

Documento técnico: análisis de brechas en ingresos entre hombres y mujeres en la población participante de los programas de la Fundación WWB Colombia.

En este documento se presentan los resultados de un ejercicio cuantitativo, usando una técnica de generación de submuestras denominada *Bootstrapping*, con el objetivo de corroborar la existencia de brechas por ingresos entre hombres y mujeres, en la población vinculada a la oferta de la Fundación WWB Colombia. Con esto, se busca corregir las dificultades que estos análisis han presentado anteriormente para la Fundación (principalmente por desbalances en la proporción de hombres y mujeres) y contribuir con evidencia a la toma de decisiones de la organización en relación con su misionalidad (el cierre de brechas de desigualdad para las mujeres).

Antecedentes y justificación

Teniendo como misión el cierre de brechas de desigualdad de género de las mujeres, la Fundación WWB Colombia, en sus acciones de analítica y generación de conocimiento, tiene como objetivo fundamental entender la formación y estructura de las brechas de género. En este sentido, se ha preocupado por estudiar y brindar evidencia de la existencia y comportamiento de las brechas que afectan diferencialmente a las mujeres y sus emprendimientos por necesidad atendidos con su oferta. En los informes de resultados de sus iniciativas, al igual que en documentos especiales y artículos de investigación, se ha identificado la existencia de una persistente brecha contra las mujeres en el desempeño financiero de sus negocios, según la cual un hombre vendería, en promedio, 2 pesos más por cada peso vendido por una mujer.

Sin embargo, múltiples dificultades limitan la demostración metódica y rigurosa de dichas desigualdades, siendo la principal la diferencia en la proporción de mujeres y hombres inscritos y certificados de los programas de la Fundación, que constituyen las bases de datos analizadas: de manera general, por cada 10 personas inscritas, 8 acostumbra ser mujeres y 2 hombres. Este desbalance puede dar lugar a resultados o interpretaciones incorrectas como se detalla a continuación:

1) Error en las pruebas de significancia estadística de la diferencia de medias

Estas pruebas se utilizan para comprobar si la diferencia de medias de dos grupos independientes son estadísticamente diferentes. Esta prueba supone normalidad y que las varianzas de ambas muestras son iguales, a pesar de que no cumplir el supuesto de normalidad no trae mayores repercusiones en los resultados, el supuesto de la varianza sí, ya que puede inducir a resultados erróneos. **El supuesto de varianzas iguales tiende a no cumplirse cuando las muestras son notablemente distintas (Welch, 1938) como en este caso donde el 85% de los datos corresponde a mujeres y el 15% a hombres .**

2) Fallas en las predicciones

Los modelos de clasificación de aprendizaje automático están diseñados para ser trabajados con bases de datos balanceadas, cuando hay mayor presencia de una clase sobre otra no se logran resultados óptimos, debido al alto porcentaje de muestras mal clasificadas en las clases menos representadas o minoritarias (Reyes, Nava, Lara & Alejo, 2018).

Como parte de los esfuerzos de la Fundación para generar evidencia que permita diseñar programas que incidan en dicho cierre de brechas, el presente documento propone un conjunto de pruebas estadísticas que comprueban la existencia de dichas desigualdades de género en la generación de ingresos de los negocios. Para esto se utilizó la base de datos del Plan Reactívale, la cual entre el 2020 y 2021, contó con más de 5.000 registros de personas emprendedoras por necesidad en más de 14 municipios entre el Valle del Cauca y el Cauca.

Para dar solución al problema de las proporciones tradicionalmente se aplican dos métodos: el primero, consiste en aumentar el número de registros de la clase que tiene menor presencia en la base de datos de forma aleatoria a partir de la muestra original hasta alcanzar un equilibrio; mientras que el segundo método, consiste en reducir el número de muestras de la clase con mayor presencia a partir de técnicas de submuestreo hasta lograr un balance con la clase menos representada. El análisis que aquí se propone se basa en la segunda metodología.

Metodología

Basado en la técnica de generación de submuestras denominada *Bootstrapping*,¹ se generaron 1,000 réplicas de muestras ($B= 1000$) sin reemplazo (todas diferentes) con el objetivo de obtener muchas submuestras seleccionadas de forma aleatoria de hombres y mujeres para diferentes tamaños de n (tamaño de la muestra).

¿Por qué 1000 muestras?

Dado que el objetivo planteado es realizar un contraste de hipótesis para determinar si el promedio de ingresos de hombres y mujeres es diferente, lo recomendado es estimar un número de réplicas superior a 500. Normalmente el valor de B se toma del orden de varias centenas o incluso millares. En los casos en los que el bootstrap se utiliza para estimar el sesgo o la varianza de un estimador, bastará tomar un número, B , de réplicas bootstrap del orden de $B = 100, 200, 500$. **Sin embargo, cuando se trata de utilizar el bootstrap para realizar contrastes de hipótesis o construir intervalos de confianza son necesarios valores mayores, del tipo $B = 500, 1000, 2000, 5000$ (Cao & Fernandez, 2021. Pág. 22).**

Pasos

- 1) Se calculan 1.000 muestras aleatorias de los ingresos de las mujeres y 1000 muestras aleatorias de los ingresos de los hombres, para tamaños de $n= 50, 100$ y 200 , esto con el objetivo de ver cómo el tamaño de la muestra afecta los resultados;
- 2) Se calcula el promedio de ingresos para cada una de las 1000 muestras de los hombres y las mujeres;
- 3) Se grafican los histogramas del promedio de ingresos de las 1000 muestras de los hombres y las mujeres en un mismo plano para observar la distribución del parámetro si hay diferencias;
- 4) Se calculan los intervalos de confianza para el promedio de ingresos de las 1000 muestras de hombres y mujeres;
- 5) Se realizan las pruebas de t (T-test) de diferencia de medias para cada par de muestras de ingresos de hombres y mujeres, identificando previamente si las varianzas de ambas muestras son iguales o no para posteriormente determinar si existen diferencias

¹ Es una técnica que consiste en generar múltiples submuestras de la población de estudio. Según los intereses del estudio, se pueden determinar diferentes réplicas (B).

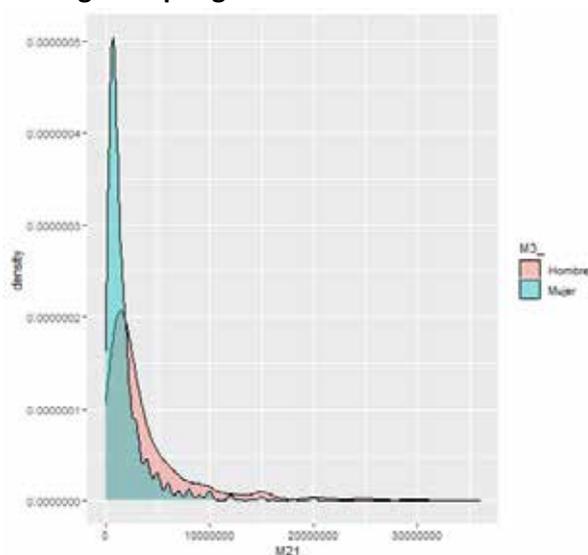
significativas y la posible existencia de brechas.

La potencia de este tipo de análisis consiste principalmente en poder comparar 1000 muestras diferentes pero del mismo tamaño para hombres y mujeres, siendo la primera ventaja que se elimina el problema de la proporción que afecta negativamente los resultados del T-test. Por otra parte, teniendo en cuenta que estas muestras son aleatorias, cualquier mujer y cualquier hombre en la base de datos tiene la misma probabilidad de ser seleccionado eliminando el problema de sesgo de selección.

Resultados

A continuación se presenta la distribución de ingresos por ventas en un mes normal por género.

Gráfico 1. Distribución de los ingresos por género



Fuente: Fundación WWB Colombia

Gráficamente se puede observar que los hombres se distribuyen a lo largo del eje X, lo que implica mayor nivel de ingresos y por ende negocios más grandes, mientras que las mujeres tienden a estar muy concentradas hacia la izquierda, lo que implica negocios más pequeños. Este resultado es consistente con otros hallazgos previos tal como se expuso en los antecedentes, al igual que con otros estudios nacionales y regionales (DANE, 2020; Observatorio de igualdad de género, 2016).

Análisis del resultado del remuestreo

Una vez calculadas las 1000 sub-muestras de los ingresos de los hombres y las mujeres, se graficaron los histogramas del promedio para cada tamaño de muestra como se puede ver a continuación:

Gráfico 2. Tamaño de muestra $n = 50$

Gráfico 3. Tamaño de muestra $n = 100$

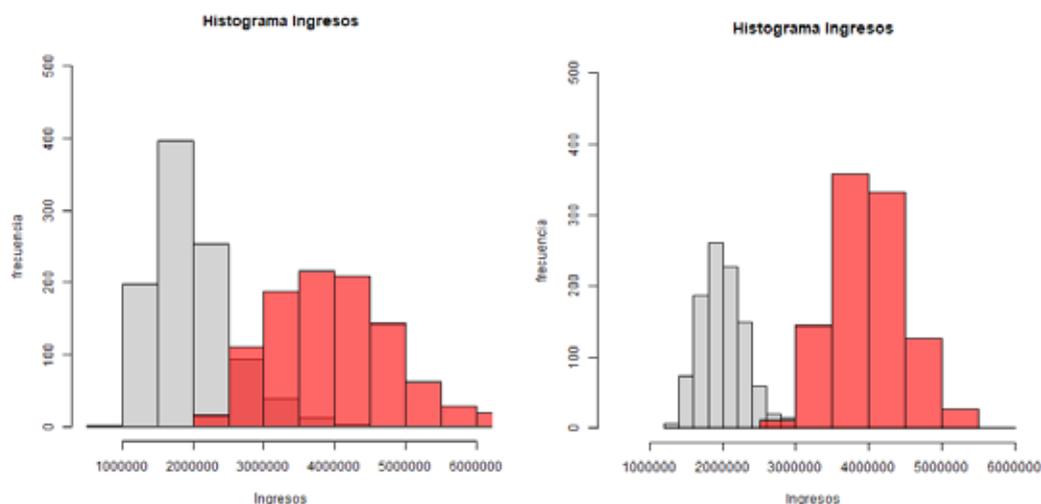
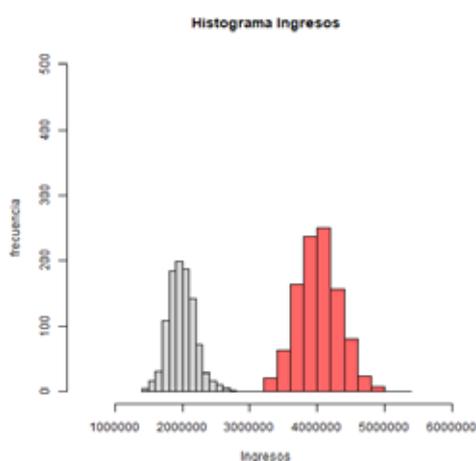


Gráfico 4. Tamaño de muestra n = 200



Fuente: Fundación WWB Colombia

Se puede observar que en los tres tamaños de muestra estimados hay diferencias notables en la distribución del promedio del ingreso de los hombres frente al de las mujeres, dando señales que el promedio que devengan ellas se ubica alrededor de los \$2.000.000 COP mientras que para ellos se ubica en alrededor de los \$4.000.000 COP, siendo un hallazgo consistente con lo encontrado en anteriores documentos elaborados por la Fundación, donde se había identificado que los hombres vendían 2 pesos por cada peso que vendía una mujer. Frente al tamaño de la muestra, se puede observar que tamaños más grandes conducen a resultados más contundentes.²

Intervalos de confianza para el promedio de ingresos de hombres y mujeres

Tabla 1. Intervalos de confianza para el promedio de ingresos de hombres y mujeres

² ¿Por qué tamaños más grandes de la muestra conducen a mejores resultados? De acuerdo con el teorema del límite central esto está relacionado con la forma de la distribución de la población principal, de modo que si se extrae la muestra de una distribución aproximadamente simétrica la aproximación normal puede ser buena, incluso para valores muy pequeños de n. Sin embargo, si la población está demasiado sesgada como en este caso donde la población tiende a concentrarse hacia el lado izquierdo de la curva, puede ser necesario una n muy grande.

Tamaño de muestra	Género	promedio	Intervalo de confianza 95%	Amplitud	Porcentaje de cobertura
n =30	Hombres	\$4.017.310	\$ 1.968.457- \$4.931.416	\$2.962.959	98%
	Mujeres	\$1.985.600	\$ 942.730- \$ 2.881.270	\$1.938.540	95%
n= 100	Hombres	\$3.968.865	\$3.027.629 - \$4.977.613	\$1.949.984	98%
	Mujeres	\$2.106.100	\$1.432.206 - \$2.527.052	\$1.094.846	95%
n= 200	Hombres	\$4.164.160	\$3.310.956 - \$4.723.985	\$1.413.029	98%
	Mujeres	\$2.072.472	\$1.580.066 - \$2.380.001	\$799.935	97%

Fuente: Fundación WWB Colombia

Los intervalos de confianza se calculan con dos objetivos: el primero para identificar a un nivel del 95% de confianza el límite inferior y superior sobre el cual se va a encontrar el parámetro de interés, en este caso el promedio de ingresos. Como se está analizando para las 1000 muestras, también es importante saber cuál es el porcentaje de cobertura, es decir cuántos de los 1000 intervalos de confianza calculados, contienen al promedio. En segundo lugar, el intervalo de confianza es una prueba estadística que permite afirmar si hay diferencias o no al comparar los promedios de ingresos de hombres y mujeres, de modo que si los intervalos se traslapan entre sí, es porque no hay diferencias, por el contrario, si estos no se solapan es porque sí hay diferencias. Finalmente, el intervalo más apropiado será aquel que contenga la mayor cantidad de veces al promedio y el que tenga menor amplitud.

Los resultados obtenidos demuestran que: 1) muestras más grandes conducen a mejores resultados para analizar si hay diferencias o no y esto queda claro en la medida que una muestra de n=50, tanto en el análisis de los histogramas como en los intervalos de confianza, dan como resultado que no hay diferencias en el promedio de ingresos, mientras que cuando n=100 o n=200, los intervalos no se solapan entre ellos, de lo que se puede concluir que efectivamente sí hay diferencias.

Contraste de hipótesis mediante test de diferencia de medias³

Para concluir si efectivamente hay una diferencia significativa en los promedios de ingresos se aplicó un test de diferencias de medias comparando cada par de muestras de hombres y mujeres (desde la número 1 hasta la número 1000). La prueba plantea como Hipótesis nula (H0) que la diferencia entre los promedios es igual a cero, contra la hipótesis alternativa (H1) que la diferencia de los promedios es diferente de cero. Se rechaza H0, si el p-valor obtenido en la prueba es inferior al valor crítico $\alpha = 0.05$, en caso de ser superior, entonces no se puede afirmar que haya diferencias significativas. A continuación se muestran los resultados obtenidos de forma gráfica:

³ Como paso previo a la aplicación del T-test se identificaron si las varianzas de las muestras de ingresos de los hombres y las mujeres eran iguales, obteniendo como resultado que no hay diferencias significativas entre ellas y por lo tanto se puede aplicar el T-test.

Resultados del test:

Gráfico 5. Tamaño de muestra n = 50

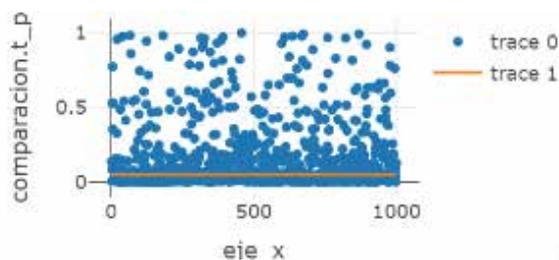


Gráfico 6. Tamaño de muestra n = 100

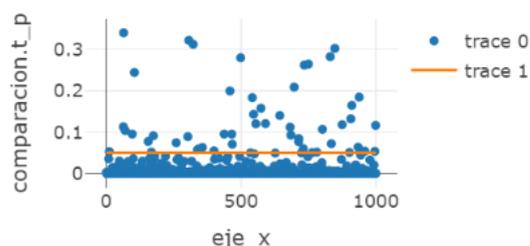
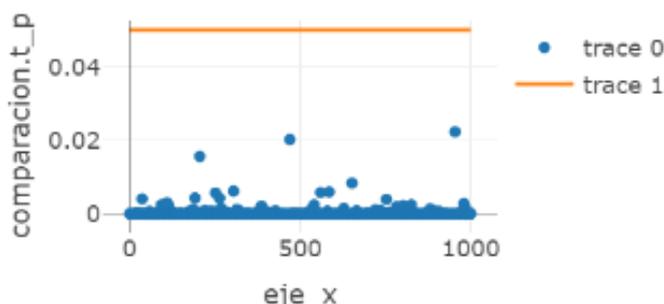


Gráfico 7. Tamaño de muestra n = 200



Fuente: Fundación WWB Colombia

Para analizar los gráficos se debe tener en cuenta que los puntos azules son los resultados de los p-valores obtenidos en cada uno de los test aplicados a cada par de muestras y la línea naranja es el valor crítico $\alpha = 0.05$. Así, todos los puntos azules por debajo de la línea naranja indican que se está rechazando la hipótesis nula que plantea que no hay diferencias significativas, mientras que aquellos que se encuentran por encima de la línea naranja no rechazan la hipótesis nula.

Los resultados son consistentes con los hallazgos anteriores, con una muestra más pequeña, en este caso cuando $n=50$, la prueba arroja que el 52% de las veces se rechaza H_0 , es decir que de cada par de muestras dentro de las 1000, 520 presentan diferencias significativas en los ingresos de los hombres en comparación con los de las mujeres. Sin embargo, cuando se incrementa el tamaño de la muestra a 100 o a 200, la prueba da resultados mucho más consistentes y contundentes. Cuando $n=100$ el 95% de las veces se rechaza H_0 , es decir que en 950 pares de muestras de 1000 se está afirmando que sí hay diferencias significativas en los promedios de ingresos que obtienen hombres frente a las mujeres. Finalmente cuando $n=200$, el porcentaje de veces que se rechaza H_0 es el 100%, es decir en todas las muestras comparadas se concluye que los ingresos de los hombres son superiores a los de las mujeres.

Análisis de brechas mediante técnicas de emparejamiento

Para profundizar los resultados mostrados anteriormente se realizó una técnica de emparejamiento conocida como Propensity Score Matching, la cual para este caso en concreto, consiste en formar parejas de hombres y mujeres con características observables similares entre sí a nivel sociodemográfico y de negocio.⁴ Este ejercicio se realizó con la información recolectada en las bases

⁴ El Propensity Score Matching (PSM), fue desarrollado por Rosenbaum y Rubin (1983) ante la limitación de que en la práctica existía la dificultad de contar con una prueba controlada aleatorizada (RCT) que permitiera realizar ejercicios de evaluación de impacto en programas sociales. Esta problemática llevó a estos investigadores a desarrollar una metodología cuasi-experimental que

de datos del Plan Reactívale entre los años 2020 y 2021, de las cuales se logró emparejar 199 pares de hombres y mujeres.

Tabla 2. Variables empleadas en el emparejamiento

Variable	Lista de valores
1. Grados de escolaridad	1. Primaria 2. Secundaria 3. Técnico 4. Superior
2. Sostenimiento del hogar	1. Yo 2. Alguien más y yo 3. Alguien más
3. Personas bajo cuidado permanente	1. Si 2. No
4. Número de personas en el hogar	Númerica
5. Número de personas personas que dependen económicamente	Númerica
6. Tenencia de hijos	1. Si 2. No
7. Antigüedad del negocio	1. Entre 6 meses y 1 año 2. Entre 1 y 2 años 3. Entre 2 y 3 años 4. Mayor a 3 años
8. Tenencia de registro mercantil	1. Si 2. No
9. Tenencia de domicilio	1. Si 2. No
10. Lugar donde funciona el negocio	1. En la vivienda 2. Fuera de la vivienda
11. Actividad-industrias manufactureras	1. Si 2. No
12. Actividad-comercio	1. Si 2. No
13. Actividad-actividades de servicios	1. Si 2. No
14. Actividad- alimentos y productos de la canasta familiar	1. Si 2. No
15. Territorio - especiales	1. Si 2. No

permitiera la construcción de un grupo de control artificial que abriera la posibilidad de evaluar los efectos asignados a un programa social.

Variable	Lista de valores
16.Territorio - desarrollo intermedio	1. Si 2. No

Fuente: Fundación WWB Colombia

Con las variables de la tabla 2 se construyeron los pares de hombres y mujeres que comparten características idénticas en cuanto a su perfil sociodemográfico como de los negocios que lideran. Los resultados obtenidos demuestran que en un mes normal, antes de la crisis generada por la pandemia del Covid-19, los hombres de esta muestra analizada generaban un promedio de ingresos de \$2.421.747 COP, mientras que las mujeres generaban un ingreso promedio de \$1.794.074 COP, observándose una brecha de ingresos de \$627.673 COP.

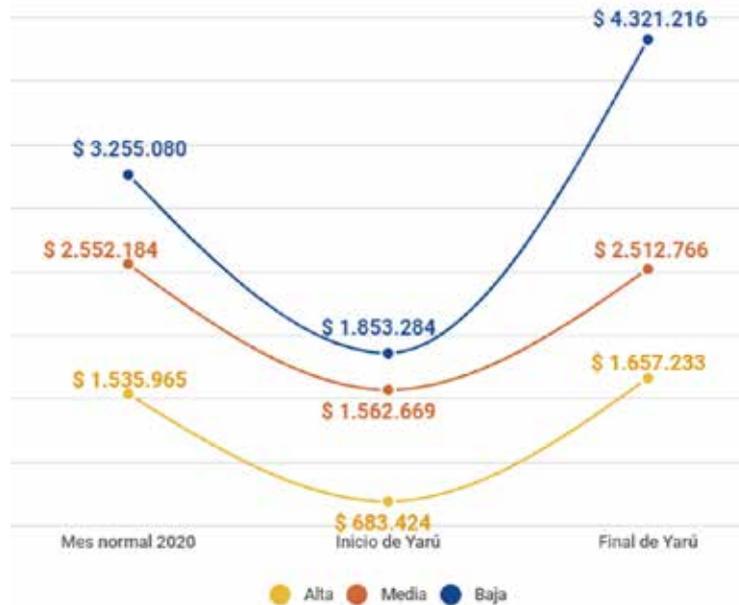
Entre las variables que resultaron significativas para explicar estas brechas se identificaron:

1. Sostenimiento del hogar
2. Personas bajo cuidado permanente
3. Número de personas en el hogar
4. Número de personas que conforman el hogar
5. Tenencia de registro Mercantil
6. Tenencia de domicilio
7. Pertenecer a actividades de industrias
8. Pertenecer a actividades de comercio
9. Pertenecer a actividades de servicios
10. Pertenecera actividades de alimentos

Estos resultados son consistentes con otros hallazgos obtenidos en diferentes análisis realizados por la Fundación, donde se construyó un índice de concentración de actividades de cuidado⁵ y se cruzó con los promedios de ventas de los negocios participantes del programa de capacitación Yarú en su versión semipresencial, en el año 2021. Como se muestra en el gráfico 8, en los tres cortes temporales, los negocios con baja concentración de actividades de cuidado (color azul) vendieron 2.1, 2.7 y 2.6 veces más que aquellos con alta concentración (color amarillo claro).

Gráfico 8. Variación en ingresos según concentración de actividades de cuidado.

⁵ Este índice se construyó a partir de cinco variables que indagan por la responsabilidad de la persona emprendedora en las siguientes actividades: preparar o servir alimentos; lavar, planchar o guardar ropa; cuidado de menores de 10 años; cuidado de personas adultas mayores y limpiar o realizar aseo. Como resultado, cada persona se clasificaba en una escala de bajo, medio y alto, según la cantidad de actividades que recayeran completamente bajo su responsabilidad.



Fuente: Fundación WWB Colombia

Conclusiones y recomendaciones

La metodología implementada mediante el remuestreo permitió identificar que la brecha de género en los ingresos no se debe a un problema de proporción, sino que efectivamente esta existe y se confirmó con las más de 1000 submuestras generadas a partir de la población original. Por su parte la metodología del emparejamiento permitió ver que incluso hombres con características sociodemográficas similares y con negocios idénticos a los de las mujeres venden 1.34 veces más que ellas.

Estos resultados son consistentes con otros estudios realizados en la región, donde se encontró que en Colombia “la brecha salarial general entre hombres y mujeres, según la media, es de 12,9% para el año 2019 (...). Esto indica que por cada 100 pesos que recibe un hombre por concepto de ingresos laborales totales, una mujer gana 87,1 pesos” (DANE, 2020, p. 11). Incluso, “las mujeres informales enfrentan brechas salariales más altas: ellas ganan 29% menos que los hombres en el ingreso laboral mensual y 2,9% menos en el ingreso por hora, mientras que para las formales estos valores son de 5,2% y -4,0% respectivamente” (DANE, 2020, p. 63). Para América Latina, el Observatorio de igualdad de género(2016) encontró que el salario de las mujeres equivale al 84% del salario de un hombre. Independientemente de los años de escolaridad, este comportamiento se mantiene para el subcontinente.

Entre las limitaciones de los métodos implementados, se debe tener en cuenta la dificultad que representa no poder incluir características no observables de los individuos como sus habilidades y competencias en el direccionamiento de sus negocios, así como más variables de contexto que pueden estar incidiendo en los resultados obtenidos. Ante esto, es necesario seguir abordando el problema desde otros puntos de vista, incluyendo metodologías cualitativas para ampliar y contrastar los resultados hasta el momento obtenidos.

A partir de estudios sobre desigualdades de género (DANE, 2020; Brunet y Santamaría, 2015) al igual que de otros ejercicios cualitativos y cuantitativos realizados por equipos de la Fundación, es posible plantear como hipótesis que las brechas de género en ingresos son consecuencia de los obstáculos presentes en la vida de las mujeres para acceder y desarrollar actividades productivas en igualdad de

condiciones que los hombres. Las principales limitaciones identificadas se corresponden con la división sexual del trabajo, que se traduce en sobrecargas de trabajo reproductivo y de cuidado para las mujeres.

Como resultado, ellas cuentan con menor tiempo para desarrollar trabajo productivo, ubicándose en sectores informales y de menor productividad que les permita conciliar dicha tensión entre producir y cuidar. Sin embargo, es necesario avanzar en estudios de mayor profundidad y rigor que permitan comprobar dichas hipótesis en los emprendimientos por necesidad. Para esto, es recomendable avanzar en metodologías mixtas que combinen pruebas estadísticas como las aquí realizadas, con abordajes cualitativos (entrevistas y grupos focales) que reconstruyan las trayectorias y funcionamientos de dichos negocios, problematizando sus dimensiones familiares, de género y otros factores no considerados.

Referencias

Departamento Administrativo Nacional de Estadística-DANE. (2020). *Brecha salarial de género en Colombia*. Colombia: DANE. Disponible en: <https://www.dane.gov.co/files/investigaciones/notas-estadisticas/nov-2020-brecha-salarial-de-genero-colombia.pdf> (Visitado el 05/08/2022).

B. L. Welch. (1938). The Significance of the Difference Between Two Means when the Population Variances are Unequal. *Biometrika*, Vol. 29, No. 3/4 (Feb., 1938), pp. 350-362.

Cao. R & Fernandez. R. (2021). Técnicas de remuestreo. Disponible en https://rubenfcasal.github.io/book_remuestreo/book_remuestreo.pdf

Observatorio de igualdad de género de América Latina y el Caribe. (2016). *Persiste la brecha salarial entre hombres y mujeres*. Nota de igualdad N° 18, 8 de marzo del 2016. Disponible en: https://www.cepal.org/sites/default/files/news/files/nota_18_brechas_salarios.pdf (Visitado el 05/08/2022).

H. Cruz-Reyes , A. Reyes-Nava , E. Rendón-Lara , R. Alejo. (2018). Estudio del desbalance de clases en bases de datos de microarrays de expresión genética mediante técnicas de Deep Learning. *Research in Computing Science* 147(5).

Rosenbaum, Paul & Rubin, Donald.(1983). *The central role of the propensity score in observational studies for causal effects*. *Biometrika*, 70,1, pp. 41-55.