Metodologías y Técnicas de la

Ciencia de Datos para el análisis de la brecha salarial de género.

Roberto León Leyva, Tecsup

Brecha salarial de género en el Estado peruano: un análisis desde la ciencia de datos

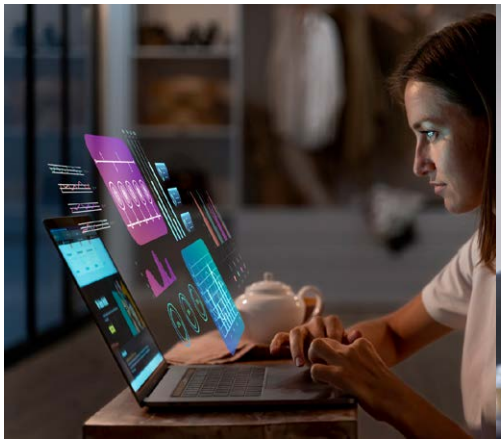
Gender Pay Gap in the Peruvian Government: an approach from Data Science

RESUMEN

El estudio tiene por objetivo analizar la existencia de la brecha salarial de género en el Estado peruano e identificar los perfiles de tales brechas a nivel de regiones con base a la información de datos abiertos proporcionada por la Autoridad Nacional del Servicio Civil (Servir), aplicando metodologías y técnicas de la ciencia de datos. La investigación fue de tipo aplicada cuantitativa, de nivel explicativo y de diseño no experimental longitudinal (2017-2021). La metodología de trabajo utilizada fue CRISP-DM (Cross Industry Standard). Durante el análisis exploratorio de datos, se encontró que a nivel nacional la participación de las mujeres está cerca de la paridad con un promedio del 47 %, en tanto que la brecha salarial de género en el Estado tenía una tendencia a la baja del 13 % (2017) a 11 % (2021) indicador mejor al promedio general de Latinoamérica que era de 14 % (2019 OIT), adicionalmente, cuando se revisan a nivel nacional pareciera que a mayor participación corresponde menor brecha, sin embargo, cuando se desagrega la información por regiones se evidencia heterogeneidad en participación y en la brecha salarial (rango intercuartílico para Apurímac de 25 % y Piura de 2 %) y de su evolución; en muchas regiones, la pandemia genera que se revierta o ralentice las mejoras en la brecha. Para determinar qué regiones tuvieron avances se utilizó un modelo de regresión lineal simple con base en el año, la pendiente negativa indicaría avances y la pendiente positiva retrocesos en el periodo bajo estudio. Las regiones que presentaron mejores avances fueron Moquegua, Huánuco y Áncash, en tanto que las que presentaron deterioro fueron Huancavelica, La Libertad y San Martín. Para determinar los perfiles de brecha, se utilizó el periodo 2021 y se añadieron variables obtenidas del INEI como el PBI per cápita y la cobertura por cada servidor público, el primer factor asociado al progreso de la región y el segundo asociado a los servicios del Estado, se determinaron dos clústeres con niveles medianos de brecha alto(12,7 %) y bajo (8,9 %). La prueba de hipótesis no paramétrica de Brown-Mood permite concluir que existe la brecha salarial de género en el Estado peruano a pesar de contar con un marco legal que prohíbe la discriminación salarial.

ABSTRACT

The objective of this study is to analyze the existence of the gender wage gap in the Peruvian government and identify the profiles of these gaps at the regional level, based on open data provided by the National Authority of the Civil Service (Servir), applying data science methodologies and techniques. The research was applied and quantitative in nature, explanatory in level, and used a non-experimental longitudinal design (2017-2021). The methodology used was CRISP-DM (Cross Industry Standard). During the Exploratory Data Analysis it was found that at the national level the female participation is closed to parity getting an average of 47 % on the other hand, the gender pay gap in the Government had a downward trend from 13 % (2017) to 11 % (2021), an indicator better than the general average for Latin America which was 14 % (2019 ILO), additionally, the correlation showed that to more participation correspond less pay gap, however, when the information is disaggregated by regions, heterogeneity of the gaps is evident (interquartile range for Apurimac of 25 % and Piura of 2 %) and its evolution; in many regions the pandemic generates that the improvements in the gap are reversed or slowed down. To determine which regions made progress, a simple linear regression model was used based on the year; the negative slope would indicate progress, and the positive slope would indicate reversals in the period under study; the regions that made the best progress were Moquegua, Huánuco and Ancash, while those that showed deterioration were Huancavelica, La Libertad and San Martin. To identify the gap profiles, the year 2021 was used and



Palabras Claves

Brecha salarial de género, participación femenina, Ley de No discriminación por Género, valoración de puestos, limpieza e imputación de datos, regresión lineal, clustering, perfiles de brecha salarial.

Key words

Gender Pay Gap, female participation, Law of non-gender discrimination, Valuation of jobs, Cleaning and imputation of data, linear regression, clustering, profiles of pay gap.

variables obtained from the INEI were added, such as GDP per capita and coverage per public servant, the first factor associated with the progress of the region and the second associated with government services, therefore two clusters were determined with median values of pay gap high (12,7 %) and low (8,9 %). The Brown-Mood nonparametric hypothesis test implies that the pay gap exists regardless of the labor laws that banning the discrimination of rewarding in the labor market.

INTRODUCCIÓN

La brecha salarial de género es un problema económico, social y laboral que se manifiesta en la diferencia de remuneraciones entre mujeres y hombres por trabajos similares. Para su cálculo, es necesario agregar valores, homogeneizar las bases de ingresos brutos (fijos, variables, gratificaciones, etc.) y ajustar las jornadas laborales (contratos a tiempo completo en comparación con los parciales). Las Naciones Unidas, a través de ONU Mujeres, aborda estos aspectos [14].

La importancia de abordar este problema abarca varias dimensiones. ONU Mujeres identifica los siguientes motivos: respeto a los derechos humanos (el principal), productividad (un salario justo contribuye a la motivación), lucha contra la pobreza y la desigualdad (la pobreza femenina prolonga el ciclo de la pobreza), y armonía social (la igualdad de oportunidades favorece el desarrollo pleno en los ámbitos personal y profesional) [14].

Según ONU Mujeres, las causas de la brecha salarial incluyen el empleo a tiempo parcial, trabajos peor remunerados, ocupaciones socialmente menos valoradas, la autoinfravaloración y la falta de conocimiento por parte de los empleados [14].

Claudia Goldin, premio Nobel de Economía 2023, revisó 200 años de historia económica de Estados Unidos y encontró hallazgos que solo pueden observarse en series de tiempo muy prolongadas. Por ejemplo, a comienzos del siglo xx, con los datos disponibles, se concluyó que una mayor participación femenina impulsa el crecimiento, lo que a su vez fomenta una mayor participación y reduce la brecha salarial. La perspectiva innovadora de Goldin permitió identificar variables que determinan la evolución de la participación y la brecha salarial, basándose en grandes cambios estructurales, como las revoluciones industriales y los cambios sociales. Esta revisión cuestiona el supuesto de la relación directa entre participación, crecimiento y brecha salarial, y se denomina la «Curva U», presentada en la figura 1. Goldin identificó que, en el inicio de la revolución industrial, las exigencias del trabajo alejaron a las mujeres casadas con hijos, quienes encontraron muy difícil equilibrar las tareas del hogar con las del trabajo en las fábricas [16].

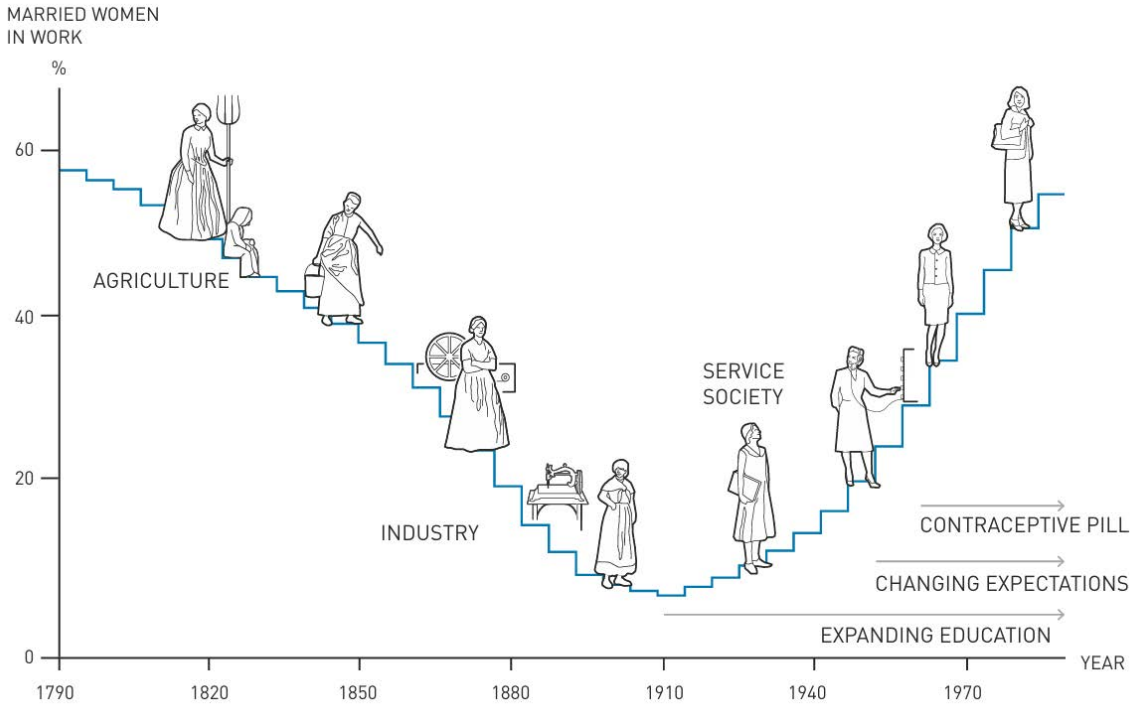


Figura 1. Curva U de la participación de la mujer en el mercado laboral americano

Fuente: [17].

Otro hallazgo de Goldin es que, a partir de la década de 1970, los cambios en las expectativas, el uso de métodos anticonceptivos y una mayor inversión en educación tuvieron un impacto significativo. Al analizar a hombres y mujeres con un MBA, Goldin pudo identificar que, en los primeros años tras la graduación, la brecha salarial era pequeña. Sin embargo, la diferencia comenzó a surgir a partir de la maternidad, cuando las mujeres asumen la mayor parte de la responsabilidad en la crianza de los hijos, lo que les resta tiempo y dificulta desempeñarse en trabajos competitivos de tiempo completo, que a menudo requieren tiempo e inversiones adicionales. Esta brecha es la que no logra cerrarse con el paso de los años, como se muestra en la figura 2.

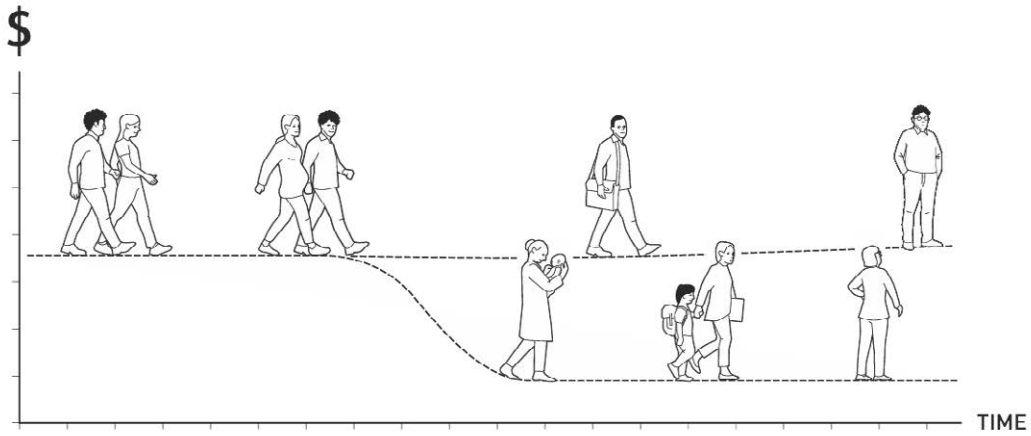


Figura 2. Efecto de la maternidad en la brecha salarial

Fuente: [17].

Trabajos posteriores basados en el estudio de Goldin, aplicados a otros países de altos ingresos con datos históricos disponibles, mostraron resultados similares, incluida la aparición de la típica curva en U de participación femenina, que relaciona el PIB per cápita (log) como variable independiente con la participación femenina como variable dependiente. La figura 3 muestra este gráfico característico en forma de U. El trabajo de Goldin proporciona un esquema que abarca la transformación estructural (industrialización, expansión de los empleos de oficina), los cambios tecnológicos que afectan la naturaleza del trabajo, las normas y expectativas sobre la responsabilidad del cuidado de los hijos, las oportunidades educativas y los cambios institucionales que eliminan barreras para el acceso y desarrollo de las mujeres en el mercado laboral. La Academia de Ciencias de Suecia menciona acertadamente que la brecha salarial no es solo un problema de equidad normativa, sino que también afecta la asignación eficiente de recursos al no aprovechar el talento de manera óptima, lo cual promueve el desarrollo de las mujeres y genera un impacto positivo en el PIB de los países [16]. El trabajo de Goldin también plantea retos para los responsables de políticas públicas, quienes deben fomentar la flexibilidad necesaria para reducir la brecha generada por la maternidad. En países en vías de desarrollo, como Perú, el desafío es aún mayor: se deben identificar variables de largo plazo en tres dimensiones —estructuras sociales, económicas y derechos— que en los países desarrollados tomaron décadas, pero que los países en desarrollo deben abordar de forma acelerada, equilibrando expectativas y realidades en estas dimensiones.

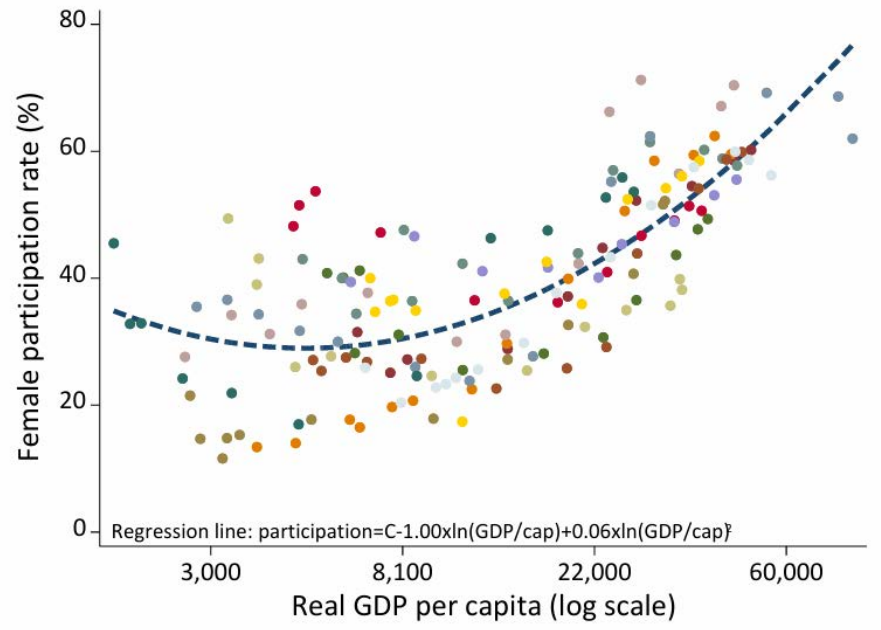


Figura 3. Participación femenina vs. PBI per cápita real (log) para 14 países industrializados en el periodo 1890-2005

Fuente: [16].

La OIT publica anualmente su Panorama Laboral. En la edición de 2020, se revisa la brecha salarial de género en los países de América Latina y el Caribe. Con base en sus datos, se ha elaborado la tabla 1, en la que se observa un avance significativo desde 2013 (18,3 %) hasta 2019 (14 %) [12].

Tabla 1
Brecha salarial en América Latina y el Caribe elaborado con base al informe OIT 2020

Año	Rem. mujer / rem. hombre	Brecha salarial
2013	84,5 %	18,3 %
2014	83,7 %	19,5 %
2015	86,2 %	16,0 %
2016	85,0 %	17,6 %
2017	86,0 %	16,3 %
2018	87,8 %	13,9 %
2019	87,7 %	14,0 %

Fuente: Elaboración propia.

En la edición 2024, la OIT [13] analiza el trabajo no remunerado e indica que, en América Latina, el número de horas semanales dedicadas a este tipo de trabajo oscila entre 22 y 43 para las mujeres, mientras que para los hombres está entre 10 y 20 horas. Este trabajo ocurre, principalmente, en los hogares. Un aspecto interesante del informe es que destaca el papel de las organizaciones privadas en la reducción de la brecha salarial, e indica acciones que deben incorporarse en la gestión de personas y en el cambio cultural, como se muestra en la figura 4.

Figura 4. Políticas empresariales para la reducción de la brecha salarial de género

Fuente: [13].

A nivel latinoamericano, el estudio de Urquidi y Chalup (2023) concluye que, según las encuestas armonizadas de hogares del Banco Interamericano de Desarrollo (BID), las mujeres deberían tener un ingreso por hora superior al de los hombres, considerando su nivel educativo, los sectores de la economía en los que trabajan y las ocupaciones. Las diferencias se atribuyen a sesgos discriminatorios y a la falta de normativa laboral. El mismo estudio calcula que, para Perú en 2019, que la participación femenina agregada fue del 43 %. En cuanto a la brecha salarial general en Perú, aplicando la metodología de Blinder-Oaxaca, se estima en -30 %, y el análisis de descomposición muestra que las variables explican casi un tercio de la brecha (-7 % explicado frente a -23 % no explicado). Al aplicar la metodología de Ñopo, la brecha se reduce a -19 %, y los factores explican casi una quinta parte (-4 % explicado frente a -15 % no explicado) [18].

En el Perú, existe un marco normativo que prohíbe la discriminación salarial entre hombres y mujeres, basado en la Ley 30709 y su reglamento. Como señala Blume, esto se traduce en obligaciones concretas para los empleadores, quienes deben implementar políticas y procedimientos, elaborar un cuadro de categorías y funciones, y comunicarlo efectivamente a los trabajadores. Esto implica que las empresas deben desarrollar proyectos para superar las barreras estructurales que han afectado a las mujeres y a algunos grupos o colectivos [4].

El Ministerio de Trabajo y Promoción del Empleo ha elaborado guías para facilitar la comprensión del tema. Por ejemplo, identifica factores objetivos (educación, experiencia laboral, horas dedicadas, sector de la economía, ocupación o puesto) y subjetivos (patrones estructurales económicos, sociales y otros mecanismos sutiles de discriminación). En cuanto a las acciones recomendadas para las organizaciones, sugieren un proceso con las siguientes fases: (1) identificar puestos de trabajo, (2) determinar el género asociado a los puestos, (3) valorizar los puestos de trabajo, (4) comparar y calcular las brechas, y (5) implementar medidas para eliminar la brecha salarial [11].

En el sector público, Servir es la entidad responsable del sistema administrativo de recursos humanos y se encarga de analizar,

proponer y controlar la correcta ejecución de los lineamientos para la gestión de personas en los organismos públicos. Anualmente, presenta un informe sobre la situación de la mujer en el servicio civil, en el que se revisan diversos aspectos, tales como (1) características (participación, sector, brecha salarial), (2) Hostigamiento sexual, (3) cargos directivos, y (4) conclusiones y referencias. En cuanto a las variables de interés para este estudio, el informe de 2024, con datos de 2022, indica que el sector público general representa el 16,7 % de la PEA asalariada, con una participación femenina del 47,3 %. Además, la brecha salarial general en el sector público mejoró sustancialmente, pasando del 18,3 % en 2017 al 2,0 % en 2022. La tabla 2 muestra las brechas por los distintos niveles del Estado peruano, elaborada con base en los datos del informe de 2024. Se observa que, para 2022, el problema parece haberse superado, persiste únicamente en el nivel de gobierno local, mientras que, con respecto al gobierno nacional, las mujeres incluso ganan más que sus pares hombres. Un dato adicional del informe revela que el nivel de calificación educativa de las mujeres es mayor que el de los hombres en todos los niveles del Estado; en promedio, un 55 % de las mujeres tiene educación universitaria o avanzada concluida, frente al 39 % de los hombres, como se muestra en la tabla 3 [3].

Tabla 2
Brecha salarial del Estado peruano por niveles 2022 vs. 2017

Segmento	2017			2022		
	Mujer	Hombre	Brecha	Mujer	Hombre	Brecha
General sector público	3,018,00	3,571,80	18,3 %	3,542,00	3,612,00	2,0 %
Gobierno local	2,273,20	2,444,20	7,5 %	2,111,00	2,345,00	11,1 %
Gobierno regional	2,204,50	2,352,70	6,7 %	2,967,00	2,984,00	0,6 %
Gobierno nacional	4,308,00	4,691,20	8,9 %	4,540,00	4,482,00	-1,3 %

Fuente: [3].

Tabla 3
Estudios universitarios y universitario avanzado promedio para el periodo 2017-2022

Segmento	Mujer	Hombre	Mujer-hombre
General sector público	55 %	39 %	16 %
Gobierno local	21 %	16 %	5 %
Gobierno regional	59 %	55 %	4 %
Gobierno nacional	56 %	35 %	21 %

Fuente: [3].

El presente trabajo investiga la brecha salarial de género en el Estado peruano. A partir de las cifras macro presentadas previamente, el problema podría parecer inexistente; sin embargo, al analizar por regiones y observar la evolución durante el periodo estudiado, se encontró evidencia de que la brecha persiste. Esto es aún más notable si consideramos que el nivel de instrucción académica de las mujeres es superior al de los hombres, lo que sugiere la presencia del fenómeno conocido como «techo de cristal». Abordar este problema no solo implica el cumplimiento y respeto de un derecho, sino que también tiene un impacto económico. En el sector público, su tratamiento está vinculado a la generación de valor público para los ciudadanos, además de convertir al Estado en un referente para la sociedad peruana.

FUNDAMENTOS

Se presentarán los conceptos y las técnicas utilizados en el desarrollo de este trabajo: definiciones sobre la brecha salarial, la discriminación, las normas laborales, y una referencia a las técnicas de ciencia de datos empleadas en el análisis.

A. Brecha salarial de género

La brecha salarial de género mide cuanto más ganan los hombres en comparación con las mujeres por un trabajo equivalente en calidad de dependiente. La fórmula propuesta por la ONU para calcular este indicador es la siguiente:

$$brecha\ salarial\ de\ género = \frac{remuneración_{hombre} - remuneración_{mujer}}{remuneración_{hombre}}$$

Los valores positivos indican una retribución a favor de los hombres, mientras que los valores negativos reflejan que las mujeres tienen una mayor retribución que los hombres. Para determinar las remuneraciones, se consideran los valores brutos, tanto fijos como variables, así como la cuantificación de los pagos en especie. Además, se debe ajustar a jornadas completas o calcular un ratio por hora trabajada para homologar las diferencias de jornada. Dado que se agrupan datos, se utiliza generalmente la media o la mediana como estimador del grupo [14].

B. Participación femenina

Es el porcentaje de mujeres que conforman la cohorte o segmento del estudio salarial. Esta agrupación puede basarse en el puesto de trabajo, nivel jerárquico (auxiliares, analistas, supervisores/jefes, gerentes, directores), área dentro de una organización, organizaciones dentro de conglomerados (como el Estado), o en criterios de geografía política (distrito, provincia, región, países), entre otros.

$$participación\ femenina = \frac{número\ de\ mujeres}{número\ de\ mujeres + número\ de\ hombres}$$

C. Principio de igual remuneración por trabajo de igual valor

Este principio amplía el criterio de igual paga por igual trabajo, establecido en la normativa peruana a través de la Ley n.º 30709, Ley que prohíbe la discriminación remunerativa entre varones y mujeres (Ley de Igualdad Salarial). Esto significa que no solo se aplica a trabajos iguales o similares, sino también a puestos distintos con un valor comparable. Es decir, se refiere a puestos con diferentes contenidos, responsabilidades, calificaciones y condiciones, pero que generan el mismo o similar valor para la organización [7] [11].

D. Política salarial o remunerativa

Es el conjunto de criterios y directrices que las organizaciones establecen para gestionar las remuneraciones (fijación o reajuste de los diferentes esquemas salariales). En el contexto de la Ley de Igualdad Salarial, estos criterios deben ser justos y evitar la discriminación salarial [7].

E. Valoración de puestos

Es un proceso sistemático que utiliza información interna y externa para determinar el valor relativo que un puesto aporta a la organización, asegurando que la compensación (remuneración) esté alineada con los perfiles del puesto y los criterios generales de evaluación. En el caso peruano, la Ley n.º 30709 y su reglamento recomiendan el uso de un modelo de puntos por factor, ya que descompone los puestos en factores y subfactores para su posterior evaluación. Los factores sugeridos por la Guía de la Ley son los siguientes: (1) calificaciones o competencias, (2) esfuerzos, (3) responsabilidades y (4) condiciones en que se realiza el trabajo. Los subfactores son los elementos que capturan los detalles de los puestos, permitiendo la asignación de puntos para su valoración. Un requisito clave es evitar sesgos, por lo que se recomienda que este proceso lo lleve a cabo un comité, y que la variable género no se utilice para asignar puntos (discriminación). Esto permitirá, a su vez, la creación de la política salarial, la estructura y las bandas salariales. La ley denomina a este proceso Cuadro de categorías y funciones [7] [11].

F. Discriminación remunerativa por género

La discriminación salarial ocurre cuando se establecen diferencias remunerativas basadas en el sexo de la persona sin criterios objetivos. La discriminación puede ser (1) directa, cuando los procedimientos o prácticas excluyen o favorecen a una persona o grupo de personas por motivos prohibidos, como el género; (2) e indirecta, cuando se implementan medidas aparentemente neutrales, pero cuya aplicación afecta de manera desproporcionada a los miembros de un grupo o colectivo protegido, generando un impacto adverso en los trabajadores [7].

G. Plataforma Nacional de Datos Abiertos

Es una plataforma tecnológica donde el Estado peruano, a través de los organismos públicos y los diferentes niveles del gobierno, pone a disposición de la ciudadanía diversas herramientas (*datasets*, páginas, documentos, *dashboards*, etc.) en el marco de las políticas de gobernanza de datos, garantizando la transparencia en la función pública. Las normas del Estado peruano establecen que las entidades públicas deben implementar, conforme a su contexto legal, tecnológico y estratégico, las herramientas necesarias para la recopilación, procesamiento, publicación, almacenamiento y apertura de los datos que administran [2].

H. Ciencia de datos

Es una práctica interdisciplinaria que integra métodos de ingeniería, estadística, minería de datos, aprendizaje automático y analítica predictiva. Similar a la investigación de operaciones, la ciencia de datos se centra en tomar decisiones basadas en datos y gestionar su implementación, con un enfoque eminentemente práctico [19].

I. Limpieza e imputación de datos

Estas técnicas conforman el análisis previo de datos y tienen como objetivo revisar las características de las variables e identificar patrones de datos perdidos para proponer un esquema de tratamiento. Para los datos faltantes, generalmente se utilizan dos técnicas, que pueden combinarse: la eliminación de los registros incompletos y la imputación de datos, aplicando métodos como la propagación hacia delante y atrás de la última observación, regresiones, entre otros [1].

J. Regresión lineal simple

Es una técnica de aprendizaje supervisado que asume que la variable independiente X influye en la variable dependiente Y, es decir, $X \rightarrow Y$. Esta relación se expresa mediante una línea recta con la siguiente fórmula:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

En la expresión anterior, β_0 es el coeficiente que representa el parámetro de intercepto poblacional, β_1 es el coeficiente que representa el parámetro de regresión poblacional asociado a la variable predictora X_i, y ε_i es la variable aleatoria no observable, conocida como término aleatorio, que contiene los efectos de otras variables predictoras no incluidas explícitamente en el modelo.

Los supuestos del modelo son (1) la variable Y_i tiene una media y varianza constantes (homocedasticidad), (2) La variable X_i es fija, es decir, no hay error en su medición, y (3) la variable $\varepsilon_i \sim N(0, \sigma^2)$, donde el valor constante de σ^2 significa que los valores de Y_i son independientes entre sí y no están correlacionados [15].

K. Estudentización de variables

Es una técnica de preprocesamiento de datos que normaliza las escalas de las variables para evitar sesgos derivados de sus medidas. Se usa con frecuencia antes de aplicar técnicas de *clustering*. Para cada dimensión de un conjunto de datos, se resta la media y se divide el resultado entre la desviación estándar muestral, logrando que la variable estandarizada tenga una media de 0 y una desviación estándar de 1. La fórmula de estandarización de la variable X es la siguiente [1].

$$x_{estudentizada} = \frac{x_i - \underline{x}}{\delta_x}$$

L. Clustering

Es una técnica de *Machine Learning* que pertenece al aprendizaje no supervisado y agrupa observaciones de manera que cada grupo sea homogéneo, es decir, que sus elementos sean similares, mientras que los grupos entre sí sean lo más distintos posible. A diferencia de las técnicas de aprendizaje supervisado, en esta técnica no se conoce de antemano el número de grupos o clústeres, lo que debe ser afinado por el analista. Un aspecto clave es el tipo de distancia a emplear y la necesidad de homogeneizar las variables. Una técnica común para este preprocesamiento es la estandarización de las variables. Entre las técnicas más habituales de *clustering* están los algoritmos aglomerativos (jerárquico) y particionados (k-means, PAM) [1] [10].

M. Medidas de validación de *clustering*

Dada la gran cantidad de algoritmos de *clustering* disponibles, es necesario contar con criterios que permitan determinar el número adecuado de clústeres. Brock propone tres esquemas: (1) interno, que utiliza la información intrínseca de los datos para evaluar la calidad del *clustering*; entre estos están la conectividad, el ancho de la silueta y el índice de Dunn. (2) Estabilidad, una variante de las medidas internas, que compara los clústeres obtenidos tras remover una columna; entre estos se encuentran la proporción de no superposición (APN), la distancia promedio (AD), la distancia promedio entre medias (ADM) y la cifra de mérito (FOM). (3) Biológico, aplicable a datos biológicos para analizar si los clústeres tienen significado. Estos tres esquemas son aplicables a varios de los algoritmos aglomerativos y particionados, y su implementación se encuentra en el paquete clValid para el R [5].

N. Perfiles de brecha salarial

Consiste en aplicar medidas de estadística descriptiva a las variables de cada conglomerado identificado mediante técnicas de *clustering*. Se utilizó la mediana debido a la asimetría de las variables, ya que es un estadístico más robusto que la media.

METODOLOGÍA

Se utilizó la metodología *Cross Industry Standard Process for Data Mining* (CRISP-DM) como marco de trabajo para el uso de herramientas de ciencia de datos en el abordaje de la brecha salarial de género en el Estado peruano [6]. Para el desarrollo de sus fases, se emplearon los paquetes indicados en la tabla 4. El estudio incluyó la ejecución de varios ciclos de la metodología CRISP-DM, aplicándola a modelos de aprendizaje supervisado mediante regresión lineal para estimar la tendencia de mejoras en las brechas regionales. Posteriormente, se añadieron variables relacionadas con el desempeño de las regiones y los servicios del Estado peruano, con el fin de aplicar técnicas de aprendizaje no supervisado e identificar perfiles de la brecha salarial en las regiones.

Tabla 4
Paquetes de RStudio empleados

Paquete	Utilización
tidyverse	Manipulación de *dataframes (datasets)*, tablas resúmenes y gráficos
readxl	Lectura y manejo de los archivos en formato Excel
janitor	Adecuación de nombres de variables
skimr	Estadísticas generales del dataset, detección de *missing values* y distribución de variables
clValid	Métricas de comparación para modelos de *clustering*
dendextend	Gráficos del *clustering* jerárquico
purr	Programación funcional, se utilizó para simplificar las regresiones lineales múltiples para los países
coin	Prueba no paramétrica de Mood para las medianas de los clústeres
broom	Obtención y manipulación de los resultados de las regresiones para cada región

Fuente: Elaboración propia.

Fase I, entendiendo el negocio

En este trabajo, se plantean dos objetivos de investigación: (1) identificar la existencia de la brecha salarial de género en el Estado peruano a nivel regional y (2) caracterizar los perfiles de brecha salarial de género presentes a nivel regional. Para el primer objetivo, se aplicarán técnicas de limpieza, imputación de datos y cálculo de las variables necesarias para el análisis (brecha y participación). Para el segundo, se añadirán variables asociadas al desarrollo regional y a los servicios del Estado, como el PBI per cápita de la región y la cobertura de personas por cada servidor público. Con los clústeres identificados, se procederá a caracterizarlos para presentar el perfil de las regiones con la información de 2021 y corroborar la existencia de la brecha salarial por medio de la prueba de hipótesis no paramétrica de diferencia de medianas de Brown-Mood.

Fase II, entendiendo datos

Los datos utilizados para el presente análisis se obtuvieron de las siguientes entidades públicas:

- Plataforma de datos abiertos y del *dataset* publicado por Servir, de aquí se obtuvieron los datos agregados por región, género, personas e ingresos promedios para los años 2017-2021, representa la principal fuente de información [2].
- Información del INEI para obtener población y PBI por regiones para 2021 [8] [9].

Durante la fase de carga y limpieza de datos, se detectó que el conjunto de datos proveniente de Servir omitió información sobre el número de servidores públicos para 2018. Se realizó la imputación utilizando la semisuma de los servidores de 2017 y 2019. Se optó por este método simple para evitar el uso de datos de 2020 o 2021, ya que estos reflejarían el impacto de la reducción de puestos de trabajo debido a la pandemia, como se muestra en la figura 5.

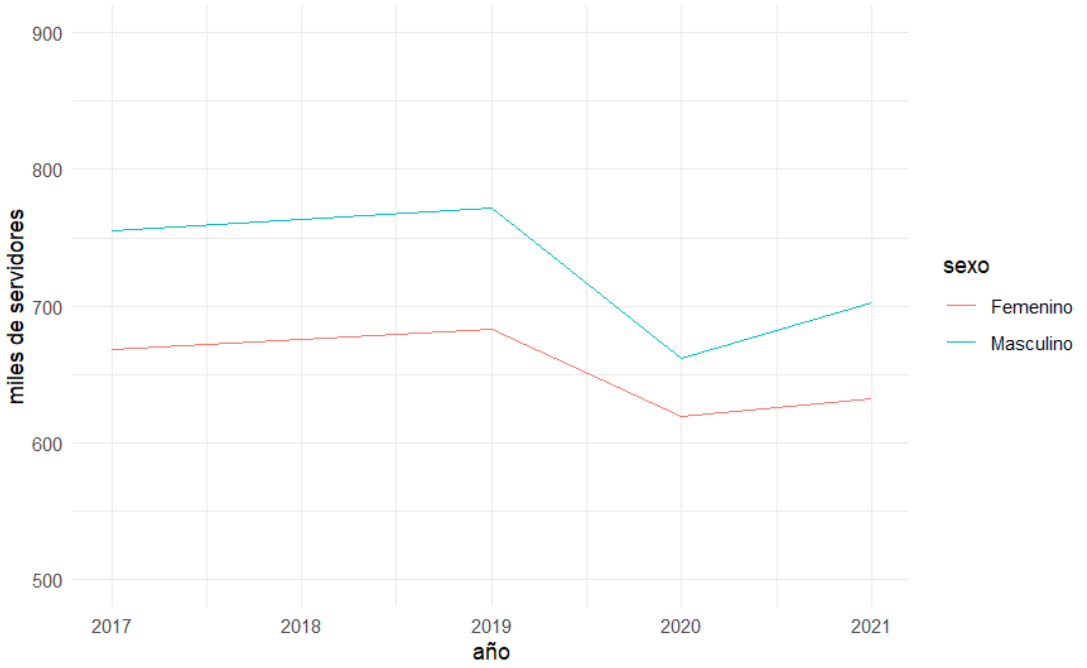


Figura 5. Evolución de los servidores públicos por sexo

Fuente: Elaboración propia.

Concluida la limpieza, la imputación, el reacomodo y la generación de las nuevas características (brecha, participación, nivel de brecha y número de servidores) se tiene disponible el *dataset* de trabajo que permitirá el análisis de los objetivos planteados, las variables se presentan en la tabla 5. Las variables disponibles para el análisis y modelado, corresponden a agregaciones por año y región.

Tabla 5
Variables disponibles

Variable	Valores únicos	Descripción
region	25	Nombre de las regiones
ubigeo	25	Código de ubigeo de las regiones, se utilizó para unir con data del INEI (PBI, población)
year	5	Año del periodo bajo estudio 2017-2021
Masculino	125	Remuneraciones promedio de hombres consolidado por regiones
Femenino	125	Remuneraciones promedio de mujeres consolidado por regiones
nro_Masculino	125	Número de hombres por región y año que laboran en el Estado peruano
nro_Femenino	125	Número de mujeres por región y año que laboran en el Estado Peruano
brecha_salarial	125	Brecha salarial de mujeres por región y año
servidores	125	Número total de servidores, es la suma de nro_Masculino y nro_Femenino
participacion	125	Participación de las mujeres por región y año
level_brecha	3	Variable ordinal (bajo, medio, alto) que agrupa los valores con base a la muestra de las brechas de regiones y años por terciles

Fuente: Elaboración propia.

a. Fase II, Entendiendo los datos por medio del análisis exploratorio de datos

Los resultados del análisis exploratorio de datos son los siguientes:

- **Nivel agregado de Estado peruano**

Revisamos la evolución de los dos indicadores clave: brecha salarial y participación. Como se aprecia en la figura 6, la participación ha alcanzado niveles cercanos a la paridad, y la brecha salarial ha ido disminuyendo de forma lenta, pero progresiva hasta antes de la pandemia. Posteriormente, ambos indicadores se deterioraron. La brecha salarial se redujo del 13,0 % en 2017 al 10,10 % en 2020, para luego aumentar al 11,16 % en 2021. De manera similar, la participación mejoró del 46,9 % en 2017 al 48,3 % en 2020, retrocediendo al 47,3 % en 2021.

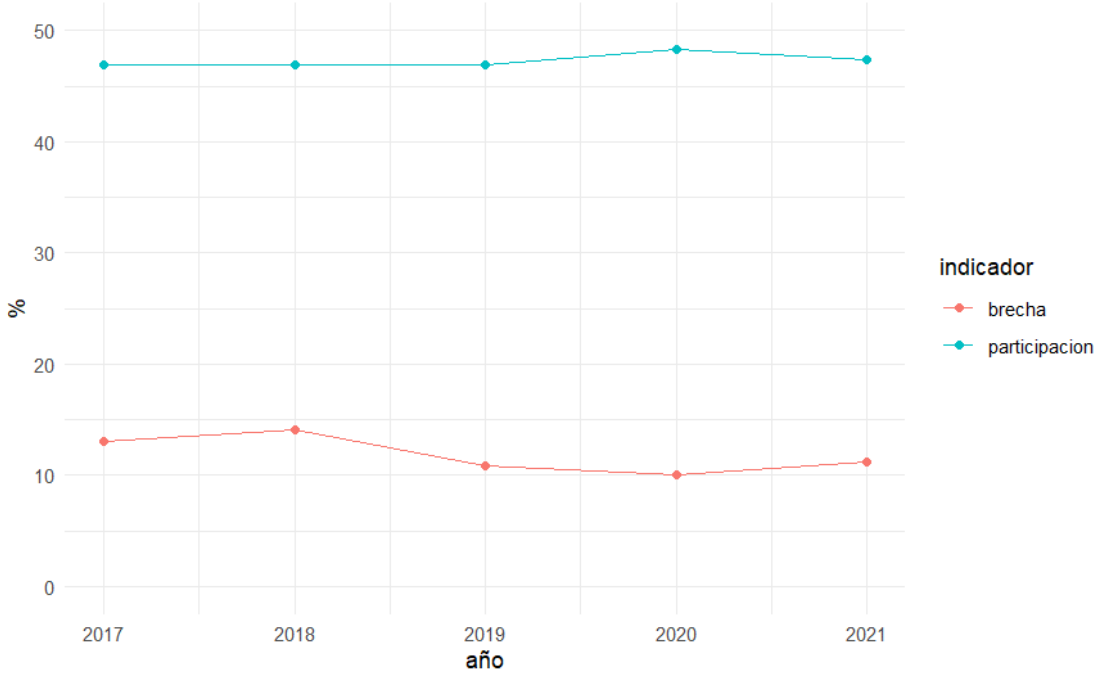


Figura 6. Evolución de brecha salarial y participación a nivel Estado

Fuente: Elaboración propia.

Analizamos la relación entre la brecha salarial y la participación a nivel agregado en el Estado peruano, corroborando la evidencia mencionada en la introducción: a mayor participación, menor es la brecha salarial, como se observa en la figura 7. La revisión a nivel agregado del Estado peruano muestra buenos indicadores, lo que podría sugerir que el problema es casi inexistente.

- **Nivel agregado de región y año**

Al revisar las variables por año y región, y utilizar diagramas de cajas para agrupar los resultados anuales, se observa una gran variabilidad, lo que propone la posible existencia de una brecha oculta al analizar los datos a nivel estatal. La figura 8 muestra la evolución de la variabilidad en la brecha salarial, mientras que la figura 9 presenta la participación por año. Los puntos corresponden a las distintas regiones.

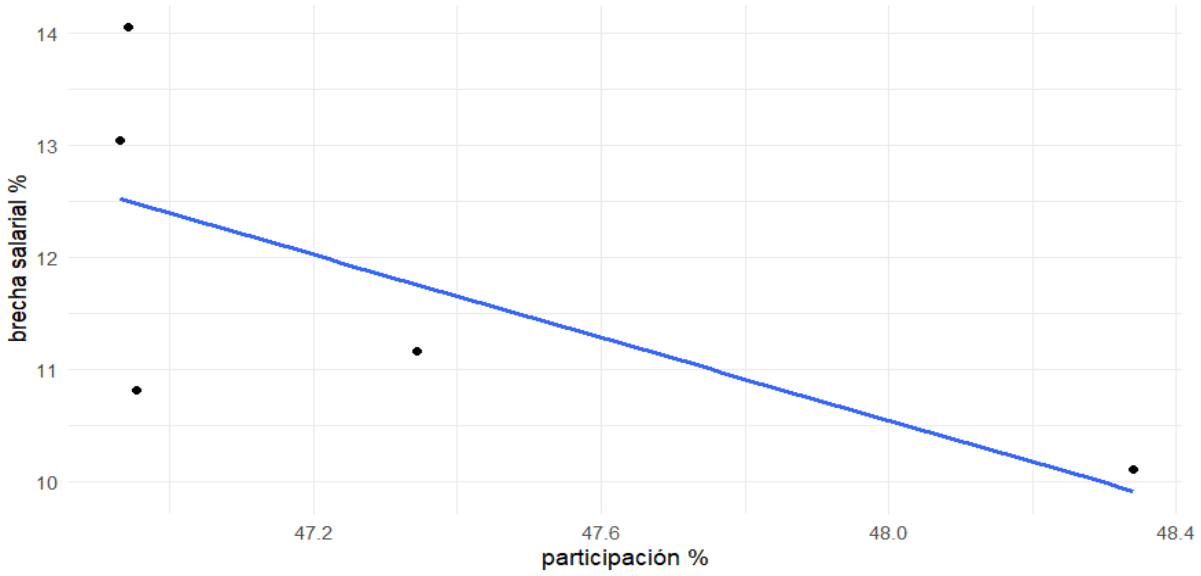


Figura 7. Relación entre brecha salarial y participación a nivel Estado

Fuente: Elaboración propia.

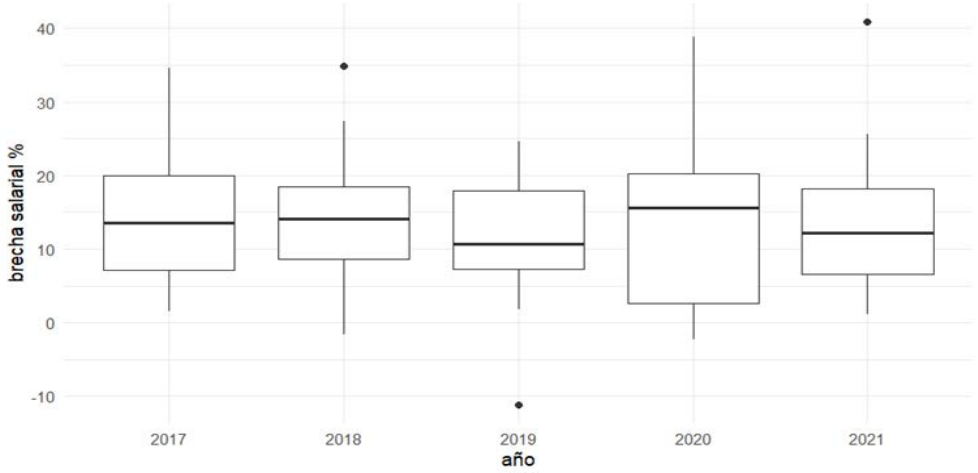


Figura 8. Brecha Salarial por año, la dispersión corresponde a las regiones

Fuente: Elaboración propia.

Al revisar el diagrama de dispersión de la brecha salarial versus la participación para todas las regiones y años, se observa una escasa relación positiva, con una correlación de 0,149, como se muestra en la figura 10. Al segmentar la brecha salarial por terciles usando la variable categórica *level* brecha (bajo, medio, alto), no se aprecia una relación significativa en los niveles bajo y medio, con correlaciones muy bajas de -0,089 y -0,015, respectivamente. Sin embargo, en el nivel alto se observa una ligera relación que contradice la teoría: a mayor participación, mayor es la brecha, con una correlación de 0,319. Esto se aprecia en la figura 11.

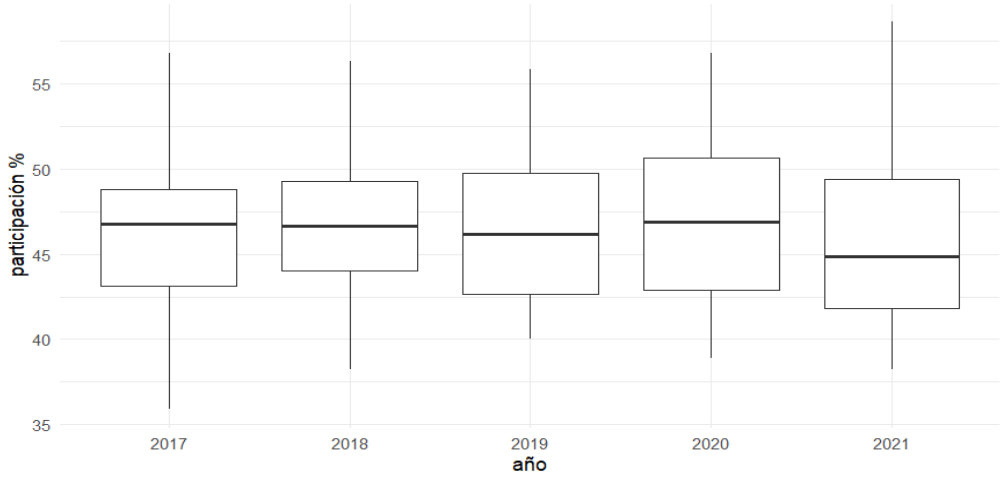


Figura 9. Participación femenina por año, la dispersión corresponde a las regiones

Fuente: Elaboración propia.

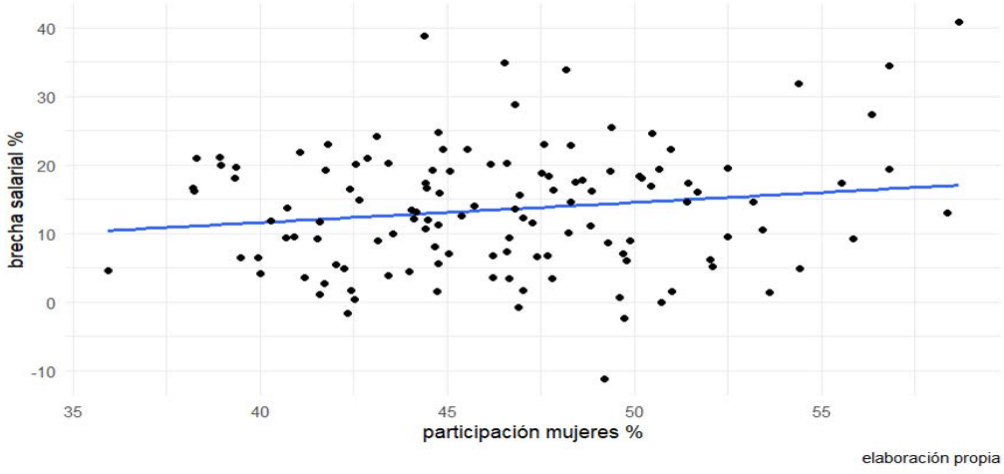


Figura 10. Brecha salarial vs. participación, cada punto representa una región por año.

Fuente: Elaboración propia.

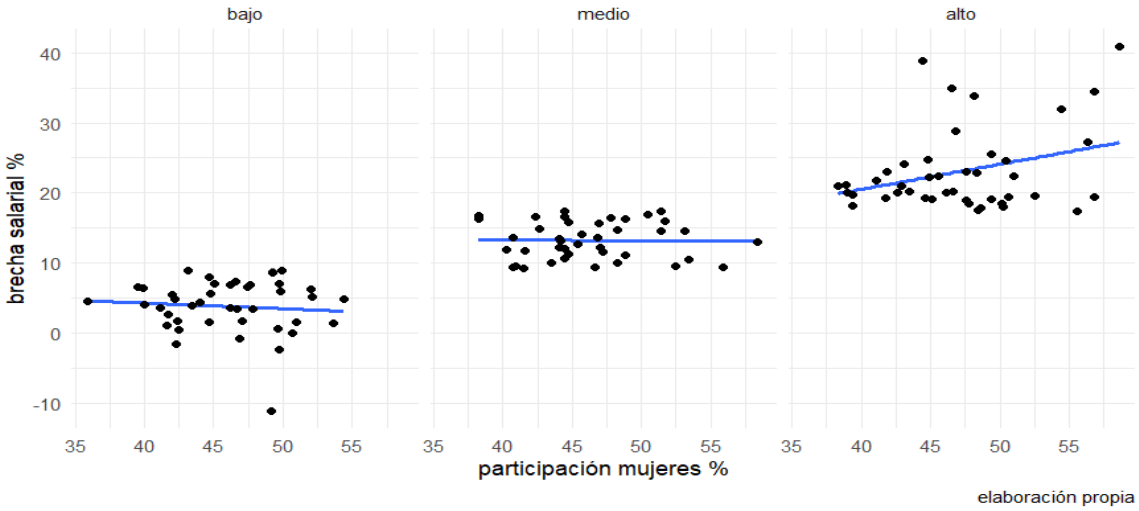


Figura 11. Brecha salarial vs. participación, con base a niveles de brecha por terciles: bajo, medio y alto.

Fuente: Elaboración propia.

- **Revisión de brecha salarial y participación por regiones**

La revisión de la brecha salarial por región muestra una gran dispersión, lo que confirma la existencia del problema al desagregar los datos a nivel estatal. La figura 12 presenta la mayor brecha salarial mediana en Pasco, con un 26 %, y la menor en Ica, con un 6,3 %. En cuanto a la dispersión, medida por el rango intercuartílico, Apurímac presenta la mayor dispersión, con un 25 %, mientras que Piura tiene la menor, con un 2 %.

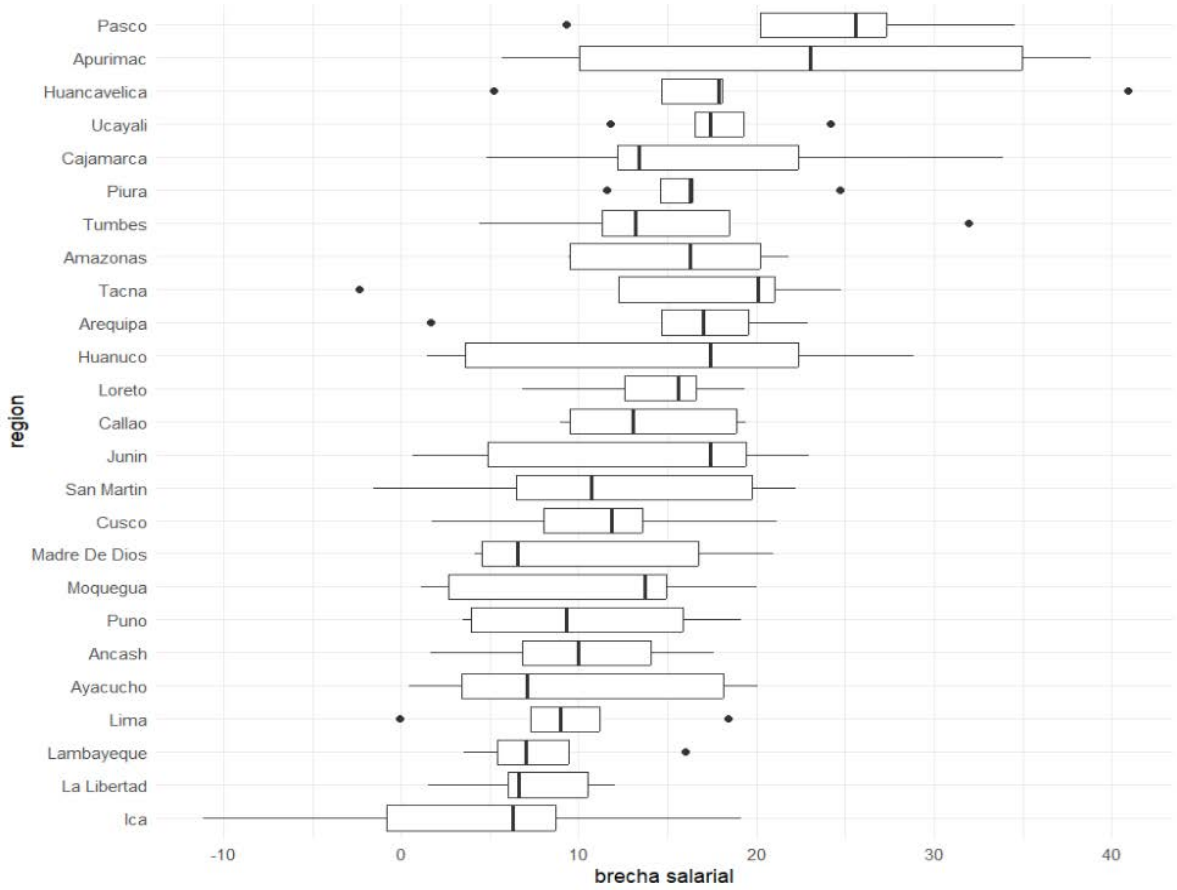


Figura 12. Diagrama de cajas por regiones para la brecha salarial

Fuente: Elaboración propia.

Al revisar la participación por regiones, también se observa una gran variabilidad, como se aprecia en la figura 13. La mayor participación mediana corresponde a Pasco, con un 56 %, mientras que la menor se registra en Madre de Dios, con un 39 %. En cuanto a la variabilidad, medida por el rango intercuartílico, Pasco presenta la mayor, con un 7 %, y la menor corresponde a Ica, con prácticamente un 0 %.

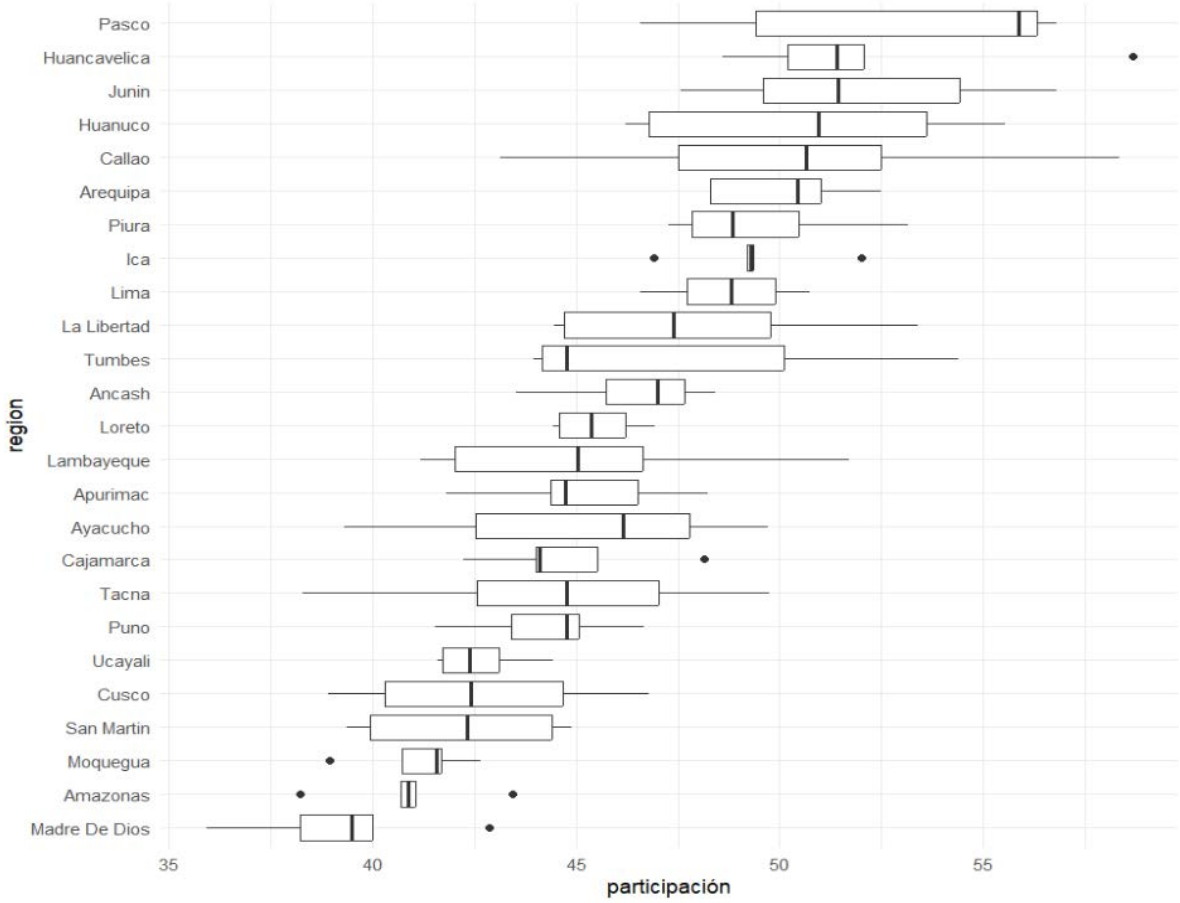


Figura 13. Diagrama de cajas por regiones para la participación femenina

Fuente: Elaboración propia.

C. Fase III, preparación de datos

La metodología CRISP-DM se ha aplicado de manera iterativa para cada uno de los objetivos planteados en este estudio y sus correspondientes modelos. Así tenemos lo siguiente:

- Para el análisis exploratorio de datos, se reorganizó la estructura original en formato tidy, es decir, con observaciones en filas y variables en columnas. Se imputaron los datos faltantes.
- Se definieron las variables necesarias para el análisis: brecha, participación y total de servidores.
- Para determinar el avance o retroceso de las regiones en el problema de la brecha salarial de género, se prepararon los datos para aplicar el modelo de regresión lineal simple. Los coeficientes de las variables y su nivel de significancia respaldaron estos hallazgos.
- Para identificar los perfiles de las regiones, se integraron los datos con la información del INEI y se definieron nuevas variables significativas para el análisis. Luego, se preprocesaron con estandarización para evitar problemas de magnitud en la aplicación de las técnicas de *clustering*.
- La caracterización de los perfiles requirió transformar los datos para aplicar estadísticas descriptivas sobre las variables originales, vinculadas con la identificación de los clústeres a los que pertenecen las regiones.

D. Fase IV, modelado y Fase V, evaluación

Se emplea el modelo de regresión lineal simple para identificar qué regiones han avanzado o retrocedido en relación con el problema de la brecha salarial de género. Se utilizó un nivel de significancia de 0,10, debido a que en los problemas sociales complejos siempre hay variables que el modelo no captura, dadas las limitaciones del muestreo, los instrumentos disponibles y los modelos teóricos.

Para la determinación de los perfiles, se emplean técnicas de *clustering* aglomerativo y particionado. La evaluación del número de clústeres se basa en la técnica del codo y se contrasta con análisis más robustos, como los criterios internos y de estabilización.

La caracterización de los perfiles se realiza mediante estadística descriptiva y una revisión gráfica entre ellos, utilizando diagramas

de violín para corroborar la existencia del problema de la brecha salarial a nivel regional.

- **Modelos para identificar los avances y retrocesos de la brecha salarial de género por región**

Se utilizó un modelo de regresión lineal donde la variable dependiente era la brecha salarial y la independiente, el año. Los resultados se presentan en la tabla 6. Esta tabla muestra modelos significativos para seis regiones, en las que se encontraron coeficientes $\hat{\beta}$ significativos. Tres regiones lograron avances en la reducción de la brecha: Huánuco, Moquegua y Áncash, mientras que tres retrocedieron: La Libertad, San Martín y Huancavelica. Para las otras 19 regiones, no se encontraron evidencias estadísticas de mejora o deterioro.

Tabla 6
Resultado de las regresiones por región, se ordena por modelo significativo y ascendentemente por $\hat{\beta}$

Modelo significativo	β significativo	Región	R^2	pvalue modelo	$\hat{\beta}$	$\hat{\beta}$ std.error	$\hat{\beta}$ statistic	$\hat{\beta}$ pvalue
sí	sí	Huánuco	0,902	0,013	-7,142	1,355	-5,270	0,013
sí	sí	Moquegua	0,884	0,017	-4,874	1,019	-4,784	0,017
sí	sí	Áncash	0,748	0,058	-3,383	1,133	-2,986	0,058
sí	sí	La Libertad	0,909	0,012	2,483	0,454	5,470	0,012
sí	sí	San Martín	0,735	0.063	5.284	1,830	2,887	0,063
sí	sí	Huancavelica	0,669	0,091	6,793	2,760	2,461	0,091
no	no	Tumbes	0,627	0,110	-5,154	2,294	-2,247	0,110
no	no	Ica	0,245	0,396	-3,530	3,575	-0,987	0,396
no	no	Pasco	0,179	0,478	-2,511	3,106	-0,808	0,478
no	no	Lima	0,295	0,344	-2,295	2,047	-1,121	0,344
no	no	Lambayeque	0,520	0,169	-2,199	1,219	-1,804	0,169
no	no	Arequipa	0,154	0,514	-2,021	2,737	-0,738	0,514
no	no	Junín	0,079	0,646	-1,733	3,406	-0,509	0,646
no	no	Ucayali	0,144	0,529	-1,077	1,517	-0,710	0,529
no	no	Tacna	0,020	0,823	-0,956	3,906	-0,245	0,823
no	no	Amazonas	0,000	0,985	-0,044	2,127	-0,021	0,985
no	no	Puno	0,000	0,991	0,033	2,563	0,013	0,991
no	no	Piura	0,068	0,671	0,805	1,717	0,469	0,671
no	no	Madre de Dios	0,028	0,788	0,819	2,787	0,294	0,788
no	no	Loreto	0.078	0,648	0,844	1,671	0,505	0,648
no	no	Callao	0.077	0,652	0,875	1,753	0,499	0,652
no	no	Cusco	0,046	0,730	0,969	2,555	0,379	0,730
no	no	Ayacucho	0,118	0,572	1,913	3,027	0,632	0,572
no	no	Cajamarca	0,248	0,393	3,517	3,534	0,995	0,393
no	no	Apurímac	0,174	0,485	3,868	4,866	0,795	0,485

Fuente: Elaboración propia.

La figura 14 muestra la evolución de las tres regiones con mejores avances, en tanto que la figura 15 muestra las regiones con mayores retrocesos.

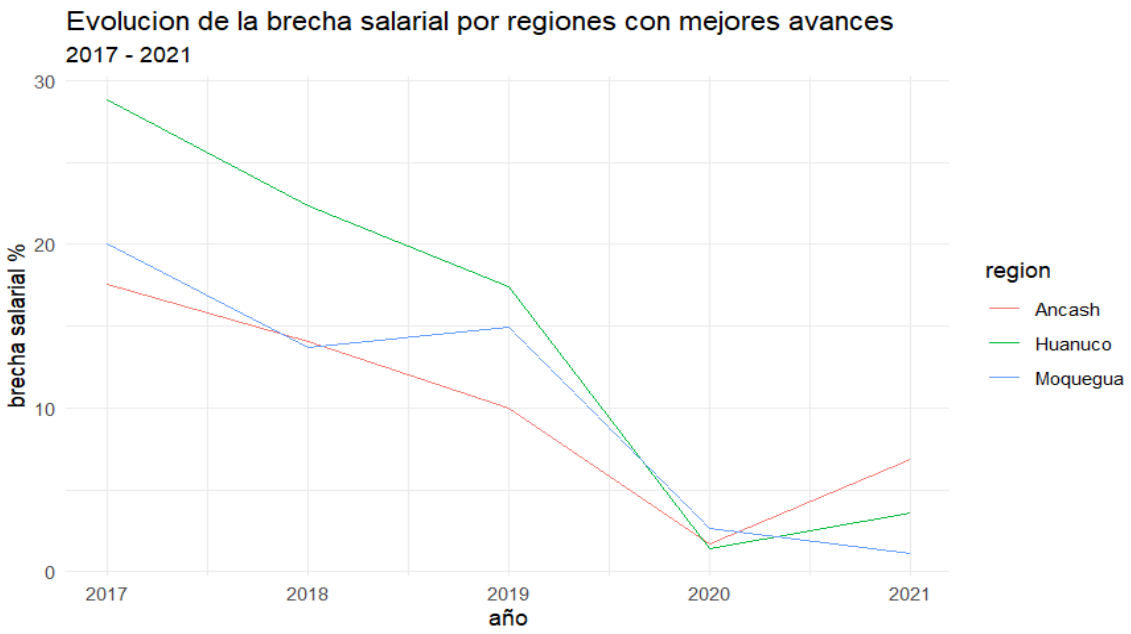


Figura 14. Regiones con los mejores avances en la reducción de la brecha salarial de género

Fuente: Elaboración propia.

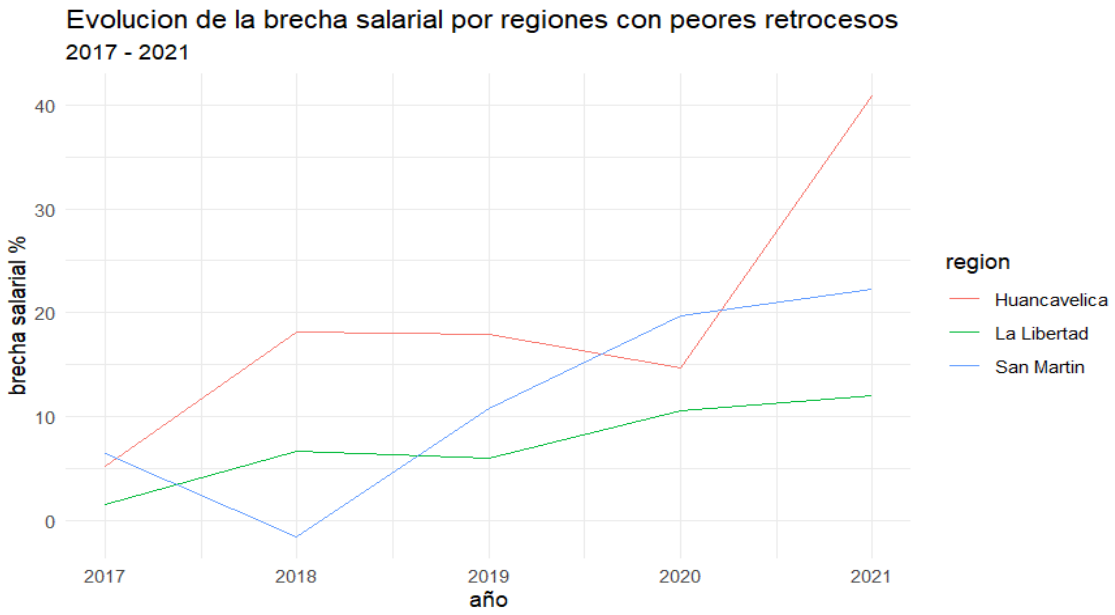


Figura 15. Regiones con los retrocesos más significativos de brecha salarial de género

Fuente: Elaboración propia.

- **Modelos de *clustering* para la determinación de los perfiles de región por brecha salarial**

Para este análisis, se consideró el último año del conjunto de datos proporcionado por Servir (2021), al cual se añadieron indicadores del INEI, como población y PBI de la región. A partir de estos, se generaron variables como el PBI per cápita (para el modelo se utilizó el logaritmo de este indicador) y la cobertura de cada servidor público.

Se realizaron varias iteraciones para determinar el número adecuado de clústeres. De acuerdo con el método del codo, se recomendarían 4 clústeres, como se muestra en la figura 16. Adicionalmente, se probó con 2, 3 y 4 clústeres utilizando los algoritmos *k-means* y jerárquico. Debido a que el número de regiones es reducido, los valores más balanceados por número de regiones se obtenían con 2 y 4 clústeres. La evaluación mediante criterios internos y de estabilidad recomendó mayoritariamente dos clústeres, como se observa en la tabla 7 y ese fue el número de clústeres que se utilizó para el agrupamiento de las observaciones (regiones) por el método jerárquico. La agrupación de las regiones se presenta en la figura 17.

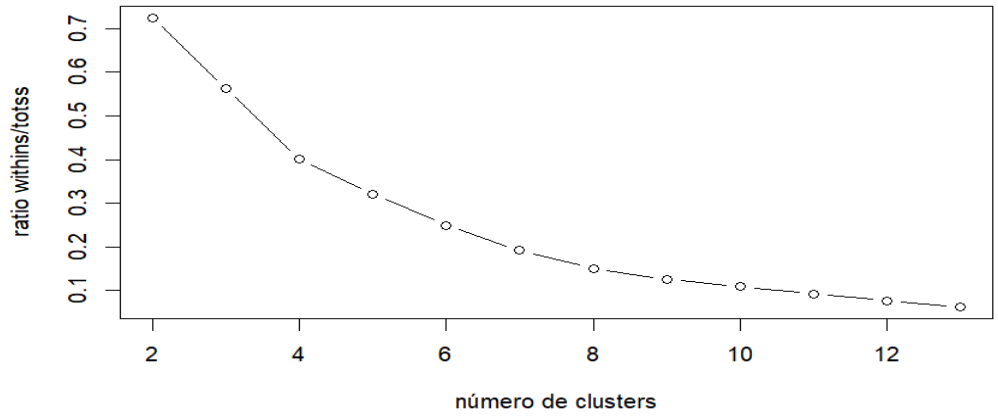


Figura 16. Método del codo para la determinación del número de clústeres

Fuente: Elaboración propia.

Tabla 7
Número de clústeres sugeridos por la validación interna y de estabilidad

Validación	Indicador	Score	Método	Número de clústeres
Interna	Connectivity	2,929	hierarchical	2
	Dunn	0,6774	hierarchical	12
	Silhouette	0,4192	kmeans	2
Estabilidad	APN	0,0383	kmeans	2
	AD	0,7795	pam	13
	ADM	0,2693	kmeans	2
	FOM	0,794	pam	11

Fuente: Elaboración propia.

En relación con la revisión del problema de la brecha salarial en los dos clústeres identificados, un diagrama de violín que muestra la mediana de ambas caracterizaciones evidencia que las distribuciones y los valores medianos (12,7 para el clúster 1 y 8,9 para el clúster 2) son claramente distintos, lo que confirma la presencia del problema de brecha salarial, como se aprecia en la figura 18. Se utilizó la prueba no paramétrica de Brown-Mood para comparar las medianas de los 2 clústeres. Con un nivel de significación de 0,05, se concluye que las medianas son diferentes, como se detalla en la tabla 8.

La categorización de los clústeres utilizando la mediana, dada la asimetría de los clústeres identificados, se presenta en la tabla 9. Se observa claramente la diferencia entre los dos clústeres: en el clúster 1, la brecha salarial es mayor, la participación es menor y el PBI per cápita de la región es más bajo. En el clúster 2, los indicadores son más favorables: la brecha salarial se reduce al 8,9 %, la participación es casi paritaria y el PBI per cápita es más del doble que en el clúster 1. Ambos clústeres comparten similar presencia promedio de servidores públicos por persona en sus respectivas regiones.

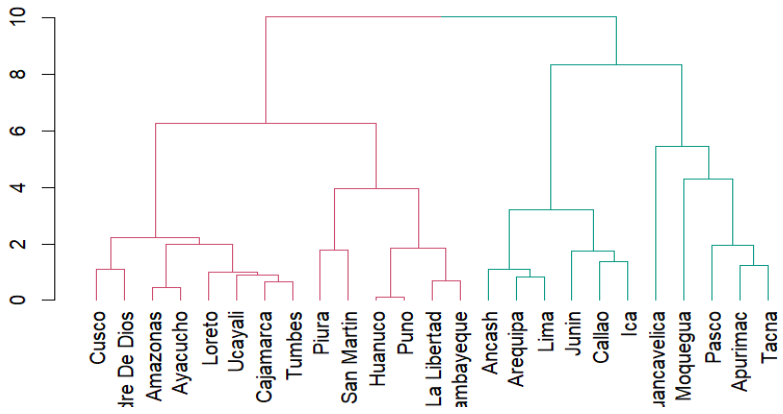


Figura 17. Regiones agrupadas en dos clústeres para analizar la brecha salarial de género

Fuente: Elaboración propia.

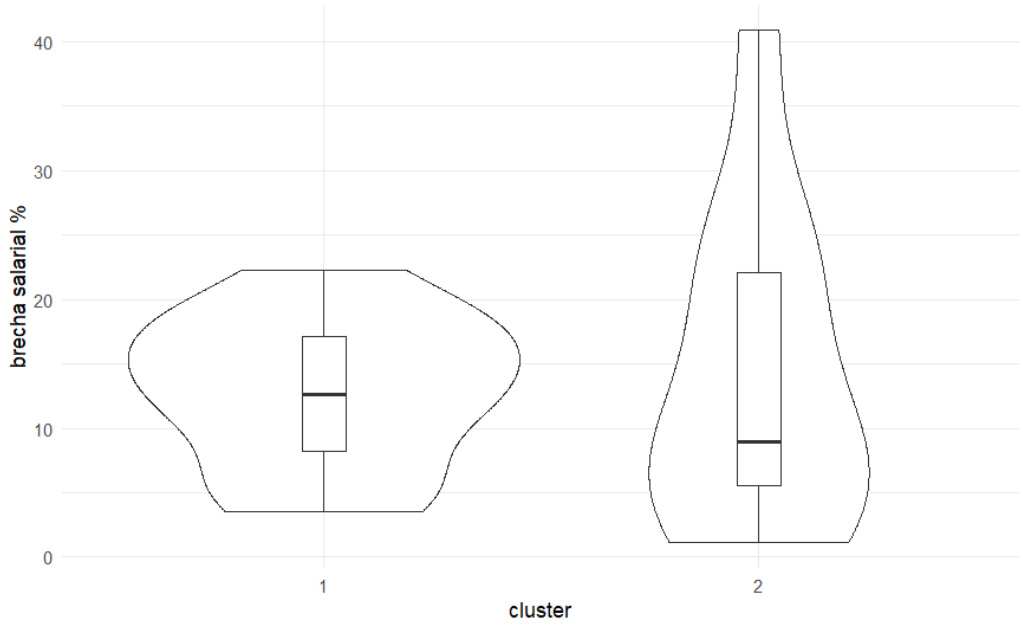


Figura 18. Diagrama de violín, se aprecia diferentes distribuciones y valores medianos para los 2 clústeres identificados

Fuente: Elaboración propia.

Tabla 8
Prueba de hipótesis no paramétrica para medianas de los clústeres

Test de medianas Brown-Mood
Estadístico Z = 2,1669
p-value = 0,03025
H0: las medianas de las brechas salarial de los clústeres 1 y 2 son iguales
H1: las medianas de las brechas salariales de los clústeres 1 y 2 son distintas

Fuente: Elaboración propia.

Tabla 9
Caracterización de los clústeres identificados

Clúster	n.º de regiones	Brecha mediana	Participación mediana	PBI per cápita en miles de soles	Cobertura servidor	Regiones
1	14	12,67	44,45	8,62	21,41	Amazonas, Ayacucho, Cajamarca, Cusco, Huánuco, La Libertad, Lambayeque, Loreto, Madre de Dios, Piura, Puno, San Martín, Tumbes, Ucayali
2	11	8,93	49,88	19,46	20,98	Áncash, Apurímac, Arequipa, Callao, Huancavelica, Ica, Junín, Lima, Moquegua, Pasco, Tacna

Fuente: Elaboración propia.

E. Fase VI, despliegue

La metodología CRISP-DM plantea seis fases, siendo la última el despliegue del modelo en un ambiente de producción. Esta fase no se desarrolla en este artículo debido a que se trata de una investigación. Sin embargo, los elementos aquí desarrollados podrían ser utilizados, en coordinación con Servir, el INEI y otras entidades públicas, para crear un *dashboard* que permita monitorear y revisar el avance de este problema social en el Estado peruano en todos sus niveles. Esta fase en sí misma constituiría un proyecto cuyo alcance excede el presente estudio.

RESULTADOS

Del análisis exploratorio de datos se observó que, a nivel estatal, los indicadores muestran mejores niveles que los promedios generales de las regiones, con una tendencia a mejorar que fue revertida por la pandemia en 2020. Sin embargo, al revisar los datos a nivel regional, se evidencia un marcado contraste. La dispersión de las variables clave, como la brecha salarial y la participación femenina, pone de manifiesto que el problema se presenta principalmente a nivel regional, según la información proporcionada por Servir. Cabe destacar que, a nivel estatal,

parece existir una relación en la que mayor participación corresponde a una menor brecha salarial. No obstante, al analizar los datos por regiones, esta relación desaparece, e incluso, en el segmento con mayor brecha salarial, se observa que a mayor participación correspondería una mayor brecha, lo que requiere una profundización del análisis con más variables.

Los modelos de regresión lineal simple mostraron que, de las 25 regiones, con un nivel de significancia del 5 %, tres avanzaron (Áncash, Huánuco y Moquegua), tres retrocedieron (Huancavelica, La Libertad y San Martín) y 19 no presentaron cambios significativos.

El análisis de *clustering* fue útil para identificar y caracterizar la naturaleza de este problema social. Se probaron varias alternativas, y finalmente se optó por definir 2 clústeres. La caracterización de estos muestra que el grupo con mejores resultados en cuanto a la brecha salarial (clúster 2, con un 9 %) presenta una participación casi paritaria, mayores ingresos per cápita (más del doble que el clúster 1) y ambos coinciden en la cobertura de servicio por trabajador del Estado, con un promedio de 21 personas por cada servidor público. El diagrama de violín y la prueba de hipótesis aplicados a los clústeres y la brecha salarial demuestran claramente que los grupos tienen una naturaleza distinta, lo que evidencia el problema a nivel regional.

CONCLUSIONES

- A nivel agregado del Estado peruano, se observaron avances en la reducción de la brecha salarial, aunque hubo un retroceso tras la pandemia. Durante el período estudiado, la brecha mejoró del 13 % en 2017 al 11,2 % en 2021, mientras que la participación femenina aumentó ligeramente del 46,9 % en 2017 al 47,3 % en 2021. La relación entre brecha salarial y participación también mostró que, a mayor participación, correspondía una menor brecha. Comparado con la brecha general en América Latina, que era del 14 % en 2019, podría considerarse que el problema en el Perú está mejorando.

- A nivel regional, la realidad era diferente: la brecha salarial y la participación mostraron valores muy diversos y alejados del promedio general a nivel estatal. Pasco presentó el mayor nivel de participación mediana (56 %), mientras que Madre de Dios tuvo el menor (39 %). En cuanto a la brecha salarial mediana, Pasco registró la mayor (26 %) e Ica la menor (6 %), lo que evidencia la presencia de este problema social. Al analizar la relación entre brecha y participación por terciles de brecha (bajo, medio, alto), los diagramas de dispersión mostraron que en los dos primeros niveles no había relación, mientras que en el nivel alto se encontró un ligero indicio contrario al observado a nivel estatal: a mayor participación, mayor brecha.

- De las 25 regiones, y con base en el modelo de regresión lineal simple, se detectaron tres con avances (Áncash, Huánuco y Moquegua), tres con retrocesos (Huancavelica, La Libertad y San Martín), y 19 sin cambios significativos. Esto muestra que existe espacio para implementar políticas públicas orientadas a reducir las brechas salariales, identificar las buenas prácticas de gobernanza que han funcionado en las regiones con avances y evitar las prácticas que están perjudicando a las regiones en retroceso.

- La caracterización de los perfiles de brechas salariales a nivel regional resultó muy útil al incorporar variables económicas y de servicio para complementar las dos variables clave del estudio (brecha y participación). El clúster 1 presenta un valor mediano de brecha del 12,7 %, mientras que el clúster 2 muestra un 8,9 %. Además, el clúster 2 presenta una participación casi paritaria (49,9 %) y un PBI per cápita más del doble que el clúster 1 (19 400 vs. 8600). Ambos clústeres tienen en común una cobertura de casi 21 personas por cada servidor civil. El clúster 2 comprende 11 regiones, pero concentra las que tienen mayor número de servidores (56 %). Este peso, junto con sus mejores indicadores, permite que a nivel agregado el Estado presente una aparente adecuada gestión del problema.

- La prueba de hipótesis no paramétrica de Brown-Mood de los valores medianos de las brechas salariales para los dos clústeres muestran evidencia estadística de la existencia del problema a nivel de regiones.

- La revisión de los grados universitarios y avanzados completos de las mujeres en cada uno de los niveles del Estado peruano (nacional, regional y local) muestra que es superior al de los hombres (55 % de mujeres frente a 39 % de hombres). Esto, en atención a las normas laborales vigentes y a lo encontrado en este estudio, ofrece más evidencia de que el problema de la brecha salarial está presente en el Estado peruano.

- Contar con más información (categorías de puesto, niveles educativos, experiencia, edad, etc.) y con un mayor nivel de granularidad permitiría mejorar el análisis a nivel de provincias, distritos, sectores e incluso organismos públicos. Poner esta información a disposición mediante plataformas en la nube y *dashboards* facilitaría a los implementadores de políticas públicas, en todos los niveles del Estado, monitorear el problema y tomar acciones para reducir la brecha salarial de género.

- Es fundamental contar con un marco metodológico que guíe tanto las revisiones del proyecto general como las revisiones internas de cada modelo, permitiendo su integración. En ese sentido, la metodología CRISP-DM fue un buen soporte para este trabajo.

REFERENCIAS

[1] Aldas, J. & Uriel, E. (2017). *Análisis multivariante aplicado con R*. Paraninfo.

[2] Autoridad Nacional del Servicio Civil (Servir). (30 de agosto de 2023). *Plataforma de datos abiertos: Información sobre la cantidad de servidores civiles y sus ingresos promedio a nivel nacional periodo 2017-2021.* https://www.datosabiertos.gob.pe/dataset/informaci%C3%B3n-sobre-la-cantidad-de-servidores-civiles-y-sus-ingresos-promedio-nivel-nacional

[3] Autoridad Nacional del Servicio Civil (Servir) (2024). *La mujer en el servicio civil peruano 2024*. Autoridad Nacional del Servicio Civil (Servir).

[4] Blume, I. (2023). Igualdad Salarial: Desarrollos actuales, perspectivas y debates. *Laborem* (27), 105-120.

[5] Brock, G., Datta, S., Pihur, V. & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*.

[6] Chapman, P. C. (2000). *CRISP-DM 1.0*.

[7] Congreso de la República. (15 de marzo de 2024). *Archivo Digital de la Legislación del Perú*. https://www.leyes.congreso.gob.pe/Documentos/2016_2021/ADLP/Normas_Legales/30709-LEY.pdf

[8] Instituto Nacional de Estadística e Informática - INEI. (15 de marzo de 2024). *Proyección población 2018-2022*. https://www.gob.pe/institucion/inei/informes-publicaciones/3464927-peru-proyecciones-de-poblacion-total-segun-departamento-provincia-y-distrito-2018-2022

[9] Instituto Nacional de Estadística e Informática - INEI. (2024, marzo 31). *PBI por región*. https://m.inei.gob.pe/estadisticas/indice-tematico/producto-bruto-interno-por-departamentos-9089/

[10] Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R*. STHDA.

[11] Loayza Pacheco, L. (2019). *Guía para igualdad. Igualdad salarial*. Ministerio de Trabajo y Promoción del Empleo.

[12] Organización Internacional del Trabajo (OIT). (2020). *Panorama Laboral 2020. América Latina y el Caribe*. OIT.

[13] Organización Internacional del Trabajo (OIT). (2024). *Panorama Laboral en América Latina y el Caribe 2024. Cerrar la brecha de género para impulsar la economía y la productividad en América Latina*. OIT.

[14] ONU Mujeres. (2023). *¿Qué es la brecha salarial?* https://lac.unwomen.org/es/que-hacemos/empoderamiento-economico/epic/que-es-la-brecha-salarial

[15] Soto, I. & Gamboa, J. (2021). *Ciencia de Datos con R. Métodos estadísticos para la investigación experimental*. Universidad Nacional Agraria La Molina.

[16] The Royal Swedish Academy of Sciences. (15 de marzo de 2024a). *Scientific background to the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2023*. https://www.nobelprize.org/uploads/2023/10/advanced-economicsciencesprize2023.pdf

[17] The Royal Swedish Academy of Sciences. (15 de marzo de 2024b). *History helps us understand gender differences in the labour market*. https://www.nobelprize.org/uploads/2023/10/popular-economicsciencesprize2023.pdf

[18] Urquidi, M. (2023). *Brecha de ingresos laborales por género en América Latina y el Caribe: un análisis de componentes*. BID.

[19] Zumel, N. & Mount, J. (2020). *Practical data science with R*. Manning Publications.

ACERCA DEL AUTOR

ROBERTO LEÓN LEYVA

Profesor de la carrera de Big Data y Ciencia de datos de Tecsup, ingeniero de sistemas por la Universidad Nacional de Ingeniería (UNI) y MBA por la Universidad del Pacífico. Cuenta con estudios concluidos en la maestría de Estadística Aplicada por la Universidad Agraria La Molina (Unalm) y especializaciones por la Universidad de Michigan en Coursera: Data Analytics in the Public Sector with R y Applied Data Science with Python, así como de Datacamp: Data Scientist y Quantitative Analyst with R.

@ robertoleon10@gmail.com

@ www.linkedin.com/in/robertoleonleyva

Recibido: 21-04-24
Revisado: 16-08-24
Aceptado: 23-09-24