



© Los autores. Publicado por la Universidad Nacional Agraria La Molina



El artículo es de acceso abierto y está bajo la licencia [CC BY](https://creativecommons.org/licenses/by/4.0/)

Recibido: 5/01/2021; Aceptado: 01/06/2021; Publicado: 30/06/2021

<http://dx.doi.org/10.21704/rtn.v15i1.907>

Caracterización de las regiones del Perú según los indicadores de género aplicando métodos de agrupamiento

Characterization of the regions of Peru according to gender indicators applying grouping methods

César Higinio Menacho Chiok^{1*}, Jorge Chue Gallardo^{1*}

* Autor de correspondencia: cmenacho@lamoliana.edu.pe

RESUMEN

En las últimas décadas muchos países están priorizando dentro de sus políticas públicas programas y proyectos para favorecer la equidad de género. Están aplicando encuestas y censos, con la finalidad de recopilar datos para generar estadísticas e indicadores que les permitan evaluar y medir la equidad de género. El objetivo es caracterizar las 24 regiones del Perú aplicando los métodos de agrupamiento jerárquico y k-medias a 29 indicadores de género recopilados del portal del INEI para el año 2019, agrupados en cuatro temas: salud y vida, educación, trabajo y empleo y violencia de género. Se evidencia una falta de equidad de género, principalmente en educación en los indicadores de la tasa de asistencia y matrícula nivel secundario y tasa de analfabetismo, en trabajo y empleo en el ingreso mensual y en la violencia familiar contra la mujer la física y sexual, psicológica.

Palabras clave: Caracterización, indicadores de género, agrupamientos, métodos jerárquicos, método no jerárquico k-medias.

ABSTRACT

In recent decades, many countries are prioritizing within their public policies programs and projects to promote gender equality. They are applying surveys and censuses, in order to collect data to generate statistics and indicators that allow them to evaluate and measure gender equity. The objective is to characterize the 24 regions of Peru by applying the hierarchical grouping and k-means methods to 29 gender indicators collected from the INEI portal for the year 2019, grouped into four themes: health and life, education, work and employment, and violence of genre. There is evidence of a lack of gender equity, mainly in education in the indicators of the attendance and enrollment rate at the secondary level and the illiteracy rate, in work and employment in monthly income and in family violence against women, physical and sexual, psychological.

Keywords: Characterization, gender indicators, groupings, hierarchical methods, non-hierarchical k-means method.

Forma de citar el artículo (Formato APA):

Menacho, C.; & Chue, J. (2021). Caracterización de las regiones del Perú según los indicadores de género aplicando métodos de agrupamiento. *Tierra Nuestra*. 15(1), 84-94. <http://dx.doi.org/10.21704/rtn.v15i1.907>

¹ Universidad Nacional Agraria La Molina, 15024, Lima, Perú. cmenacho@lamoliana.edu.pe; jchue@lamoliana.edu.pe

1. Introducción

En las últimas décadas, la equidad de género se ha convertido en una de las políticas prioritarias para muchos gobiernos y organizaciones internacionales. La equidad de género, se puede entender como la igualdad que tienen las mujeres y los hombres en las oportunidades, necesidades, beneficios, derechos y deberes como miembros de una sociedad. Para identificar las mejores políticas públicas que incidan en las mejores programas y acciones para hacer frente a las brechas de género, es necesario describir y analizar cada una de las dimensiones de educación, salud, economía y participación política (Ministerio de Educación, 2020).

En (Organización Internacional del Trabajo, 2019), la brecha salarial de género representa una de las injusticias sociales en la actualidad, es necesario aplicar métodos que permitan una evaluación crítica la formulación de políticas y monitorear la brecha salarial de género. En el estudio con 70 países y 80% de los empleados asalariados de todo el mundo, el promedio de ingreso de las mujeres sigue siendo un 20% menos que los hombres. Según un nuevo informe del Banco Mundial (2021), los países continúan avanzando lentamente hacia una mayor equidad de género; las mujeres se están enfrentando a leyes y regulaciones que limitan sus oportunidades económicas, acentuándose por la pandemia de la COVID-19 en áreas de salud, seguridad y educación.

Los países coinciden en la importancia de recopilar información estadística dirigida a evaluar la perspectiva de género. En tal sentido, se están aplicando encuestas y censos con la finalidad de generar indicadores de género que permitan medir y evaluar el impacto de los programas y proyectos sobre la equidad de género. Los indicadores de género, permiten obtener estadísticas para evaluar la magnitud y tendencias de las políticas públicas a través de cuantificar la desigualdad de género en los países de América Latina (Milosavljevic, 2007). La importancia de las estadísticas de género como una herramienta para medir y evaluar la situación de las mujeres y como instrumento para visualizar, construir políticas y avanzar hacia la igualdad de género (Guzmán Acuña, 2017).

Las estadísticas y los indicadores de género juegan un rol muy importante para la medición de las brechas de género, revelan desigualdades en la vida de las mujeres y hombres con respecto a la educación, mercado laboral, ingreso y salud (Eurostat, 2018). Se pueden conformar bases de datos con gran volumen de información estadística que permita construir y almacenar indicadores de género agrupados en diversas áreas temáticas o dimensiones.

El Observatorio de igualdad de género de CEPAL, agrupa los indicadores de género en tres áreas temática: autonomía física, autonomía en la toma de decisiones y autonomía económica; El Instituto Europeo de igualdad de género establece 10 áreas temática. El INEI-Perú, recopila indicadores históricos de género enmarcados en nueve dimensiones y desagregados a nivel departamental: 1) Índice de Desigualdad de Género, 2) Indicadores de autonomía de las mujeres, 3) Indicadores de la dinámica demográfica, 4) Indicadores de equidad en salud, 5) Indicadores de educación, 6) Indicadores acceso al trabajo, 7) Indicadores acceso a las TICs, 8) Violencia doméstica y física y 9) Indicadores de acceso a recurso agropecuarios (INEI, 2020). Estos repositorios de datos de indicadores de género pueden ser analizados, aplicando técnicas estadísticas multivariadas. Estas técnicas tienen la capacidad de descubrir perfiles y patrones a partir del aprendizaje de los datos. Un grupo de técnicas conocidos como métodos de agrupamiento, tienen como objetivo conformar grupos (clústers) de objetos con características muy similares (homogeneidad dentro del grupo) y muy disímiles entre los objetos de diferentes grupos (heterogeneidad entre grupos). En este trabajo se aplica los métodos de agrupamiento jerárquico y no jerárquico (k-medias), con la finalidad de caracterizar las 24 regiones del Perú usando los indicadores de género recopilados desde el portal del INEI para el año 2019. Se agrupan los indicadores cuatro temas: salud y vida, educación, trabajo y empleo y violencia de género. Se aplican cuatro métodos jerárquicos y el método no jerárquico k-medias, luego se selecciona con los coeficientes de agrupamiento y la distribución de cada uno de los clúster el mejor método de agrupamiento. Se identifica el número de clúster con el método del codo. Se realiza la caracterización del agrupamiento de las 24 regiones por cada tema, usando los promedios de cada uno de los indicadores.

2. Antecedentes

En el área temática de la equidad en salud, se tiene estudios como en (Parra Pera, 2018), donde se aplican técnicas de agrupamiento jerárquicos y no jerárquicos (k-medias) con la finalidad de caracterizar por grupos a los departamentos de Colombia según la tasa de mortalidad de la lista 6/67 de la OPS y establecer la posible asociación entre variables socioeconómicas usando un análisis de correspondencia. Los resultados con los clústeres de los departamentos, identificaron que las causas de la mortalidad en Colombia en el año 2018 fueron las enfermedades circulatorias y las neoplasias. Además, los departamentos más pobres están relacionados con tasas más bajas de mortalidad y altas en enfermedades consideradas. En (S., Harryson, Bolin, & Hammarstrom, 2013), se aplica el análisis de clúster con indicadores de género en búsqueda de patrones en angustia psicológica en el trabajo generando seis clústers. Los clústers con mujeres con alta angustia psicológica, se asocia a una alta desigualdad en las labores de trabajo. En (Pezer, 2018), se realiza un estudio con la finalidad de identificar grupos de países para brindar políticas de apoyo a la equidad de género y a la maternidad, niveles de fecundidad, tasa empleo de madres y brecha en desarrollo humano. Se aplican análisis de agrupamiento y análisis de componentes principales a 12 indicadores categorizados en tres dimensiones: políticas públicas, generosidad y contexto nacional. Se identifican cuatro grupos: (1) Europa occidental (menor licencia parental), (2) Europa central y oriental (bajos en economía y menor cuidado infantil), (3) Sur de Europa (escaso apoyo y pobreza femenina) y (4) Países nórdicos (generoso grupo de apoyo). Niveles altos de fecundidad se asocian a conglomerados con economías altas. En (Heinz A., Catunda C., Van Duin C., Torsheim T. and Willems H., 2020), se analiza la equidad de género entre 45 países usando 10 indicadores de salud. Se toma la encuesta de salud donde participaron 71972 estudiantes de 15 años en el año 2018. Los resultados del análisis de clúster permitieron agrupar a los países en siete grupos con similares perfiles de riesgo, mostrando un patrón sistemático de desigualdad género. Además, en ocho indicadores se evidencia una correlación negativa con el índice de desigualdad de género. En cuatro indicadores hay contraposición, cuanto mayor es la igualdad de género en un país, mayores son las diferencias entre niños y niñas en lo

que respecta a la vida satisfacción, presión escolar, múltiples problemas de salud y sensación de gordura. Se concluye que los riesgos de salud en niños y niñas son similares en los países con vinculados con geográfica.

La desigualdad salarial entre mujeres y hombres, es uno de las áreas temáticas prioritarias a nivel mundial. En (Araujo, A.; 2015), se realiza un estudio para medir la desigualdad del ingreso en el mercado laboral entre hombres y mujeres del Brasil, usando indicadores del capital humano, la integración en el trabajo, el capital social, el capital cultural y el estado civil. Los resultados de aplicar la regresión cuantílica, mostraron que las mujeres ganan en promedio 54% del salario de los hombres, notando una mayor desigualdad de género salarial en el percentil 99% con un valor de sólo 44% de las mujeres con respecto a los hombres. En (Gálvez, Rodríguez, & Domóñez, 2011), con un análisis de clustering se agrupa a 14 países de Europa por indicadores que miden el tiempo de trabajo no remunerativo en mujeres y hombres en cuatro clústers. Los países con mayor desigualdad en el tiempo de trabajo no remunerado, son los que presentan menor trabajo en las mujeres.

Respecto a la equidad de género en educación, estudios sobre la motivación de las mujeres en elegir carreras en ciencias. En (Chopra S., Mirsafian M., Khan A., and Golb L., 2019), se realiza un estudio sobre las diferencias de género en ciencia, tecnología y matemáticas con 17000 puestos de trabajo. Se aplican aprendizaje profundo, minería de textos y métodos estadísticos con los datos académicos, tiempo de admisión y puestos de trabajo de los estudiantes. Los resultados muestran que las mujeres tienen diferentes razones que los hombres en postular a programas de ingeniería, ocupar puestos gerenciales y menos interesadas en actividades empresariales.

3. Metodología

El análisis de agrupamiento (clustering), es el proceso de agrupar un conjunto de objetos en clúster de tal manera, que los objetos dentro de un clúster tengan una similitud alta entre ellos (homogeneidad dentro de grupos) y una disimilitud alta entre los objetos de diferentes clústers (heterogeneidad entre grupos). Los métodos de agrupamiento, son técnicas estadísticas

exploratorias para el aprendizaje no supervisado que permiten descubrir la estructura de distribución en conglomerados de un conjunto de datos. Estos métodos, utilizan medidas de distancia (variables cuantitativas) o índices de similitud (variables cualitativas) como métricas para evaluar y agrupar los objetos en los diferentes clústers. Existen una variedad de métodos de agrupamiento. Los métodos jerárquicos, se basan en un proceso secuencial estableciendo jerarquías entre los objetos para la formación de los clúster, pueden ser métodos aglomerativos o divisivos. Métodos no jerárquicos, se construyen los clústers por un proceso iterativo basado en particiones, el más representativo es el k-medias. Estos métodos usan métricas de distancia para medir la similitud o disimilitud entre los objetos para ir conformando cada uno de los clústers.

Métodos jerárquicos aglomerativos

En los métodos jerárquicos aglomerativos, el proceso de construcción de los clúster se inicia con el número de grupos igual al total de objetos "n". En los pasos siguientes según el método aglomerativo elegido, se van uniendo los objetos o grupos de acuerdo a medir sus similitudes con alguna métrica de distancia. Se finaliza con la conformación de un sólo clúster con todos los objetos. Los métodos aglomerativos que se usan como criterio para la conformación de los clúster son:

- **Enlace simple o vecino más cercano.** Los clúster se van uniendo de acuerdo a su menor distancia (vecino más cercano) o mayor similitud. Apropiado cuando los grupos están claramente diferenciados. Tiende a unir clúster diferentes y está afectado por la presencia de datos atípicos (forman demasiados grupos).
- **Enlace completo o vecino más alejado.** Los clúster se van uniendo de acuerdo a la distancia más grande entre un par de elementos de cada grupo o cluster (vecino más alejado). También es sensible a datos atípicos.
- **Enlace promedio.** Considera como distancia entre dos grupos, la distancia media entre todos los pares posibles de grupos. Tiende a producir cluster compactos.
- **Enlace Ward.** Utiliza el análisis de variancia para evaluar la distancia entre dos grupos (minimiza la suma de cuadrados con respecto al centroide dentro de los grupos).

Métodos no jerárquicos

Son métodos de agrupamiento basados en particiones, donde la construcción de los clústers sigue un proceso iterativo de k particiones. En cada proceso iterativo, se reorganizan los objetos asignados a cada clúster (a diferencia del jerárquico que no hay movimiento de objetos) de tal manera que se va mejorando el agrupamiento. El método no jerárquico más usado, es el k-medias. El método k-medias, basa el agrupamiento de los objetos en ir fijando en cada iteración los centroides (promedio) de cada grupo y calculando la suma de cuadrados dentro de los clúster usando una métrica de distancia (euclidiana) para ir uniendo los objetos o grupos.

Medidas de distancia

Las medidas de distancias, son métricas usadas para medir la similaridad entre los objetos en un análisis de agrupamiento. A menor valor de distancia, indica que los objetos están más cercanos. Las distancias se miden para pares de objetos y se recomienda realizar una estandarización para eliminar el efecto de las unidades de medida de las variables en estudio. Así, para el objeto (X_{ik}, X_{jk}) evaluado con p variables se definen las distancias:

- **Euclidiana.** Es depende de las unidades de medida, generalmente se aplica a variables estandarizadas. No considera la posible correlación entre las variables.

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2}$$

- **Mahalanobis.** Considera la correlación entre las variables, utiliza la inversa de la matriz de covariancias como ponderación.

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^p (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j)}$$

- **Minkowski.** Se usa cuando se desea considerar el efecto de las desviaciones al cuadrado (suavizar) extrayendo la m-ésima raíz.

$$d(x_i, x_j) = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{\frac{1}{m}}$$

- **Manhattan.** Una alternativa de la distancia euclidiana cuando se encuentra afectada por datos atípicos.

$$d(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

Determinación del número de grupos (clústers)

Existen procedimientos gráficos y cuantitativos que permiten identificar el número de grupos (k) más adecuado para el agrupamiento de los objetos.

- **Diagrama del Dendograma.** Es un gráfico que muestra la secuencia de la formación de los clúster, según el método de aglomeramiento y la métrica distancia.
- **Método de silueta.** Se elabora un gráfico a partir de calcular los índices de siluetas y el global para diferentes valores del número de clúster (k).
- **Método del codo.** Es un gráfico que se elabora a partir de calcular la suma de cuadrados del error (SCE) para diferentes valores de k. Luego se identifica en el gráfico el k óptimo, donde se presenta la mayor curvatura (codo).

Validación del agrupamiento

Existen métodos o técnicas con la finalidad de evaluar la calidad del agrupamiento. Entre los métodos de evaluación se tiene:

- **Coefficiente de agrupamiento.** Se basa en calcular la suma de cuadrados del error entre los clústers (variabilidad entre clúster). El mejor agrupamiento corresponde al mayor valor.
- **Coefficiente de correlación cofenético.** Se calculan como la correlación entre la matriz de distancias observadas y la matriz de cofenética. La matriz cofenética, se calcula con las aproximaciones de dos objetos que pertenecen al mismo clúster.

- **Evaluar la distribución de los clúster.** Se desea clúster balanceados y con una agrupación consistente de los objetos asignados a cada clúster.
- **Medidas estadísticas.** Calcular y comparar las medidas estadísticas descriptivas de cada uno de los clúster formados (promedios, desviación estándar, coeficiente de variación, asimetría, etc.). Esto también es usado para la caracterización de los clúster hallados.
- **Aplicar el análisis de variancia de un factor.** El procedimiento de prueba de hipótesis para t medias (DCA), permite validar la construcción de los clúster. Se espera rechazar la hipótesis nula de igualdad de las t medias para todas las variables para verificar el buen agrupamiento.

4. Resultados

En este estudio se usan los datos que se presentan en el portal del INEI, donde se acceden desde el rubro de Estadísticas las bases de datos de los Indicadores de Género. Se recopilan para el año 2019 y las 24 regiones del Perú, 29 indicadores de género agrupados en cuatro temas: Salud (10), Educación (7), Trabajo y empleo (7) y Violencia de género (5).

Métodos de agrupamiento jerárquicos

Se aplican cuatro métodos de enlace para el agrupamiento de las 24 regiones: simple, completo, promedio y Ward. Con la finalidad de identificar el mejor método aglomerativo, en la Tabla 1, se presenta los coeficientes de correlación cofenéticos y coeficientes de agrupamiento. Los mayores valores se muestran en los coeficientes de agrupamiento para el método Ward en los cuatro temas. Por lo tanto, para los cuatro temas, se selecciona para el método de agrupamiento jerárquico el enlace Ward.

Tabla 1. Coeficientes de correlación cofenéticos y agrupamiento

Método aglomerativo	Coeficientes de correlación cofenéticos				Coeficientes de agrupamiento			
	Tema 1	Tema 2	Tema 3	Tema 4	Tema 1	Tema 2	Tema 3	Tema 4
Simple	0,643	0,592	0,577	0,669	0,440	0,340	0,418	0,450
Compuesto	0,685	0,593	0,616	0,553	0,648	0,739	0,784	0,802
Promedio	0,786	0,629	0,632	0,717	0,534	0,577	0,642	0,674
Ward	0,505	0,583	0,570	0,585	0,754	0,832	0,835	0,862

Para determinar el número clústers, se aplicará el método del codo. En la Figura 1, se muestra el gráfico del método del codo para cada uno de los temas. Se observa que para los cuatro temas, un número de tres clústers es el más adecuado para realizar el agrupamiento de las regiones del Perú según los indicadores de género.

El resultado del agrupamiento con tres clústers de las 24 regiones del Perú con el método jerárquico y para cada uno de los temas se presenta en la Tabla 2. Se muestra distribuciones diferentes para los cuatro temas en cuanto al número y porcentajes de regiones asignadas en cada uno de los clústers.

En la Figura 2, se muestra para cada tema el agrupamiento por el método jerárquico de las 24 regiones del Perú aplicando el análisis de componentes

principales. Considerando las dos primeras componentes principales, se visualiza que los temas de Vida y salud y Educación existe una sobre posición entre los tres clústers, lo cual indicaría una falta de consistencia en el agrupamiento con el método jerárquico. Mientras, que se visualiza una consistencia en el agrupamiento de las 24 regiones con el método jerárquico con los indicadores en los temas de Trabajo y empleo y Violencia de género.

Método de agrupamiento k-medias

El resultado del agrupamiento con tres clústers de las 24 regiones del Perú con el método no jerárquico k-medias y para cada uno de los temas se presenta en la Tabla 3. Se muestra distribuciones diferentes para los cuatro temas en cuanto al número y porcentajes de regiones asignadas en cada uno de los clústers.

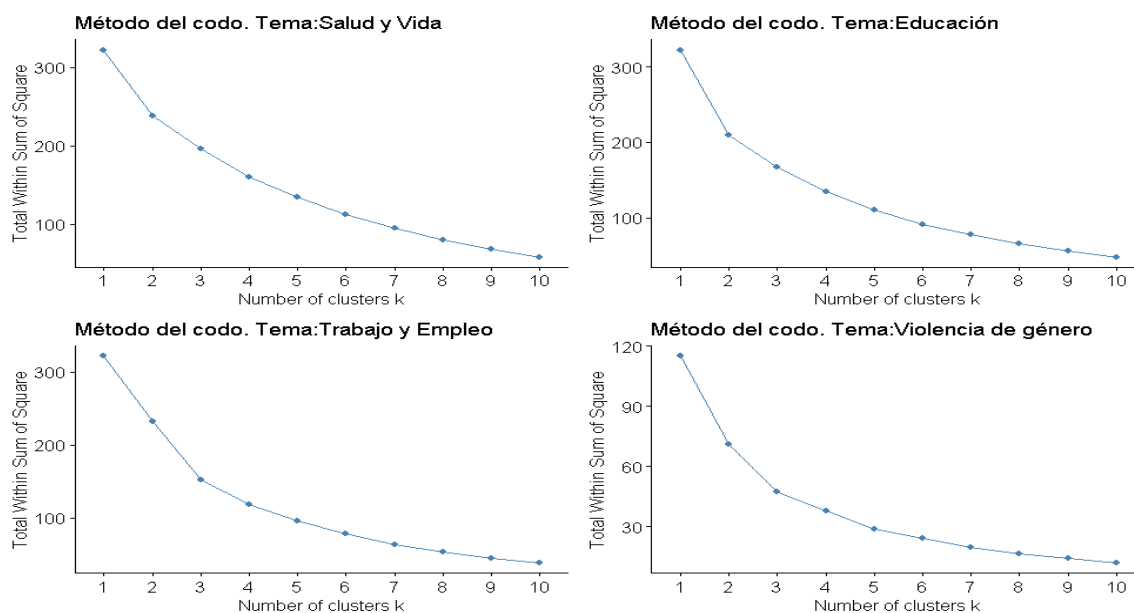


Figura 1. El método del codo para los cuatro temas

Tabla 2. Agrupamiento de las regiones del Perú con el método jerárquico

Clústers	Tema 1		Tema 2		Tema 3		Tema 4	
	Nr.	Porcentaje	Nr.	Porcentaje	Nr.	Porcentaje	Nr.	Porcentaje
1	8	33,3	11	45,8	8	33,3	12	50,0
2	10	41,7	8	33,3	7	29,2	7	29,2
3	6	25,0	5	20,8	9	37,5	5	20,8
Total	24	100,0	24	100,0	24	100,0	24	100,0

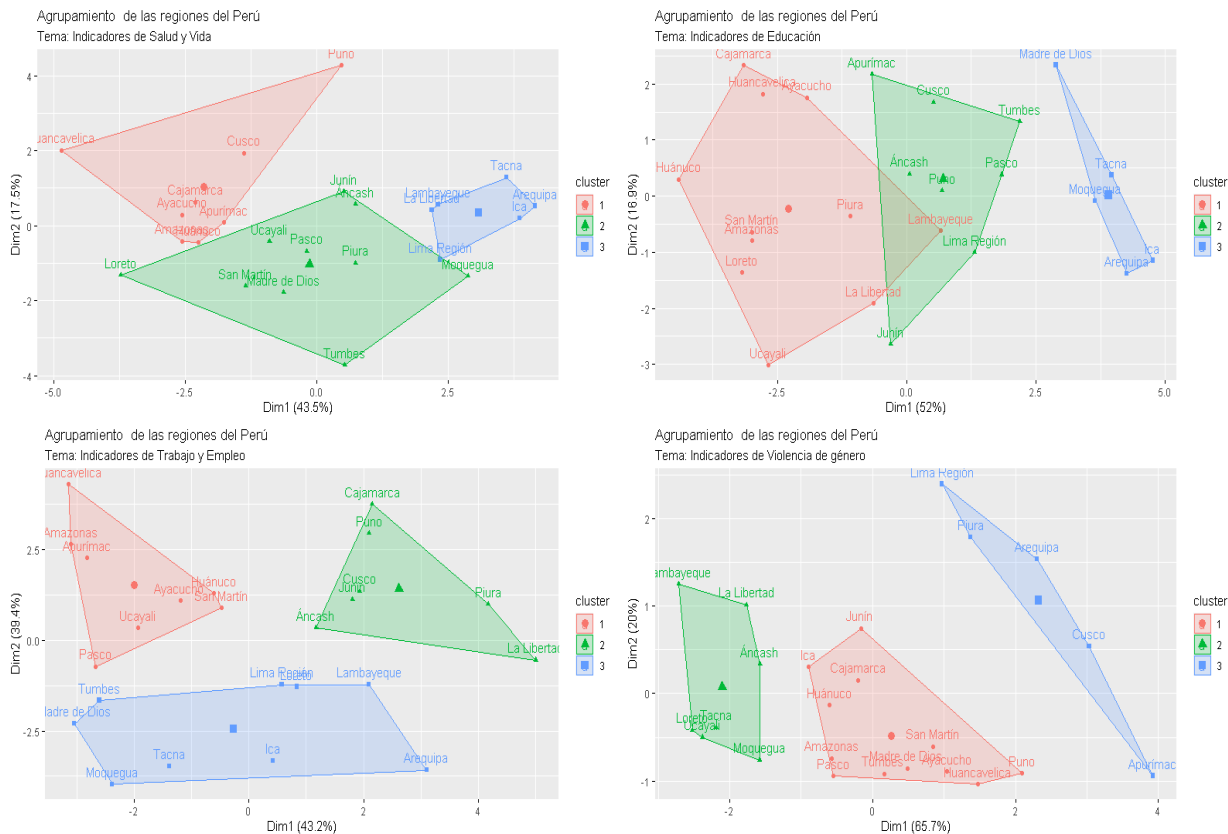


Figura 2. Agrupamiento de las regiones del Perú por tema con el método jerárquico

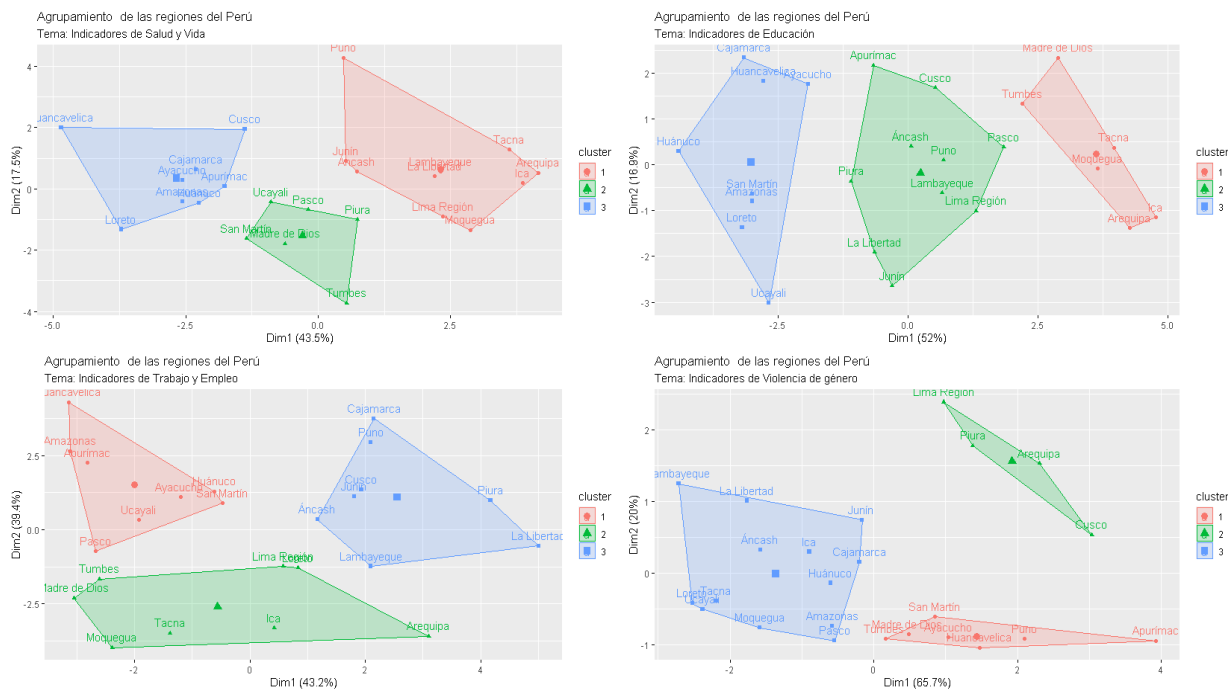


Figura 3. Agrupamiento de las regiones del Perú por tema con el método k-medias

Tabla 3. Agrupamiento de las regiones del Perú con el método k-medias

Clústers	Tema 1		Tema 2		Tema 3		Tema 4	
	Nr.	Porcentaje	Nr.	Porcentaje	Nr.	Porcentaje	Nr.	Porcentaje
1	10	41,7	6	25,0	8	33,3	7	29,2
2	6	25,0	10	10	8	33,3	4	16,7
3	8	33,3	8	8	8	33,3	13	54,2
Total	24	100,0	24	100,0	24	100,0	24	100,0

En la Figura 3, se muestra para cada tema el agrupamiento por el método k-medias para las 24 regiones del Perú aplicando el análisis de componentes principales. Considerando las dos primeras componentes principales, se visualiza en los temas de Vida y salud y Educación una sobre posición entre los tres clústers, lo cual indicaría una falta de consistencia en el agrupamiento con el método jerárquico. Mientras, que se visualiza una consistencia en el agrupamiento de las 24 regiones con el método jerárquico con los

indicadores en los temas de Trabajo y empleo y Violencia de género.

En la Tabla 4, se presenta el agrupamiento de las 24 regiones del Perú en tres clústers y por cada uno de los cuatro temas. Se observa que reordenando los clúster en los cuatro temas, existe algunas regiones que pueden coincidir el el mismo grupo. Sin embargo, para los propósitos del estudio se establece diferentes agrupamientos de las regiones en los tres clúster para los cuatro temas.

Tabla 4. Agrupamiento con tres clústers de las 24 regiones del Perú por tema

Temas	Clúster 1	Clúster 2	Clúster 3
Salud y vida	Áncash, Arequipa, Ica, Junín, La Libertad, Lambayeque, Lima Región, Moquegua, Puno, Tacna	Madre de Dios, Pasco, Piura, San Martín, Tumbes, Ucayali	Amazonas, Apurímac, Ayacucho, Cajamarca, Cusco, Huancavelica, Huánuco, Loreto
Educación	Arequipa, Ica, Madre de Dios, Moquegua, Tacna, Tumbes	Áncash, Apurímac, Cusco, Junín, La Libertad, Lambayeque, Lima Región, Pasco, Piura, Puno	Amazonas, Ayacucho, Cajamarca, Huancavelica, Huánuco, Loreto, San Martín, Ucayali
Trabajo y empleo	Amazonas, Apurímac, Ayacucho, Huancavelica, Huánuco, Pasco, San Martín, Ucayali	Arequipa, Ica, Lima Región, Loreto, Madre de Dios, Moquegua, Tacna, Tumbes	Áncash, Cajamarca, Cusco, Junín, La Libertad, Lambayeque, Piura, Puno
Violencia de género	Apurímac, Ayacucho, Huancavelica, Madre de Dios, Puno, San Martín, Tumbes	Arequipa, Cusco, Lima Región, Piura	Amazonas, Áncash, Cajamarca, Huánuco, Ica, Junín, La Libertad, Lambayeque, Loreto, Moquegua, Pasco, Tacna, Ucayali

Con la finalidad de realizar la caracterización del agrupamiento de las 24 regiones del Perú según los indicadores de género, se elaboran tablas de promedios dentro de cada clústers y por género. En la Tabla 5, se presenta los promedios de los indicadores de género para el tema de salud y vida dentro de cada uno de los tres clústers y por género. Se observa que en los tres clústers los indicadores del padecimiento de algún problema crónico, tenencia de seguro, tenencia de SIS y esperanza de vida muestran mayores porcentajes en promedio en las mujeres que en los hombres. El análisis por clúster, indica que el clúster 3 para los indicadores mencionados y para madres adolescentes o

embarazadas por primera vez y uso de métodos tradicionales de planificación familia; el clúster 1 para partos atendidos por personal de salud especializado y relación de feminidad; el clúster 2 para Uso de métodos modernos de planificación familiar por mujeres unidas en edad fértil son los que presentan los mayores porcentajes en promedio.

En la Tabla 6, se presenta los promedios de los indicadores de género para el tema de educación dentro de cada uno de los tres clústers y por género. Se puede mencionar que en los tres clústers los indicadores tasa neta de matrícula primaria, tasa neta de asistencia

primaria y tasa de analfabetismo resultan con mayores porcentajes en promedio en las mujeres que en los hombres; mientras que en los indicadores tasa neta de matrícula secundaria, tasa neta de asistencia secundaria y al menos secundaria completa los mayores porcentajes en promedio muestran los hombres que las mujeres. Para el indicador promedio de años de estudio alcanzado de 25 y más años de edad es similar en las mujeres y hombres en los tres clústers.

En la Tabla 7, se presenta los promedios de los indicadores de género para el tema de trabajo y empleo para cada uno de los tres clústers y por género. Se observa

que en los tres clústers los indicadores de edad de trabajo y personas económicamente activas y personas ocupadas el número promedio de personas es mayor en los hombres que en las mujeres, en la tasa de actividad y relación empleo-población los porcentajes en promedio son mayores en los hombres que en las mujeres, mientras que en la tasa de informalidad las mujeres muestran mayores porcentajes en promedio. En el ingreso promedio real mensual proveniente del trabajo, las mujeres tienen mucho menores ingresos que los hombres, notándose con mayor desigualdad en el clúster 2.

Tabla 5. Promedio de los indicadores de salud y vida por clústers

Indicadores de género	Clúster 1		Clúster 2		Clúster 3	
	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre
Reportaron padecer algún problema de salud crónico (%)	46,8	36,5	39,5	31,0	36,5	27,9
Tienen seguro de salud (%)	75,8	71,0	82,7	76,5	90,2	86,0
Seguro Integral de Salud (SIS) (%)	46,0	40,1	61,0	53,8	76,0	71,2
Esperanza de vida al nacer (En años quinquenales)	80,3	75,4	79,1	72,7	77,5	71,8
Partos atendidos por personal de salud especializado (%)	96,2		94,2		90,9	
Madres adolescentes o embarazadas por primera vez (%)	10,6		17,1		17,8	
Mujeres en edad fértil con necesidad insatisfecha de planificación familiar (%)	5,3		6,8		6,8	
Relación de feminidad (Cantidad de mujeres por cada 100 hombres)	100,5		89,3		96,6	
Uso de métodos modernos de planificación familiar por mujeres unidas en edad fértil (% con respecto al total)	53,2		58,3		50,4	
Uso de métodos tradicionales de planificación familiar por mujeres unidas en edad fértil (%)	24,0		15,9		25,5	

Tabla 6. Promedio de los indicadores de educación por clústers

Indicadores de género	Clúster 1		Clúster 2		Clúster 3	
	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre
Tasa neta de matrícula a educación primaria (%)	95,6	94,9	93,6	94,7	94,3	92,1
Tasa neta de matrícula a educación secundaria (%)	89,2	90,5	84,8	86,1	80,1	82,5
Tasa neta de asistencia a educación primaria (%)	95,8	94,8	93,5	94,6	94,2	91,8
Tasa neta de asistencia a educación secundaria (%)	89,1	90,2	84,3	85,3	79,2	81,8
Promedio de años de estudio alcanzado de 25 y más años de edad	10,3	10,9	9,1	9,9	8,1	8,9
Al menos secundaria completa con 17 a más años de edad (%)	65,1	73,1	52,6	62,6	40,8	50,1
Tasa de analfabetismo de 15 y más años de edad (%)	5,0	1,9	11,6	3,6	14,3	5,1

Tabla 7. Promedio de los indicadores de trabajo y empleo por clústers

Indicadores de género	Clúster 1		Clúster 2		Clúster 3	
	Mujer	Hombre	Mujer	Hombre	Mujer	Hombre
Edad de trabajo (miles de personas)	199,6	221,6	238,7	248,9	559,4	556,2
Económicamente activos (miles de personas)	137,8	189,2	147,8	197,2	378,3	458,9
Tasa de actividad económica (%)	69,8	85,6	62,7	79,8	67,9	82,5
Ocupadas/os (miles de personas)	134,5	185,5	142,9	191,6	366,8	446,3
Tasa de informalidad (%)	87,1	83,3	75,2	69,1	83,8	77,7
Relación empleo-población (%)	68,0	83,7	60,5	77,6	66,0	80,3
Ingreso promedio real mensual proveniente del trabajo (soles)	1055,9	1449,1	1333,1	1997,0	1140,3	1655,2

En la Tabla 8, se presenta los promedios de los indicadores de género para el tema de violencia de género para cada uno los tres clústers. Se observa que los mayores porcentajes en promedio, están en el clúster 1 para los indicadores violencia física contra la mujer ejercida alguna vez por parte del esposo o compañero y Violencia sexual contra la mujer ejercida alguna vez por parte del esposo o compañero, en el clúster 2 para los indicadores de la violencia psicológica o verbal contra la mujer ejercida alguna vez por parte del esposo o compañero y mujeres en edad fértil que sufrieron violencia física y/o sexual por parte del esposo; siendo para el indicador denuncias de violencia familiar con mayor número de casos registrados.

5. Conclusiones

El método no jerárquico k-medias fue más eficiente que el jerárquico para agrupar por los indicadores de género en tres clústers las 24 regiones del Perú. Para el tema salud y vida, los indicadores favorecen más a las

mujeres que a los hombres en los tres clústers, siendo las regiones del clúster 1 con mayor esperanza de vida en las mujeres y hombres. En el tema de educación, las mujeres tienen tasas menores en el nivel secundario y mayor tasa de analfabetismo en los tres clúster, siendo las regiones del clúster 3 el que muestra las mayores desigualdades. Por lo que urge realizar acciones para una equidad de género en educación. En el tema de trabajo y empleo, los indicadores desfavorecen a la mujer en los tres clúster y con mayor incidencia en el ingreso promedio mensual, siendo las regiones del clúster 2 los que muestran las mayores desigualdades de género y por lo cual se debe tomar acciones para una mejor equidad de género. Con respecto al tema de violencia de género, las regiones del clúster 1 muestran indicadores mayores en la violencia física y sexual contra las mujeres, el clúster 2 para la violencia psicológica, mujeres fértiles física y sexual y con los lo mayores casos registrados de denuncias de violencia familiar.

Tabla 8. Promedio de los indicadores de violencia de género por clústers

Indicadores de género	Clúster 1	Clúster 2	Clúster 3
Violencia física contra la mujer ejercida alguna vez por parte del esposo o compañero (%)	36,7	35,4	27,4
Violencia psicológica o verbal contra la mujer ejercida alguna vez por parte del esposo o compañero (%)	56,9	59,0	47,7
Violencia sexual contra la mujer ejercida alguna vez por parte del esposo o compañero (%)	9,5	8,8	5,7
Denuncias de violencia familiar (Casos registrados)	4056,6	19082,0	6778,3
Mujeres en edad fértil que sufrieron violencia física y/o sexual por parte del esposo o compañero en los últimos 12 meses (%)	12,0	13,3	7,8

Conflictos de intereses

Los autores firmantes del presente trabajo de investigación declaran no tener ningún potencial conflicto de interés personal o económico con otras personas u organizaciones que puedan influir indebidamente con el presente manuscrito.

Contribuciones de los autores

Preparación y ejecución; Desarrollo de la metodología; Concepción y diseño; Edición del artículo; Supervisión del estudio: MC, JC.

6. Referencias bibliográficas

- Araujo, A. (2015). La desigualdad salarial de género medida por regresión cuantílica: el impacto del capital humano, cultural y social. *Revista Mexicana de Ciencias Políticas y Sociales*. Nueva Época, Año LX(223), 287-316.
- Chopra, S., Mirsafian, M., Khan, A., & Golb, L. (2019). *Gender Differences in Science and Engineering: A Data Mining Approach*. Published in the Workshop Proceedings of the EDBT/ICDT 2019 Joint Conference. Portugal.

- Eurostat. (2018). Further information on gender gaps in each of these indicators.
- Gálvez, L., Rodríguez, P., & Domóñez, M. (2011). Use of work and time by gender: a new clustering of European welfare systems. *Economía feminista*, 17 (4).
- Guzmán, J. (2017). Los indicadores de género. La ruta hacia la igualdad. *Revista internacional de ciencias sociales y humanidades SOCIOTAM*. Vol. XXVII(2), 133-147.
- Heinz, A., Catunda, C., Van Duin, C., Torsheim, T. & Willems, H. (2020). Patterns of Health-Related Gender Inequalities: A Cluster Analysis of 45 Countries. *Journal of Adolescent Health* 66.
- Instituto Nacional de Estadística e Informática-NEI. Perú (2017). Perú brechas de género 2017. Avances hacia la igualdad de mujeres y hombres.
- Milosavljevic, V. (2007). Estadísticas para la equidad de género: Magnitudes y tendencias en América Latina. CEPAL. Fondo de desarrollo de las mujeres de las Naciones Unidas para la mujer.
- Ministerio de Educación, P. (2020). Brechas de género y generación.
- Organización Internacional del Trabajo. (2019). Informe Mundial sobre Salarios 2018 / 19.
- Parra Pera, A. (2018). Análisis multivariado para los departamentos de Colombia en el año 2018 según los criterios de mortalidad de la lista 6/67 y su relación con determinantes económicos. pp. 1-31.
- Pezer, M. (2018). Maternity support policies: a cluster analysis of 22 European Union countries. *Proceeding of the ISCCRO - International Statistical Conference in Croatia*.
- S., E., Harryson, L., Bolin, M., & Hammarstrom, A. (2013). Patterns of Gender Equality at Workplaces and Psychological Distress. pp. 1-10. higher education: A framework study of good practice. *International Review of Education*, 65(4), 605-632. <https://doi.org/10.1007/s11159-019-09792-3>