaciones realizadas por Sánchez et al. (2022) y Didier (2021) en las que obtienen una brecha de aproximadamente un 22% y 28%, respectivamente.

La brecha salarial de género depende también de características del mercado laboral como precios, rentabilidad, oferta y demanda (Blau & Kahn, 2017) . Según la OCDE (2025), Chile posee una variación porcentual del Producto Interno Bruto (PIB) de 2,1% y una tasa de desempleo del 8,3%, en comparación a la tasa mundial proporcionada por el Banco Mundial (2024), la cual es de 5,4%.

Según el Banco Mundial (2024) hay 4,01 mil millones de mujeres en el mundo, correspondiendo aproximadamente al 49,7% de la población mundial. Aun cuando las mujeres son la mitad del total de personas, la participación de ellas en la fuerza laboral ha sido inferior a la de los hombres, aunque ha ido en aumento en los últimos años (Semenza et al., 2021). Si nos centramos en Chile, hay alrededor de 50,3% de población femenina. Según el INE (2023), la participación femenina en el mercado laboral fue de 50,8%, mientras que los hombres tienen una de 70,2%.

El aumento de las mujeres en el mercado laboral permite reducir la brecha de la participación, lo que a su vez es un indicador de una disminución de las diferencias de educación, horas trabajadas ocupaciones (Goldin, 2014). Aun así, el aumento de la participación no indica necesariamente una disminución en la brecha salarial de género (Meriküll & Tverdostup, 2023). Para Goldin (2014), el salario es una muestra de cómo están otros aspectos de un país, ya sea educación, formación y ocupación. Además, posee implicancias económicas, sociales, culturales e incluso relacionadas con la violencia domestica (Ortogorin, 2024).

Goldin (2014) señala que la persistencia de la brecha salarial está asociada también a diferencias internas entre géneros, tales como que la menor capacidad de las mujeres para negociar y su menor deseo de competir. Según el OCEC udp (2022) la brecha ha persistido en Chile debido a que las mujeres tienen trabajos con menores ingresos y desarrollo profesional.

En comparación con los demás países de Latinoamérica, Chile es uno de los países con mayor apoyo hacia las mujeres, en contraste con el alto nivel de prejuicio que posee (Perticará & Tejada, 2022). Por otra parte, la falta de flexibilidad horaria en el trabajo (Goldin, 2014) dificulta que las mujeres puedan estar a cargo de los hijos y familia (Si et al., 2020). La evidencia señala que la presencia de hijos en las mujeres contribuye al aumento de la brecha salarial de género. Ejemplos de lo anterior son: Estados Unidos y Colombia (Meara et al.,

2020; Vargas Calderón et al., 2024). El INE (2023) señala que alrededor de un 33% de las mujeres que no está en el mercado laboral es debido a razones familiares.

Si et al. (2020) atribuye la brecha salarial de género a cuatro características: educación, experiencia, ocupación y discriminación. La evidencia señala que actualmente las mujeres tienen más años de educación que los hombres (Vaccaro et al., 2022). Además, Didier (2021) menciona que las mujeres “deben elegir entre estar sobreeducadas y soportar la diferencia salarial procedente del título”. Vaccaro (2022) señala que las mujeres menos educadas y en contextos de trabajo informal poseen una mayor brecha salarial.

Las mujeres han tenido que incluirse en los mercados informales, donde el sueldo es menor, debido a que en esta forma no encuentran barreras de entrada y pueden compatibilizar con sus labores del hogar (Paz, 2019). Es debido a esto que están en menor proporción en empresas grandes (Bonaccolto-Töpfer et al., 2023).

A su vez, Salazar Aramayo et al. (2025) menciona que el hecho de ser mujer disminuye en 4.71% las posibilidades de encontrar un empleo categorizado como de “buena calidad”. Por otra parte, Cabrera et al. (2024) señala que la discriminación por parte de las empresas surge debido a que contratar a hombres en comparación con mujeres es mucho más seguro y estable debido a que las mujeres cuentan con situaciones como embarazo, maternidad y familia. Asimismo, las mujeres no reciben un salario acorde incluso en la parte superior de la distribución salarial (Bonaccolto-Töpfer et al., 2023).

Vargas Calderón (2024) extiende el análisis indicando que en un comienzo y por largo tiempo se buscó explicar la brecha salarial utilizando la economía, lo cual produjo que no se pudiera comprender completamente, ya que sus factores no son solo económicos, sino que hay elementos sociales y culturales. Algunos factores son normas y actitudes persistentes arraigados a la población (Meriküll & Tverdstup, 2023). Es debido a esto, que para estimar la brecha salarial de género se debe considerar que la región de la estimación y el periodo de estudio (Rembeza & Radlińska, 2021).

En la mayoría de los estudios se estima la brecha salarial de género utilizando regresiones lineales y métodos de descomposición de medias. La descomposición más usada es la de Blinder-Oaxaca (Blinder, 1973; Oaxaca, 1973), la que permite identificar la contribución de

factores observadas y de factores no observados (para el investigador) de la brecha salarial de género. Las variables observadas habitualmente utilizadas son variables de capital humano tales como educación y experiencia, edad y características del lugar de trabajo, tales como sector económico, ocupación, jerarquía, tamaño de la empresa, ubicación geográfica, entre otras (Kristjanpoller et al., 2023; Meriküll & Tverdostup, 2023; Ortogorin, 2024).

Al estimar la brecha salarial de género, dentro de las interpretaciones que se pueden hacer, se evalúa si existe “Techo de cristal” y/o “Piso resbaladizo”. Estos fenómenos se producen cuando la brecha ajustada es mayor en el extremo superior de la distribución salarial y cuando la brecha ajustada es mayor en el extremo inferior de la distribución salarial, respectivamente (Paz, 2019).

En la Tabla 1 se presenta un resumen de estudios revisados, indicando métodos de estimación utilizados, variable explicada, variables explicativas, alcance geográfico y de tiempo, entre otros.

Tabla 1: Revisión de la literatura

| Autor (año) | Objetivo | Método de estimación | Modelo empírico | Datos | Alcance geográfico | Alcance de tiempo | Hallazgo |
|---------------------------------|--|--|---|---|---------------------------|--------------------------|---|
| Blau & Kahn (2017) | Determinar el alcance y tendencias de la brecha salarial de género | Descomposición Blinder-Oaxaca y Modelo de mínimos cuadrados ordinarios (MCO) | Se realiza una ecuación salarial para hombres y mujeres | Datos de la Encuesta de Panel de la Dinámica de Ingresos (PSID) | Estados Unidos | 1980-2010 | La brecha salarial disminuyó lentamente en la parte superior de la distribución salarial |
| Bonaccolto-Töpfer et al. (2023) | Determinar el cambio de la brecha salarial que ha tenido Alemania | Mínimos cuadrados ordinarios y regresiones cuatílicas | Analiza la relación entre la brecha salarial por género en Alemania, en un periodo determinado trabajando con un componente de características “explicado” y uno de precios “no explicado” | Datos del Panel Socioeconómico de Alemania | Alemania | 1984-2020 | La brecha salarial disminuyó en los niveles bajo y medio de la distribución salarial, mientras que en la parte superior hubo un aumento |
| Cabrera et al. (2024) | Conocer los desafíos que tiene España en relación con la brecha salarial | T-student y Chi-Cuadrado | La variable dependiente es la desigualdad por razón de género. Las variables de control tienen relación con la brecha salarial (salario, tipo de contrato, tipo de jornada laboral), techo de cristal (sector, jerarquía, entre otros) y pared de cristal | Sexta Encuesta Europea de Condiciones de Trabajo realizada por Eurofond | Europa | 2015 | Hay una brecha salarial de 409,64 euros en Europa para 2015. En España, la diferencia mensual es de 376,44 euros |
| Didier (2021) | Determinar el efecto de las credenciales en los ingresos | Regresión MCO agrupada, regresión de efectos fijos y regresión intercuartil | La variable dependiente son los ingresos, mientras que las variables de control son la existencia de credenciales. Con esta información se busca determinar cómo influye su existencia y el efecto de género que tiene en los ingresos | Encuesta de Hogares CASEN | Chile | 2013-2017 | La brecha salarial por género en Chile es de alrededor de un 28%. Las credenciales tienen un grado de discriminación hacia las mujeres |

Fuente: Elaboración Propia

(Continuación) **Tabla 1:** Revisión de la literatura

| Autor (año) | Objetivo | Método de estimación | Modelo empírico | Datos | Alcance geográfico | Alcance de tiempo | Hallazgo |
|------------------------------|--|---|--|---|---------------------------|--------------------------|---|
| Kristjanpoller et al. (2023) | Determinar la brecha salarial por género en Chile, encontrando las variables que la explican y afectan | Inferencia causal y aprendizaje automático SHAP | La variable dependiente son los ingresos, y las variables de control son sexo, educación, ocupación, tamaño de empresa y región | Ministerio de Familia y Desarrollo Social | Chile | 1990-2017 | Ha crecido la brecha salarial por género en Chile en las últimas décadas |
| Meriküll & Tverdostup (2023) | Analizar el cambio de la brecha salarial de género en Estonia en las últimas 3 décadas | Descomposición de Oaxaca-Blinder, regresión simple OLS y regresiones cuantílicas | La variable dependiente es la brecha salarial de género, las variables de control utilizadas fueron relacionadas a la demografía, educación, experiencia en el mercado laboral, factor a nivel de empresa, ocupación y factores del país | Encuesta de la Fuerza de Trabajo | Estonia | 1989-2020 | Estonia tiene una brecha salarial de aproximadamente un 25%. Luego del periodo comunista, disminuyó la brecha salarial de género |
| Ortogorin (2024) | Determinar la relación que hay entre la brecha salarial por género y la violencia doméstica | Extensión de la descomposición Blinder-Oaxaca empleando metodología de emparejamiento, además utiliza el modelo logit | La variable dependiente es la brecha salarial de género. Las variables de control son edad, región, educación, pareja, además se considera una variable de violencia doméstica | Datos de la Encuesta Nacional de Hogares (ENAH) y el informe de violencia de la Encuesta Demográfica y de Salud (ENDES) | Perú | 2007-2017 | Perú tiene una brecha salarial en 2017 de aproximadamente un 20%. Además, existe relación entre la brecha salarial y la violencia doméstica |

Fuente: Elaboración Propia

(Continuación) **Tabla 1:** Revisión de la literatura

| Autor (año) | Objetivo | Método de estimación | Modelo empírico | Datos | Alcance geográfico | Alcance de tiempo | Hallazgo |
|----------------------------|---|--|--|--|--|---|--|
| Paz (2019) | Determinar cuál es la brecha salarial por género en Argentina distinguiendo a las personas según su estado conyugal | Descomposición Blinder-Oaxaca metodología desarrollada por Junh et al. (1993), regresión cuantílica (RC) | Se estima la brecha salarial de género y una descomposición de ella, en la brecha explicada y no explicada. Las variables de control son las que tienen relación con las productividades | Datos de la Encuesta Permanente de Hogares (EPH) | Argentina | 2018 | Gran porcentaje de las mujeres tienen trabajos informales, debido a que buscan convalidar con las funciones domésticas y hay barreras de entrada |
| Perticará & Tejada (2022) | Identificar el impacto de la discriminación en el mercado laboral de nueve países de Latinoamérica | Métodos de máxima verosimilitud | La variable dependiente es el salario, mientras que en las variables de control incluye situación laboral, tiempo de desempleo, entre otras. | Encuestas de empleo y sociales para cada uno de los nueve países | Chile, Argentina, Bolivia, Colombia, Ecuador, México, Perú, Paraguay y Uruguay | 2013-2015 | La brecha salarial depende mucho del país. En el caso de Chile, se tiene un gran prejuicio hacia las mujeres. Si el prejuicio se eliminará, la brecha salarial disminuiría en un 20% |
| Rembeza & Radlińska (2021) | Determinar si las mujeres seguían siendo trabajadoras secundarias | Modelos generales de flujos de mercado laboral | La variable dependiente es el flujo en el mercado laboral, considerando variables de control las personas ocupadas, desempleadas e inactivas | Organización Internacional del Trabajo | Australia, Canadá, Dinamarca, Japón, Francia, Gran Bretaña, Corea del Sur y Estados Unidos | Variable para cada país, debido a su disponibilidad de datos. | Las mujeres tienen son trabajadoras secundarias, pero los resultados están sujetos al lugar y tiempo que se estudien |

Fuente: Elaboración Propia

(Continuación) **Tabla 1:** Revisión de la literatura

| Autor (año) | Objetivo | Método de estimación | Modelo empírico | Datos | Alcance geográfico | Alcance de tiempo | Hallazgo |
|-------------------------------|---|---|--|---|---------------------------|--------------------------|--|
| Salazar Aramayo et al. (2025) | Evidenciar la existencia de discriminación salarial hacia mujeres en Bolivia | Modelo logit de regresión y descomposición Blinder-Oaxaca | La variable dependiente es la calidad del empleo. Para las variables de control, se considera el género, edad, raza, sector económico, tamaño de la empresa, horas de trabajo, educación, AFP, contratación, relación contractual, entre otras. | Datos de la Encuesta de hogares del Instituto Nacional de Estadística (INE) | Bolivia | 2011-2021 | Los ingresos obtenidos del trabajo son sensibles al género de la persona |
| Sánchez et al. (2022) | Analizar el poder de las empresas en el mercado laboral y sus diferencias de género, estimando las elasticidades de la oferta laboral | Modelo monopsonio dinámico | La variable dependiente es la elasticidad de la oferta laboral de la empresa. Las variables de control son las elasticidades de separación del empleo y desempleo, proporción de reclutamiento, la prima por la búsqueda de empleo. Todo esto separado para cada género. | Base de Datos Administrativa de Desempleo de Chile y Ministerio de la Educación | Chile | 2010-2019 | La brecha salarial en Chile es de aproximadamente un 22%. Además, las elasticidades laborales masculinas son mayores que las femeninas, siendo 0,7 y 0,51, respectivamente. Esto explica un tercio de la brecha salarial de género |
| Semenza et al. (2021) | Analizar la segregación de género que hay en 2 contextos diferentes, Italia y Chile | Método de disimilitud, Índice de Mujeres y Empleo (WEI) | La variable dependiente es el WEI, que mide la presencia diferenciada entre hombre y mujeres dentro de una clasificación específica de grupos ocupacionales | Datos equivalentes a la Encuesta de Fuerza Laboral | Italia y Chile | 1992-2017 | Ambos países tienen una alta segregación de las mujeres. Mas del 60% de las mujeres se centran en trabajos de calificación media a baja |

Fuente: Elaboración Propia

(Continuación) Tabla 1: Revisión de la literatura

| Autor (año) | Objetivo | Método de estimación | Modelo empírico | Datos | Alcance geográfico | Alcance de tiempo | Hallazgo |
|-------------------------------|--|---|---|--|---|-------------------|--|
| Si et al. (2020) | Determinar la brecha salarial por género en países en desarrollo | Descomposición Blinder-Oaxaca, regresión MCO, regresión cuantílica, | Se realiza una ecuación salarial para ambos géneros, luego se determina la brecha de género con la regresión cuantílica y finalmente se determina la brecha salarial de género con la descomposición Blinder-Oaxaca | Datos de la Encuesta para Empleo y la Productividad elaborada por el Banco Mundial | Colombia, Ghana, Kenia, Sri Lanka, Ucrania, Armenia, Republica democrática Popular Lao, Macedonia, Vietnam, China, Bolivia, Georgia | 2012-2013 | La brecha salarial se encuentra en todos los niveles salariales. Por otra parte, la educación y maternidad tienen estrecha relación con la brecha salarial |
| Vaccaro et al. (2022) | Proporcionar una visión general de cómo ha ido evolucionando la brecha salarial por género en Perú | Descomposición Blinder-Oaxaca | Se realiza una ecuación salarial para ambos géneros, se utilizan variables utilizadas en la literatura, como edad, educación, sector de trabajo, entre otros. | Datos de la Encuesta Nacional de Hogares de Perú (ENAH) | Perú | 2007-2018 | Perú es de los países con mayor desigualdad salarial, por otra parte, deben potenciar la educación ya que mitiga la brecha salarial |
| Vargas Calderón et al. (2024) | Analizar la brecha salarial que tiene la ciudad de Cali en Colombia. | Descomposición Blinder-Oaxaca y corrección de sesgo de Heckman | Se realiza una ecuación salarial para ambos géneros. Donde la variable dependiente es la brecha salarial por género. | Para determinar la brecha salarial que existe entre géneros, se utilizaron los datos de la Gran Encuesta Integrada de Hogares (GEIH) | Colombia | 2023 | La variable que más contribuye a la brecha salariales son las horas trabajadas y el cuidado de los niños. |

Fuente: Elaboración Propia

3. Metodología

3.1 Método empírico

Para la estimación de la brecha salarial de género se utilizan ecuaciones salariales, donde se determina el impacto del género en el salario (Meriküll & Tverdostup, 2023). En este informe se presentan dos métodos distintos para la estimación de la brecha salarial de género. El primero corresponde al método de Mínimos Cuadrados Ordinarios en la regresión lineal múltiple. Posteriormente, se aplica la metodología de regresión simbólica para la identificación de una expresión funcional interpretable que relacione el salario con las variables explicativas, sin asumir una forma estructural previa y explorando combinaciones de operadores matemáticas mediante programación genética (Wang et al., 2022).

3.2 Método de estimación

3.2.1 Mínimos Cuadrados Ordinarios

La regresión lineal permite determinar de forma estadística la relación de variables. Los componentes para considerar son: una variable dependiente y , variables independientes x , coeficientes de la regresión β y la constante de error aleatorio ε . β es el valor que cambia y , por un cambio en una unidad de x_i , manteniendo a todas las variables constantes (Montgomery, 2004). La ecuación (1) muestra la forma general de la regresión lineal.

$$y = \beta_0 + \sum_{i=1}^N \beta_i x_i + \varepsilon \quad (1)$$

Si N solo tiene un elemento se considera una regresión lineal simple, en caso contrario es una regresión lineal múltiple, es decir, hay más de una variable independiente. Para determinar el valor de los coeficientes de la regresión, se utiliza el método de los mínimos cuadrados ordinarios (MCO), cuyo objetivo es minimizar la suma de los cuadrados de los errores. Una de las consideraciones para utilizar este método es tener t observaciones, y k coeficientes de la regresión, siendo $t > k$.

Las variables dependientes, tendrán la siguiente notación: x_i^t , donde $i = 1, \dots, N$ representa cada variable en $t = 1, \dots, T$ observaciones. De esta forma la ecuación (1) se desglosa para cada observación, como se muestra en la ecuación (2).

$$y_t = \beta_0 + \sum_{i=1}^N \beta_i x_i^t + \varepsilon_t \quad t=1, \dots, T \quad (2)$$

Para luego realizar la minimización de los errores al cuadrado, es decir, se realizan la función de MCO, que se muestra en la ecuación (3). A su vez, es equivalente a la ecuación (4).

$$\text{MCO} = \sum_{t=1}^T \varepsilon_t^2 \quad (3)$$

$$\text{MCO} = \sum_{t=1}^T \left(y_t - \beta_0 + \sum_{i=1}^N \beta_i x_i^t \right)^2 \quad (4)$$

Esta memoria de título utiliza este método para obtener el logaritmo natural del salario por hora, variable dependiente, considerando el objetivo de esta investigación y las variables independientes que se encuentran en los datos de la Encuesta CASEN para el periodo 1992, 2000, 2011 y 2022.

3.2.2 Metodología Regresión Simbólica

El objetivo es lograr identificar expresiones matemáticas que puedan modelar de una forma precisa e interpretable la relación funcional entre variables independientes con la variable dependiente a partir de datos observados. A diferencia de la regresión lineal clásica, la regresión simbólica no asume una forma predeterminada, lo que busca es estructurar y ajustar la forma desde expresiones ya establecidas (Makke & Chawla, 2024).

Se inicia de un conjunto de datos $D = \{x_i, y_i\}_{i=1}^n$, donde $x_i \in R^d$ y $y_i \in R^d$. El objetivo es encontrar una función simbólica $f \in F$ que minimice el error cuadrático medio.

$$f = \min \sum_{i=1}^N (y_i - f(x_i))^2 \quad (5)$$

Las expresiones candidatas son construidas en base a la combinación de elementos de una biblioteca de operadores establecidos, los cuales pueden incluir operaciones aritméticas simples como la suma, resta, división y multiplicación, también son consideradas las funciones logarítmicas, exponenciales, seno, coseno y además pueden ser incluidas

constantes. Al incorporar funciones específicas se responde a la necesidad de restringir el espacio de búsqueda, lo cual favorece la interpretabilidad. Para abordar la regresión simbólica se pueden utilizar metodologías, dentro de las cuales se encuentra la programación genética y los métodos basados en redes neuronales.

3.2.2.1 Programación Genética

La programación genética es un algoritmo evolutivo (Kronberger et al.,2025). Se inicializa una población de árboles aleatorios (expresiones matemáticas) y se iteran generaciones aplicando operadores de cruce, mutación y selección según una función de aptitud que en este caso es el error cuadrático medio entre las predicciones del modelo y los valores observados. Se evalúa a cada individuo la función de aptitud mencionada anteriormente y son seleccionados los mejores individuos para reproducirse en generaciones futuras (Makke & Chawla, 2024).

3.3 Pruebas estadísticas

Se presentan las pruebas estadísticas utilizados en la investigación. Se utilizaron basándose en mencionadas por Montgomery (2004) y Hair et al. (1999).

3.3.1 R^2 y R^2 ajustado

El R^2 consiste en el coeficiente de determinación de la regresión lineal múltiple. Se utiliza como medida de precisión, la cual determina la relación entre la variable dependiente y las variables independientes. Su magnitud se interpreta como la variabilidad explicada por el modelo. Sus valores son entre de 0 a 1.

$$R^2 = \frac{SSR}{SST} \quad (6)$$

$$SST = SSR + SSE \quad (7)$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 \quad (9)$$

El término SST , ecuación (7), es la suma de cuadrados total. Mientras que SSE , ecuación (8), y SSR , ecuación (9), son la suma de cuadrados de los errores y de la regresión, respectivamente. La variable y_i corresponde a los valores originales, \hat{y}_i es el valor estimado e \bar{y}_i es el promedio de la variable independiente. Cabe destacar que, a mayor número de variables, más alto será el valor del coeficiente, lo cual puede no indicar que sea mejor que el que tiene menos variables. Por otra parte, el R^2 ajustado, ecuación (10), es una variante del R^2 , el cual no es tan sensible a la adición de variables independientes como R^2 y puede ocuparse para expresiones no lineales. Por otra parte, n es el número de observaciones y p el número de variables en la regresión.

$$R^2_{ajustado} = 1 - \frac{n-1}{n-p}(1 - R^2) \quad (10)$$

3.3.2 Error cuadrático medio

El error cuadrático medio (MSE, por sus siglas en inglés) corresponde a la diferencia entre el valor original (y_i) y la estimación (\hat{y}_i) al cuadrado, dividido el número de observaciones (n) menos el número de variables (p). A mayor valor del MSE, significa menor precisión en la regresión lineal. El MSE, ecuación (11), es considerado como criterio para comparar los resultados de la regresión lineal múltiple y la regresión simbólica.

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{n - p} \quad (11)$$

3.3.3 Prueba F

La prueba F permite identificar si la regresión lineal es significativa o no. Esta consiste en determinar si la variable dependiente es explicada por las variables independientes. Las hipótesis, ecuación (12), que se utilizan para el análisis son las siguientes:

$$\begin{aligned} H_0 &= \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 &= \beta_k \neq 0 \text{ para alguna } k \end{aligned} \quad (12)$$

Al rechazar la hipótesis nula (H_0) indica que al menos una variable independiente afecta significativamente en la regresión y con ello, la regresión es significativa. La prueba F permite determinar ello. Se calcula según la ecuación (13).

$$F = \frac{MSR}{MSE} \quad (13)$$

Donde MSE es el error cuadrático medio y MSR es el cuadrado medio de la regresión, cuya fórmula se presenta en la ecuación (14).

$$MSR = \frac{SSR}{k} \quad (14)$$

La hipótesis nula se rechaza si el estadístico F es mayor que $F_{\alpha, k, n-p}$, donde α es el grado de confianza, n el número de observaciones y p el número de variables en general consideradas en la regresión o $k+1$.

4. Datos

4.1 Fuente de datos

Los datos se obtuvieron de la Encuesta de Caracterización Socioeconómica Nacional, CASEN (2022). Los responsables son el Ministerio de Desarrollo Social y Familia de Chile, mientras que los ejecutores son el Instituto Nacional de Estadística (INE), Centro de Microdatos (CMD) y Cadem. La Tabla 2 muestra las características de los datos que se presentan en la encuesta.

Tabla 2: Características de la encuesta CASEN 2022

| Aspectos de la encuesta | Características |
|-------------------------------|---|
| Tamaños logrados | Vivienda: 70.751 viviendas Personas: 202.231 personas |
| Periodo de obtención de datos | 01 de noviembre del 2022 – 02 de febrero del 2023 |
| Representatividad | Nacional, zonas rurales y urbanas. Abarcando las 16 regiones del país |
| Número de comunas | 335 |

Fuente: CASEN 2022

4.2 Creación de la muestra de estimación

Para construir la muestra de estimación se utiliza el libro de códigos de la base de datos de la encuesta CASEN 2022. En ella se pudo encontrar los códigos de las variables que se utilizaron para estimar la brecha salarial. Para determinar las variables se utilizó de referencia la literatura.

Luego de seleccionar las variables que se utilizarían, se realizó una limpieza de los datos. Se eliminaron las respuestas: “No responde”, “No sabe”, “Glosa no permite clasificación” o “.” debido a que no entregan información a la investigación. Al realizar ello, la muestra inicial disminuyó su tamaño a $N= 48.868$.

Posterior a ellos, se realizaron tres filtros de datos. El primero de ellos, considera solo a las personas que hayan tenido al menos una hora de trabajo la semana anterior a ser encuestados. Al aplicar el filtro el tamaño de la muestra descendió a $N= 46.131$. El segundo filtro, tuvo por objetivo utilizar solo los datos de salario válidos. El tamaño de muestra resultante fue de $N= 45.959$. El último filtro, considerara solo a los trabajadores mayores a quince años, al igual que en la literatura existente. Finalmente, al incorporar el tercer filtro queda un tamaño de muestra de $N= 45.959$.

Tabla 3: Evolución del tamaño de muestra para 2022

| Muestra | Tamaño de muestra |
|---|--------------------------|
| Muestra original | 202.321 |
| Muestra inicial | 48.868 |
| Muestra aplicando filtro 1 “Al menos una hora trabajada la semana pasada” | 46.131 |
| Muestra aplicando filtro 2 “Valores de salario validos” | 45.959 |
| Muestra aplicando filtro 3 “Trabajadores mayores de 15 años” | 45.959 |
| Muestra de la investigación | 45.959 |

Fuente: Elaboración propia

4.2.1 Modificación de las variables

Con el objetivo de entregar información valiosa para la investigación, las variables tuvieron que ser sometidas a modificaciones en su presentación. Se inicio, considerando la definición de salario por horas, definiéndolo como la proporción entre salario bruto obtenido y la cantidad de horas mensuales trabajadas. La evidencia señala que este valor se considera aplicando logaritmo natural, debido a que suaviza los extremos de los valores.

Luego, el nivel de escolaridad se clasifico en cuatro categorías considerando las principales: menor a educación media (educación básica), educación media completa, título técnico y título universitario o más. La cuatro variables son dicotómicas, es decir tendrán valor de 1 si es que se cumple la condición y 0 en otro caso.

El estado civil también tuvo que ser clasificado en tres categorías, lo que corresponde a tres variables dicotómicas. Las categorías son: casado, conviviente y soltero. La primera considera a las personas unidas por el matrimonio y por acuerdo de unión civil, mientras que la tercera categoría abarca a las personas anuladas, separadas, divorciadas, viudas y solteras.

Para el tipo de contrato se consideró tres variables dicotómicas, las cuales son: contrato indefinido, contrato a plazo fijo y contrato por obra. El tamaño de la empresa también se agrupo en categorías, basadas en la cantidad de trabajadores que posee, lo que corresponde a variables dicotómicas. La clasificación se rigió de acuerdo con el Código del Trabajo Chileno (2002), que señala que las empresas que poseen entre 1 a 9 trabajadores se denominan microempresas, de 10 a 49 trabajadores se consideran pequeñas empresas, de 50 a 199 trabajadores son denominadas medianas empresas y si cuentan con más de 200 trabajadores es una gran empresa.

Para el sector de la industria se consideraron doce variables dicotómicas, dentro de las que se destacan si corresponde a sector primario, extracción de recursos naturales, manufactura,

servicios básicos, servicios de agua y descontaminación, construcción, comercio, transporte, hospitalidad, información, inmobiliaria y otros servicios. Esta última variable incluye actividades profesionales, científicas y técnicas, además de actividades de servicios administrativos y de apoyo, administración pública y defensa, enseñanza, atención de salud, actividades artísticas, actividades de organizaciones y órganos extraterritoriales, actividades de los hogares y otros servicios. Al igual que para la industria, en el área de trabajo también se realizaron variables dicotómicas.

En el sector laboral, se utilizó dos variables dicotómicas, público o privado. Asimismo, se realizó con la presencia de hijos y área de la ciudad, urbana o rural. Finalmente, el territorio geográfico fue dividido en cinco macrozonas, según lo señalado por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. La macrozona norte considera las regiones de Arica y Parinacota, Antofagasta y Atacama. La macrozona centro abarca las regiones de Coquimbo, Valparaíso y Metropolitana. Las regiones de O'Higgins, Maule, Ñuble y Biobío pertenecen a la macrozona centro-sur, mientras que las regiones de La Araucanía, Los Lagos y Los ríos a la macrozona sur. Finalmente, las regiones de Aysén y Magallanes son consideradas en la macrozona austral.

4.3 Descripción de la muestra de estimación

Se consideraron para la investigación un grupo de 13 categorías en la cual se agrupan las 53 variables modificadas. En la Tabla 4 se muestra las áreas y variables asociadas a cada una de ellas, junto a el significado de cada una de ellas.

Tabla 4: Variables de la investigación CASEN 2022

| Categoría | Variable | Observaciones |
|------------------|----------------------------|---|
| Salario | Salario por hora | Salario bruto dividido horas trabajadas en el mes |
| | Logaritmo salario | Se aplica logaritmo natural a salario por hora |
| Sexo | Mujer | Sea 1 si es mujer, 0 en otro caso. |
| | Hombre | Sea 1 si es hombre, 0 en otro caso. |
| Edad | Edad | Valor entre 15 a 120 años |
| | Edad2 | Valor al cuadrado de Edad |
| Educación | Educación media incompleta | Sea 1 si posee la educación media incompleta, 0 en otro caso |
| | Educación media completa | Sea 1 si solo posee la educación media completa, 0 en otro caso |
| | Título técnico | Sea 1 si solo posee un título técnico, 0 en otro caso |
| | Título universitario o mas | Sea 1 si posee título universitario o más, 0 en otro caso |

Fuente: Elaboración propia

(Continuación) Tabla 4: Variables de la investigación CASEN 2022

| Categoría | Variable | Observaciones |
|-------------------------|--|--|
| Estado civil | Casado | Sea 1 si se encuentra casado, 0 en otro caso |
| | Conviviente | Sea 1 si es conviviente, 0 en otro caso |
| | Soltero | Sea 1 si es soltero, anulante, separado, divorciado o viudo, 0 en otro caso |
| Tipo de contrato | Contrato indefinido | Sea 1 si posee contrato indefinido, 0 en otro caso |
| | Contrato a plazo fijo | Sea 1 si posee contrato a plazo fijo, 0 en otro caso |
| | Contrato de obra | Sea 1 si posee contrato de obra, 0 en otro caso |
| Tamaño de la empresa | Microempresa | Sea 1 si es una microempresa, 0 en otro caso |
| | Pequeña empresa | Sea 1 si es una pequeña empresa, 0 en otro caso |
| | Mediana empresa | Sea 1 si es una mediana empresa, 0 en otro caso |
| | Gran empresa | Sea 1 si es una gran empresa, 0 en otro caso |
| Sector de la industria | Sector primario | Sea 1 si pertenece al sector primario, 0 en otro caso |
| | Extracción de recursos | Sea 1 si pertenece al sector de la extracción de recursos, 0 en otro caso |
| | Manufactura | Sea 1 si pertenece al sector de manufactura, 0 en otro caso |
| | Servicios básicos | Sea 1 si pertenece al sector de servicios básicos, 0 en otro caso |
| | Servicios de agua y descontaminación | Sea 1 si pertenece al sector de servicios de agua y descontaminación, 0 en otro caso |
| | Construcción | Sea 1 si pertenece al sector de la construcción, 0 en otro caso |
| | Comercio | Sea 1 si pertenece al sector de comercio, 0 en otro caso |
| | Transporte | Sea 1 si pertenece al sector de transporte, 0 en otro caso |
| | Hospitalidad | Sea 1 si pertenece al sector de hospitalidad, 0 en otro caso |
| | Información | Sea 1 si pertenece al sector de la información, 0 en otro caso |
| Área de trabajo | Inmobiliaria | Sea 1 si pertenece al sector de inmobiliario, 0 en otro caso |
| | Otros servicios | Sea 1 si pertenece a otros servicios, 0 en otro caso |
| | Fuerzas Armadas | Sea 1 si pertenece a las fuerzas armadas, 0 en otro caso |
| | Ejecutivos | Sea 1 si es del área ejecutiva, 0 en otro caso |
| | Académicos | Sea 1 si es del área académica, 0 en otro caso |
| | Técnicos | Sea 1 si es del área técnica, 0 en otro caso |
| | Administrativos | Sea 1 si es del área administrativa, 0 en otro caso |
| | Vendedores | Sea 1 si es vendedor, 0 en otro caso |
| | Agricultores | Sea 1 si es del área de la agricultura, 0 en otro caso |
| | Artesanos | Sea 1 si es artesano, 0 en otro caso |
| Operadores | Sea 1 si es operador, 0 en otro caso | |
| Ocupaciones elementales | Sea 1 si tiene una ocupación elemental, 0 en otro caso | |
| Sector laboral | Público | Sea 1 si trabaja en el sector público, 0 en otro caso |
| | Privado | Sea 1 si trabaja en el sector privado, 0 en otro caso |
| Macrozona | Macrozona norte | Sea 1 si pertenece a la macrozona norte, 0 en otro caso |
| | Macrozona centro | Sea 1 si pertenece a la macrozona centro, 0 en otro caso |
| | Macrozona centro sur | Sea 1 si pertenece a la macrozona centro-sur, 0 en otro caso |
| | Macrozona sur | Sea 1 si pertenece a la macrozona sur, 0 en otro caso |
| | Macrozona austral | Sea 1 si pertenece a la macrozona austral, 0 en otro caso |
| Hijos | Con hijos | Sea 1 si tiene al menos 1 hijo, 0 en otro caso |
| | Sin hijos | Sea 1 si no tiene hijos, 0 en otro caso |
| Jefe de hogar | Jefe de hogar | Sea 1 si es jefe de hogar, 0 en otro caso |
| Área | Urbano | Sea 1 si se encuentra en un área urbana, 0 en otro caso |
| | Rural | Sea 1 si se encuentra en un área rural, 0 en otro caso |

Fuente: Elaboración Propia

La Tabla 5 muestra las estadísticas descriptivas de las variables correspondientes a las 13 categorías, es decir la cantidad de observaciones, media, desviación estándar, valor máximo y mínimo.

Tabla 5: Estadísticas descriptivas de las variables de CASEN 2022

| Categoría | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|------------------------|--------------------------------------|----------------------|--------------|----------------------------|---------------|---------------|
| Salario | Logaritmo salario | 45.959 | 8,2064 | 0,6735 | 4,715 | 14,22 |
| Sexo | Mujer | 45.959 | 0,4421 | 0,4966 | 0 | 1 |
| | Hombre | 45.959 | 0,5579 | 0,4966 | 0 | 1 |
| Edad | Edad | 45.959 | 42,6930 | 13,0437 | 17 | 92 |
| | Edad2 | 45.959 | 1992,828 | 1168,943 | 289 | 8464 |
| Educación | Educación media incompleta | 45.959 | 0,1513 | 0,3584 | 0 | 1 |
| | Educación media completa | 45.959 | 0,3261 | 0,4688 | 0 | 1 |
| | Título técnico | 45.959 | 0,2663 | 0,4420 | 0 | 1 |
| | Título universitario o mas | 45.959 | 0,2563 | 0,4366 | 0 | 1 |
| Estado civil | Casado | 45.959 | 0,3192 | 0,4662 | 0 | 1 |
| | Conviviente | 45.959 | 0,2560 | 0,4364 | 0 | 1 |
| | Soltero | 45.959 | 0,4247 | 0,4943 | 0 | 1 |
| Tipo de contrato | Contrato a plazo fijo | 45.959 | 0,1768 | 0,3815 | 0 | 1 |
| | Contrato indefinido | 45.959 | 0,7169 | 0,4505 | 0 | 1 |
| | Contrato de obra | 45.959 | 0,1063 | 0,3082 | 0 | 1 |
| Tamaño de la empresa | Microempresa | 45.959 | 0,2584 | 0,4378 | 0 | 1 |
| | Pequeña empresa | 45.959 | 0,2307 | 0,4213 | 0 | 1 |
| | Mediana empresa | 45.959 | 0,2072 | 0,4053 | 0 | 1 |
| | Gran empresa | 45.959 | 0,3037 | 0,4599 | 0 | 1 |
| Sector de la industria | Sector primario | 45.959 | 0,1123 | 0,3157 | 0 | 1 |
| | Extracción de recursos | 45.959 | 0,0492 | 0,2162 | 0 | 1 |
| | Manufactura | 45.959 | 0,0856 | 0,2797 | 0 | 1 |
| | Servicios básicos | 45.959 | 0,0072 | 0,0848 | 0 | 1 |
| | Servicios de agua y descontaminación | 45.959 | 0,0087 | 0,0927 | 0 | 1 |
| | Construcción | 45.959 | 0,0762 | 0,2653 | 0 | 1 |
| | Comercio | 45.959 | 0,1371 | 0,3439 | 0 | 1 |
| | Transporte | 45.959 | 0,0507 | 0,2195 | 0 | 1 |
| | Hospitalidad | 45.959 | 0,0447 | 0,2066 | 0 | 1 |
| | Información | 45.959 | 0,0176 | 0,1314 | 0 | 1 |
| | Inmobiliaria | 45.959 | 0,0058 | 0,0761 | 0 | 1 |
| | Otros servicios | 45.959 | 0,4050 | 0,4909 | 0 | 1 |

Fuente: Elaboración Propia

(Continuación) **Tabla 5:** Estadísticas descriptivas de las variables CASEN 2022

| Categoría | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|-------------------|-------------------------|---------------|--------|---------------------|--------|--------|
| Área de trabajo | Fuerzas Armadas | 45.959 | 0,0042 | 0,0645 | 0 | 1 |
| | Ejecutivos | 45.959 | 0,0277 | 0,1642 | 0 | 1 |
| | Académicos | 45.959 | 0,1670 | 0,3730 | 0 | 1 |
| | Técnicos | 45.959 | 0,1240 | 0,3296 | 0 | 1 |
| | Administrativos | 45.959 | 0,0629 | 0,2427 | 0 | 1 |
| | Vendedores | 45.959 | 0,1753 | 0,3802 | 0 | 1 |
| | Agricultores | 45.959 | 0,0304 | 0,1717 | 0 | 1 |
| | Artesanos | 45.959 | 0,1097 | 0,3126 | 0 | 1 |
| | Operadores | 45.959 | 0,1010 | 0,3014 | 0 | 1 |
| | Ocupaciones elementales | 45.959 | 0,1978 | 0,3983 | 0 | 1 |
| Sector laboral | Público | 45.959 | 0,2037 | 0,4027 | 0 | 1 |
| | Privado | 45.959 | 0,7963 | 0,4027 | 0 | 1 |
| Macrozona | Macrozona norte | 45.959 | 0,1669 | 0,3729 | 0 | 1 |
| | Macrozona centro | 45.959 | 0,3447 | 0,4753 | 0 | 1 |
| | Macrozona centro sur | 45.959 | 0,2811 | 0,4495 | 0 | 1 |
| | Macrozona sur | 45.959 | 0,1527 | 0,3597 | 0 | 1 |
| | Macrozona austral | 45.959 | 0,0547 | 0,2273 | 0 | 1 |
| Hijos | Con hijos | 45.959 | 0,7296 | 0,4442 | 0 | 1 |
| | Sin hijos | 45.959 | 0,2704 | 0,4442 | 0 | 1 |
| Jefatura de hogar | Jefatura de hogar | 45.959 | 0,5142 | 0,4998 | 0 | 1 |
| Área | Urbano | 45.959 | 0,8326 | 0,3734 | 0 | 1 |
| | Rural | 45.959 | 0,1674 | 0,3734 | 0 | 1 |

Fuente: Elaboración Propia

4.4 Muestra para distintos períodos

Con el fin de profundizar el análisis de la brecha salarial por género en Chile, se realiza una comparación entre las brechas obtenidas de distintos periodos, debido a que permite observar su evolución en los últimos 30 años. Para ello, se utilizan muestras de la encuesta CASEN para 1992, 2000 y 2011.

Al igual que para la encuesta CASEN 2022, fueron sometidas a los mismos cuatro filtros. Cabe destacar que, al ser encuestas distintas, con diferencias de aproximadamente 10 años, la cantidad de información presente es distinta. Las diferencias no son de gran magnitud. Se considera como base las variables utilizadas para 2022. Dependiendo de los datos disponibles disminuyen la cantidad de variables utilizadas para cada año. La Tabla 6, se muestra de forma comprimida el tamaño de muestra y el número de variables consideradas.

Tabla 6 : Muestras CASEN para distintos períodos

| Año | N | Número de variables consideradas |
|------------|----------|---|
| 1992 | 45.238 | 42 |
| 2000 | 70.147 | 46 |
| 2011 | 69.618 | 48 |
| 2022 | 45.959 | 52 |

Nota: (a) El número de variables, no considera la variable de salario por hora

Fuente: Elaboración Propia

Las Tablas 17, 18 y 19 son presentadas en el anexo y muestran las variables para 1992, 2000 y 2011, respectivamente. Las estadísticas descriptivas para la muestra de 1992, 2000 y 2011, se muestran en la Tabla 7, 8 y 9, respectivamente.

Tabla 7: Estadísticas descriptivas de las variables CASEN 1992

| Categoría | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|------------------------|----------------------------|----------------------|--------------|----------------------------|---------------|---------------|
| Salario | Logaritmo salario | 45.238 | 5,9654 | 0,8769 | 4,715 | 14,22 |
| Sexo | Mujer | 45.238 | 0,2828 | 0,4504 | 0 | 1 |
| | Hombre | 45.238 | 0,7172 | 0,4504 | 0 | 1 |
| Edad | Edad | 45.238 | 37,1888 | 13,1774 | 17 | 92 |
| | Edad2 | 45.238 | 1556,644 | 1107,432 | 289 | 8464 |
| Educación | Educación media incompleta | 45.238 | 0,6540 | 0,4757 | 0 | 1 |
| | Educación media completa | 45.238 | 0,2562 | 0,4366 | 0 | 1 |
| | Título universitario o mas | 45.238 | 0,0898 | 0,2859 | 0 | 1 |
| Estado civil | Casado | 45.238 | 0,5540 | 0,4971 | 0 | 1 |
| | Conviviente | 45.238 | 0,0672 | 0,2503 | 0 | 1 |
| | Soltero | 45.238 | 0,3788 | 0,4851 | 0 | 1 |
| Tamaño de la empresa | Microempresa | 45.238 | 0,5395 | 0,4984 | 0 | 1 |
| | Pequeña empresa | 45.238 | 0,2402 | 0,4272 | 0 | 1 |
| | Mediana empresa | 45.238 | 0,1159 | 0,3201 | 0 | 1 |
| | Gran empresa | 45.238 | 0,1044 | 0,3057 | 0 | 1 |
| Sector de la industria | Sector primario | 45.238 | 0,2808 | 0,4494 | 0 | 1 |
| | Extracción de recursos | 45.238 | 0,0331 | 0,1789 | 0 | 1 |
| | Manufactura | 45.238 | 0,1348 | 0,3415 | 0 | 1 |
| | Servicios básicos | 45.238 | 0,0055 | 0,0743 | 0 | 1 |
| | Construcción | 45.238 | 0,0777 | 0,2677 | 0 | 1 |
| | Comercio | 45.238 | 0,1535 | 0,3604 | 0 | 1 |
| | Transporte | 45.238 | 0,0636 | 0,2440 | 0 | 1 |
| | Otros servicios | 45.238 | 0,2511 | 0,4336 | 0 | 1 |

Fuente: Elaboración Propia

(Continuación) **Tabla 7:** Estadísticas descriptivas de las variables CASEN 1992

| Categoría | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|------------------|-------------------------|----------------------|--------------|----------------------------|---------------|---------------|
| Área de trabajo | Fuerzas Armadas | 45.238 | 0 | 0 | 0 | 1 |
| | Ejecutivos | 45.238 | 0,0485 | 0,2149 | 0 | 1 |
| | Académicos | 45.238 | 0,0587 | 0,2351 | 0 | 1 |
| | Técnicos | 45.238 | 0,0438 | 0,2047 | 0 | 1 |
| | Administrativos | 45.238 | 0,0586 | 0,2350 | 0 | 1 |
| | Vendedores | 45.238 | 0,1199 | 0,3249 | 0 | 1 |
| | Agricultores | 45.238 | 0,1029 | 0,3038 | 0 | 1 |
| | Artesanos | 45.238 | 0,16 | 0,3666 | 0 | 1 |
| | Operadores | 45.238 | 0,0767 | 0,2662 | 0 | 1 |
| | Ocupaciones elementales | 45.238 | 0,3308 | 0,4705 | 0 | 1 |
| Macrozona | Macrozona norte | 45.238 | 0,1192 | 0,3240 | 0 | 1 |
| | Macrozona centro | 45.238 | 0,4143 | 0,4926 | 0 | 1 |
| | Macrozona centro sur | 45.238 | 0,3258 | 0,4687 | 0 | 1 |
| | Macrozona sur | 45.238 | 0,1019 | 0,3025 | 0 | 1 |
| | Macrozona austral | 45.238 | 0,0388 | 0,1932 | 0 | 1 |
| Jefe de hogar | Si es jefe de hogar | 45.238 | 0,5383 | 0,4985 | 0 | 1 |
| | No es jefe de hogar | 45.238 | 0,4617 | 0,4985 | 0 | 1 |
| Área | Urbano | 45.238 | 0,6789 | 0,4669 | 0 | 1 |
| | Rural | 45.238 | 0,3211 | 0,4669 | 0 | 1 |

Fuente: Elaboración Propia

Tabla 8: Estadísticas descriptivas de las variables CASEN 2000

| Categoría | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|------------------------|----------------------------|----------------------|--------------|----------------------------|---------------|---------------|
| Salario | Logaritmo salario | 70.147 | 6,7235 | 0,8168 | 2,1203 | 11,9469 |
| Sexo | Mujer | 70.147 | 0,3086 | 0,4619 | 0 | 1 |
| | Hombre | 70.147 | 0,6914 | 0,4619 | 0 | 1 |
| Edad | Edad | 70.147 | 39,2295 | 12,9676 | 16 | 95 |
| | Edad2 | 70.147 | 1707,11 | 1123,63 | 256 | 9025 |
| Educación | Educación media incompleta | 70.147 | 0,6045 | 0,4890 | 0 | 1 |
| | Educación media completa | 70.147 | 0,2941 | 0,4556 | 0 | 1 |
| | Título técnico | 70.147 | 0,0401 | 0,1961 | 0 | 1 |
| | Título universitario o mas | 70.147 | 0,0613 | 0,2399 | 0 | 1 |
| Estado civil | Casado | 70.147 | 0,5263 | 0,4993 | 0 | 1 |
| | Conviviente | 70.147 | 0,1223 | 0,3276 | 0 | 1 |
| | Soltero | 70.147 | 0,3514 | 0,4774 | 0 | 1 |
| Tipo de empleo | Empleo a plazo fijo | 70.147 | 0,1833 | 0,4412 | 0 | 1 |
| | Empleo indefinido | 70.147 | 0,7352 | 0,3869 | 0 | 1 |
| | Empleo de obra | 70.147 | 0,0361 | 0,1866 | 0 | 1 |
| Tamaño de la empresa | Microempresa | 70.147 | 0,5348 | 0,4988 | 0 | 1 |
| | Pequeña empresa | 70.147 | 0,1942 | 0,3956 | 0 | 1 |
| | Mediana empresa | 70.147 | 0,1141 | 0,3179 | 0 | 1 |
| | Gran empresa | 70.147 | 0,1568 | 0,3636 | 0 | 1 |
| Sector de la industria | Sector primario | 70.147 | 0,2830 | 0,4505 | 0 | 1 |
| | Extracción de recursos | 70.147 | 0,0197 | 0,1389 | 0 | 1 |
| | Manufactura | 70.147 | 0,1090 | 0,3117 | 0 | 1 |
| | Servicios básicos | 70.147 | 0,0071 | 0,0839 | 0 | 1 |
| | Construcción | 70.147 | 0,0792 | 0,2701 | 0 | 1 |
| | Comercio | 70.147 | 0,1607 | 0,3672 | 0 | 1 |
| | Transporte | 70.147 | 0,0613 | 0,2398 | 0 | 1 |
| | Otros servicios | 70.147 | 0,280 | 0,4490 | 0 | 1 |
| Área de trabajo | Fuerzas Armadas | 70.147 | 0,0038 | 0,0616 | 0 | 1 |
| | Ejecutivos | 70.147 | 0,0691 | 0,2537 | 0 | 1 |
| | Académicos | 70.147 | 0,0577 | 0,2332 | 0 | 1 |
| | Técnicos | 70.147 | 0,0486 | 0,2151 | 0 | 1 |
| | Administrativos | 70.147 | 0,0633 | 0,2435 | 0 | 1 |
| | Vendedores | 70.147 | 0,1128 | 0,3163 | 0 | 1 |
| | Agricultores | 70.147 | 0,1386 | 0,3455 | 0 | 1 |
| | Artesanos | 70.147 | 0,1345 | 0,3412 | 0 | 1 |
| | Operadores | 70.147 | 0,0936 | 0,2913 | 0 | 1 |
| | Ocupaciones elementales | 70.147 | 0,2779 | 0,4480 | 0 | 1 |

Fuente: Elaboración Propia

Tabla 9: Estadísticas descriptivas de las variables CASEN 2011

| Categoría | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|------------------------|----------------------------|---------------|----------|---------------------|--------|---------|
| Salario | Logaritmo salario | 69.618 | 7,4679 | 0,7808 | 1,2701 | 13,3851 |
| Sexo | Mujer | 69.618 | 0,4047 | 0,4908 | 0 | 1 |
| | Hombre | 69.618 | 0,5953 | 0,4908 | 0 | 1 |
| Edad | Edad | 69.618 | 39,5633 | 12,7552 | 16 | 99 |
| | Edad2 | 69.618 | 1727,951 | 1069,206 | 256 | 9801 |
| Educación | Educación media incompleta | 69.618 | 0,3716 | 0,4832 | 0 | 1 |
| | Educación media completa | 69.618 | 0,4360 | 0,4959 | 0 | 1 |
| | Título universitario o mas | 69.618 | 0,1924 | 0,3942 | 0 | 1 |
| Estado civil | Casado | 69.618 | 0,3846 | 0,4872 | 0 | 1 |
| | Conviviente | 69.618 | 0,1987 | 0,3990 | 0 | 1 |
| | Soltero | 69.618 | 0,4137 | 0,4925 | 0 | 1 |
| Tipo de contrato | Contrato a plazo fijo | 69.618 | 0,2645 | 0,4411 | 0 | 1 |
| | Contrato indefinido | 69.618 | 0,7355 | 0,4411 | 0 | 1 |
| Tamaño de la empresa | Microempresa | 69.618 | 0,2595 | 0,4384 | 0 | 1 |
| | Pequeña empresa | 69.618 | 0,1805 | 0,3846 | 0 | 1 |
| | Mediana empresa | 69.618 | 0,1321 | 0,3387 | 0 | 1 |
| | Gran empresa | 69.618 | 0,4278 | 0,4948 | 0 | 1 |
| Sector de la industria | Sector primario | 69.618 | 0,1382 | 0,3451 | 0 | 1 |
| | Extracción de recursos | 69.618 | 0,0546 | 0,2272 | 0 | 1 |
| | Manufactura | 69.618 | 0,0828 | 0,2756 | 0 | 1 |
| | Servicios básicos | 69.618 | 0,0095 | 0,0968 | 0 | 1 |
| | Construcción | 69.618 | 0,0806 | 0,2726 | 0 | 1 |
| | Comercio | 69.618 | 0,1596 | 0,3663 | 0 | 1 |
| | Transporte | 69.618 | 0,0660 | 0,2482 | 0 | 1 |
| | Hospitalidad | 69.618 | 0,0396 | 0,1950 | 0 | 1 |
| | Inmobiliaria | 69.618 | 0,0532 | 0,2245 | 0 | 1 |
| | Otros servicios | 69.618 | 0,3156 | 0,4648 | 0 | 1 |
| Área de trabajo | Fuerzas Armadas | 69.618 | 0,0060 | 0,0772 | 0 | 1 |
| | Ejecutivos | 69.618 | 0,0082 | 0,0904 | 0 | 1 |
| | Académicos | 69.618 | 0,1147 | 0,3186 | 0 | 1 |
| | Técnicos | 69.618 | 0,0703 | 0,2557 | 0 | 1 |
| | Administrativos | 69.618 | 0,0918 | 0,2887 | 0 | 1 |
| | Vendedores | 69.618 | 0,1540 | 0,3610 | 0 | 1 |
| | Agricultores | 69.618 | 0,0496 | 0,2171 | 0 | 1 |
| | Artesanos | 69.618 | 0,1193 | 0,3241 | 0 | 1 |
| | Operadores | 69.618 | 0,1028 | 0,3037 | 0 | 1 |
| | Ocupaciones elementales | 69.618 | 0,2832 | 0,4506 | 0 | 1 |
| Sector laboral | Público | 69.618 | 0,1622 | 0,3686 | 0 | 1 |
| | Privado | 69.618 | 0,8378 | 0,3686 | 0 | 1 |

Fuente: Elaboración Propia

(Continuación) Tabla 9: Estadísticas descriptivas de las variables CASEN 2011

| Área | Variable | Observaciones | Media | Desviación estándar | Mínimo | Máximo |
|-------------------|----------------------|---------------|--------|---------------------|--------|--------|
| Macrozona | Macrozona norte | 69.618 | 0,2220 | 0,4156 | 0 | 1 |
| | Macrozona centro | 69.618 | 0,2694 | 0,4436 | 0 | 1 |
| | Macrozona centro sur | 69.618 | 0,2446 | 0,4298 | 0 | 1 |
| | Macrozona sur | 69.618 | 0,1871 | 0,39 | 0 | 1 |
| | Macrozona austral | 69.618 | 0,0770 | 0,2666 | 0 | 1 |
| Jefatura de hogar | Si es jefe de hogar | 69.618 | 0,4803 | 0,4996 | 0 | 1 |
| | No es jefe de hogar | 69.618 | 0,5197 | 0,4996 | | |
| Área | Urbano | 69.618 | 0,8203 | 0,3839 | 0 | 1 |
| | Rural | 69.618 | 0,1797 | 0,3839 | 0 | 1 |

Fuente: Elaboración Propia

5. Resultados

5.1 Regresión lineal múltiple

La regresión lineal múltiple es aplicada para cada muestra. Manteniendo la variable dependiente, logaritmo natural del salario por hora, pero disminuyendo la cantidad de variables independientes si es que la información no estaba disponible. Las categorías de sexo, estado civil, área de trabajo, macrozonas y área no tuvieron modificaciones en ninguna de las muestras.

Para realizar la regresión, por criterio se elimina en cada categoría la variable que tuviera mayor frecuencia. El criterio es aplicado a cada muestra del estudio. En la muestra del año 2022, había un 55,79% de hombres, 32,61% de personas con educación media completa, superando a los 15,13% de educación media incompleta, 26,63% de título técnico y a los 25,63% de título universitario o más. Además, había 42,47% de personas solteras, valor mayor a los de las otras variables de su categoría.

En tamaño de empresa, el 30,37% es gran empresa, valor muy superior a las otras variables, lo mismo para sector de la industria y área de trabajo, donde las variables de más frecuencias correspondían a 37,26% otros servicios y 19,78% ocupaciones elementales, respectivamente. Para sector laboral y macrozonas, la diferencia fue muy significativa, donde la variable privada tuvo un 79,63% de presencia, mientras que macrozona tuvo una de 34,47%. Finalmente, para hijos y área, se eliminó con un 72,96% a la variable con hijos y urbano con 83,26%. La Tabla 10 desglosa las variables independientes eliminadas para cada muestra de diferente año.

Tabla 10: Variable eliminada en cada categoría en base a la mayor frecuencia para cada muestra.

| Categoría | 1992 | 2000 | 2011 | 2022 |
|---------------------------|----------------------------|----------------------------|--------------------------|--------------------------|
| Sexo | Hombre | Hombre | Hombre | Hombre |
| Educación | Educación media incompleta | Educación media incompleta | Educación media completa | Educación media completa |
| Estado civil | Casado | Casado | Soltero | Soltero |
| Tipo de Contrato / Empleo | - | Empleo indefinido | Contrato indefinido | Contrato indefinido |
| Tamaño de la empresa | Microempresa | Microempresa | Gran empresa | Gran empresa |
| Sector de la industria | Sector primario | Sector primario | Otros servicios | Otros servicios |
| Área de trabajo | Ocupaciones elementales | Ocupaciones elementales | Ocupaciones elementales | Ocupaciones elementales |
| Sector laboral | - | - | Privado | Privado |
| Macrozonas | Macrozona centro | Macrozona centro | Macrozona centro | Macrozona centro |
| Hijos | - | - | - | Con hijos |
| Jefe de hogar | Si es jefe de hogar | Si es jefe de hogar | No es jefe de hogar | - |
| Área | Urbano | Urbano | Urbano | Urbano |

(a) No se considera jefe de hogar para el año 2022, debido a que es una única variable binaria, a diferencia de en los demás años.

Fuente: Elaboración Propia

5.1.1 Regresión lineal múltiple para 2022

Se realizó la regresión lineal en el programa STATA, cuyo modelo es estadísticamente significativo al 95% debido a la prueba *F*. Por otra parte, tiene un MSE de 0,2816 y un R^2 ajustado de 0,3785. En base al valor de MSE, se puede mencionar que la predicción es precisa, debido a que se tiene un valor bastante pequeño, lo que indica que los errores son menores.

La brecha salarial de género para 2022 es de 11,76% al 99,9% de confianza a favor de los hombres. Dentro de las variables con mayor impacto en la regresión lineal, se destaca a título universitario (coeficiente de 0,3141), microempresa (coeficiente de -0,1618), extracción de recursos (coeficiente de 0,2543), ejecutivos (coeficiente de 0,8190), académicos (coeficiente de 0,5928) y técnicos (coeficiente de 0,2329). La Tabla 11 muestra los coeficientes y error estándar robusto de cada variable independiente que participo en la regresión.

Tabla 11: Resultados regresión lineal múltiple para el año 2022

| <i>Variable dependiente:</i> Salario por horas (en logaritmo natural) | | |
|---|-------------------|-------------------------------|
| <i>Variables independientes</i> | Coficiente | Error estándar robusto |
| Mujer (Brecha salarial de género) | -0,1176*** | 0,0059 |
| Edad | 0,0164*** | 0,0014 |
| Edad2 | -0,0002*** | 0,00002 |
| Educación media incompleta | -0,0934*** | 0,0077 |
| Título técnico | 0,0673*** | 0,0067 |
| Título universitario o mas | 0,3143*** | 0,0105 |
| Casado | 0,1008*** | 0,0066 |
| Conviviente | 0,0461*** | 0,0064 |
| Contrato a plazo fijo | -0,0711*** | 0,0068 |
| Contrato de obra | -0,0292** | 0,0094 |
| Microempresa | -0,1618*** | 0,0073 |
| Pequeña empresa | -0,1251*** | 0,0070 |
| Mediana empresa | -0,0695*** | 0,0071 |
| Sector primario | -0,0395*** | 0,0098 |
| Extracción de recursos | 0,2543*** | 0,0147 |
| Manufactura | -0,0339** | 0,0105 |
| Servicios básicos | 0,0729* | 0,0295 |
| Servicios de agua y descontaminación | -0,0751** | 0,0247 |
| Construcción | 0,0368** | 0,0121 |
| Comercio | -0,0319*** | 0,0088 |
| Transporte | 0,0272* | 0,01365 |
| Hospitalidad | -0,0553*** | 0,1340 |
| Información | 0,0666** | 0,0224 |
| Inmobiliaria | 0,0835* | 0,0383 |
| Fuerzas Armadas | 0,3637*** | 0,0406 |
| Ejecutivos | 0,8190*** | 0,0227 |
| Académicos | 0,5928*** | 0,0131 |
| Técnicos | 0,2329*** | 0,0101 |
| Administrativos | 0,1032*** | 0,0114 |
| Vendedores | 0,0308*** | 0,0083 |
| Agricultores | -0,0056 | 0,014 |
| Artesanos | 0,1095*** | 0,0099 |
| Operadores | 0,1140*** | 0,0102 |
| Publico | 0,0067 | 0,0078 |
| Macrozona norte | -0,0317*** | 0,0082 |
| Macrozona centro sur | -0,1545*** | 0,0063 |
| Macrozona sur | -0,1101*** | 0,0077 |
| Macrozona austral | 0,042*** | 0,0109 |
| Sin hijos | 0,0096 | 0,0068 |
| Jefatura de hogar | 0,0759*** | 0,0053 |
| Rural | -0,0338*** | 0,0068 |
| Constante | 7,6739*** | 0,031 |
| <i>N</i> | | 45.959 |
| <i>R² ajustado</i> | | 0,3791*** |
| <i>MSE</i> | | 0,5309 |
| <i>Prob > F</i> | | 0,0000 |

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Fuente: Elaboración Propia

5.1.1 Regresión lineal múltiple para 1992, 2000 y 2011

Para 1992, la brecha salarial de género fue de 17,52%. La regresión lineal múltiple tiene un MSE de 0,4823 y un R^2 ajustado de 0,3724. Valores que hacen referencia a una predicción no muy exacta debido a que el MSE tiene un valor alto, mientras que el R^2 ajustado un valor bajo. Analizando la prueba F , se señala que el modelo es significativo al 95%.

Tabla 12: Resultados regresión lineal múltiple para el año 1992

| <i>Variable dependiente:</i> Salario por horas (en logaritmo natural) | | |
|---|---------------------|-------------------------------|
| <i>Variables independientes</i> | Coefficiente | Error estándar robusto |
| Mujer (Brecha salarial de género) | -0,1752*** | 0,0096 |
| Edad | 0,0228*** | 0,0016 |
| Edad2 | -0,0002*** | 0,00002 |
| Educación media completa | 0,2815*** | 0,0090 |
| Título universitario o mas | 0,7374*** | 0,0186 |
| Conviviente | -0,1379*** | 0,0129 |
| Soltero | -0,1164*** | 0,0084 |
| Pequeña empresa | -0,1031*** | 0,0081 |
| Mediana empresa | -0,0457*** | 0,010 |
| Gran empresa | 0,0562*** | 0,0109 |
| Extracción de recursos | 0,2415*** | 0,0195 |
| Manufactura | 0,0102 | 0,0131 |
| Servicios básicos | 0,0960* | 0,0416 |
| Construcción | 0,1017*** | 0,0148 |
| Comercio | -0,0123 | 0,015 |
| Transporte | 0,1267*** | 0,0174 |
| Otros servicios | -0,0787*** | 0,0118 |
| Fuerzas Armadas | 0 (Omitida) | |
| Ejecutivos | 1,1605*** | 0,0239 |
| Académicos | 0,8101*** | 0,0221 |
| Técnicos | 0,6441*** | 0,0190 |
| Administrativos | 0,3476*** | 0,0148 |
| Vendedores | 0,2414*** | 0,0134 |
| Agricultores | 0,2875*** | 0,0135 |
| Artesanos | 0,2184*** | 0,0111 |
| Operadores | 0,1920*** | 0,0136 |
| Macrozona norte | -0,0487*** | 0,0110 |
| Macrozona centro sur | -0,2104*** | 0,0080 |
| Macrozona sur | -0,1502*** | 0,0116 |
| Macrozona austral | 0,0377* | 0,0168 |
| No es jefe de hogar | -0,1233*** | 0,0091 |
| Rural | -0,0595*** | 0,0089 |
| Constante | 5,2980*** | 0,0347 |
| <i>N</i> | | 45.238 |
| <i>R² ajustado</i> | | 0,3728 |
| <i>MSE</i> | | 0,69472 |
| <i>Prob>F</i> | | 0,0000 |

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Fuente: Elaboración Propia

En el caso del año 2000, la brecha salarial de género es de un 16,77%. El modelo tiene un MSE de 0,4266, R^2 ajustado de 0,3602 y F que señala que es significativo al 95%. Las variables que mayor contribuyen a la regresión son título técnico, título universitario o más, extracción de recursos, fuerzas armadas, ejecutivos, entre otros.

Tabla 13: Resultados regresión lineal múltiple para el año 2000

| <i>Variable dependiente:</i> Salario por horas (en logaritmo natural) | | |
|---|---------------------|-------------------------------|
| <i>Variables independientes</i> | Coefficiente | Error estándar robusto |
| Mujer (Brecha salarial de género) | -0,1677*** | 0,0070 |
| Edad | 0,0208*** | 0,0013 |
| Edad2 | -0,0002*** | 0,00002 |
| Educación media completa | 0,2353*** | 0,0067 |
| Título técnico | 0,4484*** | 0,0147 |
| Título universitario o mas | 0,8425*** | 0,0188 |
| Conviviente | -0,0950*** | 0,0080 |
| Soltero | -0,1008*** | 0,0063 |
| Empleo a plazo fijo | -0,1014*** | 0,0068 |
| Empleo de obra | -0,1507*** | 0,0118 |
| Pequeña empresa | -0,1096*** | 0,0063 |
| Mediana empresa | -0,0795*** | 0,0073 |
| Gran empresa | -0,0112 | 0,0071 |
| Extracción de recursos | 0,2482*** | 0,0188 |
| Manufactura | 0,0447*** | 0,0103 |
| Servicios básicos | 0,1093*** | 0,0267 |
| Construcción | 0,1031*** | 0,0108 |
| Comercio | -0,0507*** | 0,0109 |
| Transporte | 0,0909*** | 0,0140 |
| Otros servicios | 0,0390*** | 0,0086 |
| Fuerzas Armadas | 0,3563*** | 0,0342 |
| Ejecutivos | 0,8717*** | 0,0157 |
| Académicos | 0,7774*** | 0,0192 |
| Técnicos | 0,6093*** | 0,0143 |
| Administrativos | 0,3292*** | 0,0107 |
| Vendedores | 0,2061*** | 0,01 |
| Agricultores | 0,2066*** | 0,0088 |
| Artesanos | 0,2349*** | 0,0091 |
| Operadores | 0,1780*** | 0,0098 |
| Macrozona norte | -0,0166 | 0,0098 |
| Macrozona centro sur | -0,1528*** | 0,0059 |
| Macrozona sur | -0,1102*** | 0,0076 |
| Macrozona austral | 0,1583*** | 0,0171 |
| No es jefe de hogar | -0,1287*** | 0,0064 |
| Rural | -0,0494*** | 0,0060 |
| Constante | 6,1077*** | 0,0282 |
| <i>N</i> | | 70.147 |
| R^2 ajustado | | 0,3606 |
| MSE | | 0,6533 |
| <i>Prob>F</i> | | 0,0000 |

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Fuente: Elaboración Propia

La regresión lineal para 2011 tiene un MSE de 0,3650 y un R^2 ajustado de 0,4009. Considerando F global, se tiene que el modelo es significativo al 95%. La brecha salarial de género para 2011 es de 16,29% perjudicando a las mujeres.

Tabla 14: Resultados regresión lineal múltiple para el año 2011

| <i>Variable dependiente:</i> Salario por horas (en logaritmo natural) | | |
|---|---------------------|-------------------------------|
| <i>Variables independientes</i> | Coefficiente | Error estándar robusto |
| <i>Mujer (Brecha salarial de género)</i> | -0,1629*** | 0,0060 |
| Edad | 0,01379*** | 0,0012 |
| Edad2 | -0,0001*** | 0,00001 |
| Educación media incompleta | -0,1734*** | 0,0059 |
| Título universitario o mas | 0,334*** | 0,0093 |
| Casado | 0,0764*** | 0,0058 |
| Conviviente | 0,0262*** | 0,0063 |
| Contrato a plazo fijo | -0,1585*** | 0,0059 |
| Microempresa | -0,2163*** | 0,0065 |
| Pequeña empresa | -0,1232*** | 0,0064 |
| Mediana empresa | -0,0710*** | 0,0070 |
| Sector primario | -0,0594*** | 0,0097 |
| Extracción de recursos | 0,2227*** | 0,0137 |
| Manufactura | 0,0007 | 0,0101 |
| Servicios básicos | 0,0660** | 0,0237 |
| Construcción | 0,0438*** | 0,0106 |
| Comercio | -0,080*** | 0,0083 |
| Transporte | -0,0296* | 0,0116 |
| Hospitalidad | -0,0182 | 0,0131 |
| Inmobiliaria | 0,0807*** | 0,0116 |
| Fuerzas Armadas | 0,4359*** | 0,0313 |
| Ejecutivos | 0,8149*** | 0,0407 |
| Académicos | 0,7693*** | 0,0125 |
| Técnicos | 0,3452*** | 0,0116 |
| Administrativos | 0,1856*** | 0,0091 |
| Vendedores | 0,1043*** | 0,0084 |
| Agricultores | 0,0417*** | 0,0118 |
| Artesanos | 0,1219*** | 0,0084 |
| Operadores | 0,1126*** | 0,0089 |
| Publico | 0,0164* | 0,0080 |
| Macrozona norte | 0,0583*** | 0,0069 |
| Macrozona centro sur | -0,1130*** | 0,0064 |
| Macrozona sur | -0,0967*** | 0,0070 |
| Macrozona austral | 0,1053*** | 0,0093 |
| Si es jefe de hogar | 0,1260*** | 0,0052 |
| Rural | 0,0027 | 0,0067 |
| Constante | 7,1066*** | 0,0254 |
| <i>N</i> | | 69,618 |
| <i>R² ajustado</i> | | 0,4013 |
| <i>MSE</i> | | 0,6043 |
| <i>Prob>F</i> | | 0,0000 |

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Fuente: Elaboración Propia

Las variables que más perjudican a la regresión lineal para la muestra del 2011 son ejecutivos, académicos, fuerzas armadas, entre otros. Finalmente, la Tabla 15, muestra de forma compacta la brecha salarial de género para cada año, además de los valores estadísticos relevantes. Mientras la figura 1, muestra la evolución de brecha salarial de género utilizando regresión lineal múltiple.

Tabla 15: Resultados de la brecha salarial de género en Chile para los últimos 30 años utilizando regresión lineal múltiple.

| Años | <i>N</i> | Brecha salarial de género | <i>R</i>² ajustado | MSE | <i>Prob>F</i> |
|-------------|-----------------|----------------------------------|--------------------------------------|------------|-------------------------|
| 1992 | 45.238 | 17,5100%*** | 0,3728*** | 0,4823 | 0,0000 |
| 2000 | 70.147 | 16,7700%*** | 0,3606*** | 0,4266 | 0,0000 |
| 2011 | 69.618 | 16,2800%*** | 0,4013*** | 0,3650 | 0,0000 |
| 2022 | 45.959 | 11,7600%*** | 0,3791*** | 0,2816 | 0,0000 |

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Fuente: Elaboración Propia

5.2 Regresión simbólica

La regresión simbólica (RS) permite obtener una expresión que se ajuste a la información de una base de datos, debido a ello intenta explicar la variable dependiente de forma completa. Para aplicar la RS, se utiliza “PySR” que es un módulo de Python realizado por Cranmer (2023) que permite trabajar de forma ágil y sencilla.

Al inicio, se considera para realizar la regresión simbólica con los siguientes hiperparámetros, 100 iteraciones de 31 poblaciones cada una. Además, se fijó una semilla. Los operadores utilizados fueron seno, coseno, exponencial, logaritmo en base 10, suma, resta, multiplicación y división. Después solo se consideró la suma, resta y multiplicación para mayor comprensión de la ecuación. Como criterio de selección se utiliza el método accuracy, el cual consiste en obtener la ecuación con menor pérdida o error en la predicción, valor relacionado con el error cuadrático medio (MSE). Por otra parte, se utiliza valores de entrenamiento y de prueba, los cuales fueron escogidos de forma aleatoria, teniendo una relación de 80% y 20%, respectivamente. Los primeros intentos de regresión simbólica obtenida para el año 2022, entregaban ecuaciones donde las variables se dividían, multiplicaban y no se podía interpretar de forma sencilla. Además de que no estaba presente la variable mujer.

Generalmente la mejor ecuación obtenido con RS no incorpora aquella variable dentro de las ecuaciones. Debido a esto se modificó la complejidad de la variable de género, disminuyendo su valor desde 1 (valor por defecto) a 0,0001. Con ello, se buscaba que la variable sea considerada.

Aun con esta medida, hay ecuaciones incluyendo la “mejor ecuación”, que no consideran la variable de género. Es producto de ello, que al conjunto de “buenas ecuaciones”, es decir, las ecuaciones finales fueron sometidas a un filtro. El cual consiste en discriminar a aquellas variables que no poseen la variable de sexo en su composición. De esta forma, se buscó que la SR esté dentro de las ecuaciones factibles que contienen la variable independiente de interés. Se selecciono la ecuación que posee menor pérdida.

Además, se modificó los parámetros para cada año, con el objetivo de obtener una ecuación que fuera precisa e interpretable. Considerando como fijos los parámetros, método de selección “accuracy”, semilla, complejidad de las variables y operadores.

5.2.1 Primeros experimentos de regresión simbólica

Se realizaron varios experimentos para determinar los hiper parámetros adecuados para cada año, En un comienzo no se trabajó con una semilla, lo que provocó que los resultados fueran diferentes cada vez que se solicitara una ecuación. Este trabajo permitió ver el comportamiento de los datos, y las ecuaciones que la regresión simbólica iba entregando.

En base a lo observado en esas instancias, se presionó al programa a considerar la variable de género. En el caso que la ecuación escogida por la regresión simbólica no la considerara, se escogía aquella factible con mejor MSE que cumpliera la restricción.

Los parámetros en las instancias iniciales eran moderadamente distintos, pero se mantuvo la condición de las operaciones permitidas. Se consideraba una mayor complejidad permitida para cada caso. Se permitieron complejidades máximas entre 70 y 100. Aun así, cuando la complejidad máxima era alta, las ecuaciones obtenidas no alcanzaban esos valores.

Las complejidades de las ecuaciones rondaban entre 30 y 50. Una característica que se observó de las ecuaciones con mayor complejidad corresponde a que algunas variables se multiplicaban entre ellas.

Luego, se aplicó la semilla, para obtener un resultado estable en el tiempo. Algunos hiper parámetros fueron modificados. El cambio principal, fue la complejidad máxima. Se disminuyó a 30 para 2022, 2011 y 2000, y a 40 para 1992. Debido a que a valores mayores no se obtenía resultado. Incluso, se utilizó tiempo máximo para facilitar el trabajo, pero aun con ello no se obtenía una ecuación. Otro hiper parámetro modificado fue el número de iteraciones para 1992, donde bajo de 100 a 80. Esto debido a que sucedía el mismo efecto que con las complejidades.

Finalmente, si se comparan los MSE de las ecuaciones obtenidas inicialmente con las ecuaciones finales, no son muy diferentes, siendo de aproximadamente 0,01. Además, las variables utilizadas son similares, pero la cantidad difiere en aproximadamente 1.

5.2.2 Regresión simbólica para 2022

Considerando los ajustes mencionados, se busca determinar la ecuación que mejor se ajustaba a los datos. Para ello se considera, además de los parámetros fijos, el número de iteraciones fijado en 100 y que la máxima complejidad permitida por la ecuación es fijada en tamaño 30.

Ese valor fue determinado debido a que mayores valores no consideran a la variable mujer dentro de sus ecuaciones y valores menores generan ecuaciones donde el error cuadrático medio es mayor. Cabe destacar que la muestra utilizada es la misma que la de la regresión lineal. Debido a que permite ver la variación de los coeficientes y el comportamiento de las variables.

La ecuación (15) presenta una complejidad de 26 y una pérdida de 0,3122. Tiene ocho variables independientes. La brecha salarial de género para el 2022 utilizando RS, con enfoque en balance entre precisión e interpretación es de 13,05%. La RS resultante, se muestra en (15), donde y es la variable dependiente, logaritmo del salario por hora.

$$\begin{aligned}
 y = & -0,1305 \times \text{mujer} + 0,8490 \times \text{academia} - 0,1305 \times \text{contrato fijo} \\
 & - 0,1305 \times \text{educación básica} + 0,1305 \times \text{extracción recursos} \\
 & - 0,1305 \times \text{macrozona centrosur} + 0,2609 \times \text{tecnicos} + \text{ejecutivos} + 8,1297
 \end{aligned}
 \tag{15}$$

El MSE de la RS para 2022 utilizando los datos de prueba fue de 0,3089. Mientras que el R^2 ajustado utilizando los mismos datos es de 0,3123. Se puede observar que las variables mujer, contrato fijo, educación básica y macrozona centro sur afecta negativamente al salario. Por el contrario, academia, extracción de recursos, técnicos y ejecutivos benefician el salario de las personas.

Comparando con respecto a la regresión lineal, el efecto de cada una de las variables es el mismo. La magnitud es distinta, siendo mayor en la regresión simbólica. En la regresión lineal el coeficiente de mujer es de 11,76% mientras que acá aumenta en más de un punto porcentual.

5.2.3 Regresión simbólica para 1992, 2000 y 2011

Se determinaron tres ecuaciones para ajustar los datos para cada año. Basado en el enfoque de precisión e interpretabilidad. En primer lugar, para 1992, se consideró los parámetros fijos y , además 80 iteraciones y una complejidad máxima de 40. Debido a que valores más pequeños consideraban muy pocas variables y un aumento de MSE. Se obtuvo la RS que se muestra en la ecuación (16), donde y es la variable dependiente, correspondiente al logaritmo del salario por hora.

$$y = -0,2047 \times \text{mujer} + 0,2047 \times \text{educación media} + \text{ejecutivos} + 0,2047 \times \text{gran empresa} - 0,2047 \times \text{macrozona centro sur} - 0,2047 \times \text{no es jefe de hogar} + \text{título universitario} + 5,9744 \quad (16)$$

La SR para 1992 tiene una pérdida de 0,5471 y una complejidad de 22. A diferencia de lo obtenido para 2022, la mejor ecuación no considero a la variable “mujer”. Debido a ello, se considera la mejor ecuación que tuviera a la variable de estudio. La ecuación contiene siete variables independientes, además de mostrar una brecha salarial de género del 20,47% a desventaja de las mujeres. El MSE y R^2 ajustado evaluados con respecto a los datos de prueba, son de 0,5477 y 0,2879, respectivamente.

Al igual que en 2022, los valores de la RS son mayores en relación con la regresión lineal. Además, los efectos de las variables son iguales en ambos métodos. Para el 2000 y 2011 se considera un número de iteraciones de 100 y una complejidad máxima de 30, debido a que valores mayores dificultaban el análisis. La RS para 2000 tiene una pérdida es de 0,4946, complejidad de 27 y considera siete variables independientes. La ecuación (17) fue la mejor obtenida por la RS. La brecha salarial de género es de 16,29%. El MSE y R^2 ajustado utilizando los datos de prueba son de 0,4946 y 0,2545, respectivamente.

$$y = -0,1629 \times \text{mujer} + 0,77512 \times \text{academia} + 0,9141 \times \text{ejecutivos} - 0,1629 \times \text{empleo fijo} + 0,7512 \times \text{fuerzas armadas} + 0,7512 \times \text{técnicos} + 0,7512 \times \text{título universitario} + 6,6103 \quad (17)$$

La regresión simbólica para el año 2011, tuvo una pérdida de 0,4196 y una complejidad de 16. Considera cinco variables independientes, debido a que algunas de las variables estuvieron repetidas. La brecha salarial es género es de 20,50% y tuvo un MSE de 0,4196, mientras que un R^2 ajustado de 0,2990, ambos fueron calculados utilizando los datos de prueba. La ecuación (18) muestra la RS.

$$y = -0,2050 \times \text{mujer} + 0,6631 \times \text{academia} - 0,2050 \times \text{contrato fijo} - 0,2050 \times \text{microempresa} + 0,6631 \times \text{título universitario} + 7,4018 \quad (18)$$

Finalmente, los efectos de las variables en la regresión simbólica son los mismo que en la regresión lineal. En su mayoría, la regresión simbólica tiene valores más altos en los coeficientes. Además, presenta para todos los años mayores valores de MSE y menores valores de R^2 ajustado.

5.3 Regresión lineal vs Regresión simbólica

El enfoque que se buscó con la RS es una ecuación que estime la brecha salarial de género, además posea un MSE pequeño y se pueda interpretar. Considerando solo esas características se logró el objetivo.

Aun así, la RS aumenta el MSE en todos los años estudiados en comparación con la regresión lineal. Para 2022, aumenta el MSE en un 9,695%, mientras que para 2011, 2000 y 1992, aumenta 14,96%, 17,11% y 13,56%, respectivamente. Para 1992 y 2022, el aumento no fue tan grande como los demás años.

Si se analiza el R^2 ajustado, la RS entrega peores valores, esto se produce principalmente, por que las expresiones no buscan aumentar aquel estadístico, sino que tener un mejor ajuste considerando los residuos. Por otra parte, el valor de este estadístico puede ser debido a la pequeña cantidad de variables consideradas por la RS.

Si nos enfocamos en la brecha salarial de género, los valores se asemejan bastante. La mayor diferencia se produce en 2011 y 1992, donde es de aproximadamente 3 puntos porcentuales. Hay evidencia diversa con respecto a valores de la brecha salarial de género. Para el año 2002, Meller et al. (2011) señala una brecha salarial de género de un 22,97%. (Salce Díaz, 2021) señala en sus resultados un aumento en la brecha salarial de género alrededor del 2010, además señala que en la década entre 1990-2000 las brechas salariales eran por sobre el 30%. Aguayo (2023) señala que la brecha salarial en Chile en 2011 fue de 32,1% y para 2022 fue de 16%.

La regresión simbólica, pudo haber tenido peor desempeño debido a las restricciones rígidas exigidas. Una de ellas, es la limitación de los operadores básicos y unitarios, buscando una mayor interpretación de la ecuación. Aun así, la regresión simbólica intento ser similar a la regresión lineal múltiple, pero no pudo ser mejor que la original. Considerando que la regresión lineal tiene por objetivo entregar una ecuación lineal.

Finalmente, la regresión lineal múltiple entrega mejores resultados, basándose en el MSE. Disminuye los errores de la predicción y genera ecuaciones con mayor porcentaje predictivo. La Tabla 16, presenta la comparación de los resultados de la metodología.

Tabla 16: Comparación información entrega por la regresión lineal y regresión simbólica

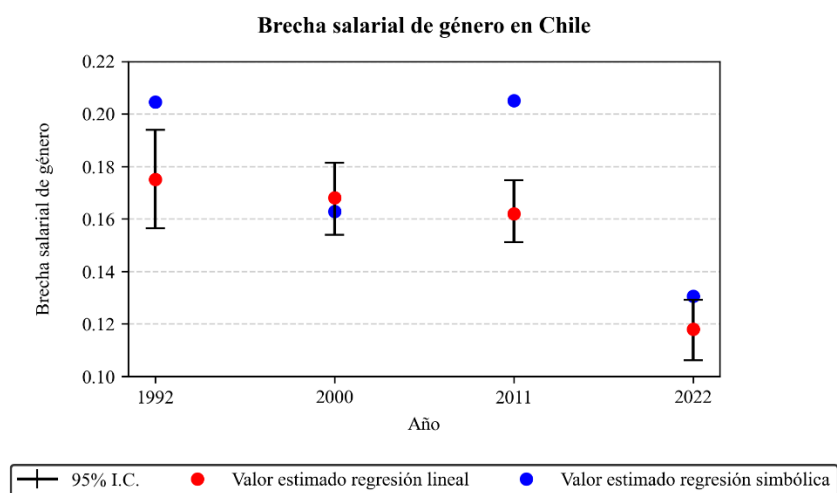
| Característica | Modelo | 2022 | 2011 | 2000 | 1992 |
|------------------------------------|--------------------|----------|----------|----------|----------|
| Brecha salarial de género | RL | 11,7600% | 16,2900% | 16,7200% | 17,5200% |
| | RS | 13,0500% | 20,5000% | 16,2800% | 20,4600% |
| MSE | RL | 0,2816 | 0,3650 | 0,4266 | 0,4823 |
| | RS (prueba) | 0,3089 | 0,4196 | 0,4946 | 0,5477 |
| | RS (entrenamiento) | 0,3122 | 0,4196 | 0,4946 | 0,5471 |
| R ² Ajustado | RL | 0,3791 | 0,4013 | 0,3606 | 0,3728 |
| | RS (prueba) | 0,3123 | 0,2990 | 0,2542 | 0,2913 |
| | RS (entrenamiento) | 0,3121 | 0,3147 | 0,2591 | 0,2860 |
| Número de variables independientes | RL | 41 | 36 | 35 | 32 |
| | RS | 8 | 5 | 7 | 7 |

Nota: (a) RL significa regresión lineal y RS regresión simbólica (b) El número de variables independientes considera las que se encuentran en la regresión, en el caso de 2000, la variable “Fuerzas armadas”, fue omitida, aunque en este caso igual se considera dentro del número.

Fuente: Elaboración Propia

La Figura 1, muestra la brecha salarial de género utilizando regresión lineal múltiple y RS. Para la regresión lineal, muestra el intervalo de confianza de cada uno de los resultados, mientras que para la RS muestra en los puntos azules el resultado. Se puede observar la diferencia en los resultados de los métodos.

Figura 1: Resultados de la regresión lineal múltiple y regresión simbólica



Fuente: Elaboración Propia

6. Conclusión

La brecha salarial de género es un tema ampliamente estudiado en la literatura. Posee gran variedad de resultados, dependiendo de la fuente de datos, variables y/o metodología utilizada. El caso de Chile no es la excepción. Diversos autores han estimado la brecha salarial de género, obteniendo valores distintos para un mismo periodo.

La literatura generalmente realiza una estimación de salario utilizando la ecuación de Mincer en una regresión lineal múltiple para estimar la brecha salarial de género. En esta memoria se propone un nuevo método de estimación, regresión simbólica que es basada en técnicas de Machine Learning. Su objetivo es entregar en base a los datos una ecuación que se ajuste a ellos. Posteriormente, ambos métodos fueron comparados, con el fin de determinar cuál estima de mejor forma la brecha salarial de género.

Los datos se obtuvieron de la Encuesta de Caracterización Socioeconómica de Chile (CASEN) para 1992, 2000, 2011 y 2022, con el fin de determinar cómo ha evolucionado la brecha salarial de género en Chile. Los datos disponibles en cada una de las encuestas varía, debido a que las preguntas realizadas en cada uno de los años tenían diferencias, las cuales no eran de gran magnitud. Es producto de ello que no todas las muestras presentan la misma cantidad de categorías y/o variables. Se diferencian principalmente en las categorías de educación, tipo de contrato, sector laboral e hijos. Además, las muestras fueron sometidas a filtraciones de datos, considerando solo a trabajadores, personas mayores de 15 años y aquellos que posean datos válidos.

Para realizar la estimación, se aplicó como criterio a todas las muestras la eliminación en cada categoría de la variable con mayor frecuencia. Además de identificar a la variable dependiente como el logaritmo natural del salario por hora. Cabe destacar, que el valor estudiado, brecha salarial de género, es el valor que acompaña a la variable mujer.

La regresión lineal múltiple para cada año fue significativa al 95% considerando la prueba F . Las brechas salariales de género fueron 11,76%, 16,28%, 16,8% y 17,5% para 2022, 2011, 2000 y 1992, respectivamente. El error cuadrático medio, tuvo su valor más alto para 1992 con 0,4823 y el valor más bajo para 2022 con 0,2816. Tuvieron valores de R^2 ajustado pequeños, siendo de 0,3785 para 2022.

Mientras, para la regresión simbólica, se utilizó un enfoque basado en la precisión, buscando disminuir el error cuadrático medio. Se utilizó muestra de entrenamiento y prueba, considerando la relación 80% y 20%.

Los hiper parámetros utilizados para estimar la brecha salarial de género para cada año fueron principalmente los mismos, a excepción de los utilizados en 1992. El número de iteraciones fue 100, mientras que el máximo valor de complejidad fue de 30. Por otra parte, para el caso de 1992, el número de iteraciones fue de 80 y el valor máximo de complejidad fue de 40.

Para todos los años, se buscó obtener resultados permanentes con los parámetros utilizados. Es debido a ello, que se especificó que se resolviera de forma determinística y serial. Se destaca, que las operaciones permitidas para la construcción de las ecuaciones se limitaron. Las operaciones podían ser multiplicación, suma y/o resta, debido a que mayor cantidad de operaciones aumentaban la complejidad de la ecuación y con ello, la compresión de la variable.

Además, se utilizó penalizaciones para poder incorporar la variable mujer dentro de la ecuación. Aun así, en el caso de 1992, la mejor expresión obtenida no considero esa variable. Es por ello, que se escogió dentro de las “buenas soluciones”, filtrando las que tuvieran la variable de sexo y escogiendo aquella con mejor precisión.

La regresión simbólica, obtuvo una brecha salarial de género para 2022 de 13,05%, mientras que para 2011, 2000 y 1992 fueron de 20,50%, 16,28% y 20,46%, respectivamente. Considerando el error cuadrático medio utilizando los datos de prueba se obtuvo 0,3089 para 2022, 0,4196 para 2011, 0,4946 para 2000 y 0,5477 para 1992. Observando el R^2 ajustado utilizando los datos de prueba se tienen valores de 0,3123, 0,2990, 0,2542, 0,2913, para 2022, 2011, 2000 y 1992. Además, las ecuaciones consideraron entre cinco a ocho variables independientes en su ecuación.

El criterio para comparar los dos métodos es el error cuadrático medio. Considerando ello, para todos los años la regresión lineal posee un menor valor, lo que indica una mayor precisión. Por otra parte, las brechas salariales de la regresión simbólica son mayores en la mayoría de los casos, menos para el 2000. Es por ello, que se puede considerar que la

regresión simbólica enfatiza en una mayor brecha salarial de género, lo que podría ser debido a la menor cantidad de variables en la ecuación.

La regresión lineal múltiple fue mejor que la regresión simbólica, considerando el error cuadrático medio. Una de las posibles razones de ello, fue la reducción de las operaciones disponibles en la regresión simbólica. Al solo considerar multiplicación, suma y resta, se pretendió buscar una ecuación lineal, lo que produce que la regresión simbólica no pueda llegar a ser mejor que la regresión lineal múltiple. El fin de la regresión lineal es encontrar una ecuación lineal, mientras que la regresión simbólica intenta realizarlo, acercándose a ella, pero siendo estadísticamente peor.

Si no se limitan los operadores, la regresión simbólica podría llegar a estimar mejor considerando algún criterio. Futuras investigaciones, pueden abordar esta temática. Destacando que un mejor ajuste de los datos puede ser una expresión no lineal, en la cual para obtener la brecha salarial de género se deban realizar operaciones matemáticas. Por otro lado, en el futuro se puede ampliar el estudio al incorporar la decisión de participar en el mercado por género utilizando regresión simbólica.

La investigación estimo la brecha salarial de género en Chile utilizando la regresión lineal múltiple y regresión simbólica para 1992, 2000, 2011 y 2022. De esta forma, se cumple el objetivo principal de la presente memoria de título. Además, se presentó el estado del arte del tema, se comparó y discutió los resultados obtenidos, cumpliendo los tres objetivos específicos.

La brecha salarial de género no es un tema trivial, debido a que perjudican a un gran sector de la población, ya sea en Chile o en el mundo. Se recomienda, considerar estos valores para realizar políticas de disminución de ella.

7. Referencias

- Aguayo, I. (2023). Brecha Salarial de género en Chile utilizando datos de CASEN 2017 y CASEN 2022, Asesoría Técnica Parlamentaria. https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/35574/1/BCN_112023_Brecha_salarial.pdf
- Banco Mundial. (2024). Indicadores del desarrollo mundial: Países y economías [Conjunto de datos] <https://datos.bancomundial.org/>
- Blau, F. D., & Kahn, L. M. (1980). *The Gender Wage Gap: Extent, Trends, and Explanations*. <https://doi.org/10.1257/jel.20160995>
- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436. <https://doi.org/10.2307/144855>
- Bonaccolto-Töpfer, M., Castagnetti, C., & Rosti, L. (2023). Changes in the gender pay gap over time: the case of West Germany. *Journal for Labour Market Research*, 57(1), 1–17. <https://doi.org/10.1186/S12651-023-00338-7/FIGURES/3>
- Cabrera, G.-, Mercedes, A., & Cedres-Hernández, Y. (2024). Igualdad de género en el mercado laboral: Retos y cambio institucional. *Revista de Ciencias Sociales (RCS)*, XXX(1), 77–96.
- Cranmer, M. (2023) Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. arXiv: <https://arxiv.org/abs/2305.01582>
- Código del Trabajo en Chile (2002). Ley no. 20.416 ART 505 bis. Última modificación, Ley 21. 690. 24 de agosto de 2024. https://www.dt.gob.cl/legislacion/1624/articles-95516_recurso_1.pdf
- Díaz, F. S. (2021). Evolution and analysis of salary discrimination by gender in Chile. *Trimestre Económico*, 88(349), 39–75. <https://doi.org/10.20430/ETE.V88I349.984>
- Didier, N. (2021). Does credentialism affect the gender wage gap? Evidence from Chile. *Latin American Policy*, 12(1), 69–96. <https://doi.org/10.1111/lamp.12209>
- Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4), 1091–1119. <https://doi.org/10.1257/aer.104.4.1091>

- Hair, J., Anderson, R., Tatham, R., Black, W. (1999) *Análisis Multivariante*. (pp. 160-240). Pretence Hall
- Instituto Nacional de Estadística. (2023). Género y mercado laboral en Chile [Presentación] Mes del trabajador, Chile.
https://www.ine.gob.cl/docs/default-source/genero/infograf%C3%ADas/autonomia-economica/inf-ond22-empleo-sexo.pdf?sfvrsn=aa849346_3
- Instituto Nacional de Estadística. (2023). Género y desigualdad de ingresos en Chile [Presentación] Día Internacional de las mujeres trabajadoras, Chile.
https://www.ine.gob.cl/docs/default-source/prensa-y-comunicacion/g%C3%A9nero-y-desigualdad-de-ingresos-en-chile.pdf?sfvrsn=64af7d60_2
- Kristjanpoller, W., Michell, K., & Olson, J. E. (2023). Determining the gender wage gap through causal inference and machine learning models: evidence from Chile. *Neural Computing and Applications*, 35(13), 9841–9863.
<https://doi.org/10.1007/s00521-023-08221-9>
- Kronberger, G., Burlacu, B., Kommenda, M., Winkler, S. M., & Affenzeller, M. (2025). *Symbolic Regression*.
- Kunze, A. (2017). The Gender Wage Gap in Developed Countries. *The Oxford Handbook of Women and the Economy*, 369–394.
<https://doi.org/10.1093/OXFORDHOB/9780190628963.013.11>
- Makke, N., & Chawla, S. (2024). Interpretable scientific discovery with symbolic regression: a review. *Artificial Intelligence Review*, 57(1), 1–38.
<https://doi.org/10.1007/S10462-023-10622-0/TABLES/6>
- Meara, K., Pastore, F., & Webster, A. (2020). The gender pay gap in the USA: a matching study. *Journal of Population Economics*, 33(1), 271–305.
<https://doi.org/10.1007/S00148-019-00743-8/TABLES/10>
- Meller, P., Valdés, G., & Lara, B. (2011). Female Discrimination at the professional level in Chile (Vol. 36).
- Meriküll, J., & Tverdostup, M. (2023). The gap that survived the transition: The gender wage gap in Estonia over three decades. *Economic Systems*, 47(3), 101127. <https://doi.org/10.1016/J.ECOSYS.2023.101127>
- Mincer, J. A. (1974). The human capital earnings function. In *Schooling, experience, and earnings* (pp. 83-96). NBER.

- Ministerio de desarrollo social y familia. (1992). La Encuesta de Caracterización Socioeconómica Nacional [Conjunto de datos]. Encuesta Nacional del Gobierno de Chile.
<https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-1992>
- Ministerio de desarrollo social y familia. (2000). La Encuesta de Caracterización Socioeconómica Nacional [Conjunto de datos]. Encuesta Nacional del Gobierno de Chile.
<https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2000>
- Ministerio de desarrollo social y familia. (2011). La Encuesta de Caracterización Socioeconómica Nacional [Conjunto de datos]. Encuesta Nacional del Gobierno de Chile.
<https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2011>
- Ministerio de desarrollo social y familia. (2022). La Encuesta de Caracterización Socioeconómica Nacional [Conjunto de datos]. Encuesta Nacional del Gobierno de Chile.
<https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2022>
- Ministerio de Economía, Fomento y Turismo. (2014). Antecedentes para la revisión de los criterios de clasificación del Estatuto Pyme. División de estudios Gobierno de Chile. Montgomery, D. C., & Runger, G. C. (2004). Proba
- Montgomery, D. C., & Runger, G. C. (2004). Probabilidad y estadística aplicadas a la ingeniería (2.a ed.). Editorial Limusa.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693. <https://doi.org/10.2307/2525981>
- Observatorio del Contexto económico UDP (OCEC udep), Gobierno de Chile. (2022). Estudio Anual de Zoom de género [Informe anual]
<https://www.chilemujeres.cl/wp-content/uploads/2023/04/Estudio-Anual-Zoom-de-Genero-2023.pdf>
- Oficina Internacional del Trabajo. (2020). Panorama Laboral 2020 [Informe Anual]. Oficina Regional para América Latina y el Caribe.
- Oficina Internacional del Trabajo. (2024). La brecha salarial de género.
<https://www.ilo.org/resource/other/gender-pay-gap>

- Olivetti, C., & Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, 8(Volume 8, 2016), 405–434. <https://doi.org/10.1146/ANNUREV-ECONOMICS-080614-115329/1>
- Organización para la Cooperación y Desarrollos Económicos. (2025, enero, 15). OECD Economic Surveys: Chile 2025. https://www.oecd.org/en/publications/oecd-economic-surveys-chile-2025_efad96ce-en.html
- Organización para la Cooperación y Desarrollos Económicos. (2023). Gender wage gap [Conjunto de datos] <https://www.oecd.org/en/data/indicators/gender-wage-gap.html?oecdcontrol6c2e197565-var3=1988>
- Ortogonal, D. Q. (2024). The gender wage gap and domestic violence against women: evidence from Peru. *Desarrollo y Sociedad*, 2024(96), 39–63. <https://doi.org/10.13043/DYS.96.2>
- Paz, J. A. (2019). The gender gap in Argentina: An analysis on labor market segregation. *Sociedade e Cultura*, 22(1), 1–27. <https://doi.org/10.5216/sec.v22i1.57894>
- Perticará, M., & Tejada, M. (2022). Sources of gender wage gaps for skilled workers in Latin American countries. *Journal of Economic Inequality*, 20(2), 439–463. <https://doi.org/10.1007/s10888-021-09500-y>
- Rembeza, J., & Radlińska, K. (2021). Labor market discrimination – are women still more secondary workers? *Journal of Business Economics and Management*, 22(1), 77–97. <https://doi.org/10.3846/jbem.2020.13648>
- Salazar Aramayo, J. Á., Iraizos Quintanilla, D., & Pereira Delgadillo, M. A. A. (2025). Evidencia sobre discriminación salarial por género en Bolivia. *Investigación & Desarrollo*, 24(3). <https://doi.org/10.23881/idupbo.024.3-4e>
- Salce Díaz, F. (2021). Evolución y análisis de la discriminación salarial por género en Chile. *El Trimestre Económico*, 88(349), 39–75. <https://doi.org/10.20430/ETE.V88I349.984>
- Sánchez, R., Finot, J., & Villena, M. G. (2022). Gender wage gap and firm market power: evidence from Chile. *Applied Economics*, 54(18), 2109–2121. <https://doi.org/10.1080/00036846.2021.1985070>

- Semenza, R., Boccardo, G., & Sarti, S. (2021). So Far, so Similar? Labour Market Feminization in Italy and Chile. *Social Indicators Research*, 154(3), 917–942. <https://doi.org/10.1007/s11205-020-02551-0>
- Si, C., Nadolnyak, D., & Hartarska, V. (2020). The Gender Wage Gap in Developing Countries. *Applied Economics and Finance*, 8(1), 1. <https://doi.org/10.11114/aef.v8i1.5082>
- Vaccaro, G., Basurto, M. P., Beltrán, A., & Montoya, M. (2022). The Gender Wage Gap in Peru: Drivers, Evolution, and Heterogeneities. *Social Inclusion*, 10(1), 19–34. <https://doi.org/10.17645/si.v10i1.4757>
- Vargas Calderón, D. F., Rios Obando, J. F., Rengifo Ortiz, C. I., & Serna Muñoz, K. G. (2024). Gender Wage Gap: A Labor Market Analysis of Cali-Colombia and its Metropolitan Area. *Revista Venezolana de Gerencia*, 29(11 Especial), 667–683. <https://doi.org/10.52080/rvgluz.29.e11.40>
- Wang, C., Zhang, Y., Wen, C., Yang, M., Lookman, T., Su, Y., & Zhang, T. Y. (2022). Symbolic regression in materials science via dimension-synchronous-computation. *Journal of Materials Science & Technology*, 122, 77–83. <https://doi.org/10.1016/J.JMST.2021.12.052>

8. Anexos

Tabla 17: Variables de la investigación CASEN 1992

| Área | Variable | Observaciones |
|-------------------------|----------------------------|---|
| Salario | Salario por hora | Salario bruto dividido horas trabajadas semanalmente por 4 |
| | Logaritmo salario | Se aplica logaritmo natural a salario por hora |
| Sexo | Mujer | Sea 1 si es mujer, 0 en otro caso. |
| | Hombre | Sea 1 si es hombre, 0 en otro caso. |
| Edad | Edad | Valor entre 15 a 120 años |
| | Edad2 | Valor al cuadrado de Edad |
| Educación | Educación media incompleta | Sea 1 si posee la educación media incompleta, 0 en otro caso |
| | Educación media completa | Sea 1 si solo posee la educación media completa, 0 en otro caso |
| | Título universitario o mas | Sea 1 si posee título técnico, universitario o más, 0 en otro caso |
| Estado civil | Casado | Sea 1 si se encuentra casado, 0 en otro caso |
| | Conviviente | Sea 1 si es conviviente, 0 en otro caso |
| | Soltero | Sea 1 si es soltero, anulante, separado, divorciado o viudo, 0 en otro caso |
| Tamaño de la empresa | Microempresa | Sea 1 si es una microempresa, 0 en otro caso |
| | Pequeña empresa | Sea 1 si una pequeña empresa, 0 en otro caso |
| | Mediana empresa | Sea 1 si es una mediana empresa, 0 en otro caso |
| | Gran empresa | Sea 1 si es una gran empresa, 0 en otro caso |
| Sector de la industria | Sector primario | Sea 1 si pertenece al sector primario, 0 en otro caso |
| | Extracción de recursos | Sea 1 si pertenece al sector de la extracción de recursos, 0 en otro caso |
| | Manufactura | Sea 1 si pertenece al sector de manufactura, 0 en otro caso |
| | Servicios básicos | Sea 1 si pertenece al sector de servicios básicos, 0 en otro caso |
| | Construcción | Sea 1 si pertenece al sector de la construcción, 0 en otro caso |
| | Comercio | Sea 1 si pertenece al sector de comercio, 0 en otro caso |
| | Transporte | Sea 1 si pertenece al sector de transporte, 0 en otro caso |
| | Otros servicios | Sea 1 si pertenece a otros servicios, 0 en otro caso |
| | Área de trabajo | Fuerzas Armadas |
| Ejecutivos | | Sea 1 si es del área ejecutiva, 0 en otro caso |
| Académicos | | Sea 1 si es del área académica, 0 en otro caso |
| Técnicos | | Sea 1 si es del área técnica, 0 en otro caso |
| Administrativos | | Sea 1 si es del área administrativa, 0 en otro caso |
| Vendedores | | Sea 1 si es vendedor, 0 en otro caso |
| Agricultores | | Sea 1 si es del área de la agricultura, 0 en otro caso |
| Artesanos | | Sea 1 si es artesano, 0 en otro caso |
| Operadores | | Sea 1 si es operador, 0 en otro caso |
| Ocupaciones elementales | | Sea 1 si tiene una ocupación elemental, 0 en otro caso |
| Macrozona | Macrozona norte | Sea 1 si pertenece a la macrozona norte, 0 en otro caso |
| | Macrozona centro | Sea 1 si pertenece a la macrozona centro, 0 en otro caso |
| | Macrozona centro sur | Sea 1 si pertenece a la macrozona centro-sur, 0 en otro caso |
| | Macrozona sur | Sea 1 si pertenece a la macrozona sur, 0 en otro caso |
| | Macrozona austral | Sea 1 si pertenece a la macrozona austral, 0 en otro caso |
| Jefe de hogar | Si es jefe de hogar | Sea 1 si es jefe de hogar, 0 en otro caso |
| | No es jefe de hogar | Sea 1 si no es jefe de hogar, 0 en otro caso |
| Área | Urbano | Sea 1 si se encuentra en un área urbana, 0 en otro caso |
| | Rural | Sea 1 si se encuentra en un área rural, 0 en otro caso |

Fuente: Elaboración Propia

Tabla 18: Variables de la investigación CASEN 2000

| Área | Variable | Observaciones |
|-------------------------|--|---|
| Salario | Salario por hora | Salario bruto dividido horas trabajadas |
| | Logaritmo salario | Se aplica logaritmo natural a salario por hora |
| Sexo | Mujer | Sea 1 si es mujer, 0 en otro caso. |
| | Hombre | Sea 1 si es hombre, 0 en otro caso. |
| Edad | Edad | Valor entre 15 a 120 años |
| | Edad2 | Valor al cuadrado de Edad |
| Educación | Educación media incompleta | Sea 1 si posee la educación media incompleta, 0 en otro caso |
| | Educación media completa | Sea 1 si solo posee la educación media completa, 0 en otro caso |
| | Título técnico | Sea 1 si solo posee un título técnico, 0 en otro caso |
| | Título universitario o mas | Sea 1 si posee título técnico, universitario o más, 0 en otro caso |
| Estado civil | Casado | Sea 1 si se encuentra casado, 0 en otro caso |
| | Conviviente | Sea 1 si es conviviente, 0 en otro caso |
| | Soltero | Sea 1 si es soltero, anulante, separado, divorciado o viudo, 0 en otro caso |
| Tipo de empleo | Empleo a plazo fijo | Sea 1 si posee contrato a plazo fijo, 0 en otro caso |
| | Empleo indefinido | Sea 1 si posee contrato indefinido, 0 en otro caso |
| | Empleo de obra | Sea 1 si posee contrato de obra, 0 en otro caso |
| Tamaño de la empresa | Microempresa | Sea 1 si es una microempresa, 0 en otro caso |
| | Pequeña empresa | Sea 1 si una pequeña empresa, 0 en otro caso |
| | Mediana empresa | Sea 1 si es una mediana empresa, 0 en otro caso |
| | Gran empresa | Sea 1 si es una gran empresa, 0 en otro caso |
| Sector de la industria | Sector primario | Sea 1 si pertenece al sector primario, 0 en otro caso |
| | Extracción de recursos | Sea 1 si pertenece al sector de la extracción de recursos, 0 en otro caso |
| | Manufactura | Sea 1 si pertenece al sector de manufactura, 0 en otro caso |
| | Servicios básicos | Sea 1 si pertenece al sector de servicios básicos, 0 en otro caso |
| | Construcción | Sea 1 si pertenece al sector de la construcción, 0 en otro caso |
| | Comercio | Sea 1 si pertenece al sector de comercio, 0 en otro caso |
| | Transporte | Sea 1 si pertenece al sector de transporte, 0 en otro caso |
| | Otros servicios | Sea 1 si pertenece a otros servicios, 0 en otro caso |
| Área de trabajo | Fuerzas Armadas | Sea 1 si pertenece a las fuerzas armadas, 0 en otro caso |
| | Ejecutivos | Sea 1 si es del área ejecutiva, 0 en otro caso |
| | Académicos | Sea 1 si es del área académica, 0 en otro caso |
| | Técnicos | Sea 1 si es del área técnica, 0 en otro caso |
| | Administrativos | Sea 1 si es del área administrativa, 0 en otro caso |
| | Vendedores | Sea 1 si es vendedor, 0 en otro caso |
| | Agricultores | Sea 1 si es del área de la agricultura, 0 en otro caso |
| | Artesanos | Sea 1 si es artesano, 0 en otro caso |
| | Operadores | Sea 1 si es operador, 0 en otro caso |
| Ocupaciones elementales | Sea 1 si tiene una ocupación elemental, 0 en otro caso | |

Fuente: Elaboración Propia

(Continuación) Tabla 18: Variables de la investigación CASEN 2000

| Área | Variable | Observaciones |
|---------------|----------------------|--|
| Macrozona | Macrozona norte | Sea 1 si pertenece a la macrozona norte, 0 en otro caso |
| | Macrozona centro | Sea 1 si pertenece a la macrozona centro, 0 en otro caso |
| | Macrozona centro sur | Sea 1 si pertenece a la macrozona centro-sur, 0 en otro caso |
| | Macrozona sur | Sea 1 si pertenece a la macrozona sur, 0 en otro caso |
| | Macrozona austral | Sea 1 si pertenece a la macrozona austral, 0 en otro caso |
| Jefe de hogar | Si es jefe de hogar | Sea 1 si es jefe de hogar, 0 en otro caso |
| | No es jefe de hogar | Sea 1 si no es jefe de hogar, 0 en otro caso |
| Área | Urbano | Sea 1 si se encuentra en un área urbana, 0 en otro caso |
| | Rural | Sea 1 si se encuentra en un área rural, 0 en otro caso |

Fuente: Elaboración Propia

Tabla 19: Variables de la investigación CASEN 2011

| Área | Variable | Observaciones |
|------------------------|----------------------------|---|
| Salario | Salario por hora | Salario bruto dividido horas trabajadas semanalmente por 4 |
| | Logaritmo salario | Se aplica logaritmo natural a salario por hora |
| Sexo | Mujer | Sea 1 si es mujer, 0 en otro caso. |
| | Hombre | Sea 1 si es hombre, 0 en otro caso. |
| Edad | Edad | Valor entre 15 a 120 años |
| | Edad2 | Valor al cuadrado de Edad |
| Educación | Educación media incompleta | Sea 1 si posee la educación media incompleta, 0 en otro caso |
| | Educación media completa | Sea 1 si solo posee la educación media completa, 0 en otro caso |
| | Título universitario o mas | Sea 1 si posee título técnico, universitario o más, 0 en otro caso |
| Estado civil | Casado | Sea 1 si se encuentra casado, 0 en otro caso |
| | Conviviente | Sea 1 si es conviviente, 0 en otro caso |
| | Soltero | Sea 1 si es soltero, anulante, separado, divorciado o viudo, 0 en otro caso |
| Tipo de contrato | Contrato indefinido | Sea 1 si posee contrato indefinido, 0 en otro caso |
| | Contrato a plazo fijo | Sea 1 si posee contrato a plazo fijo, 0 en otro caso |
| Tamaño de la empresa | Microempresa | Sea 1 si es una microempresa, 0 en otro caso |
| | Pequeña empresa | Sea 1 si una pequeña empresa, 0 en otro caso |
| | Mediana empresa | Sea 1 si es una mediana empresa, 0 en otro caso |
| | Gran empresa | Sea 1 si es una gran empresa, 0 en otro caso |
| Sector de la industria | Sector primario | Sea 1 si pertenece al sector primario, 0 en otro caso |
| | Extracción de recursos | Sea 1 si pertenece al sector de la extracción de recursos, 0 en otro caso |
| | Manufactura | Sea 1 si pertenece al sector de manufactura, 0 en otro caso |
| | Servicios básicos | Sea 1 si pertenece al sector de servicios básicos, 0 en otro caso |
| | Construcción | Sea 1 si pertenece al sector de la construcción, 0 en otro caso |
| | Comercio | Sea 1 si pertenece al sector de comercio, 0 en otro caso |
| | Transporte | Sea 1 si pertenece al sector de transporte, 0 en otro caso |
| | Hospitalidad | Sea 1 si pertenece al sector de hospitalidad, 0 en otro caso |
| | Inmobiliaria | Sea 1 si pertenece al sector de inmobiliario, 0 en otro caso |
| | Otros servicios | Sea 1 si pertenece a otros servicios, 0 en otro caso |
| Área de trabajo | Fuerzas Armadas | Sea 1 si pertenece a las fuerzas armadas, 0 en otro caso |
| | Ejecutivos | Sea 1 si es del área ejecutiva, 0 en otro caso |
| | Académicos | Sea 1 si es del área académica, 0 en otro caso |
| | Técnicos | Sea 1 si es del área técnica, 0 en otro caso |
| | Administrativos | Sea 1 si es del área administrativa, 0 en otro caso |
| | Vendedores | Sea 1 si es vendedor, 0 en otro caso |
| | Agricultores | Sea 1 si es del área de la agricultura, 0 en otro caso |
| | Artesanos | Sea 1 si es artesano, 0 en otro caso |
| | Operadores | Sea 1 si es operador, 0 en otro caso |
| | Ocupaciones elementales | Sea 1 si tiene una ocupación elemental, 0 en otro caso |
| Macrozona | Macrozona norte | Sea 1 si pertenece a la macrozona norte, 0 en otro caso |
| | Macrozona centro | Sea 1 si pertenece a la macrozona centro, 0 en otro caso |
| | Macrozona centro sur | Sea 1 si pertenece a la macrozona centro-sur, 0 en otro caso |
| | Macrozona sur | Sea 1 si pertenece a la macrozona sur, 0 en otro caso |
| | Macrozona austral | Sea 1 si pertenece a la macrozona austral, 0 en otro caso |

Fuente: Elaboración Propia

(Continuación) Tabla 19: Variables de la investigación CASEN 2011

| Área | Variable | Observaciones |
|---------------|---------------------|---|
| Jefe de hogar | Si es jefe de hogar | Sea 1 si es jefe de hogar, 0 en otro caso |
| | No es jefe de hogar | Sea 1 si no es jefe de hogar, 0 en otro caso |
| Área | Urbano | Sea 1 si se encuentra en un área urbano, 0 en otro caso |
| | Rural | Sea 1 si se encuentra en un área rural, 0 en otro caso |

Fuente: Elaboración Propia