

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA
ESCUELA DE POSGRADO
MAESTRÍA EN ESTADÍSTICA APLICADA**



**"CLASIFICACIÓN DE LA EFICIENCIA DEL GASTO
PÚBLICO EN LAS REGIONES DEL PERÚ APLICANDO
CONGLOMERADOS DE SERIES TEMPORALES, 2007 -
2019"**

Presentada por:

ITALO BRAYAN ROMERO CUADROS

**TESIS PARA OPTAR EL GRADO DE MAESTRO
MAGISTER SCIENTIAE EN ESTADÍSTICA APLICADA**

Lima - Perú

2022

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN ESTADÍSTICA APLICADA**

**“CLASIFICACIÓN DE LA EFICIENCIA DEL GASTO
PÚBLICO EN LAS REGIONES DEL PERÚ APLICANDO
CONGLOMERADOS DE SERIES TEMPORALES, 2007 - 2019”**

**TESIS PARA OPTAR EL GRADO DE MAESTRO
MAGISTER SCIENTIAE**

Presentada por:

ITALO BRAYAN ROMERO CUADROS

Sustentada y aprobada ante el siguiente jurado:

Dr. César Menacho Chiok
PRESIDENTE

Mg. Jesús Gamboa Unsihuay
ASESOR

Dr. Rino Sotomayor Ruiz
MIEMBRO

M.Sc. Grimaldo Febres Huamán
MIEMBRO

DEDICATORIA

A Dios, por darme una segunda oportunidad de vida.

A mis padres Alejandro Romero y Sonia Cuadros,

mi hermana Dana Romero (que en paz descanse),

mi hermano David Romero,

mi novia Ibi Quispe,

quienes han depositado su confianza,

brindándome su apoyo constante,

quienes son mi motivo para seguir adelante,

este logro es de ustedes.

AGRADECIMIENTOS

Quiero estar agradecido con Dios, por darme un segundo aliento de vida y que dio su parte para el desarrollo de esta tesis y a las personas que me ayudaron a concluirla, en especial a las siguientes:

- Jesús Gamboa, mi asesor de tesis, por sus orientaciones, dedicación, revisión exhaustiva, motivación y determinación para sacar esta investigación a flote.
- Cesar Menacho, Grimaldo Febres y Rino Sotomayor, miembros distinguidos del jurado, por su colaboración en la finalización de la tesis.
- William Postigo, a pesar de compartir poco tiempo me brindo consejos y enfoque muy importante para mí desarrollo e impulso para sacar adelante la tesis de posgrado.
- Juan Pichihua, por mostrarme como debe ser un profesional desde lo académico, profesional y humano, todos sus consejos me sirvieron para concluir la tesis.
- Ibi Quispe, por brindarme toda la confianza, aliento, colaboración y cariño para sacar adelante esta investigación.
- Paco Cuadros, por darme la reflexión y guía para mantener la vida con una perspectiva honrada y justa.
- Promoción “TIKVA – 2009” por brindarme su apoyo de forma inesperada en un momento muy delicado que me hizo creer nuevamente en las personas.
- Edgar Olivera, por darme su amistad por tantos años y ser un ejemplo para mi persona en el aspecto académico y humano.

ÍNDICE GENERAL

RESUMEN	i
ABSTRACT	ii
I. INTRODUCCIÓN	1
II. REVISIÓN DE LITERATURA	5
2.1. ESTUDIOS DE AGRUPACIÓN DE SERIES DE TIEMPO	5
2.2. FRONTERA DE POSIBILIDADES DE PRODUCCIÓN	6
2.3. BENCHMARKING	7
2.4. MÉTODOS PARA MEDIR LA EFICIENCIA DEL GASTO PÚBLICO	7
2.4.1. Análisis coste – beneficio	7
2.4.2. Técnicas no paramétricas	8
2.5. ANÁLISIS DE CONGLOMERADOS	10
2.5.1. Medidas de distancia	10
2.5.2. Agrupación jerárquica	10
2.6. MODELOS DE SERIES DE TIEMPO	12
2.6.1 Prueba de cox - stuart	13
2.6.2 Prueba de dickey- fuller	14
2.7 CONGLOMERADOS DE SERIES DE TIEMPO	14
2.7.1 Conglomerado jerárquico de series de tiempo	15
2.7.2 Prototipos de series de tiempo	18
2.7.3 Criterios de comparación	20
III. MATERIALES Y MÉTODOS	22
3.1. MATERIALES	22

3.2.	DESCRIPCIÓN DEL CASO	22
3.2.1	Población	22
3.2.2	Muestra	22
3.2.3	Identificación de Variables	22
3.2.4	Tipo de Investigación	23
3.2.5	Diseño de la Investigación	23
3.2.6	Formulación de Hipótesis	23
3.3.	METODOLOGÍA APLICADA	23
3.4.	CONSTRUCCIÓN DE DATOS	24
IV.	RESULTADOS Y DISCUSIÓN	27
4.1	ANÁLISIS EXPLORATORIO	27
4.1.1	Análisis de tendencia y estacionariedad	27
4.2	EVALUACIÓN DE CLÚSTERS	39
4.2.1	Índice de la silueta	39
4.2.2	Índice de calinski y harabasz	40
4.2.3	Perfil de los conglomerados óptimos	41
V.	CONCLUSIONES	49
VI.	REFERENCIAS BIBLIOGRÁFICAS	50
VII.	ANEXOS	53

ÍNDICE DE CUADROS

Cuadro 1: Clasificación del índice de la silueta	21
Cuadro 2: P-value de las pruebas estadísticas de insumo de educación	28
Cuadro 3: P-value de las pruebas del índice resultados de educación	30
Cuadro 4: P-value de las pruebas estadísticas de insumos de salud	32
Cuadro 5: P-value de las pruebas estadísticas del índice resultados de salud	34
Cuadro 6: P-value de las pruebas estadísticas del índice insumos general	36
Cuadro 7: P-value de las pruebas estadísticas del índice resultados general	38
Cuadro 8: Comparación de los métodos de distancias y función de enlace	40
Cuadro 9: Comparación de los métodos de distancias y función de enlace	41
Cuadro 10: Características de los prototipos	42

ÍNDICE DE FIGURAS

Figura 1: Tasa de crecimiento del PBI, periodo 2007-2019	1
Figura 2: Enfoque de la agrupación de series de tiempo	15
Figura 3: Descripción gráfica del DTW	17
Figura 4: Representación visual de la ruta óptima encontrada	18
Figura 5: Frontera de posibilidades de producción	25
Figura 6: Visualización gráfica del índice de insumos educación	27
Figura 7: Visualización gráfica del índice de resultados de educación	29
Figura 8: Visualización gráfica del índice de insumos de salud	31
Figura 9: Visualización gráfica del índice de resultados de salud	33
Figura 10: Visualización gráfica del índice de insumos general	35
Figura 11: Visualización gráfica del índice resultados general	37
Figura 12: Departamentos por conglomerados por índice de eficiencia insumos educación	43
Figura 13: Departamentos por conglomerados por índice de eficiencia resultados educación	44
Figura 14: Departamentos por conglomerados por índice de eficiencia insumos salud	45
Figura 15: Departamentos por conglomerados por índice de eficiencia resultados salud	46
Figura 16: Departamentos por conglomerados por índice de eficiencia insumos general	47
Figura 17: Departamentos por conglomerados por índice de eficiencia resultados general	48

ÍNDICE DE ANEXOS

Anexo 1: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia insumos educación	53
Anexo 2: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia resultados educación	53
Anexo 3: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia insumos salud	53
Anexo 4: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia resultados salud	54
Anexo 5: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia insumo general	54
Anexo 6: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia resultado general	54
Anexo 7: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia insumos educación	55
Anexo 8: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia resultados educación	55
Anexo 9: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia insumos salud	55
Anexo 10: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia resultados salud	56
Anexo 11: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia insumo general	56
Anexo 12: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia resultado general	56
Anexo 13: Gasto público por PBI en las regiones para el periodo 2007 – 2019	57
Anexo 14: Índice de calidad de la infraestructura en las regiones para el periodo 2007 – 2019	58
Anexo 15: Índice de salud en las regiones para el periodo 2007 – 2019	59
Anexo 16: Índice de educación en las regiones para el periodo 2007 – 2019	60
Anexo 17: Índice de distribución en las regiones para el periodo 2007 – 2019	61
Anexo 18: Índice de estabilidad económica en las regiones para el periodo 2007 – 2019	62
Anexo 19: Índice de performance económico en las regiones para el periodo 2007 – 2019	63
Anexo 20: Índice de desempeño general en las regiones para el periodo 2007 – 2019	64

Anexo 21: Índice de eficiencia insumos salud por regiones en el periodo 2007 – 2019	65
Anexo 22: Índice de eficiencia resultados salud por regiones en el periodo 2007 – 2019	66
Anexo 23: Índice de eficiencia insumos educación por regiones en el periodo 2007 – 2019	67
Anexo 24: Índice de eficiencia resultados educación por regiones en el periodo 2007 - 2019	68
Anexo 25: Índice de eficiencia insumos general por regiones en el periodo 2007 – 2019	69
Anexo 26: Índice de eficiencia resultados general por regiones en el periodo 2007 – 2019	70
Anexo 27: Funciones y procedimientos en R	71

RESUMEN

Esta investigación tiene como objetivo de determinar los patrones de clasificación del desempeño del sector público mediante la técnica de análisis conglomerados de series de tiempo a las regiones del Perú para el periodo 2007 – 2019. El desarrollo se realizó con los indicadores creados bajo la metodología de la frontera de posibilidades de producción y se evaluó la asociación a través del tiempo con el método de distancia de deformación del tiempo. El indicador utilizado para decidir el número de los conglomerados fue el de Silueta y Calinski. Entre los resultados más importantes se encontró que se mantiene un patrón que predomina entre las regiones que mejor usan los recursos y obtienen resultados idóneos donde resalta Moquegua, Ica y Lima mientras que otro grupo que mantiene fuertes tendencias a ser ineficientes en comparación como Ayacucho y Huancavelica. Finalmente, se observó una predominancia en las regiones que subdivide entre dos conglomerados y se mantiene tanto sectorial y global.

Palabras claves: Análisis de conglomerados, Series de tiempo, indicadores socioeconómicos, sector público, eficiencia.

ABSTRACT

The objective of this research is to determine the classification patterns of the performance of the public sector through the technique of conglomerate analysis of time series to the regions of Peru for the period 2007 - 2019. The development was carried out with the indicators created under the methodology of the production possibility frontier and the association was evaluated over time with the time warp distance method. The indicator used to decide the number of clusters was that of Silueta and Calinski. Among the most important results, it was found that a pattern is maintained that prevails among the regions that best use resources and obtain ideal results, where Moquegua, Ica and Lima stand out, while another group that maintains strong tendencies to be inefficient compared to Ayacucho and Huancavelica. Finally, a predominance was observed in the regions that are subdivided into two conglomerates and remain both sectoral and global.

Keywords: Cluster analysis, Time Series, Socioeconomic indicators, public sector, efficiency.

I. INTRODUCCIÓN

En los últimos años el papel del Estado se ha desplazado hacia evaluaciones empíricas de la eficiencia y utilidad de las actividades del sector público. La eficiencia del gasto público es un aspecto esencial de la política fiscal, y un elemento indispensable para alcanzar los objetivos de desarrollo económico, social e institucional de los países, tales como la aceleración del crecimiento, la reducción de la pobreza y el fortalecimiento de la legitimidad del Estado (Machado 2006).

A nivel macroeconómico, la economía peruana presenta buenos resultados para el periodo 2007 - 2019, tanto en nivel absoluto como en comparación con otras economías de la región y del mundo como se muestra en la Figura 1, pero los referidos al bienestar de los hogares de ingresos medios o bajos muestran mejoras menores y en algunos casos incluso estancamiento. Los resultados son menos alentadores en términos de productividad, empleo y distribución (del ingreso, de accesos y fundamentalmente de oportunidades) y determinan diferencias en la calidad de vida y en las propias posibilidades de desarrollo futuro de la población.

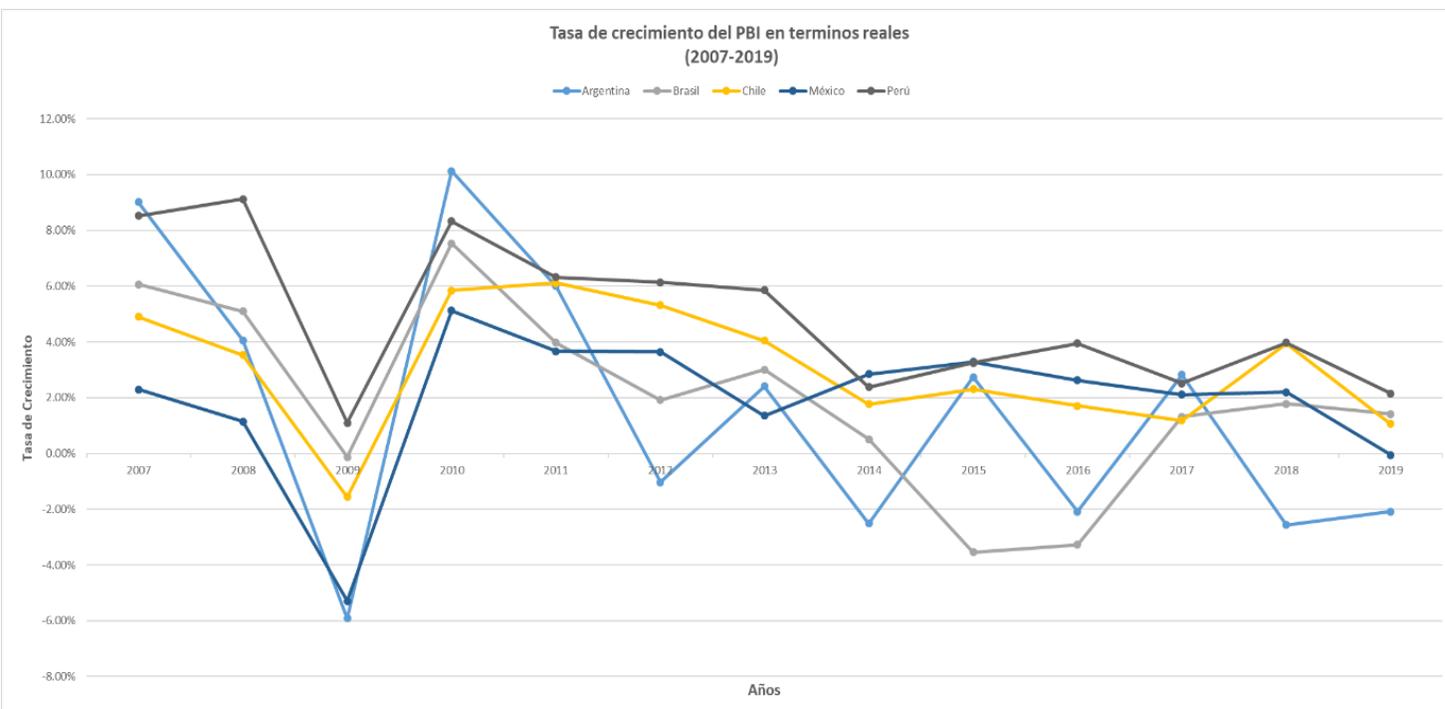


Figura 1: Tasa de Crecimiento del PBI, periodo 2007-2019

Fuente: CEPAL (Comisión Económica para América Latina y el Caribe, 2019)

El desempeño desigual puede ser confirmado por la persistencia de grandes deficiencias en la prestación de los servicios públicos las cuales terminan incidiendo negativamente en la percepción ciudadana sobre la gestión pública y el desempeño del Estado en el Perú, de las cuales se puede mencionar la ausencia de un sistema eficiente de planeamiento y problemas de articulación con el sistema de presupuesto público debido a objetivos desvinculados de las demandas de la población. Además, la deficiencia en el diseño de la estructura de organización y funciones con inadecuados procesos de producción de bienes y servicios públicos, infraestructura, equipamiento y gestión logística insuficientes agudiza la situación del sector público.

Otro aspecto resaltante es la limitada evaluación de resultados e impactos, así como de seguimiento y monitoreo de los insumos, procesos, productos y resultados de proyectos y actividades. Se debe considerar que la dirección de la política inadecuada, gestión de recursos humanos, carencia de sistemas, métodos de gestión de la información y la débil articulación intergubernamental e intersectorial son las posibles causas que el desempeño del sector público no sea equitativo en todas las regiones del país (Secretaría de Gestión Pública 2016).

Diversos autores, entre los que se encuentran Mueller (1997), Persson y Tabellini (2001), Shleifer y Vishny (1998), Strauch y Von Hagen (2000), Tanzi y Schuknecht (1997, 2000), Rodrik (2000), Gwartney *et al.* (2002), Galvis – Aponte (2015), Fontalvo *et al.* (2020) han estado investigando los efectos de estabilización, asignación y distribución del gasto público. No obstante, la mayoría de estos estudios concluyen que el gasto público podría ser mucho menor y más eficiente que en la actualidad. Sin embargo, para que esto suceda, los gobiernos deberían adoptar mejores instituciones y deberían transferir muchas actividades complementarias al sector privado. Además, la medición del desempeño del sector público, definida como el resultado de las actividades del sector público y la eficiencia, como el resultado relativo a los recursos empleados, son todavía limitados (Afonso *et al.* 2003).

Por lo mencionado anteriormente, se deriva que es de interés no solo conocer la oferta de servicios públicos de la que pueden disponer, sino también la relación que existe entre ésta y el gasto público que genera la producción de la misma y que este gasto sea sostenible en el tiempo, de tal forma que el sector público no incurra en gastos innecesarios, que ahorre cuando sea posible y que, en definitiva, este agregado macroeconómico se debe efectuar de un modo racional, evitando la mala gestión de los mismos (Rueda 2011).

Diversos estudios han realizado el análisis de la eficiencia mediante diversas técnicas de benchmarking, frontera de posibilidades de producción empleando la técnica de *Free Disposal Hull* (FDH) que a su vez permitirá la creación de índices de eficiencia insumos y resultados para las regiones del país a fin de evaluar cómo se emplean el gasto público y los resultados obtenidos.

Para la presente investigación, es de interés conocer la clasificación de los índices de eficiencia tanto por insumos y resultados general y sectorial en las regiones a partir del periodo del 2007 - 2019 a través de la herramienta *clustering* de series de tiempo a fin de poder de responder esta interrogante: ¿Es posible determinar algún patrón de segmentación del desempeño en el sector público de las regiones mediante la agrupación de indicadores?

Justificación

La investigación parte de la importancia de generar agrupaciones de indicadores longitudinales para cada región del país a fin de comprender los avances en el desarrollo económico y social de las regiones del Perú. Sin embargo, en el Perú no se ha constatado una fuente predeterminada que cumpla con el diseño de indicadores y genere un perfil de las regiones a partir de estos.

Por ello, esta investigación buscar identificar un patrón de clasificación del desempeño de las regiones del sector público mediante una técnica de agrupación de series de tiempo a fin de comprender las diferentes dinámicas locales y generar un estudio comparativo entre las regiones que permite identificar el nivel de competitividad y posibles factores que la determinan.

De esta manera, es posible contar con un panorama claro y detallado por región, el cual resulta fundamental para la discusión y la toma de decisiones de políticas públicas que impulsen el desarrollo a nivel regional. Como referencia se puede señalar al índice de competitividad regional, que elabora el Instituto Peruano de Economía de manera anual, cuyo objetivo es analizar los avances en el desarrollo económico y social de las regiones del Perú. Por otro lado, es de importancia porque aporta a la línea de investigación hacia el uso de herramientas de agrupaciones de series temporales, que se basa en temas de sector público, gerencia social y administración pública.

Finalmente, se puede indicar que esta investigación es viable en términos financieros porque no habrá la necesidad de hacer encuestas ya que la información se encuentra disponible en las plataformas de las instituciones públicas como el INEI y MEF.

Objetivo general:

Identificar los patrones de clasificación del desempeño del sector público en las regiones del Perú para el periodo del 2007 – 2019 mediante la agrupación temporal.

Objetivos específicos:

- Desarrollar a partir del esquema de datos, un análisis exploratorio de los índices de eficiencia insumo y resultado.
- Determinar la arquitectura de los conglomerados que indiquen la forma de gestión de las regiones tanto por insumos y resultados.
- Describir el perfil de los conglomerados obtenidos.

II. REVISIÓN DE LITERATURA

2.1. ESTUDIOS DE AGRUPACIÓN DE SERIES DE TIEMPO

Respecto al uso de la técnica de agrupación de Series de Tiempo existen diversos estudios entre los cuales encontramos los siguientes:

Cano *et al.* (2019) realizó un análisis estadístico para la administración del inventario en empresas comerciales aplicando el método de conglomeración jerárquico a las series de datos y clasificar tanto los artículos como puntos de venta, analizando el comportamiento de las ventas mediante los modelos SARIMA. El estudio considera primero extraer una serie modelo dentro de cada clúster para luego proyectar la variable de interés. Se procedió a trabajar 2187 productos clasificados en ocho grupos para proyectar las ventas a fin de mejorar los sistemas de control interno y gestión de inventarios capturando el comportamiento estacional en las ventas durante los 24 meses lo cual permite establecer un modelo de series de tiempo.

Solorzano (2016) identificó patrones climáticos en la ciudad de Manizales por estaciones meteorológicas a partir de datos hidrometeorológicos para el periodo 2009 -2011 como la temperatura, radiación solar, humedad y precipitación empleando el algoritmo de conglomerados jerárquicos aglomerativos, donde cada variable muestra patrones por estaciones. Este análisis ayudó a simplificar el manejo y mejorar el conocimiento de las variables de interés, se logró obtener de ellos información caracterizadora del clima de la ciudad siendo efectivo para la identificación de patrones climáticos. A partir de las correlaciones identificadas y patrones se puede plantear un modelo climático para Manizales, permitiendo entre otras cosas hacer regresiones que mejoren la predicción del clima y gestión de riesgos asociados a este evento.

Vasilios *et al.* (2020) utilizaron la agrupación de países con respecto a casos activos por población y casos activos por área y por población en base de datos epidemiológicos respecto al desarrollo del COVID - 19 a fin de distinguir objetivamente países con diferentes resultados y diseminación de esta enfermedad. Se procedió aplicar la técnica de análisis jerárquico temporal con enlace completo y con el método de distancia euclidiana por días a los países y se encontró diversos comportamientos dependiendo del punto de partida del

análisis. Considerando solo los casos activos se observó que China formaba un solo grupo al igual que Irán e Italia, un cuarto grupo que tiene a Israel, España y Corea del Sur y el resto de los países en otro grupo, por lo que se procedió a quitar estos países y se generó tres clústeres, mientras que al considerar la población del país, la situación cambia drásticamente al igual que considerar la superficie terrestre del cual al final se concluye que se puede observar mediante este último índice el impacto de la enfermedad que se propaga más fácilmente en áreas densamente pobladas, esto quiere decir que los países con grandes ciudades son más vulnerables al virus.

Wijk y Selow (1999) introdujeron una nueva visión para analizar las series de tiempo a través del clúster jerárquico con la finalidad de identificar patrones y tendencias en múltiples escalas en el tiempo simultáneamente aplicados al consumo de energía con la finalidad de encontrar patrones similares entre los proveedores de energía y gas y de esta manera mejorar la negociación en la tarifa optima en Holanda. Esto permitió generar una segmentación dentro del mercado y orientar el consumo de energía respecto a los tipos de clientes, mediante este estudio se pudo observar la correlación entre los proveedores a fin de establecer un punto de partida para generar un modelo que puede ser estimado por un método de regresión a fin de predecir el consumo de energía. Por último, consideran como una herramienta de análisis general para visualizar series univariados de gran longitud.

Yoshihide *et al.* (1998) emplearon la agrupación jerárquica y particionada para clasificar los terremotos y explosiones mineras a través de un análisis multivariante de series de tiempo. Esta metodología emplea las diferencias de medias por diferencias en la estructura de covarianzas a su vez emplearon las mediciones de Kullback-Leibler y Chernoff permite clasificar y discriminar entre los registros generados por los terremotos y explosiones mineras. Emplearon como criterio el tiempo de duración de estos dos eventos para clasificarlos y establecieron un perfil de los conglomerados que usualmente eran dos grupos de resultados que es el resultado esperado. Sin embargo, se puede emplear disparidades no paramétricas espectrales que en si pueden incluir estimadores autorregresivos para tener una nueva perspectiva de agrupación.

2.2. FRONTERA DE POSIBILIDADES DE PRODUCCIÓN

La frontera de posibilidades de producción (FPP) se define como una muestra de las cantidades máximas de producción que puede obtener una economía, dados sus conocimientos tecnológicos y la cantidad disponible de insumos que se emplean eficientemente.

Actualmente la FPP se emplea en diferentes tipos de análisis para medir eficiencia técnica en los diversos sectores como salud, agricultura, educación, social entre otros, los análisis se realizan con el fin de conocer la productividad de los recursos empleados en cada sector dado, los resultados obtenidos permiten tomar decisiones como incrementar los insumos o mejorar la administración de los recursos a fin de alcanzar un mayor nivel de eficiencia (Nicholson 2008).

En las distintas economías a nivel mundial como el caso de África, Asia, Europa y América Latina están empleando el análisis de la FPP para medir la eficiencia relativa a través de indicadores y su nivel de gasto recurrido con distintas economías de su región para realizar medidas de política económica y alcanzar a la economía más eficiente (Machado, 2006).

2.3. BENCHMARKING

Se denomina *benchmarking* al proceso continuo de medir nuestros productos, servicios y prácticas con la competencia o con compañía líderes, encontrando las mejores prácticas y, posteriormente, ejecutarlas. Aplicado en primera instancia en el sector privado, el benchmarking cuenta con investigaciones dentro del sector público con la finalidad de tener un horizonte hacia donde deben fijar sus objetivos las economías del mundo (Camp 1991).

A inicios de 1990, los primeros países en adaptarse al *benchmarking* público fueron los países de la OECD (Organización para la Cooperación y el Desarrollo Económicos) donde evalúan a sus empleados respecto al sector privado para alcanzar mejores resultados, su respectivo presupuesto frente a otros países y efectividad sociopolítica entre las administraciones regionales. Como se observa, el mencionado análisis se realiza desde una perspectiva interna y externa. Estas experiencias en tópicos de formulación, ejecución y monitoreo de políticas públicas en los países industrializados han sido sistematizadas dando proceso al benchmarking político (Cleary 2006).

2.4. MÉTODOS PARA MEDIR LA EFICIENCIA DEL GASTO PÚBLICO

2.4.1. Análisis coste – beneficio

El método del análisis coste beneficio es un enfoque que permite tomar constantemente decisiones sobre la conveniencia de emprender determinadas inversiones, primero se debe identificar el conjunto de proyectos posibles que debe examinarse, reconocer las consecuencias de estas posibilidades, luego asignarles un valor a cada uno de los factores y de los resultados y por último sumar los costes y los beneficios para hallar la rentabilidad

total del proyecto teniendo en cuenta el costo de oportunidad, es decir el rendimiento de los fondos invertidos en otra parte (Stiglitz 1988).

El gobierno presenta costes y beneficios sociales a diferencia del sector privado, que solo se preocupa por su rentabilidad, mientras que el gobierno debe evaluar varias posibilidades sobre la ejecución de un proyecto además se presenta la dificultad que no puede usar el mecanismo del mercado, debido que no existen precios para esos productos que desean brindar a la sociedad, por lo tanto esta metodología presenta métodos sistemáticos para analizar costes y beneficios cuando los precios de mercado no reflejan los costes y beneficios sociales.

Este modelo emplea el excedente del consumidor para garantizar la viabilidad de un proyecto. Si este excedente más los ingresos son superiores al coste (en otras ocasiones brindan el cociente entre los beneficios y los costes) el gobierno debe decidir sobre el criterio que debe considerar añadiendo los costes y beneficios no monetizados como el tiempo y la vida de las personas, estos son reflejados en los precios sombra que incluyen la imperfección del mercado en donde se encuentra operando el proyecto.

Por último, este análisis incluye una tasa social de descuento que refleja la preferencia temporal de los consumidores (la tasa a la que pueden pedir préstamos los consumidores), o también puede ser el coste de oportunidad del capital (la tasa a la que pueden pedir préstamo los productores) y como a su vez es posible que no refleje los dos puntos de vista que se presenta cuando los proyectos afectan a diferentes generaciones.

2.4.2. Técnicas no paramétricas

Para medir eficiencia de las unidades de producción mediante una frontera de posibilidades de producción se emplean los enfoques paramétricos y no paramétricos. Respecto al método paramétrico emplea modelos econométricos como frontera determinísticas y estocásticas. Para estimar estos modelos se establece a priori la forma funcional de la frontera y permite corregir el error de predicción y otros errores aleatorios no relacionados a los datos de entrada. Sin embargo, presentan algunas deficiencias como el caso de la regresión lineal que puede no ajustarse a los datos observados, el incumplimiento de los supuestos en la ejecución del modelo (Tan 2006).

Las técnicas no paramétricas son metodologías que utilizan la base de datos para crear una frontera de posibilidades de producción mediante la libre determinación de los datos de la muestra o el uso de un proceso de optimización de programación lineal determinístico siendo

los métodos más frecuentes para la medición de la eficiencia en el sector público, esto se debe a que no es necesario plantear supuestos a priori sobre una forma funcional de la frontera de eficiencia, para esto se debe tener información sobre el nivel de producción y los insumos necesarios para alcanzar un determinado nivel de producto. Para este análisis los productos son los índices socioeconómicos y los insumos son el gasto público por regiones dentro de cada área a evaluar y por medio de la frontera se pueda visualizar la eficiencia de cada región analizada. Entre los modelos más destacados se tiene el Análisis Envoltente de Datos (DEA) y el *Free Disposal Hull* (FDH).

Ambos modelos permiten estimar los puntajes de eficiencia mediante la orientación al insumo u orientación al producto, los investigadores que han aplicado sugieren que estas son más adecuadas para el análisis de la eficiencia del sector público. Las razones son diversas como que no requieren la especificación de precios a los que se valoren los costos o beneficios de la producción, trabajan con multiplicidad de productos, no requieren supuestos sobre la tecnología de la producción, la cual afecta a los métodos de técnicas paramétricas (Tan 2006).

La diferencia del DEA Y FDH es que, en este último, la frontera de eficiencia no incluye a las unidades virtuales, aquellas que se forman de la combinación de las unidades de decisión eficientes, el análisis FDH es menos exigente que el DEA, debido a que el DEA requiere de un mayor número de unidades de decisión eficientes.

El único supuesto del FDH es que las entradas y/o salidas pueden disponer libremente de ellos, por lo que es posible con la misma tecnología de producción reducir las salidas manteniendo el mismo nivel de insumos o tener la posibilidad de aumentar los insumos manteniendo el nivel de las salidas. Este supuesto garantiza la existencia de un FDH continua, o frontera de posibilidades de producción, para cualquier muestra de los resultados de producción. Por lo tanto, el análisis FDH proporciona una herramienta intuitiva que se puede utilizar para identificar las mejores prácticas en el gasto público y para evaluar cómo los gobiernos de los países van avanzando en comparación con estas mejores prácticas.

Varios estudios han utilizado el análisis FDH para evaluar la eficiencia del gasto público, donde uno de los estudios de Vanden Eeckhaut *et al.* (1993) han tratado de establecer la eficiencia relativa del gasto de los municipios en Bélgica, en esta investigación se realizó una comparación de los resultados del análisis FDH con los resultados de DEA, y concluyen

que el supuesto de convexidad impuesta por DEA distorsiona los resultados de análisis de la eficiencia.

2.5. ANÁLISIS DE CONGLOMERADOS

El objetivo del análisis de conglomerados es segmentar unidades (objetos) en agrupaciones (conglomerados) de modo que las unidades dentro de cada grupo son similares entre sí con respecto a las variables observadas, y los conglomerados mismos se distinguen entre sí. En general, los métodos de agrupación se clasifican como jerárquicos y no jerárquicos. Respecto al primer tipo de método se puede realizar mediante agrupamiento aglomerativo y divisivo. Estos dos tipos de enfoque organizan los datos en la estructura jerárquica basada en medidas de proximidad adecuadas. La agrupación no jerárquica divide directamente los puntos de datos en un número predeterminado de agrupaciones sin la estructura jerárquica (Maharaj *et al.* 2019).

2.5.1. Medidas de distancia

Sea $X = \{x_{ij}: 1, \dots, I; j = 1, \dots, J\} = \{x_i = (x_{i1}, \dots, x_{ij}, \dots, x_{iJ})': i = 1, \dots, I\}$ la matriz de datos donde x_{ij} representa la j -ésima variable observada en el i -ésimo objeto y x_i representa el vector de la i -ésima observación. La clase de medida de distancia más común utilizada en el análisis de conglomerados es la clase de distancia de Minkowski (Everitt *et al.* 2011):

$${}_r d_{il} = \left[\sum_{j=1}^J |x_{ij} - x_{lj}|^r \right]^{\frac{1}{r}}, r \geq 1 \quad (1)$$

Para $r = 1$, tenemos la distancia de la cuadra de la ciudad (o la distancia de Manhattan):

$${}_1 d_{il} = \sum_{j=1}^J |x_{ij} - x_{lj}| \quad (2)$$

y para $r = 2$, tenemos la distancia euclidiana, probablemente la medida de distancia más utilizada en el análisis de conglomerados:

$${}_2 d_{il} = \left[\sum_{j=1}^J (x_{ij} - x_{lj})^2 \right]^{\frac{1}{2}} \quad (3)$$

2.5.2. Agrupación jerárquica

Este método produce una serie de particiones de los datos: La agrupación aglomerativa empieza señalando una cantidad predeterminada de agrupaciones a partir del conjunto de

observaciones, cada una de las cuales incluye exactamente un punto de datos. Luego se sigue una serie de operaciones de fusión que eventualmente fuerza a todos los objetos al mismo grupo. El agrupamiento aglomerativa general se puede resumir mediante el siguiente procedimiento:

- Comience con l clústers únicos y se procede a calcular la matriz de proximidad.
- En la matriz de distancia, busque la distancia mínima ($d(C_c, C_{c'}) = \min_{1 \leq p, q \leq l} d(C_p, C_q)$), donde $d(.,.)$ es la función de distancia, y combine el clúster C_c y $C_{c'}$ para formar un nuevo clúster $C_{cc'}$.
- Actualice la matriz de distancia calculando las distancias entre el clúster $C_{cc'}$ y los otros clústeres;
- Repetir los pasos 2 y 3 hasta que solo quede un grupo.

La fusión de un par de grupos o la formación de uno nuevo depende de la definición de la función de distancia entre dos grupos. Existe un gran número de definiciones de funciones de distancia entre un clúster C_q y un nuevo clúster $C_{cc'}$ formado por la fusión de dos clústers C_c y $C_{c'}$. A continuación, se muestra brevemente algunos métodos para definir funciones de distancia:

- Método de enlace único (método vecino más cercano): la distancia entre un par de grupos está determinada por las dos unidades más cercanas a los diferentes grupos. La agrupación de enlaces individuales tiende a generar grupos alargados, lo que provoca el efecto de encadenamiento. Como resultado, se pueden conectar dos grupos con propiedades bastante diferentes debido a la existencia de ruido. Sin embargo, si los grupos están separados uno del otro, el método de enlace único funciona bien.
- Método de enlace completo: en contraste con la agrupación de enlaces individuales, el método de enlace completo utiliza la distancia más lejana de un par de objetos para definir la distancia entre grupos.
- Método de vinculación promedio grupal (promedio de método de grupo de pares no ponderado): la distancia entre dos grupos se define como el promedio de las distancias entre todos los pares de puntos de datos, cada uno de los cuales proviene de un grupo diferente.
- Método de vinculación promedio ponderado (promedio ponderado del método de grupo de pares): el enlace promedio también se usa para calcular la distancia entre dos grupos.

La diferencia es que las distancias entre el grupo recién formado y el resto se ponderan en función del número de puntos de datos en cada grupo.

- Método de enlace centroide (centroide de método de grupo de pares no ponderado) dos grupos se fusionan en función de la distancia de sus centroides (medias).
- Método de Ward (método de varianza mínima): el objetivo del método de Ward es minimizar el aumento de la llamada suma de los errores al cuadrado dentro de la clase.

El enlace único, el enlace completo y el enlace promedio consideran todos los puntos de un par de grupos al calcular su distancia entre grupos, y también se les llama métodos de gráficos. Los otros se llaman métodos geométricos porque usan centros geométricos para representar grupos y determinar sus distancias. Los resultados de la agrupación jerárquica generalmente se representan mediante un gráfico particular llamado dendrograma. El nodo raíz del dendrograma representa todo el conjunto de datos, y cada nodo hoja se considera un punto de datos. Los nodos intermedios describen así la medida en que los objetos son proximales entre sí; y la altura del dendrograma generalmente expresa la distancia entre cada par de puntos de datos o grupos, o un punto de datos y un grupo. Los resultados finales de agrupamiento se pueden obtener cortando el dendrograma a diferentes niveles. Esta representación proporcional descripciones muy informativas y una visualización de las posibles estructuras de agrupación de datos, especialmente cuando existen relaciones jerárquicas reales en los datos (Everitt *et al.* 2011).

En particular, con respecto a la elección de la partición óptima (número óptimo de grupos) tenemos diferentes criterios de validez de grupo para métodos jerárquicos y de partición. Para los métodos de agrupamiento jerárquico, la partición óptima se logra seleccionando una de las soluciones en la secuencia anidada de agrupamientos que comprende la jerarquía, equivalente a cortar un dendrograma a una altura particular (a veces denominado el mejor corte). Esto define una partición tal que los grupos por debajo de esa altura están distantes entre sí al menos en esa cantidad, y la apariencia del dendrograma puede sugerir informalmente el número de grupos. Se toman grandes cambios en los niveles de fusión para indicar el mejor corte (Xu y Wunsch 2009).

2.6. MODELOS DE SERIES DE TIEMPO

Una serie de tiempo es un conjunto de observaciones, donde cada una de las cuales se registra en un tiempo específico “t”, mediante el cual se especifica un modelo a partir de datos

observados $\{x_t\}$ es una especificación de las distribuciones conjuntas de una secuencia de variables aleatorias $\{X_t\}$.

Para generar un modelo de series de tiempo es necesario evaluar la estacionariedad y tendencia. Una serie es estacionario si para todo el periodo de la serie (t_1, t_2, \dots, t_m) y todo valor entero del vector de variables $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$ tiene la misma probabilidad conjunta que el vector que cumple con esa distribución $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k})$, sin embargo esta forma de estacionariedad es exigente motivo por el cual se emplea la estacionariedad débil, que se produce cuando todos los momentos de primer (esperanza matemática) y segundo (varianza) orden del proceso estocástico son invariantes en el tiempo (Brockwell y Davis 2016).

Respecto a la tendencia es el cambio que existe en el largo plazo en el nivel para el primer momento. Esta condición viola el comportamiento de una serie estacionaria que es requerido para estimar un modelo de series de tiempo.

2.6.1 Prueba de cox - stuart

Esta prueba apareja números de límites o posiciones opuestas, y realiza una prueba de comparación de signos. En caso de presentar una tendencia en la muestra tomada, un miembro de cada par tendría una mayor probabilidad a ser mayor que el otro miembro.

Para realizar esta prueba tiene como supuestos que el vector de variables $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$ tomadas al azar, son mutuamente independientes y la medida de las variables deben tener al menos un nivel ordinal.

Los datos obtenidos se disponen en el orden que son observados. Se procede a agrupar en pares a las variables según el siguiente patrón: $(y_{t_1}, y_{t_1+c}), (y_{t_2}, y_{t_2+c}), \dots, (y_{t_{m-c}}, y_{t_m})$ donde m = número total de datos, y $c = m/2$, si m es par, o $c = (m+1) / 2$ si m es impar (en caso de ser m impar, la variable que caiga en la mitad de la prueba es eliminada).

Se procede a ordenar y aparear los datos, y se asignara a cada par un valor de “+” si $y_{t_i} < y_{t_{i+c}}$; y si $y_{t_i} > y_{t_{i+c}}$ se le asignará al par el valor “-“, eliminándose los empates. Si el propósito es resolver una situación en la cual la hipótesis nula es que no exista tendencia, para que no haya tendencias la probabilidad de que haya un valor “+” debe ser igual al de que haya uno “-“.

Así, si llamamos a “p” la probabilidad de que exista un valor “+”, la hipótesis puede presentarse de la siguiente manera:

$$H_0: \rho = 1/2 \text{ (No existen tendencias)}$$

$$H_1: \rho \neq 1/2 \text{ (Existe tendencia, sea positiva o negativa)}$$

2.6.2 Prueba de dickey- fuller

Para decidir si un proceso es o no estacionario se puede realizar un contraste de raíces unitarias. Estos contrastes indican si debemos tomar una diferencia adicional en una serie para convertirla en estacionaria. El test de Dickey-Fuller, mide la presencia de raíz unitaria mediante el siguiente modelo:

$$Y_t = \rho Y_{t-1} + \epsilon_t \quad (4)$$

H_0 : Existe raíz unitaria en la serie o $\rho = 1$ (serie no es estacionaria)

H_1 : No existe raíz unitaria en la serie o $\rho \neq 1$ (serie es estacionaria)

La fórmula del estadístico, la cual se aproxima a la distribución *t-student* es la siguiente:

$$t = \frac{\rho-1}{sd(\rho)} \sim t \left(n - 1; 1 - \frac{\alpha}{2} \right) \quad (5)$$

Cuanto mayor sea el valor del *t-student*, más probable es que la hipótesis alterna sea la correcta. De la misma forma, cuanto más se aproxime a cero el valor del *t-student* calculado se ajustará al distribuido (Peña 2010).

2.7 CONGLOMERADOS DE SERIES DE TIEMPO

Esta técnica al igual que la agrupación de datos estáticos, requiere un algoritmo o procedimiento de agrupación para formar agrupaciones dado un conjunto de objetos de datos no etiquetados y la elección del algoritmo de agrupación depende tanto del tipo de datos disponibles como del propósito particular y la aplicación.

Existen tres enfoques para generar esta técnica: basado en datos sin procesar, basado en características y basado en modelos como se observa en la Figura 2. El primer enfoque generalmente funciona directamente con datos de series de tiempo que no requieren un proceso de transformación, modelamiento y/o parametrización respecto a los datos y su uso es cambiando la medida distancia/ similitud que se le conoce como el método de datos sin procesar (a).

Los otros dos enfoques convierten los datos brutos de una serie temporal en un vector de características de menor dimensión o en una serie de parámetros del modelo, y luego aplica

un algoritmo de agrupación convencional a los vectores de características extraídos o parámetros del modelo, lo que se denomina enfoque basado en características (b) y el enfoque basado en modelos que emplea los parámetros obtenidos (c) para la presente investigación se usó el método de datos sin procesar.

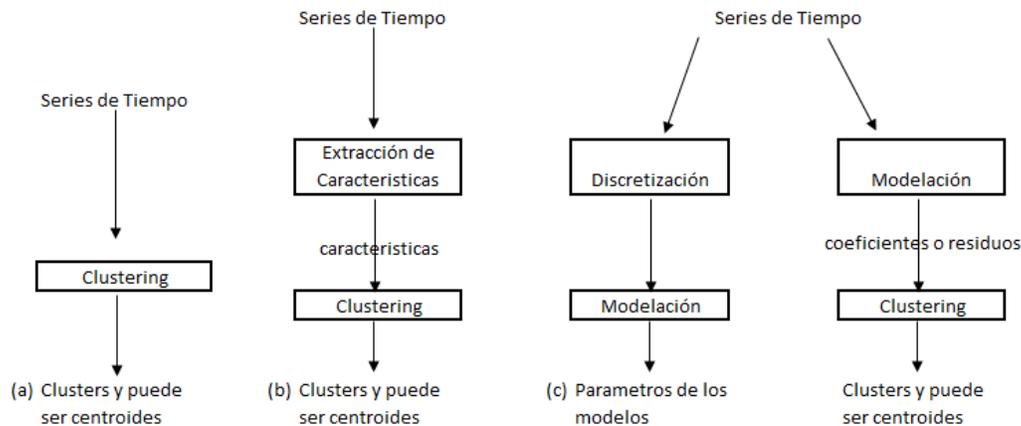


Figura 2: Enfoque de la agrupación de series de tiempo

Fuente: Warren (2005)

2.7.1 Conglomerado jerárquico de series de tiempo

Este método funciona agrupando series de tiempo en un árbol de agrupaciones. El método de agrupamiento jerárquico aglomerativo es más popular que el método divisivo (Sarda 2019). Los criterios de elección para formar los *clusters* que comúnmente se emplean son los siguientes:

El algoritmo de enlace único (completo) mide la similitud entre dos grupos como la similitud del par más cercano (más alejado) de puntos de datos que pertenecen a diferentes grupos, combina los dos grupos que tienen la distancia mínima, repite el proceso de fusión hasta que todos los objetos se fusionen para formar un clúster.

El algoritmo de varianza mínima de Ward fusiona los dos grupos que darán como resultado el aumento más pequeño en el valor de la varianza de la suma de cuadrados. En cada paso de agrupación, se intentan todas las fusiones posibles de dos agrupaciones. La varianza de la suma de los cuadrados se calcula para cada uno y se selecciona el que tiene el valor más pequeño. Sean C_i y C_j dos clústers con n_i y n_j elementos y medias m_i y m_j . Sea m la media del clúster que se formaría al unir C_i y C_j . En cada paso, la distancia o función objetivo a

minimizar es $I_{C_i C_j}$, que representa el incremento en la suma de cuadrados dentro del clúster cuando se unen los dos clústers.

$$I_{C_i C_j} = \sum_{l \in C_i \cup C_j} \|x_l - \mathbf{m}\|^2 - \left\{ \sum_{l \in C_i} \|x_l - \mathbf{m}_i\|^2 + \sum_{l \in C_j} \|x_l - \mathbf{m}_j\|^2 \right\} = \frac{n_i n_j}{n_i + n_j} \|\mathbf{m}_i - \mathbf{m}_j\|^2 \quad (6)$$

Sean C_i y C_j dos clústers con n_i y n_j elementos. Sean \mathbf{m}_i y \mathbf{m}_j y m vectores con elementos m_{ik}, m_{jk} y m_k , $k = 1, \dots, p$, que representan la media del clúster C_i , C_j y del clúster que se formaría al unir C_i y C_j , respectivamente.

Entonces, sean $x_{lp}^{(i)}$ la componente p -ésima del vector l -ésimo del clúster C_i y $x_{lp}^{(i,j)}$ la componente p -ésima del vector l -ésimo del clúster $C_i \cup C_j$, entonces:

$$\sum_{l \in C_i \cup C_j} \|x_l - \mathbf{m}\|^2 - \left\{ \sum_{l \in C_i} \|x_l - \mathbf{m}_i\|^2 + \sum_{l \in C_j} \|x_l - \mathbf{m}_j\|^2 \right\} = \sum_{l=1}^{n_i+n_j} \sum_{s=1}^p \left(x_{ls}^{(i,j)} - m_s \right)^2 - \left\{ \sum_{l=1}^{n_i} \sum_{s=1}^p \left(x_{ls}^{(i)} - m_{is} \right)^2 + \sum_{l=1}^{n_j} \sum_{s=1}^p \left(x_{ls}^{(j)} - m_{js} \right)^2 \right\} \quad (7)$$

Respecto a los métodos de distancia, se debería incorporar a lo mencionado en el apartado anterior, la deformación dinámica del tiempo (DTW) que es un algoritmo de programación dinámica que compara dos series y trata de encontrar la ruta de deformación óptima entre ellas bajo ciertas restricciones, como la monotonicidad. Comenzó a ser utilizado por la comunidad de minería de datos para superar algunas de las limitaciones asociadas con la distancia euclidiana (Ratanamahatana y Keogh 2004).

En la Figura 3 muestra la alineación entre dos series temporales de muestra x (línea continua negro) e y (línea discontinua roja). En este caso, los puntos iniciales y finales de la serie deben coincidir, pero otros puntos pueden deformarse a tiempo para encontrar mejores coincidencias, donde las líneas azules discontinuas ejemplifican cómo algunos puntos se asignan entre sí, lo que muestra cómo se pueden deformar a tiempo (Berndt y Clifford 1994).

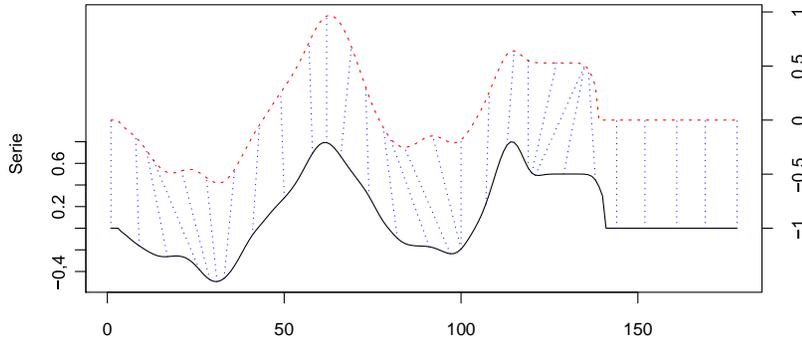


Figura 3: Descripción gráfica del DTW

Fuente: Sardá (2019)

El rendimiento de un método de agrupamiento jerárquico aglomerativa a menudo sufre de su incapacidad para ajustarse una vez que se ha ejecutado una decisión de fusión. Lo mismo es cierto para los métodos de agrupamiento jerárquico divisivo. La agrupación jerárquica no está restringida a series de tiempo de agrupación con la misma longitud. También es aplicable a series de longitud desigual si se usa una medida de distancia apropiada, como la deformación dinámica del tiempo, para calcular la distancia / similitud.

El primer paso en DTW consiste en crear una matriz de costo local (LCM, *Local Cost Matrix*), que tiene dimensiones $n \times m$. Dicha matriz debe crearse para cada par de series comparadas, lo que significa que los requisitos de memoria pueden crecer rápidamente a medida que crece el tamaño del conjunto de datos. Considerando x e y como la serie de entrada, para cada elemento (i, j) del LCM, se debe calcular la norma L_p entre x_i e y_j que denota explícitamente que las series multivariadas son compatibles siempre que tengan el mismo número de variables (tenga en cuenta que para las series univariadas, el LCM será idéntico independientemente de la norma utilizada). Por lo tanto, tiene sentido hablar de una distancia de deformación dinámica del tiempo (DTW), donde p corresponde a la norma L_p que se usó para construir el LCM. Esto se define en la siguiente ecuación:

$$lcm(i, j) = (\sum_v |x_i^v - y_j^v|^p)^{1/p} \quad (8)$$

Sin embargo, esta norma también juega un papel importante en el próximo paso de DTW. En un corto periodo de tiempo, el algoritmo DTW encuentra la ruta que minimiza la alineación entre x e y al pasar iterativamente por el LCM, comenzando en LCM $(1,1)$ y terminando en LCM (n, m) , y agregando el costo. En cada paso, el algoritmo encuentra la dirección en la cual el costo aumenta menos bajo las restricciones elegidas; en la Figura 4 se

observa una representación visual de la ruta correspondiente a la Figura 3. Si definimos $\varphi = \{(1,1), \dots, (n, m)\}$ como el conjunto que contiene todos los puntos que caen en la ruta óptima, entonces la distancia final se calcularía con la siguiente ecuación (9), donde m_φ es un coeficiente de ponderación por paso y M_φ es la constante de normalización correspondiente (Giorgino 2009).

$$DTW_p(x, y) = \left(\sum_v \frac{m_\varphi lcm(k)^p}{M_\varphi} \right)^{1/p}, \forall k \in \varphi \quad (9)$$

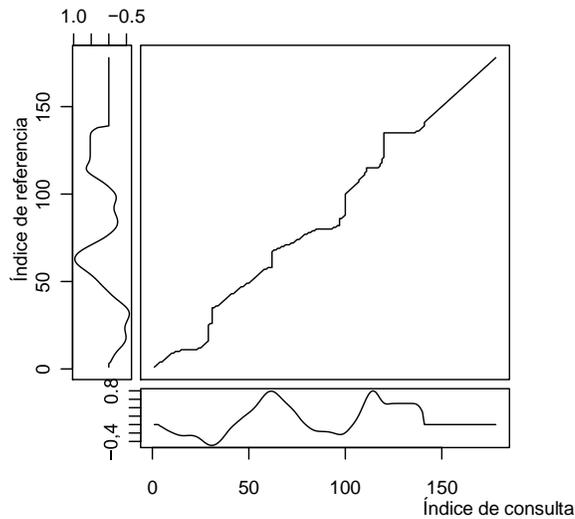


Figura 4: Representación visual de la ruta óptima encontrada.

Fuente: Sardá (2019)

2.7.2 Prototipos de series de tiempo

Un paso muy importante de la agrupación de series temporales tiene que ver con el cálculo de los llamados prototipos de series temporales. Se espera que todas las series dentro de un grupo sean similares entre sí, y uno puede estar interesado en tratar de definir una serie temporal que resuma efectivamente las características más importantes de todas las series en un grupo dado. Esta serie a veces se conoce como una serie promedio, y los prototipos también se denominan promedios o centroides de series temporales.

La computación de prototipos se realiza comúnmente como una subrutina necesaria para obtener la representatividad por cada conglomerado. La creación de prototipos podría incluso ser un paso de modelamiento, por el cual se pueden resumir diferentes muestras de la misma fuente antes de la agrupación (por ejemplo, para el conjunto de datos de trayectorias de características, se pueden resumir todas las trayectorias de la misma característica y luego

se pueden buscar grupos similares), por lo tanto, reduciendo la cantidad de datos y el tiempo de ejecución (Sardá-Espinoza 2019).

Otro ejemplo es la clasificación de series de tiempo basada en los vecinos más cercanos, que puede optimizarse considerando solo los prototipos grupales como vecinos en lugar de la unión de todos los grupos. Sin embargo, es importante tener en cuenta que la distancia utilizada en la tarea general debe ser congruente con el centroide elegido, por ejemplo, utilizando la distancia DTW para prototipos basados en DTW.

La elección de la función de creación de prototipos está estrechamente relacionada con la medida de distancia elegida y, de manera similar, no es simple saber qué tipo de prototipo será mejor a priori. Existen varias estrategias disponibles para la creación de prototipos de series temporales, aunque debido a su alta dimensionalidad, lo que constituye exactamente una serie temporal promedio es discutible, y algunas nociones podrían empeorar significativamente el rendimiento.

La media aritmética se usa muy a menudo junto con la distancia euclidiana, y en muchas aplicaciones esta combinación es muy competitiva, incluso con datos multivariados. Sin embargo, debido a la estructura de las series de tiempo, la media es posiblemente una mala elección de prototipos, e incluso podría perturbar la convergencia de un algoritmo de agrupamiento, a fin de evaluar la eficiencia de esta forma de generar el prototipo se requerirá una evaluación mediante los indicadores de validación del clúster como Calinski y Silueta (Petitjean *et al.* 2011).

Matemáticamente, la media simplemente toma el promedio de cada punto de tiempo i en todas las variables de las series de tiempo consideradas. Para un grupo C de tamaño N , la media de la serie temporal (posiblemente multivariada) μ puede calcularse con la siguiente ecuación (10), donde $x_{c,i}^v$ es el elemento i -ésimo de la variable v -ésima de la serie c -ésima grupo C .

$$\mu_i^v = \frac{1}{N} \sum x_{c,i}^v, \forall c \in C \quad (10)$$

Otro enfoque muy común, usar la partición alrededor de medoides (PAM). Un medoide es simplemente un objeto representativo de un grupo, en este caso también una serie de tiempo, cuya distancia promedio a todos los demás objetos en el mismo grupo es mínima. Dado que el medoide es siempre un elemento de los datos originales, a veces se prefiere PAM sobre la media o la mediana para que la estructura de la serie de tiempo no se altere.

Otra posible ventaja de PAM es que, dado que los datos no cambian, es posible calcular previamente toda la matriz de distancia una vez y reutilizarla en cada iteración, e incluso en diferentes números de grupos y repeticiones aleatorias. Sin embargo, esto no es adecuado para grandes conjuntos de datos, ya que toda la matriz de distancia debe asignarse de una vez. Para generar el proceso se selecciona “k” series de los datos que se eligen aleatoriamente como centroides iniciales. Luego, se calcula la distancia entre todas las series y los centroides (o se recupera de la matriz de distancia completa si se calculó previamente), y cada serie se asigna al grupo de su centroide más cercano. Para cada grupo creado, se calcula la distancia entre todas las series de miembros (si es necesario) y la serie con la suma mínima de distancias se elige como el nuevo centroide. Esto continúa de forma iterativa hasta que ninguna serie cambie los clústeres o se haya superado el número máximo de iteraciones permitidas.

Como alternativa se puede usar la distancia DTW para la creación de prototipos que puede admitir series con diferentes longitudes directamente, con la advertencia de que la longitud del prototipo resultante será la misma que la longitud de la serie de referencia.

2.7.3 Criterios de comparación

Los criterios útiles de validez de grupo para determinar el número de grupos para conglomerados de series de tiempo son los siguientes (Maharaj *et al.* 2019):

Calinski y Harabasz (1987) sugieren tomar el valor de g , el número de grupos, que corresponde al valor máximo de CH_C :

$$CH_C = \frac{\text{trace}(B)}{(C-1)} / \frac{\text{trace}(W)}{(I-1)} \quad (11)$$

Donde B es la matriz de dispersión entre grupos y W es la matriz de dispersión dentro del grupo. Al igual que con todas las técnicas para determinar el número de grupos, la evaluación de este criterio en un número dado de grupos, C , requiere el conocimiento de la membresía del grupo para determinar las matrices B y W . Observe que $T = W + B$, donde T indica la matriz de dispersión total. En general, el número de grupos elegidos depende del método de clúster (y la implementación) utilizado.

El criterio de silueta considera una i -ésima unidad que cumple con lo siguiente: $i \in (1, \dots, I)$ que pertenece al grupo $p \in (1, \dots, C)$. Por ejemplo, mediante un algoritmo de agrupación c -medias, esto significa que la unidad i -ésima está más cerca del centroide de la agrupación p -ésima que de cualquier otro centroide. Supongamos que la distancia promedio (al cuadrado

euclidiana) de la unidad i -ésima a todas las demás unidades pertenecientes al grupo p se denota por a_{ip} . Además, supongamos que la distancia promedio de esta unidad a todas las unidades que pertenecen a otro grupo q , $q \neq p$, se llama d_{iq} . Finalmente, dejemos que b_{ip} sea la distancia mínima calculada sobre $q = 1, \dots, c$, $q \neq p$, que representa la disimilitud de la unidad i -ésima con respecto a su grupo vecino más cercano. Luego, la silueta del i -ésimo objeto se define de la siguiente manera:

$$S_i = \frac{b_{ip} - a_{ip}}{\max\{a_{ip}, b_{ip}\}} \quad (12)$$

Donde el denominador es un término de normalización. Evidentemente, cuanto mayor sea el valor de S_i , mejor será la asignación de la unidad i -ésima al grupo c -ésimo. La silueta definida como el promedio de S_i sobre $i = 1, \dots, I$ es:

$$S_i = \frac{b_{ip} - a_{ip}}{\max\{a_{ip}, b_{ip}\}} \quad (13)$$

La mejor partición se logra cuando la silueta se maximiza, lo que implica minimizar la distancia dentro del grupo (a_{ip}) mientras se maximiza la distancia entre grupos (b_{ip}).

Kaufman y Rousseeuw (1990) señalan que valores cercanos a uno de dicho coeficiente indican una estructura clara; por otro lado, valores muy bajos indican que se debe aplicar métodos alternativos para analizar el conjunto de datos. Se muestra una interpretación según estos autores de los valores de los índices de silueta.

Cuadro 1: Clasificación del índice de la silueta

Índice de la silueta	Interpretación propuesta
0,71 - 1,00	Se ha encontrado una estructura sólida.
0,51 - 0,70	Se ha encontrado una estructura razonable.
0,26 - 0,50	La estructura es débil y posiblemente artificial. Se debe probar métodos adicionales en este conjunto de datos.
$\leq 0,25$	No existe estructura subyacente.

Fuente: Kaufman y Rosseeuw (1987)

III. MATERIALES Y MÉTODOS

3.1. MATERIALES

Para la presente investigación se utilizaron los siguientes materiales:

- Una computadora Intel core i7.
- Ms. Office 2016.
- Dos tintas color negro
- Dos millares de hojas bond A4.
- R, Rstudio versión 3.6.0.
- Impresora multifunción HP LaserJet Pro M15.
- Paquetes de R: dtwclust, cluster, tidyverse, gghighlight, tseries, TSclust, TSdist, Kendall, TSA, aTSA, ggplot2 y trend.

3.2. DESCRIPCIÓN DEL CASO

3.2.1 Población

Para este estudio comprenderán las 24 regiones del Perú que ejecutan el gasto público desde el 2007 hasta 2019.

3.2.2 Muestra

En la investigación no se utilizó una muestra, debido que la aplicación de las técnicas no involucró un proceso de inferencia.

3.2.3 Identificación de Variables

En la presente investigación se tienen grupos de variables para construir el esquema de datos previos para la ejecución de la investigación, mediante estas variables se procederá a calcular los índices de eficiencia insumo y resultados:

- Gasto público.
- Porcentaje de hogares que tienen servicio de internet.
- Promedio de años de estudio alcanzado por la población de 15 y más años.
- Tasa neta de secundaria de 12 a 16 años.
- Esperanza de vida al nacer.
- Tasa de mortalidad infantil por 1000 nacidos vivos.
- Población con al menos una necesidad básica insatisfecha.
- Índice general de precios del consumidor.

- Producto Bruto Interno a precios constantes de 2007.
- Tasa de desempleo de la población.

3.2.4 Tipo de Investigación

Para el presente estudio se realizó una investigación exploratoria cuantitativa de series de tiempo. Es exploratoria porque se aplicó la técnica de conglomerados, agrupando a los índices de insumos y eficiencia mediante la recopilación de datos. Es cuantitativa por la naturaleza de la información de las variables que son de carácter cuantificable y comparable. Es de series de tiempo porque se diseñó la investigación considerando los efectos del cambio del tiempo con la finalidad de identificar agrupaciones de desempeño del sector público en las regiones del Perú en base al uso del insumo (gasto público) y sus resultados a través del tiempo.

3.2.5 Diseño de la Investigación

La clasificación del presente estudio o investigación es cuantitativo no experimental de tipo temporal. Es no experimental porque las variables empleadas para la creación de indicadores no fueron controladas para la investigación. Es de tipo temporal porque los datos se recolectaron en diferentes momentos del tiempo cuya fuente de información es secundaria debido a que se extrajo de diversas fuentes de páginas oficiales.

3.2.6 Formulación de Hipótesis

La hipótesis general del presente trabajo de investigación fue la siguiente:

La segmentación del desempeño del sector público muestra un patrón diferenciado por el uso de los índices insumos y los índices resultados en las regiones del Perú para el periodo del 2007 – 2019.

Las hipótesis secundaria fueron las siguientes:

- El desempeño de los sectores de salud y educación generan la eficiencia relativa de las regiones según su segmentación del desempeño general.
- Los conglomerados se pueden clasificar mediante una relación directa de recursos y resultados formando agrupaciones predeterminados.

3.3. METODOLOGÍA APLICADA

Los pasos que se realizaron para llevar a cabo este trabajo se detallan a continuación:

- Búsqueda y estructuración del conjunto de datos en base de información secundaria.

- Creación de los indicadores empleados como input para el desarrollo de la investigación.
- Análisis descriptivo de las variables utilizando indicadores de tendencia y estacionariedad.
- Aplicación del método de conglomerado jerárquico de series de tiempo.

En esta etapa se identifica los clusters mediante el algoritmo del conglomerado jerárquico a partir de la ecuación 6 se identifica a I como un vector de los valores de los índices a través del tiempo que puede ser dinámico y puede trasladarse de clúster dependiendo de los valores y el método de distancia.

$$I_{C_i C_j} = \sum_{l \in C_i \cup C_j} \|x_l - \mathbf{m}\|^2 - \left\{ \sum_{l \in C_i} \|x_l - \mathbf{m}_i\|^2 + \sum_{l \in C_j} \|x_l - \mathbf{m}_j\|^2 \right\} = \frac{n_i n_j}{n_i + n_j} \|\mathbf{m}_i - \mathbf{m}_j\|^2 \quad (6)$$

- Identificar la cantidad de grupos formados.
- Evaluación de los clústeres encontrados mediante índices internos.
- Identificación de los prototipos de cada clúster y brindar conclusiones.

3.4. CONSTRUCCIÓN DE DATOS

En la presente investigación se trabajó como variables de estudio a los índices de eficiencia insumos que identifican los mínimos niveles de gasto necesarios para alcanzar cada nivel de desempeño y los índices de eficiencia resultados que identifican los máximos niveles de desempeño posibles para cada nivel de gasto en cada región del país a partir del índice de desempeño del sector público (IDSP). Para esto se crea la frontera de posibilidades de producción mediante la técnica de *Free Disposal Hull* (FDH). El primer paso para construir esta FPP es graficar los pares ordenados IDSP-gasto de cada región en un plano que tenga al IDSP en el eje de ordenadas y al gasto público como porcentaje del PBI en el eje de abscisas. Luego identificar a la región con el máximo IDSP de las 24 regiones (región A), el cual estará automáticamente sobre la FPP. Todas las regiones que tengan un nivel de gasto superior a este, estarán por debajo de la frontera. Luego se identifica a la región con el máximo IDSP que tenga un menor gasto que A (región B). Todas las regiones que tengan un nivel de gasto superior a B e inferior a A, también estarán por debajo de la FPP. A continuación, se identifica a la región con el máximo IDSP con menor gasto que B, y así sucesivamente. Se continúa con este proceso hasta llegar a la región con menor nivel de gasto (región G), que por definición también estará sobre la frontera, pues no existirá

ninguno que logre un mayor IDSP con un menor nivel de gasto. Finalmente, la FPP unirá todas estas observaciones y tendrá una forma escalonada de modo que:

$$IDSP = f(gasto) = \left\{ \begin{array}{l} IDSP_A; gasto_A \leq gasto \\ IDSP_B; gasto_B \leq gasto < gasto_A \\ \dots \\ IDSP_G; gasto_G \leq gasto < gasto_F \\ 0; gasto < gasto_G \end{array} \right\}$$

Este método permite calcular el índice de eficiencia-insumos que se obtiene de la división de la región que menor gasta y está sobre la frontera, y el índice de eficiencia-resultados que se obtiene de la división de la región que mejor resultado obtiene y está sobre la frontera en cada año analizado. Este cálculo se realizó para el desempeño general, educación y salud¹.

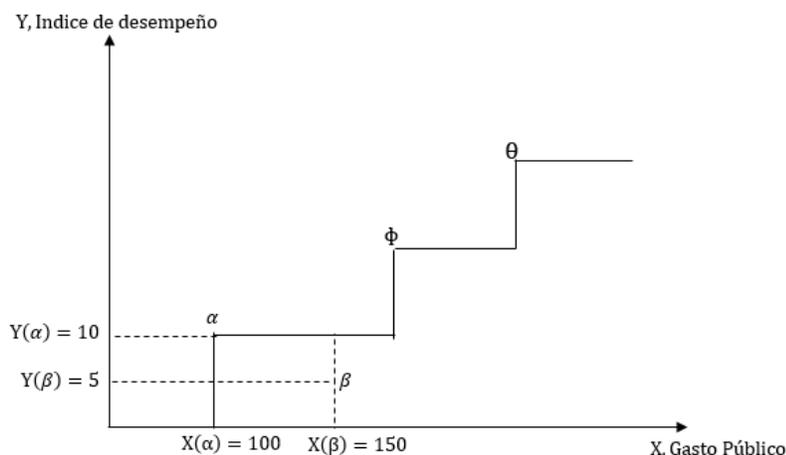


Figura 5: Frontera de Posibilidades de Producción

Fuente: Elaboración Propia.

La metodología de FDH puede ser ilustrada gráficamente asumiendo para cuatro regiones α , β , ϕ y θ que utilizan una determinada cantidad de gasto público, medido en el eje horizontal en unidades monetarias. Se asume entonces que las regiones lograrán un cierto nivel de desempeño del gasto público, medido en el eje vertical. La eficiencia de los cuatro regiones es obviamente diferente. Por ejemplo, el país β usa más insumos que el país α [$X(\beta) > X(\alpha)$], pero produce menos productos [$Y(\beta) < Y(\alpha)$]. Por lo tanto, la región β es relativamente ineficiente en comparación con la región α . Por otro lado, la región α es eficiente en relación a la región β y está ubicado en la frontera de posibilidades de producción. Esto significa que no hay otros países además del país α que entreguen el mismo nivel de producción con un nivel más bajo de insumos. Similarmente, las regiones de ϕ y θ son eficientes y también

¹ Los resultados de la frontera de posibilidades de producción se muestran en los anexos.

están en la frontera de posibilidades de producción. Ninguna otra región es ineficiente comparado con ellos.

IV. RESULTADOS Y DISCUSIÓN

4.1 ANÁLISIS EXPLORATORIO

4.1.1 Análisis de tendencia y estacionariedad

Para la presente investigación se procedió a realizar un análisis de tendencia y estacionariedad con la finalidad de encontrar algún comportamiento que puede señalar alguna asociación entre las regiones a nivel sectorial y general.

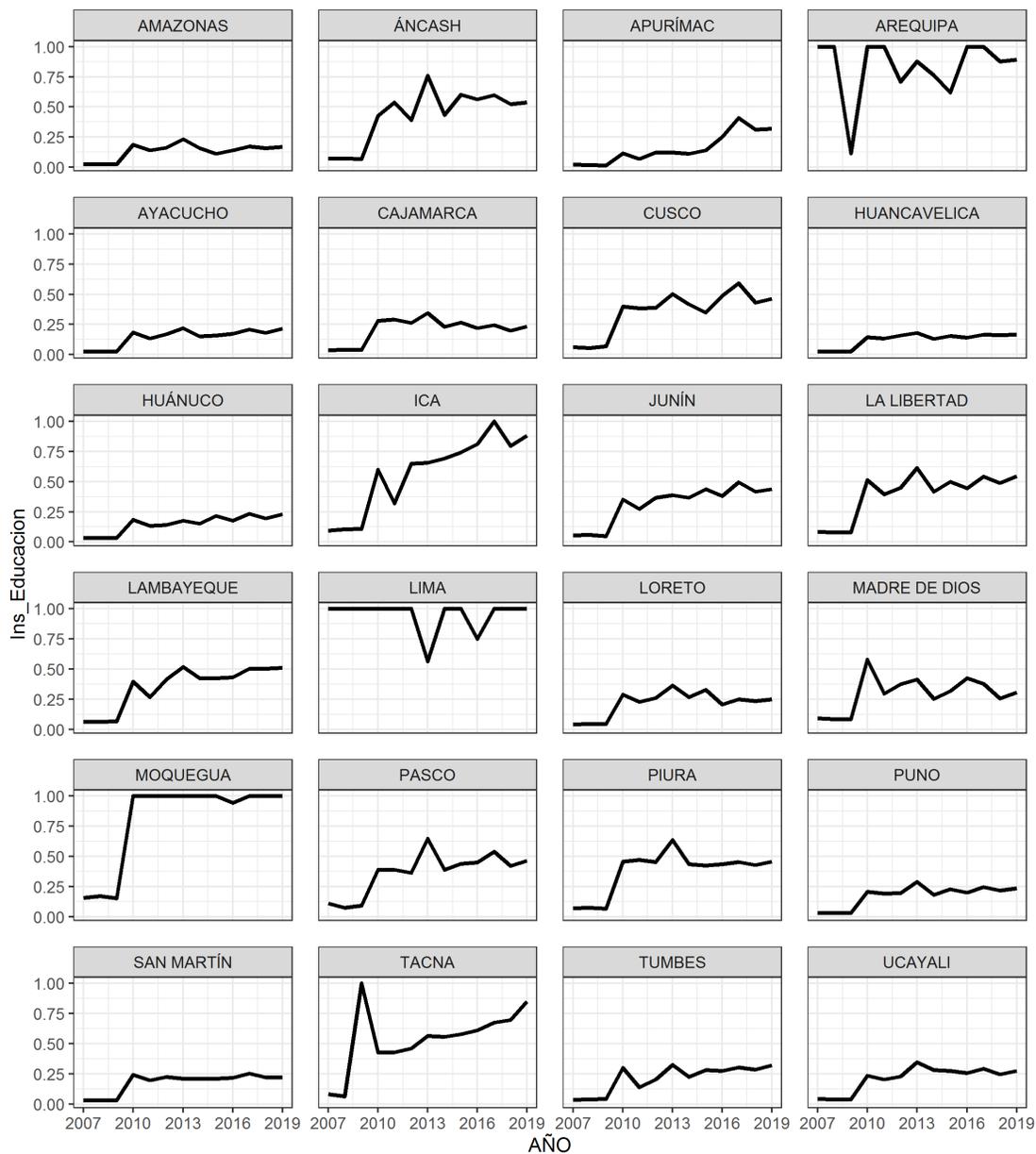


Figura 6: Visualización gráfica del índice de insumos educación

Cuadro 2: P-value de las pruebas estadísticas de insumo de educación

Departamentos	Cox Stuart - Decreciente	Cox Stuart - Creciente	Prueba ADF
Amazonas	0.9844	0.1094	0.2031
Áncash	0.9844	0.1094	0.6601
Apurímac	1.0000	0.0156	0.7587
Arequipa	0.5000	0.8125	0.4801
Ayacucho	1.0000	0.0156	0.0499
Cajamarca	0.6562	0.6562	0.0100
Cusco	1.0000	0.0156	0.2517
Huancavelica	1.0000	0.0156	0.0100
Huánuco	1.0000	0.0156	0.0100
Ica	1.0000	0.0156	0.8940
Junín	1.0000	0.0156	0.0133
La Libertad	1.0000	0.0156	0.0115
Lambayeque	1.0000	0.0156	0.2774
Lima	0.5000	1.0000	0.9355
Loreto	0.8906	0.3438	0.1800
Madre De Dios	0.6562	0.6562	0.0100
Moquegua	0.9375	0.3125	0.0100
Pasco	1.0000	0.0156	0.6556
Piura	0.8906	0.3438	0.0617
Puno	1.0000	0.0156	0.1041
San Martín	0.9844	0.1094	0.0100
Tacna	0.9844	0.1094	0.2885
Tumbes	1.0000	0.0156	0.0100
Ucayali	1.0000	0.0156	0.5240

A partir de la visualización gráfica a niveles del índice insumo de educación mostrada en la Figura 6 y las pruebas estadísticas de Cox Stuart y Dickey-Fuller aplicados a los índices insumos de educación del Cuadro 2 se encontró lo siguiente:

Las regiones de Amazonas, Áncash, Arequipa, Lima, Loreto y Tacna son regiones con comportamiento no estacionario y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que las regiones de Apurímac, Cusco, Ica, Lambayeque, Pasco, Puno y Ucayali no son estacionarios, pero presentan tendencia creciente a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo con estas características tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

Las regiones de Cajamarca, Madre de Dios, Moquegua, Piura y San Martín son regiones estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que las regiones de Ayacucho, Huancavelica, Huánuco, Junín, La Libertad y Tumbes son estacionarios, pero presentan tendencia creciente a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

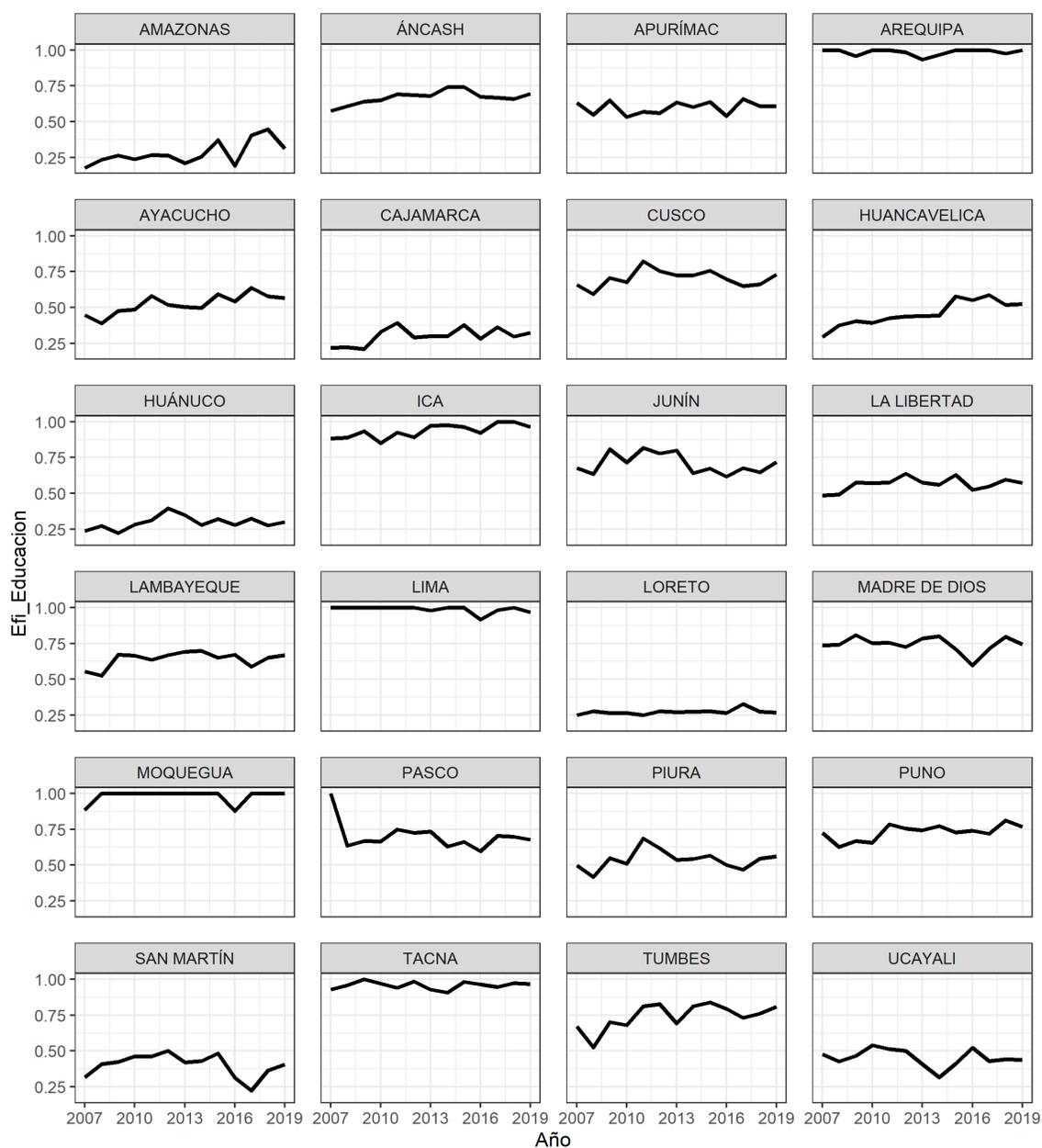


Figura 7: Visualización gráfica del índice de resultados de educación.

Cuadro 3: P-value de las pruebas del índice resultados de educación

Departamentos	Cox Stuart - Decreciente	Cox Stuart - Creciente	Prueba ADF
Amazonas	0.9844	0.1094	0.9900
Áncash	0.9844	0.1094	0.8799
Apurímac	0.8906	0.3438	0.0484
Arequipa	0.6875	0.6875	0.5924
Ayacucho	0.9844	0.1094	0.0956
Cajamarca	0.9844	0.1094	0.0503
Cusco	0.3437	0.8906	0.5187
Huancavelica	1.0000	0.0156	0.4960
Huánuco	0.8906	0.3438	0.4671
Ica	0.9844	0.1094	0.2154
Junín	0.1094	0.9844	0.6503
La Libertad	0.6562	0.6562	0.4836
Lambayeque	0.8906	0.3438	0.6610
Lima	0.1250	1.0000	0.4665
Loreto	0.6562	0.6562	0.3783
Madre De Dios	0.6562	0.6562	0.2376
Moquegua	0.7500	0.7500	0.5187
Pasco	0.3437	0.8906	0.0818
Piura	0.3437	0.8906	0.5326
Puno	1.0000	0.0156	0.3359
San Martín	0.3437	0.8906	0.4636
Tacna	0.3437	0.8906	0.9220
Tumbes	0.8906	0.3438	0.0915
Ucayali	0.1094	0.9844	0.5776

A partir de la visualización gráfica a niveles del índice resultado de educación mostrada en la Figura 7 y las pruebas estadísticas de Cox Stuart y Dickey-Fuller aplicados a los índices insumos de educación del Cuadro 3 se registró lo siguiente:

Las regiones de Amazonas, Áncash, Arequipa, Cusco, Huánuco, Ica, Junín, La Libertad, Lambayeque, Lima, Loreto, Madre de Dios, Moquegua, Piura, San Martín, Tacna y Ucayali son regiones con un comportamiento no estacionario y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que las regiones de Huancavelica y Puno no son estacionarios, pero presentan tendencia creciente a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo con estas características tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

Las regiones de Apurímac, Ayacucho, Cajamarca, Pasco y Tumbes son regiones estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

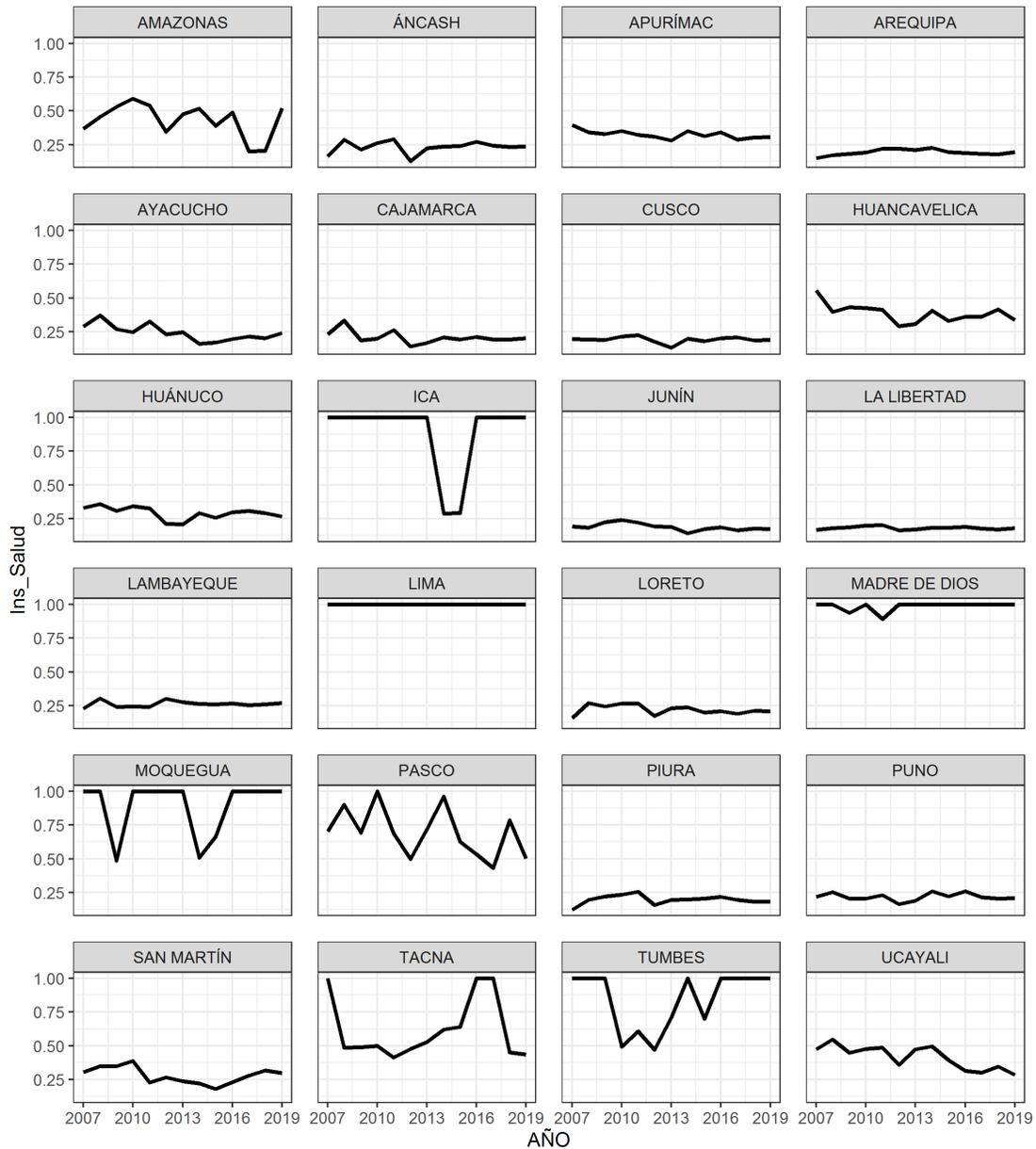


Figura 8: Visualización gráfica del índice de insumos de salud

Cuadro 4: P-value de las pruebas estadísticas de insumos de salud

Departamentos	Cox Stuart – Decreciente	Cox Stuart - Creciente	Prueba ADF
Amazonas	0.3437	0.8906	0.2769
Áncash	0.6562	0.6562	0.7154
Apurímac	0.1094	0.9844	0.6044
Arequipa	0.6562	0.6562	0.6348
Ayacucho	0.1094	0.9844	0.9735
Cajamarca	0.3437	0.8906	0.8192
Cusco	0.6562	0.6562	0.6071
Huancavelica	0.3437	0.8906	0.7641
Huánuco	0.1094	0.9844	0.8913
Ica	0.2500	1.0000	0.8825
Junín	0.0156	1.0000	0.5862
La Libertad	0.8906	0.3438	0.0709
Lambayeque	0.8906	0.3438	0.1089
Lima	0.5000	1.0000	0.6842
Loreto	0.3437	0.8906	0.1730
Madre De Dios	1.0000	0.2500	0.6618
Moquegua	0.5000	0.8750	0.5088
Pasco	0.6562	0.6562	0.0100
Piura	0.6562	0.6562	0.2118
Puno	0.8906	0.3438	0.7481
San Martín	0.3437	0.8906	0.8885
Tacna	0.8906	0.3438	0.1120
Tumbes	0.9375	0.3125	0.0401
Ucayali	0.1094	0.9844	0.6655

A partir de la visualización gráfica a niveles del índice insumo de salud mostrada en la Figura 8 y las pruebas estadísticas de Cox Stuart y Dickey-Fuller aplicados a los índices insumos de salud del Cuadro 4 se observó lo siguiente:

Las regiones de La Libertad, Pasco y Tumbes son regiones estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que Junín es no estacionario con tendencia decrecientes y todas las demás regiones tienen un comportamiento no estacionario y no poseen tendencia a un nivel de significancia del 10 por ciento.

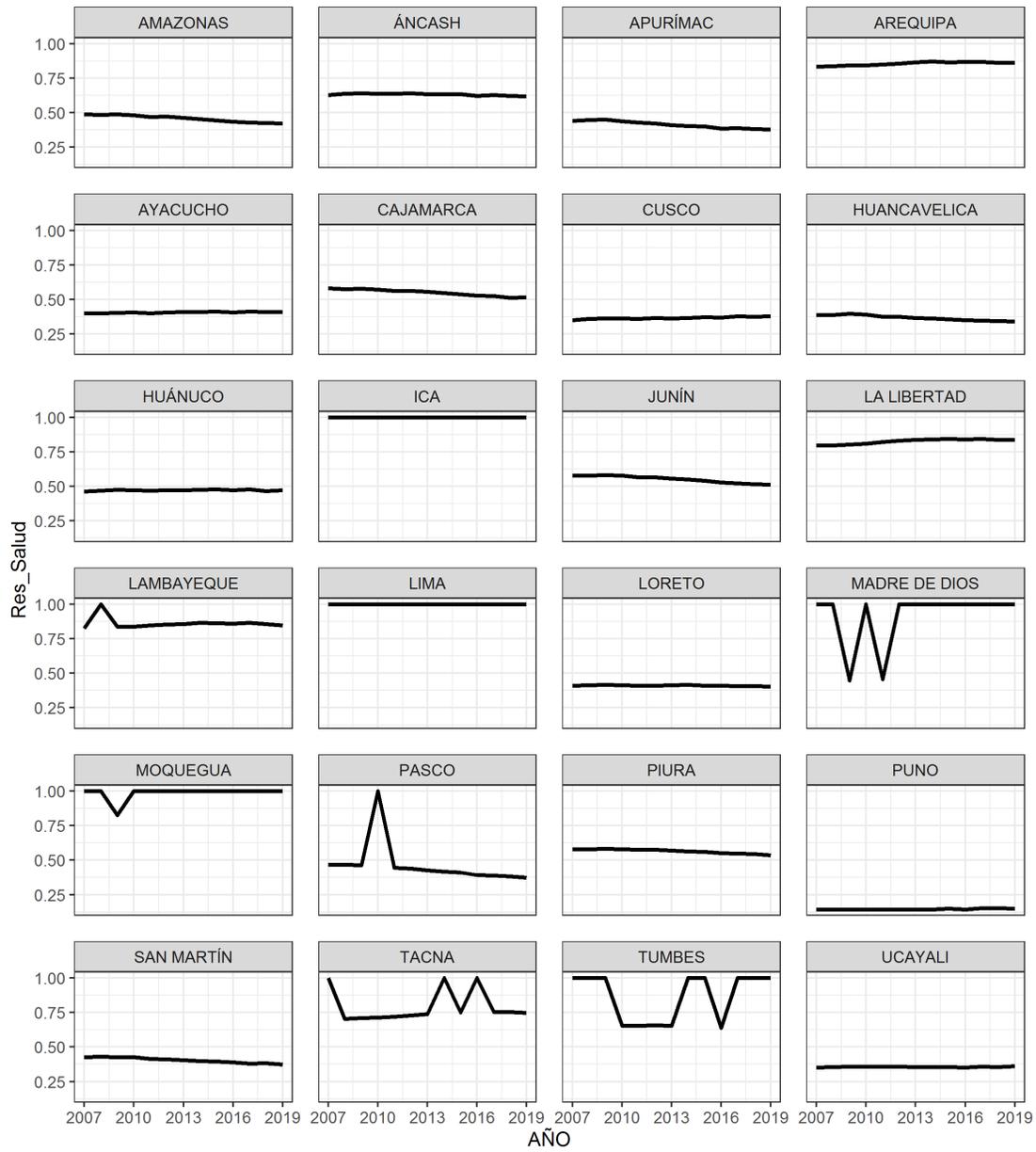


Figura 9: Visualización gráfica del índice de resultados de salud

Cuadro 5: P-value de las pruebas estadísticas del índice resultados de salud

Departamentos	Cox Stuart - Decreciente	Cox Stuart - Creciente	Prueba ADF
Amazonas	0.0156	1.0000	0.6775
Áncash	0.1094	0.9844	0.4954
Apurímac	0.0156	1.0000	0.9543
Arequipa	1.0000	0.0156	0.9900
Ayacucho	1.0000	0.0156	0.8598
Cajamarca	0.0156	1.0000	0.6247
Cusco	1.0000	0.0156	0.8246
Huancavelica	0.0156	1.0000	0.8527
Huánuco	0.6562	0.6562	0.9643
Ica	1.0000	1.0000	0.0000
Junín	0.0156	1.0000	0.5833
La Libertad	1.0000	0.0156	0.8537
Lambayeque	0.8906	0.3438	0.9840
Lima	1.0000	1.0000	0.0000
Loreto	0.1094	0.9844	0.8148
Madre De Dios	1.0000	0.2500	0.8771
Moquegua	1.0000	0.5000	0.0100
Pasco	1.0000	0.5000	0.0101
Piura	0.0156	1.0000	0.6473
Puno	1.0000	0.0313	0.9683
San Martín	0.0156	1.0000	0.3277
Tacna	1.0000	0.0313	0.9119
Tumbes	0.9375	0.3125	0.4050
Ucayali	0.6562	0.6562	0.6734

A partir de la visualización gráfica a niveles del índice resultados de salud mostrada en la Figura 9 y las pruebas estadísticas de Cox Stuart y Dickey-Fuller aplicados a los índices resultados de salud del Cuadro 5 se encontró lo siguiente:

Las regiones de Ica, Lima, Moquegua y Pasco son regiones estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que las regiones de Amazonas, Apurímac, Cajamarca, Huancavelica, Junín, Piura y San Martín no son estacionarias, pero presentan tendencia decreciente a un nivel de significancia del 10 por ciento. Las regiones de Áncash, Huánuco, Lambayeque, Loreto, Madre de Dios, Tumbes y Ucayali son regiones no estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento y se espera que las regiones dentro de cada grupo tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

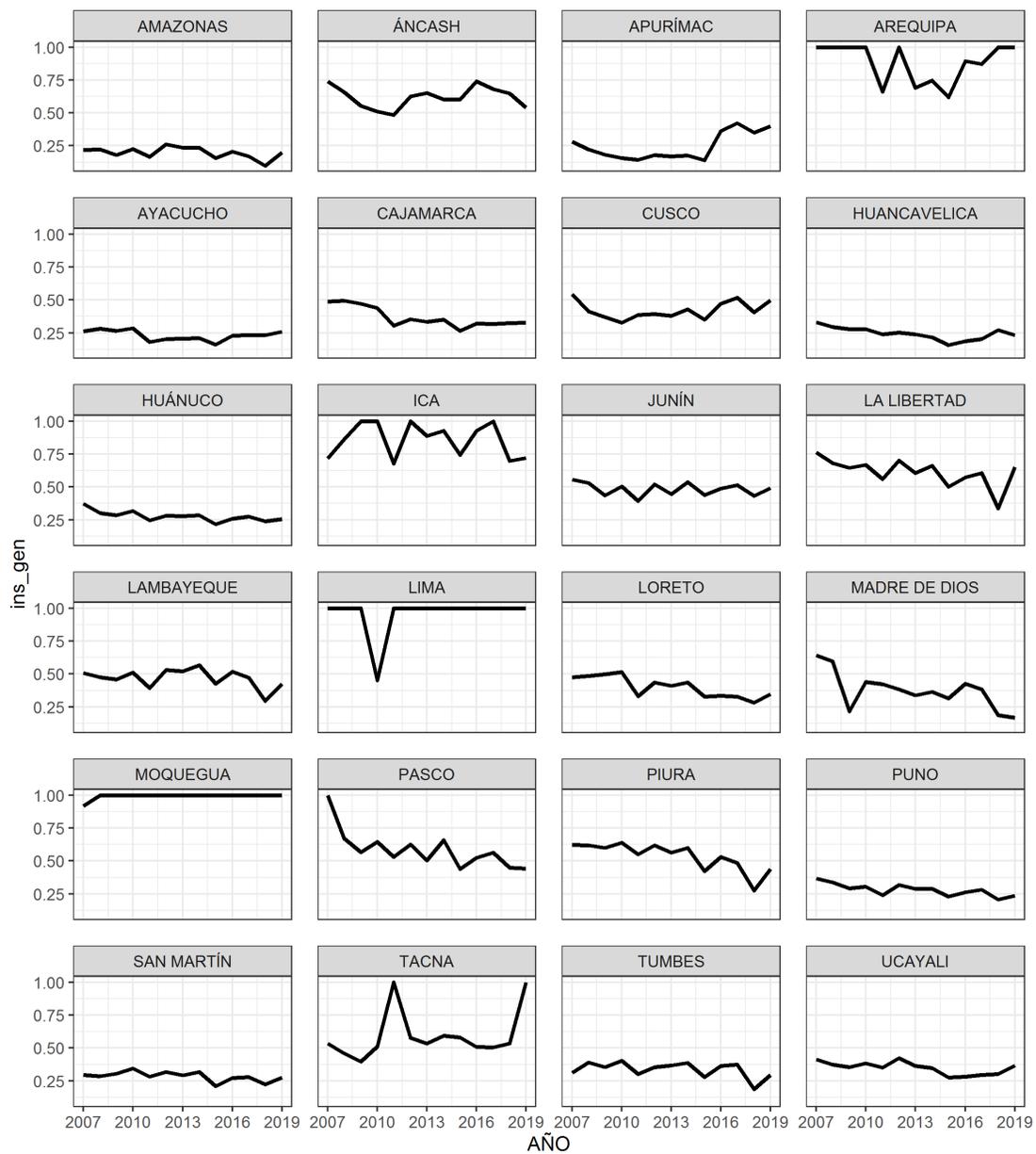


Figura 10: Visualización gráfica del índice de insumos general

Cuadro 6: P-value de las pruebas estadísticas del índice insumos general

Departamentos	Cox Stuart - Decreciente	Cox Stuart - Creciente	Prueba ADF
Amazonas	0.3437	0.8906	0.8640
Áncash	0.6562	0.6562	0.4963
Apurímac	0.8906	0.3438	0.8361
Arequipa	0.3125	0.9375	0.9472
Ayacucho	0.3437	0.8906	0.9577
Cajamarca	0.1094	0.9844	0.8019
Cusco	0.8906	0.3438	0.0244
Huancavelica	0.1094	0.9844	0.5492
Huánuco	0.0156	1.0000	0.6703
Ica	0.5000	0.8125	0.0367
Junín	0.6562	0.6562	0.5282
La Libertad	0.0156	1.0000	0.8005
Lambayeque	0.3437	0.8906	0.9726
Lima	0.6875	0.6875	0.0100
Loreto	0.0156	1.0000	0.0433
Madre De Dios	0.1094	0.9844	0.8698
Moquegua	1.0000	0.5000	0.0000
Pasco	0.0156	1.0000	0.1200
Piura	0.0156	1.0000	0.9773
Puno	0.0156	1.0000	0.2973
San Martín	0.1094	0.9844	0.2699
Tacna	0.9844	0.1094	0.4549
Tumbes	0.3437	0.8906	0.5020
Ucayali	0.0156	1.0000	0.3123

A partir de la visualización gráfica a niveles del índice insumo general mostrada en la Figura 10 y las pruebas estadísticas de Cox Stuart y Dickey-Fuller aplicados a los índices insumos general del Cuadro 6 se observó lo siguiente:

Las regiones de Cusco, Ica, Lima y Moquegua son regiones estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que Loreto es estacionario, pero presentan tendencia decreciente a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo con estas características tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

Las regiones de Huánuco, La Libertad, Pasco, Piura, Puno y Ucayali no son estacionarios, pero presentan tendencia decreciente a un nivel de significancia del 10 por ciento. Mientras que el resto de las regiones son no estacionarios y no poseen tendencia a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

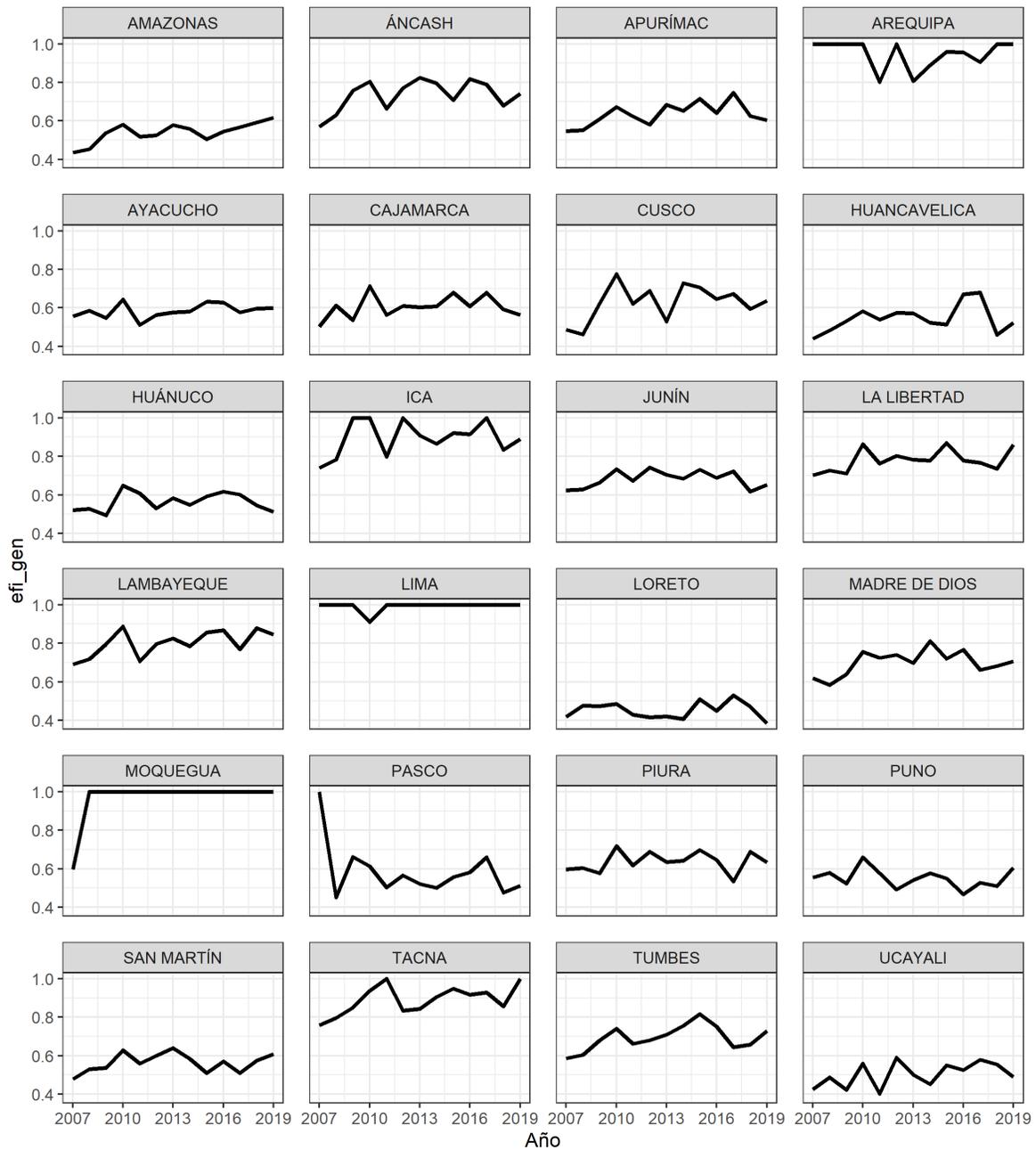


Figura 11: Visualización gráfica del índice resultados general

Cuadro 7: P-value de las pruebas estadísticas del índice resultados general

Departamentos	Cox Stuart - Creciente	Cox Stuart - Decreciente	Prueba ADF
Amazonas	0.9844	0.1094	0.6673
Áncash	0.3438	0.8906	0.5378
Apurímac	0.0156	1.0000	0.7589
Arequipa	0.9688	0.1875	0.9401
Ayacucho	0.1094	0.9844	0.6419
Cajamarca	0.3438	0.8906	0.5398
Cusco	0.6562	0.6562	0.0146
Huancavelica	0.3438	0.8906	0.3099
Huánuco	0.6562	0.6562	0.2094
Ica	0.5000	0.8125	0.0100
Junín	0.6562	0.6562	0.8961
La Libertad	0.3438	0.8906	0.0553
Lambayeque	0.1094	0.9844	0.0217
Lima	0.5000	1.0000	0.0149
Loreto	0.6562	0.6562	0.5502
Madre De Dios	0.6562	0.6562	0.4238
Moquegua	0.5000	1.0000	0.0000
Pasco	0.8906	0.3437	0.5081
Piura	0.3438	0.8906	0.3535
Puno	0.8906	0.3437	0.0100
San Martín	0.3438	0.8906	0.4477
Tacna	0.3438	0.8906	0.0218
Tumbes	0.3438	0.8906	0.7680
Ucayali	0.1094	0.9844	0.3712

A partir de la visualización gráfica a niveles del índice resultados general mostrada en la Figura 11 y las pruebas estadísticas de Cox Stuart y Dickey-Fuller aplicados a los índices resultados general del Cuadro 7 se registró lo siguiente:

Las regiones de Cusco, Ica, La Libertad, Lambayeque, Lima, Moquegua, Puno y Tacna son regiones estacionarias y no poseen tendencia a un nivel de significancia del 10 por ciento. Mientras que Apurímac es no estacionario, pero presenta tendencia creciente a un nivel de significancia del 10 por ciento, se espera que las regiones dentro de cada grupo con estas

características tengan alguna asociación entre ellas y puedan pertenecer a un clúster determinado.

El resto de las regiones son no estacionarios y no poseen tendencia a un nivel de significancia del 10 por ciento.

4.2 EVALUACIÓN DE CLÚSTERS

4.2.1 Índice de la silueta

Se aplicó el método jerárquico de series de tiempo al conjunto de datos con una cantidad de 2 hasta 10 conglomerados con los métodos de distancia de deformación dinámica del tiempo, euclidiana y método de enlace Ward y centroide. Se eligió hasta 10 conglomerados para trabajar con un número no muy alto de clases. Para el caso del índice de la silueta, el mejor valor será aquel que se acerque más a 1. Por ello, se observa en los anexos 1 al 6 con un ancho de silueta promedio de 0.630, el mejor número de conglomerados a realizar es dos para índice de eficiencia insumos y resultados general aplicando como método de distancia a la deformación dinámica del tiempo y función de enlace a Ward cuyo resultado se repite empleando como función de enlace a centroide, sin embargo, Ward presenta una menor variabilidad.

Para el índice eficiencia insumos de educación se obtiene un valor de 0.6408 aplicando como método de distancia a la deformación dinámica del tiempo y función de enlace a Ward que recomienda realizar dos conglomerados, mientras que la función de enlace de centroide señala un índice de silueta de 0.6648 empleando dos conglomerados. Cuando se decide optar por el método de distancia euclidiana provee un valor de 0.6088 señalando dos conglomerados para la función de enlace de Ward y un índice de 0.6437 usando como función de enlace al centroide que a su vez selecciona dos conglomerados, la selección del número de clústers mediante los indicadores se resume en el siguiente cuadro.

Cuadro 8: Comparación de los métodos de distancias y función de enlace

Tipo de datos	Método de distancia	Índice de eficiencia - insumos educación	Índice de eficiencia - resultados educación	Índice de eficiencia - insumos salud	Índice de eficiencia - resultados salud	Índice de eficiencia - insumos general	Índice de eficiencia - resultados general
Deformación dinámica del tiempo	Ward	2	3	2	2	2	2
	Centroide	2	2	2	2	2	2
Euclidiana	Ward	2	3	2	2	2	2
	Centroide	2	3	2	2	2	2

4.2.2 Índice de calinski y harabasz

Se aplicó el método jerárquico de series de tiempo al conjunto de datos con una cantidad de 2 hasta 10 conglomerados con los métodos de distancia de deformación dinámica del tiempo, euclidiana y método de enlace Ward y centroide. Se eligió hasta 10 conglomerados para poder trabajar con un número no muy alto de clases. Para el caso del índice de calinski y harabasz, el mejor valor será el más alto. Por ello, se observa en los anexos 7 al 12, con un ancho de silueta promedio de 21.49, el mejor número de conglomerados a realizar es dos para índice de eficiencia insumos y resultados general aplicando como método de distancia a la deformación dinámica del tiempo y función de enlace a Ward cuyo resultado se repite empleando como función de enlace a centroide, sin embargo, Ward presenta una menor variabilidad.

Para la índice eficiencia insumos de educación se obtiene un valor de 17.01 aplicando como método de distancia a la deformación dinámica del tiempo y función de enlace a Ward que recomienda realizar cuatro conglomerados, mientras que la función de enlace de centroide

señala un índice de calinski y harabasz de 13.34 empleando dos conglomerados. Cuando se decide optar por el método de distancia euclidiana provee un valor de 23.84 señalando dos conglomerados para la función de enlace de Ward y un índice de 26.79 usando como función de enlace al centroide que a su vez selecciona dos conglomerados, la selección de los Clústers mediante los indicadores se resume en el siguiente cuadro.

Cuadro 9: Comparación de los métodos de distancias y función de enlace

Tipo de datos	Método de distancia	Índice de eficiencia - insumos educación	Índice de eficiencia - resultados educación	Índice de eficiencia - insumos salud	Índice de eficiencia - resultados salud	Índice de eficiencia - insumos general	Índice de eficiencia - resultados general
Deformación dinámica del tiempo	Ward	2	3	2	2	2	2
	Centroide	2	2	2	2	2	2
Euclidiana	Ward	2	3	2	2	2	2
	Centroide	2	3	2	2	2	2

4.2.3 Perfil de los conglomerados óptimos

Tras evaluar los métodos de distancia (2), considerando la función de enlace Ward y el índice de silueta, se tiene dos como el valor óptimo de número de conglomerados a excepción para el índice de eficiencia resultados en educación que genera tres conglomerados. Este orden de datos no muestra una estructura de los datos y requiere de un análisis adicional, mientras tanto el índice de calinski y harabasz muestra un mismo comportamiento a pesar de que tiene valores relativamente bajos pues puede tener valores que tienden a infinito de igual manera este índice.

Cuadro 10: Características de los prototipos

Indicador	Región	Valor del índice promedio	Clúster
Índice de eficiencia insumos – educación	Apurímac	0.16	1
	Tacna	0.54	2
Índice de eficiencia resultados – educación	San Martín	0.40	1
	Cusco	0.70	2
	Tacna	0.96	3
Índice de eficiencia insumos – salud	Cajamarca	0.21	1
	Moquegua	0.90	2
Índice de eficiencia resultados – salud	Cajamarca	0.55	1
	Moquegua	0.99	2
Índice de eficiencia insumos – general	Cajamarca	0.37	1
	Ica	0.87	2
Índice de eficiencia resultados – general	Cajamarca	0.61	1
	Tacna	0.89	2

Según lo observado, el conglomerado 1 para el índice de eficiencia insumos de educación, se caracteriza por departamentos que están operando de forma óptima respecto al uso de sus recursos debido a que emplean los mínimos niveles de gastos necesarios para alcanzar un desempeño donde estos resultados se mantiene a través del tiempo. Esto se debe a que estas regiones mantienen una alta competitividad en el uso de los recursos debido a que lideran en asistencia escolar y rendimiento escolar en primaria y secundaria mientras que el conglomerado 2 muestra que existen regiones que persisten en usar los recursos de manera inadecuada como el caso de Ayacucho, Apurímac, Huancavelica y Cusco que se mantiene en el trapecio andino de la pobreza y no pueden desarrollarse, como se observa en la Figura 12.

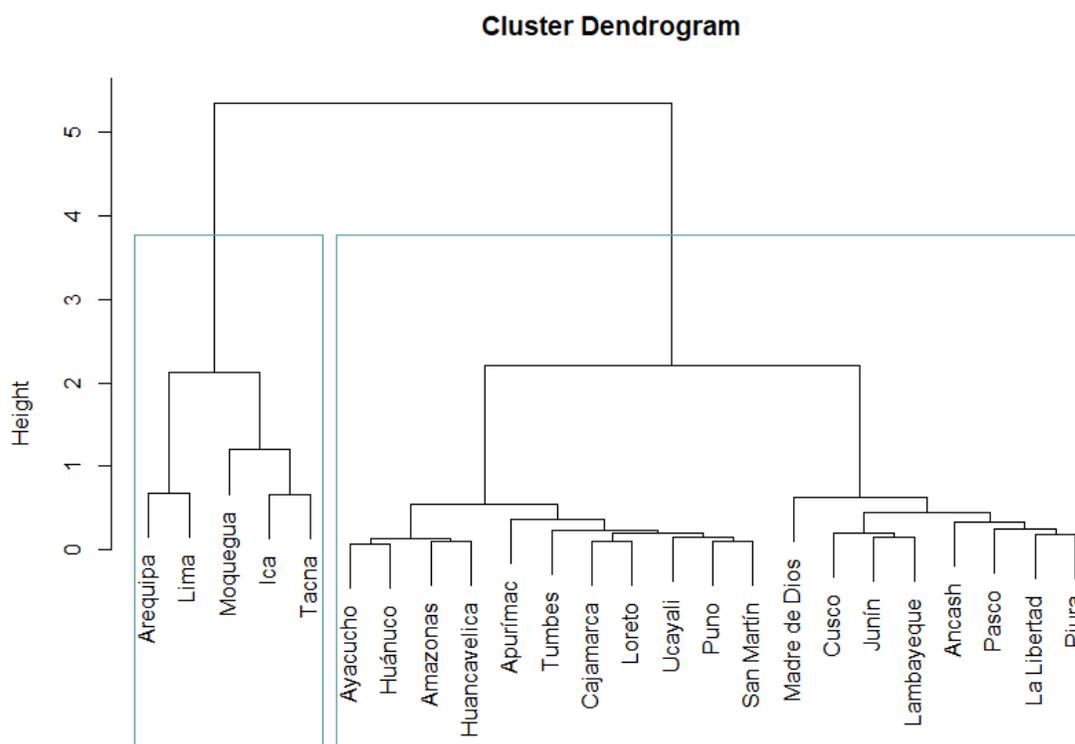


Figura 12: Departamentos por conglomerados por índice de eficiencia insumos educación

Respecto al índice de eficiencia resultados en educación, se tiene tres escalas diferente de nivel de resultados se observa que el grupo que está representado por la región San Martín muestra los peores resultados y que tienen oportunidades de mejora debido que una parte de su población aún no asiste a las instituciones educativas y se debe inyectar un mayor presupuesto, mientras que las regiones que son representadas por Cusco se encuentran en un nivel intermedio sobre los resultados a obtener, requiere solo mejorar en la gestión administrativa pues se encontraron problemas de corrupción y el último grupo que corresponde a Tacna que obtienen los mejores resultados respecto a todas las regiones. Este clúster mantiene el mismo comportamiento respecto al índice insumos pues tienen mejores accesos a servicios, rendimiento en matemáticas, lectura, mayor tasa de matrículas en inicial, primaria y secundaria, esto se observa en la Figura 13.

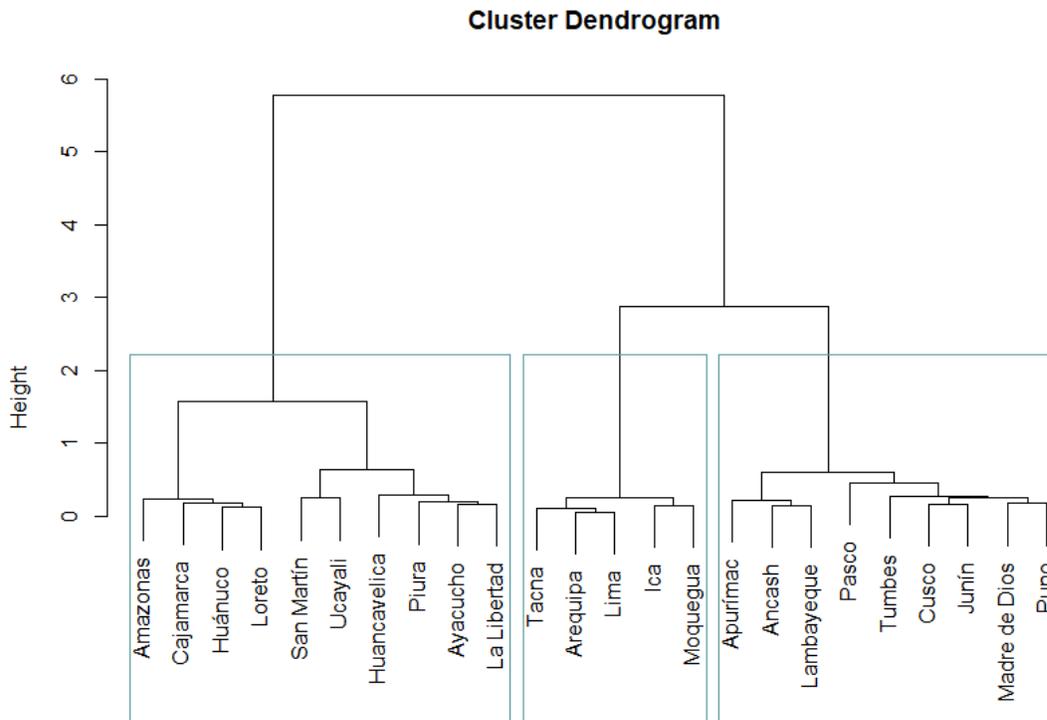


Figura 13: Departamentos por conglomerados por índice de eficiencia resultados educación

Respecto al índice de eficiencia resultados en salud, se muestra un comportamiento ya repetitivo que las regiones de mayor presupuesto emplean de manera inadecuada su presupuesto y que están representado por Cajamarca que no cuenta con el equipamiento adecuado para ofrecer mejor servicio y no son prioridad del Estado peruano, mientras que existe un grupo reducido que está obteniendo mejores resultados representado por Moquegua, que son regiones donde se realizaron obras importantes como los hospitales regionales y un aumento en la cantidad de profesionales de la salud como se observa en la Figura 14.

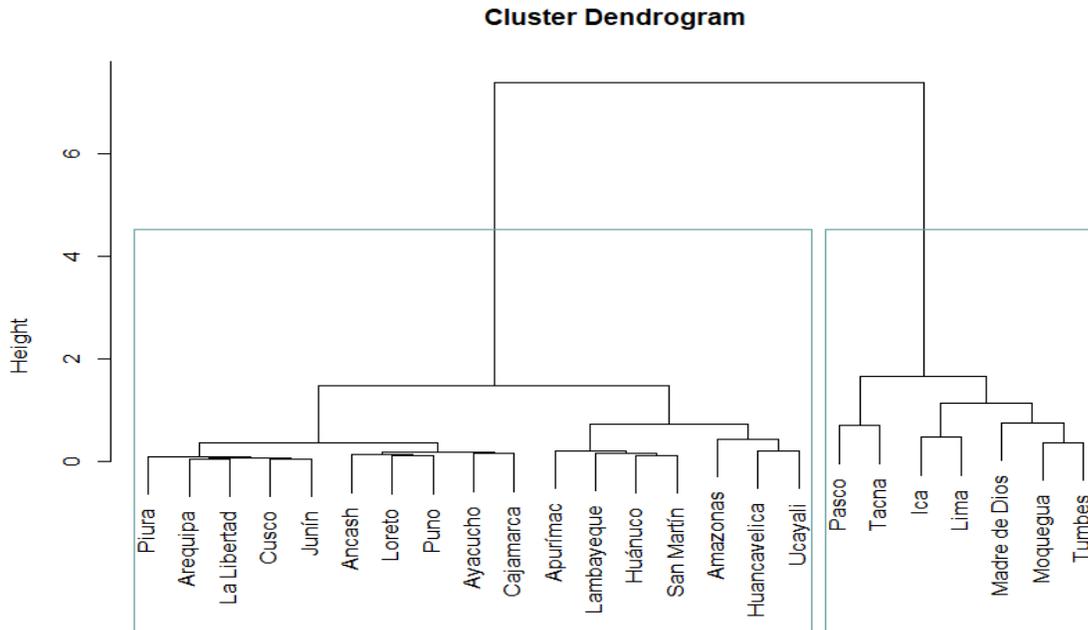


Figura 14: Departamentos por conglomerados por índice de eficiencia insumos salud

Respecto a los índices de eficiencia insumos en salud, las regiones como Madre de Dios, Moquegua y Tumbes mantienen el menor Gasto Público respecto al Producto Bruto Interno, mientras que Ica y Lima respecto al producto interno que generan son relativamente mayor al gasto público mientras que todas estas regiones muestran buenos resultados en cobertura hospitalaria, desnutrición crónica, partos institucionales y cobertura del personal médico. Las otras regiones requieren un mayor incremento en el gasto público por PBI para mejorar sus resultados pues una parte importante se pierde por burocracia o corrupción, debido a esto se tiene planificado hacer un seguimiento a las actividades que realiza los gobiernos regionales como se observa en la Figura 15.

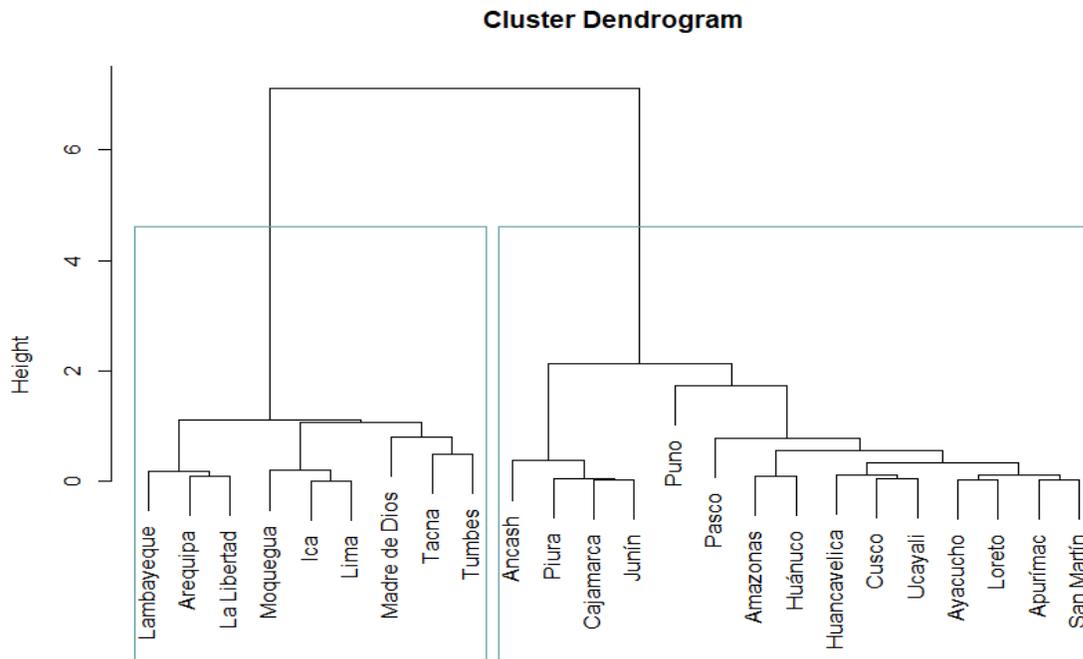


Figura 15: Departamentos por conglomerados por índice de eficiencia resultados salud

Se observa que el clúster que obtiene los mejores resultados con el uso de los recursos de manera óptima destaca principalmente por los tener buenos indicadores en salud como la esperanza de vida al nacer, cobertura hospitalaria, prevalencia de anemia, vacunación, desnutrición crónica. Esto se puede observar debido a las necesidades que cubre cada región como el caso de Moquegua que pidió presupuesto para equipar y construir el hospital regional como Lima que tiene la mayor población y por ende requiere de una mejor infraestructura hospitalaria, a su vez se observa regiones tanto para costa y selva que resaltan en los resultados en el sector salud mientras que la sierra aún no está dentro de sus prioridades mejorar este índice.

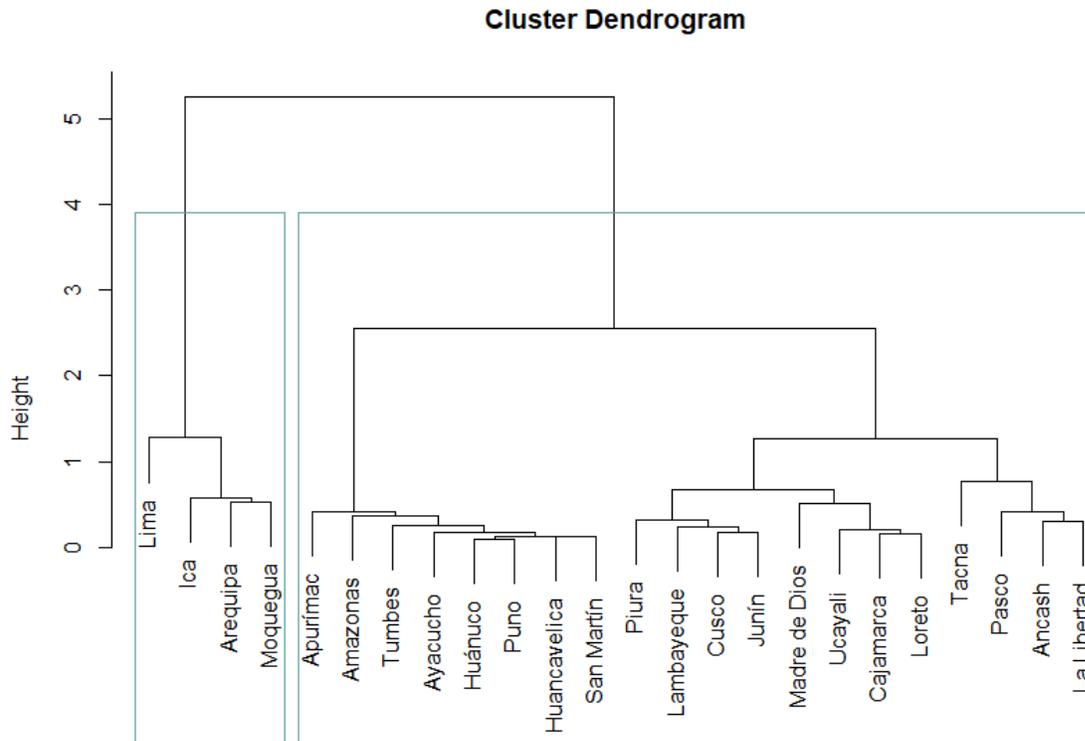


Figura 16: Departamentos por conglomerados por índice de eficiencia insumos general

Por último, se observa que de manera global en el índice de eficiencia insumos tiene el mismo comportamiento del índice de eficiencia insumos en salud y educación que son los sectores más importantes y se puede señalar que se mantiene una hegemonía por parte de las regiones que presentan los mejores resultados en salud y educación como el caso de Lima que centraliza el uso de los recursos, Moquegua que es la región que presentó mejores resultados en educación según los resultados de los exámenes PISA y la incorporación de infraestructura y equipamiento en salud con un entorno económico debido a un presupuesto público per cápita y stock de capital elevado que puede solventar su economía regional.

Se observa que dentro de Ica no solo la intervención del sector público ayudó a mejorar la región, brinda oportunidades impulsando el desarrollo laboral pues se preocupa en la educación de la fuerza laboral y que tengan un empleo adecuado a su vez tienen acceso a cobertura eléctrica, agua y desagüe. Mientras que Lima tiene mayor acceso a servicios públicos, un buen entorno económico como el producto interno, gasto por hogar, presupuesto per cápita, acceso a crédito y disponibilidad de servicios financieros. Por último, Arequipa

es la región que mayor desarrollo tiene en el sector salud debido a que tiene menor mortalidad infantil, desnutrición crónica y partos institucionales.

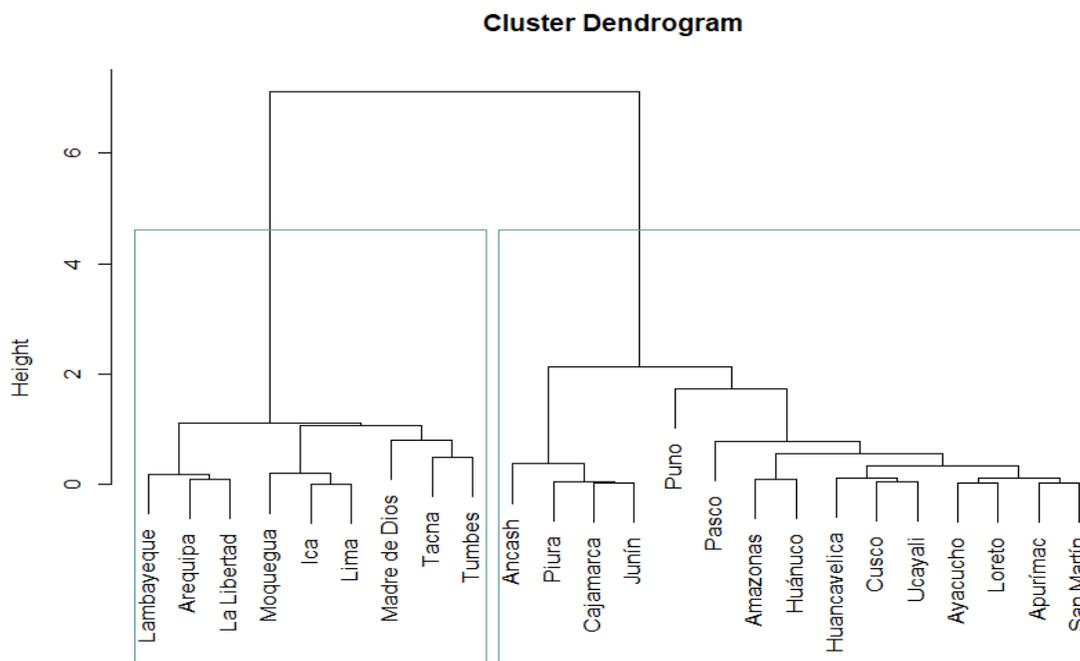


Figura 17: Departamentos por conglomerados por índice de eficiencia resultados general

Por último, se observa que de manera global en el índice de eficiencia resultados tiene un comportamiento homogéneo al índice de insumos pues las regiones de la costa muestran un resultado superior por su relativa posición en los ámbitos de calidad de infraestructura, desempeño económico, estabilidad, salud, educación que tienen mayor desarrollo y crecimiento respecto al segundo grupo. Se observa que el primer grupo donde se encuentra Lambayeque, La Libertad, Tacna, Tumbes y las demás regiones mencionadas anteriormente obtienen mejores resultados a lo largo del periodo por aprovechar los recursos respecto al otro clúster enfocándose en los sectores de salud y educación que al final generan un crecimiento que tendrá un impacto para el mediano y largo plazo.

V. CONCLUSIONES

1. El número óptimo de conglomerados para clasificar a las a las regiones del país según el índice de eficiencia insumos y resultados es de dos con excepción al índice de eficiencia resultados de educación con tres conglomerados de acuerdo con el índice silueta y empleando conglomerados de series de tiempo jerárquico.
2. Este algoritmo de series de temporales muestra un patrón que se mantiene para los índices de eficiencia insumos y resultados de manera sectorial y global, de esta manera permite visualizar un comportamiento estático a pesar de los cambios en la gestión del gasto público
3. Se observa que existen regiones que deben mejorar sus resultados, pues la mayoría de los grupos se subdividen en solos dos conglomerados que muestra regiones con buen resultado y peores resultados.
4. Se muestra una predominancia de las regiones de la costa por el uso de los recursos y resultados respecto a las regiones de la sierra y selva y esto se mantiene a lo largo del periodo analizado y que requiere un cambio estructural.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Afonso, A., Schuknecht, L., y Tanzi, V. 2003. Public Sector Efficiency: An International Comparison. *Working Paper No. 242*, 1-27.
- Berndt, D., y Clifford, J. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. *In KDD Workshop, volume 10*, pp. 359 - 370.
- Brockwell, P., y Davis, R. 2016. *Introduction to Time Series and Forecasting*. USA: Springer.
- Cano, J., Sanchez, L., y Puerta, E. 2019. Análisis estadístico para la administración del inventario en empresas comerciales aplicando métodos multivariados. *Comunicaciones en Estadística, volume 12*, pp. 71 - 89.
- Camp, R. 1991. Benchmarking, "la evaluación comparativa. Cómo analizar las practicas de las mejoras empresas para convertirse en la primera". *Milan*.
- Cleary, E. 2006. El Benchmarking desde el sector privado al sector público: "Apuntes sobre su aplicación en el sector público en América Latina.". *Centro de Desarrollo y Asistencia Técnica en Tecnología para la Organización Pública*.
- Desgraupes, B. 2013 Clustering Indices. University Paris Ouest. FR. 34p.
- Everitt, B., Landau, S., Leese, M., y Stahl, D. 2011. Cluster Analysis. *Wiley*.
- Fontalvo Herrera, T., Mendoza Mendoza, A., y De la Hoz Dominguez, E. 2020. Eficiencia del gasto público y privado en salud en los países de América Latina: 2000 - 2015. *Prospectiva Vol 18 2*.
- Galvis - Aponte, L. A. 2015. La Eficiencia del Gasto Público en Educación en Colombia. *Documentos de Trabajo sobre Economía Regional*, 75-98.
- Giorgino, T. 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, 1-24.
- Gwartney, J., Lawson, R., Park, W., Wagh, S., Edwards, C., y Ruggy, V. 2002. Economic Freedom of the World: 2002 Annual Report. *The Fraser Institute*.

- Kaufman, L. y Rousseeuw, P. 1990. Finding Groups in DATA. An introduction to cluster analysis. Wiley ed. BE. 335p.
- Liao, T. W. 2005. Clustering of time series data. *Pattern Recognition* 38, 1857 - 1874.
- Machado, R. 2006. ¿Gastar más o Gastar mejor? La eficiencia del Gasto Público en América Central y República Dominicana. *Banco Interamericano de Desarrollo*, 1-31.
- Maharaj, E., D'Urso, P., y Caiado, J. 2019. *Time Series Clustering and Classification*. Miami: CRC Press.
- Mueller, D. 1997. Perspectives on Public Choice. *A Handbook*, Cambridge.
- Nicholson, W. 2008. Teoría Macroeconómica: Principios Básicos y Aplicaciones 9. ed. Santa Cruz. *Cengage Learning*.p., 335-347.
- Peña, D. 2010. *Análisis de series temporales*. Madrid: Alianza Editorial.
- Persson, T., y Tabellini, G. 2001. Political Institutions and Policy Outcomes: What are the Stylized Facts? *Journal of Policy Modelling* 23, 433-467.
- Petitjean, F., Ketterlin, A., y Gançarski, P. 2011. A Global Averaging Method for Dynamic Time Warping , with Applications to Clustering. *Patter Recognition*, 678 - 693.
- Ratanamahatana, C., y Keogh, E. 2004. Everything you know about dynamic time warping is wrong. *In Third workshop on mining temporal and sequential data*.
- Rodrik, D. 2000. Institutions for High-Quality Growth: What they are and how to Acquire them. *NBER Working Paper 7540*.
- Rueda, N. 2011. La eficiencia y su importancia en el sector público. *eXtoikos*, 38 - 46.
- Sardá-Espinoza, A. 2019. Comparing Time-Series Clustering Algorithms in Using the dtwclust Package. *Reference Manual*.
- Secretaría de Gestión Pública, 2. 2016. Identificación y Análisis de las tendencias en materia de modernación de la gestión pública con una visión a 2030. *Presidencia del Consejo de Ministros*, 18-30.
- Shleifer, A., y Vishny, R. 1998. The Grabbing Hand: Government Pathologies and their Cures. *Cambridge: Harvard University Press*.

- Solórzano, E. 2016 Identificación y caracterización de patrones climáticos en la ciudad de Manizales, usando técnicas de tiempo y de conglomerados. *Universidad Nacional de Colombia: Departamento de Matemáticas y Estadística*.
- Stiglitz, J. 1988. Economía del Sector Público . 3. ed. *Barcelona. A. Bosch*, 181-217, 313 - 341.
- Strauch, R., y Hagen, J. 2000. Institutions, Politics and Fiscal . *Boston: Kluwer Academic Publishers*.
- Tan, M. 2006. Una aproximación a la eficiencia técnica del Gasto Público en Educación en las regiones del Perú. *Consortio de investigación económica y social* , 1-147.
- Tanzi, V., y Schuknecht, L. 2000. Public Spending in the 20th Century: A Global Perspective. *Cambridge: Cambridge University Press*.
- Tanzi, V., & Schuknecht, L. 2000. Reconsidering the Fiscal Role of Government: The International Perspective. *American Economic Review*, 87(2), 164-168.
- Vanden Eeckhaut, P., Tulkens, H., & M., J. 1993. Cost efficiency in Belgian municipalities. *LIDAM Reprints CORE*, 300-334.
- Vasilios, Z., Stavros G. Pouloupoulos, Zoe, G. y Efthimios, Z. 2020. *Clustering analysis of countries using the COVID-19 cases dataset. Science Direct*.
- Xu, R., & Wunsch, D. 2009. Clustering Hobok. *IEEE Computational Intelligence Magazine Press*.
- Yoshihide, K., Shumway, R. y Taniguchi, M. 1998. Discrimination and Clustering for Multivariate Time Series. *Journal of the American Statistical Association*, March 328-340, 1998.
- Wijik, J., & Selow, E. 1999. Cluster and Calendar based Visualization of Time Series Data. *Symposium on Information Visualization, San Francisco, October 25-26, 1999*.

VII. ANEXOS

Anexo 1: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia de insumos educación

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - insumos educación	Deformación dinámica del tiempo	Ward	0.641	0.515	0.549	0.546	0.504	0.480	0.378	0.306	0.241
		Centroide	0.665	0.599	0.521	0.479	0.429	0.480	0.378	0.315	0.187
	Euclidiana	Ward	0.609	0.500	0.513	0.508	0.491	0.486	0.357	0.247	0.249
		Centroide	0.644	0.570	0.553	0.472	0.443	0.486	0.367	0.329	0.244

Anexo 2: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia de resultados educación.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados educación	Deformación dinámica del tiempo	Ward	0.558	0.601	0.594	0.544	0.460	0.465	0.432	0.388	0.293
		centroide	0.578	0.553	0.594	0.483	0.460	0.387	0.350	0.274	0.316
	Euclidiana	Ward	0.486	0.570	0.462	0.455	0.410	0.406	0.395	0.374	0.350
		centroide	0.518	0.570	0.494	0.440	0.415	0.449	0.386	0.340	0.287

Anexo 3: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia de insumos salud.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	2	7	8	9	10
Índice de eficiencia - insumos salud	Deformación dinámica del tiempo	Ward	0.753	0.613	0.456	0.455	0.442	0.389	0.371	0.339	0.338
		centroide	0.753	0.613	0.565	0.576	0.570	0.538	0.440	0.403	0.360
	Euclidiana	Ward	0.661	0.540	0.474	0.451	0.453	0.439	0.424	0.296	0.286
		centroide	0.700	0.568	0.558	0.538	0.520	0.515	0.506	0.414	0.361

Anexo 4: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia de resultados salud.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados salud	Deformación dinámica del tiempo	Ward	0.697	0.554	0.595	0.511	0.575	0.576	0.568	0.514	0.505
		centroide	0.697	0.630	0.595	0.468	0.459	0.447	0.536	0.557	0.527
	Euclidiana	Ward	0.674	0.508	0.539	0.487	0.515	0.535	0.547	0.545	0.478
		centroide	0.674	0.589	0.455	0.443	0.312	0.463	0.547	0.545	0.511

Anexo 5: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia de insumo general.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados salud	Deformación dinámica del tiempo	Ward	0.630	0.451	0.428	0.379	0.380	0.330	0.281	0.276	0.262
		centroide	0.630	0.549	0.299	0.244	0.239	0.294	0.233	0.321	0.287
	Euclidiana	Ward	0.496	0.434	0.352	0.350	0.343	0.319	0.269	0.266	0.250
		centroide	0.595	0.545	0.491	0.474	0.394	0.330	0.209	0.173	0.108

Anexo 6: Comparación de los métodos de distancias y función de enlace para el índice de silueta para índice de eficiencia de resultado general.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados general	Deformación dinámica del tiempo	Ward	0.630	0.451	0.428	0.379	0.380	0.330	0.281	0.276	0.262
		centroide	0.630	0.549	0.299	0.244	0.239	0.294	0.233	0.321	0.287
	Euclidiana	Ward	0.496	0.434	0.352	0.350	0.343	0.319	0.269	0.266	0.250
		centroide	0.595	0.545	0.491	0.474	0.394	0.330	0.209	0.173	0.108

Anexo 7: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia de insumos educación.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - insumos educación	Deformación dinámica del tiempo	Ward	11.448	15.588	17.008	15.064	14.032	12.885	11.853	11.544	10.907
		centroide	13.341	7.086	7.885	6.816	5.584	12.885	11.853	11.079	10.158
	Euclidiana	Ward	23.841	16.276	15.559	13.725	13.105	13.235	11.207	10.884	11.286
		centroide	26.792	14.101	10.139	9.619	8.352	13.235	12.223	10.882	10.264

Anexo 8: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia de resultados educación.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados educación	Deformación dinámica del tiempo	Ward	29.578	31.681	25.413	22.720	19.475	16.706	14.808	13.366	12.140
		centroide	26.192	20.683	25.413	20.174	19.479	16.358	14.436	12.970	12.490
	Euclidiana	Ward	24.676	22.731	21.535	17.338	14.082	12.399	11.295	10.294	9.633
		centroide	25.798	22.731	15.536	12.079	9.712	12.860	11.761	10.330	9.549

Anexo 9: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia de insumos salud.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - insumos salud	Deformación dinámica del tiempo	Ward	49.427	25.105	21.785	18.990	17.322	15.892	15.613	15.382	14.887
		centroide	49.427	25.105	17.341	14.689	13.143	11.541	14.278	13.138	12.448
	Euclidiana	Ward	31.774	21.058	16.872	14.534	13.250	12.121	11.730	10.995	11.385
		centroide	41.992	22.004	15.503	12.652	11.255	10.135	9.280	11.538	10.792

Anexo 10: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia de resultados salud.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados salud	Deformación dinámica del tiempo	Ward	59.919	36.621	29.947	26.540	23.613	21.709	21.769	20.369	22.520
		centroide	59.919	31.505	29.947	23.478	20.128	15.581	21.566	22.741	22.522
	Euclidiana	Ward	44.497	25.275	19.801	20.649	18.885	17.740	18.208	18.430	18.178
		centroide	44.497	25.196	18.048	14.250	11.989	13.150	18.208	18.430	17.919

Anexo 11: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia de insumo general.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - insumo general	Deformación dinámica del tiempo	Ward	21.487	18.298	13.967	11.386	9.717	9.954	9.072	8.986	8.652
		centroide	21.487	11.955	8.710	6.515	5.574	7.223	6.793	7.939	7.435
	Euclidiana	Ward	25.805	20.040	12.378	10.177	8.546	8.727	8.080	7.872	7.390
		centroide	22.244	11.608	7.892	6.184	9.359	8.325	7.391	6.974	6.449

Anexo 12: Comparación de los métodos de distancias y función de enlace para el índice de Calinski - Harabasz para índice de eficiencia de resultado general.

Tipo de datos	Método de distancia	Función de enlace	Número de conglomerados								
			2	3	4	5	6	7	8	9	10
Índice de eficiencia - resultados general	Deformación dinámica del tiempo	Ward	29.578	31.681	25.413	22.720	19.475	16.706	14.808	13.366	12.140
		centroide	26.192	20.683	25.413	20.174	19.479	16.358	14.436	12.970	12.490
	Euclidiana	Ward	24.676	22.731	21.535	17.338	14.082	12.399	11.295	10.294	9.633
		centroide	25.798	22.731	15.536	12.079	9.712	12.860	11.761	10.330	9.549

Anexo 13: Gasto Público por PBI en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.37	0.43	0.57	0.51	0.59	0.49	0.55	0.61	0.70	0.63	0.82	0.46	0.27
Ancash	0.11	0.15	0.18	0.23	0.20	0.21	0.20	0.24	0.18	0.18	0.21	0.07	0.10
Apurímac	0.29	0.44	0.56	0.74	0.69	0.72	0.75	0.80	0.77	0.36	0.33	0.13	0.13
Arequipa	0.09	0.11	0.13	0.14	0.15	0.16	0.19	0.19	0.18	0.15	0.16	0.06	0.07
Ayacucho	0.31	0.34	0.39	0.40	0.55	0.63	0.62	0.68	0.68	0.57	0.61	0.19	0.21
Cajamarca	0.17	0.20	0.22	0.26	0.32	0.36	0.38	0.40	0.41	0.41	0.44	0.14	0.16
Cusco	0.15	0.23	0.28	0.35	0.25	0.33	0.34	0.33	0.31	0.28	0.27	0.11	0.11
Huancavelica	0.25	0.33	0.37	0.41	0.41	0.51	0.54	0.66	0.70	0.70	0.69	0.17	0.23
Huánuco	0.22	0.32	0.36	0.36	0.40	0.45	0.46	0.50	0.51	0.51	0.51	0.19	0.21
Ica	0.11	0.11	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.14	0.15	0.06	0.07
Junín	0.15	0.18	0.23	0.23	0.25	0.25	0.29	0.26	0.25	0.27	0.27	0.10	0.11
La Libertad	0.11	0.14	0.16	0.17	0.17	0.18	0.21	0.21	0.22	0.23	0.23	0.13	0.08
Lambayeque	0.16	0.20	0.22	0.23	0.25	0.24	0.25	0.25	0.26	0.25	0.30	0.15	0.13
Lima	0.24	0.26	0.26	0.25	0.24	0.24	0.25	0.28	0.29	0.30	0.32	0.28	0.28
Loreto	0.17	0.20	0.20	0.22	0.30	0.29	0.31	0.33	0.33	0.39	0.43	0.16	0.15
Madre de Dios	0.13	0.16	0.47	0.26	0.23	0.33	0.38	0.39	0.35	0.31	0.36	0.24	0.32
Moquegua	0.09	0.10	0.10	0.12	0.10	0.13	0.13	0.14	0.11	0.13	0.14	0.05	0.05
Pasco	0.08	0.14	0.18	0.18	0.18	0.20	0.25	0.22	0.25	0.25	0.25	0.10	0.12
Piura	0.13	0.16	0.17	0.18	0.18	0.21	0.23	0.24	0.26	0.25	0.29	0.16	0.12
Puno	0.22	0.29	0.35	0.38	0.41	0.40	0.44	0.49	0.48	0.50	0.50	0.22	0.23
San Martín	0.27	0.34	0.33	0.34	0.35	0.40	0.44	0.45	0.52	0.48	0.51	0.20	0.20
Tacna	0.15	0.21	0.26	0.23	0.20	0.22	0.24	0.24	0.19	0.26	0.28	0.08	0.07
Tumbes	0.26	0.25	0.29	0.29	0.32	0.36	0.35	0.37	0.39	0.36	0.38	0.24	0.18
Ucayali	0.20	0.26	0.29	0.30	0.28	0.30	0.35	0.41	0.40	0.46	0.48	0.15	0.15

**Anexo 14: Índice de calidad de la infraestructura en las regiones para el periodo
2007-2019**

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	1.21	1.35	1.22	1.35	1.32	1.58	1.40	2.03	1.76	2.05	2.51	2.65	2.18
Ancash	2.33	2.16	2.37	2.97	3.29	3.14	3.60	3.36	3.18	3.03	2.79	2.75	3.11
Apurímac	1.17	1.25	1.08	1.05	1.16	1.48	1.45	1.36	1.30	1.00	1.57	1.48	1.38
Arequipa	4.77	4.44	5.15	5.82	5.01	5.14	4.95	5.01	5.53	5.51	5.65	4.94	5.30
Ayacucho	1.51	1.39	1.64	1.73	1.48	1.80	1.39	1.63	1.46	1.34	1.39	1.54	1.48
Cajamarca	1.47	1.74	1.98	2.16	1.86	1.50	1.65	1.42	1.59	1.52	1.66	1.80	1.64
Cusco	1.77	2.05	1.81	2.01	1.92	2.35	2.05	2.40	2.33	2.15	2.07	2.36	2.23
Huancavelica	1.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.06	1.00	1.00	1.00
Huánuco	1.69	1.56	1.53	1.86	1.78	1.89	1.73	1.97	2.09	1.66	1.74	1.78	1.81
Ica	2.67	2.58	3.10	3.37	4.09	4.20	4.35	4.46	4.85	4.44	4.89	4.77	4.67
Junín	2.03	1.98	1.93	2.31	2.38	2.15	2.47	2.42	3.06	3.29	3.33	3.27	3.03
La Libertad	4.04	3.77	3.27	3.77	3.35	4.15	3.89	3.93	4.07	4.15	3.85	3.68	3.91
Lambayeque	2.89	3.35	3.19	3.75	3.41	3.77	3.94	4.27	4.53	4.39	4.33	4.10	4.23
Lima	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00
Loreto	1.64	1.67	1.45	1.53	1.64	1.56	1.13	1.75	2.14	2.55	3.14	3.02	2.53
Madre de Dios	1.00	1.14	1.17	1.68	1.78	2.64	3.33	2.70	2.54	2.06	2.17	3.23	2.74
Moquegua	2.59	4.09	3.64	4.66	4.17	4.81	4.14	3.99	4.36	4.61	4.19	4.42	4.40
Pasco	1.64	1.63	1.53	1.83	1.32	1.44	1.29	1.45	1.55	1.33	1.47	1.44	1.41
Piura	2.16	2.19	2.18	2.44	2.73	3.03	3.07	2.76	2.73	2.46	2.56	3.18	2.85
Puno	1.13	1.00	1.34	1.25	1.44	1.51	1.52	1.64	1.52	1.29	1.58	1.78	1.57
San Martín	1.81	2.12	1.73	2.36	1.88	2.23	1.95	2.07	2.01	1.96	2.14	2.41	2.14
Tacna	3.91	4.86	4.50	5.36	4.59	4.73	4.83	5.07	4.89	5.55	5.67	6.04	5.46
Tumbes	1.60	1.53	1.84	2.46	2.95	3.72	3.62	3.93	4.44	3.62	3.56	3.28	3.75
Ucayali	2.16	1.98	2.15	2.24	2.08	1.94	2.05	1.93	2.16	2.20	2.41	2.49	2.24

Anexo 15: Índice de salud en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	3.41	3.38	3.41	3.37	3.28	3.27	3.22	3.15	3.12	3.04	3.00	2.98	2.95
Ancash	4.40	4.45	4.48	4.46	4.46	4.44	4.41	4.41	4.43	4.36	4.39	4.36	4.33
Apurímac	3.08	3.13	3.14	3.07	3.00	2.94	2.85	2.81	2.80	2.70	2.71	2.68	2.65
Arequipa	5.82	5.86	5.91	5.91	5.95	5.95	6.01	6.04	6.06	6.08	6.07	6.04	6.03
Ayacucho	2.80	2.79	2.83	2.85	2.82	2.82	2.85	2.86	2.89	2.85	2.90	2.87	2.87
Cajamarca	4.07	4.02	4.05	4.00	3.93	3.92	3.86	3.81	3.77	3.70	3.68	3.59	3.61
Cusco	2.45	2.51	2.55	2.54	2.53	2.54	2.52	2.55	2.61	2.58	2.64	2.64	2.65
Huancavelica	2.73	2.71	2.79	2.74	2.63	2.61	2.54	2.52	2.49	2.46	2.43	2.42	2.39
Huánuco	3.24	3.29	3.32	3.30	3.29	3.28	3.29	3.31	3.35	3.30	3.35	3.27	3.30
Ica	6.51	6.56	6.63	6.65	6.69	6.73	6.76	6.78	6.72	6.70	6.64	6.58	6.53
Junín	4.04	4.05	4.07	4.04	3.96	3.93	3.87	3.82	3.78	3.69	3.65	3.62	3.59
La Libertad	5.58	5.57	5.63	5.66	5.74	5.79	5.82	5.83	5.91	5.89	5.90	5.85	5.85
Lambayeque	5.78	5.79	5.85	5.86	5.93	5.93	5.94	6.01	6.04	6.02	6.07	6.00	5.94
Lima	7.00	7.00	7.00	7.00	7.00	6.95	6.95	6.95	7.00	7.00	7.00	7.00	7.00
Loreto	2.88	2.89	2.91	2.90	2.87	2.86	2.87	2.89	2.87	2.87	2.85	2.85	2.83
Madre de Dios	3.02	3.05	3.12	3.13	3.20	3.23	3.26	3.26	3.34	3.31	3.34	3.38	3.33
Moquegua	5.70	5.72	5.78	5.80	5.86	5.89	5.90	5.96	6.01	5.98	6.02	5.96	5.91
Pasco	3.27	3.26	3.24	3.21	3.12	3.04	2.97	2.90	2.87	2.75	2.73	2.67	2.62
Piura	4.05	4.04	4.07	4.04	4.04	4.00	3.96	3.92	3.92	3.86	3.82	3.81	3.75
Puno	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.05	1.00	1.05	1.05	1.05
San Martín	2.99	3.00	2.99	2.98	2.90	2.85	2.80	2.76	2.76	2.71	2.66	2.67	2.62
Tacna	4.87	4.93	4.98	4.99	5.04	5.07	5.12	5.20	5.25	5.22	5.28	5.27	5.23
Tumbes	4.55	4.56	4.60	4.58	4.58	4.56	4.54	4.49	4.52	4.46	4.42	4.40	4.35
Ucayali	2.45	2.48	2.51	2.51	2.51	2.49	2.47	2.47	2.49	2.47	2.51	2.49	2.52

Anexo 16: Índice de educación en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	1.20	1.65	1.79	1.67	1.77	1.76	1.36	1.73	2.43	1.30	2.72	2.94	2.11
Ancash	3.86	4.25	4.36	4.54	4.56	4.58	4.40	5.05	4.84	4.51	4.48	4.31	4.69
Apurímac	4.25	3.84	4.42	3.73	3.75	3.74	4.10	4.09	4.15	3.62	4.42	3.99	4.09
Arequipa	6.71	7.00	6.51	6.99	6.40	6.57	6.03	6.56	6.38	6.70	6.64	6.41	6.74
Ayacucho	3.00	2.73	3.25	3.39	3.83	3.47	3.26	3.38	3.87	3.64	4.29	3.80	3.81
Cajamarca	1.47	1.56	1.44	2.30	2.59	1.94	1.95	2.04	2.46	1.90	2.44	1.95	2.19
Cusco	4.41	4.16	4.81	4.73	5.43	5.04	4.68	4.92	4.94	4.67	4.38	4.34	4.92
Huancavelica	1.98	2.64	2.76	2.76	2.80	2.92	2.85	3.01	3.77	3.69	3.95	3.41	3.53
Huánuco	1.60	1.92	1.52	1.98	2.07	2.65	2.25	1.91	2.10	1.87	2.19	1.81	2.03
Ica	5.92	6.22	6.34	5.95	6.11	5.95	6.29	6.63	6.29	6.18	6.72	6.56	6.48
Junín	4.54	4.45	5.50	5.01	5.40	5.20	5.16	4.35	4.40	4.13	4.54	4.25	4.84
La Libertad	3.26	3.44	3.91	4.00	3.80	4.25	3.72	3.81	4.10	3.52	3.69	3.92	3.86
Lambayeque	3.73	3.68	4.56	4.66	4.19	4.46	4.48	4.74	4.24	4.49	3.96	4.27	4.50
Lima	6.43	6.71	6.79	6.44	6.60	6.67	6.33	6.80	6.52	6.15	6.59	6.48	6.50
Loreto	1.68	1.94	1.80	1.86	1.65	1.86	1.75	1.86	1.80	1.77	2.20	1.80	1.81
Madre de Dios	4.94	5.17	5.49	5.25	4.98	4.84	5.07	5.43	4.62	4.01	4.79	5.22	5.00
Moquegua	5.93	6.45	6.67	6.22	6.36	6.43	6.47	6.57	5.82	5.87	6.60	6.50	6.48
Pasco	4.16	4.45	4.55	4.64	4.95	4.85	4.75	4.28	4.33	4.00	4.74	4.58	4.56
Piura	3.34	2.93	3.74	3.56	4.53	4.11	3.47	3.69	3.69	3.36	3.15	3.59	3.78
Puno	4.87	4.40	4.55	4.59	5.17	5.04	4.81	5.25	4.75	4.96	4.83	5.33	5.16
San Martín	2.13	2.87	2.88	3.23	3.05	3.35	2.72	2.92	3.16	2.10	1.50	2.40	2.74
Tacna	6.23	6.70	6.80	6.79	6.20	6.57	6.00	6.17	6.40	6.45	6.36	6.39	6.51
Tumbes	4.50	3.69	4.77	4.75	5.35	5.52	4.49	5.53	5.46	5.32	4.92	5.00	5.46
Ucayali	3.21	2.99	3.17	3.77	3.39	3.35	2.65	2.16	2.68	3.50	2.90	2.89	2.95

Anexo 17: Índice de distribución en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	3.18	3.34	3.30	3.22	3.64	3.76	3.23	3.00	3.61	3.62	3.60	3.59	2.82
Ancash	4.84	5.56	5.86	5.92	6.16	5.89	6.10	5.88	5.91	6.19	6.22	5.67	5.17
Apurímac	4.42	5.44	5.08	5.59	5.47	5.32	5.45	5.94	6.43	5.91	6.14	6.53	4.57
Arequipa	5.99	6.49	6.38	6.28	6.25	6.54	6.26	6.58	6.75	6.76	6.62	6.75	6.37
Ayacucho	3.50	4.21	4.91	5.15	4.77	4.46	4.64	4.72	5.10	5.25	5.53	5.64	3.84
Cajamarca	3.54	4.47	4.41	4.92	4.81	4.74	4.85	5.28	5.08	5.17	5.38	5.40	3.96
Cusco	4.00	4.40	5.26	5.20	5.55	5.54	5.37	6.00	6.19	5.89	5.74	6.07	4.42
Huancavelica	1.87	3.01	3.61	3.77	4.35	4.28	4.30	4.35	4.78	5.06	5.45	5.23	2.47
Huánuco	3.42	4.68	5.12	5.28	5.02	4.93	4.80	4.68	4.95	4.74	4.40	4.93	3.53
Ica	6.04	5.15	6.14	6.15	6.29	6.58	6.27	6.78	6.60	6.68	7.00	7.00	6.24
Junín	4.33	4.54	4.78	5.21	5.00	5.11	4.80	5.28	4.85	4.81	4.79	4.88	4.48
La Libertad	5.92	6.00	6.35	6.18	6.73	6.26	6.43	6.66	6.33	6.45	6.41	6.68	5.95
Lambayeque	6.04	6.42	6.52	6.19	5.82	5.96	6.22	6.49	6.25	6.73	6.31	6.79	6.09
Lima	7.00	7.00	6.98	7.00	6.89	7.00	6.68	6.90	7.00	7.00	6.96	6.85	7.00
Loreto	1.00	1.16	1.00	1.47	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Madre de Dios	4.41	4.35	4.74	4.84	4.63	4.74	4.55	4.46	4.41	4.53	5.21	4.71	4.02
Moquegua	5.83	6.51	6.82	6.43	6.81	6.60	6.61	7.00	6.83	6.39	6.64	6.57	6.33
Pasco	1.76	1.95	2.59	2.96	2.82	2.44	2.03	2.34	2.76	3.36	3.94	3.64	1.98
Piura	4.49	4.49	4.73	4.91	4.84	4.81	4.45	4.93	5.02	4.84	4.67	4.81	4.06
Puno	4.28	4.52	4.94	5.01	5.08	4.94	4.52	4.97	4.45	4.83	5.03	5.43	4.26
San Martín	3.04	3.14	3.48	3.61	3.37	3.46	2.96	3.23	3.06	3.39	3.74	3.72	2.89
Tacna	6.56	6.93	7.00	6.66	7.00	6.77	7.00	6.89	6.83	6.93	7.00	6.85	6.94
Tumbes	3.71	4.03	4.89	4.78	4.71	4.71	4.36	4.45	4.87	4.59	4.85	4.60	3.90
Ucayali	1.15	1.00	1.09	1.00	1.99	2.25	1.66	1.94	2.37	2.53	2.71	2.68	1.85

Anexo 18: Índice de estabilidad económica en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	3.49	3.08	3.75	5.58	3.86	4.17	6.13	5.24	2.08	5.14	3.91	4.96	6.82
Ancash	2.72	3.41	4.40	6.15	2.30	6.33	6.03	5.21	3.43	5.59	5.57	3.88	4.63
Apurímac	4.03	4.15	3.20	6.26	4.55	3.05	4.81	3.69	5.30	4.40	4.53	2.80	5.43
Arequipa	2.56	4.08	2.88	6.14	1.83	3.33	2.11	3.85	4.80	4.46	2.51	4.59	4.30
Ayacucho	6.13	6.24	1.09	5.07	1.58	4.26	4.19	3.69	5.00	4.98	2.61	3.81	4.96
Cajamarca	5.89	6.47	1.31	6.32	3.27	5.15	4.66	5.44	6.97	5.19	6.42	5.02	4.31
Cusco	2.19	1.00	2.86	7.00	2.08	3.88	1.00	4.56	4.17	3.61	4.84	2.95	4.86
Huancavelica	5.87	4.51	3.50	5.26	4.70	5.60	4.78	3.37	1.00	6.80	7.00	1.19	4.38
Huánuco	6.37	4.47	1.03	6.34	5.50	2.47	4.08	3.05	3.88	5.50	5.73	3.72	3.57
Ica	3.37	4.64	4.15	6.13	1.54	2.36	4.35	1.00	3.18	3.10	3.61	1.00	3.66
Junín	4.93	5.09	2.09	6.44	3.62	5.86	3.92	4.50	4.58	3.41	4.76	2.55	3.36
La Libertad	4.13	4.73	1.00	6.51	3.78	4.68	4.21	4.00	5.99	3.13	3.00	3.22	6.94
Lambayeque	4.31	4.65	2.50	6.12	2.13	4.32	4.75	2.99	4.64	4.16	2.51	7.00	5.63
Lima	7.00	6.69	2.10	1.00	4.41	4.34	4.53	3.83	3.62	4.35	4.29	6.22	4.67
Loreto	5.28	6.76	5.23	5.76	4.58	4.83	4.20	3.37	5.94	4.46	5.84	4.72	1.00
Madre de Dios	5.82	4.60	2.95	6.33	7.00	6.33	5.58	7.00	7.00	7.00	3.11	4.91	6.82
Moquegua	1.00	4.57	2.87	5.52	3.35	3.63	4.10	4.15	6.79	2.86	4.93	5.21	4.71
Pasco	3.70	3.43	7.00	5.59	3.69	5.93	4.21	4.20	4.84	5.55	6.90	2.09	5.08
Piura	5.18	6.08	1.36	6.56	2.15	4.62	4.56	3.75	4.58	4.41	1.00	6.33	4.29
Puno	5.86	7.00	1.43	6.43	3.81	1.00	3.20	3.36	3.78	1.00	2.48	1.60	5.29
San Martín	4.19	4.57	3.05	5.12	4.46	5.35	7.00	5.69	2.15	5.28	4.12	5.23	7.00
Tacna	3.84	3.79	2.87	5.73	2.80	4.08	3.29	3.97	5.41	3.16	4.34	2.72	4.97
Tumbes	4.32	5.09	2.42	5.54	2.46	3.85	4.49	5.05	5.03	4.76	1.66	3.68	4.94
Ucayali	4.16	6.45	1.61	6.19	1.00	7.00	4.72	3.92	5.60	3.41	6.07	6.01	4.05

Anexo 19: Índice de performance económico en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	3.79	3.70	3.76	3.92	4.12	3.58	4.43	3.23	3.79	2.67	2.78	3.58	4.20
Ancash	3.13	3.04	2.83	2.36	2.25	2.30	3.61	2.32	1.71	3.03	2.32	2.79	3.39
Apurímac	3.56	2.26	2.64	2.39	3.65	3.53	4.71	3.64	3.78	3.27	5.01	4.38	2.48
Arequipa	3.39	2.72	2.34	1.71	2.35	2.16	2.24	1.26	2.30	1.70	2.10	2.41	1.84
Ayacucho	3.85	3.89	3.78	2.91	3.29	2.69	3.39	2.86	2.70	2.46	2.17	3.19	3.46
Cajamarca	2.39	3.96	4.01	3.72	3.06	3.88	3.61	2.06	2.66	2.32	2.62	2.92	3.56
Cusco	3.40	2.72	2.68	4.01	4.02	4.42	2.45	3.57	3.23	2.19	2.35	2.42	2.69
Huancavelica	2.98	3.69	3.42	3.61	3.17	3.46	4.06	2.95	4.03	2.83	2.42	2.82	4.07
Huánuco	3.18	3.25	3.29	2.52	3.41	3.13	3.74	3.14	3.28	3.04	2.24	3.53	3.22
Ica	3.17	3.31	3.13	2.26	2.93	2.67	3.01	2.88	2.97	2.72	2.29	3.23	2.79
Junín	3.47	2.73	2.91	1.05	2.94	3.41	3.81	2.14	3.59	3.12	2.56	3.03	3.02
La Libertad	3.33	2.89	2.63	2.23	2.99	2.61	2.67	1.39	2.50	2.24	2.26	2.39	2.90
Lambayeque	3.14	2.17	2.86	2.61	3.02	3.13	2.83	1.37	2.73	2.54	2.00	2.58	2.49
Lima	2.99	1.96	2.17	1.43	2.76	2.63	2.67	1.48	2.07	1.14	0.86	1.41	2.00
Loreto	3.11	2.88	2.81	2.46	3.10	2.23	3.37	2.51	3.18	1.98	2.30	3.11	3.98
Madre de Dios	3.94	2.90	3.00	3.57	3.55	3.80	2.01	3.89	2.00	4.11	3.04	2.43	2.21
Moquegua	1.29	2.46	2.46	1.78	1.37	0.64	2.07	2.29	2.08	1.67	1.66	2.25	1.63
Pasco	3.02	1.68	2.33	1.86	1.53	1.89	2.56	1.34	2.12	2.01	1.79	2.24	1.85
Piura	3.12	2.18	2.38	2.10	3.14	3.24	2.21	2.07	3.20	2.18	2.29	2.36	2.93
Puno	3.65	3.15	3.50	3.39	3.47	3.52	3.46	2.80	2.71	2.16	2.31	2.66	3.40
San Martín	3.75	3.56	3.08	3.34	3.75	3.51	4.40	2.54	3.76	3.16	2.53	3.71	3.38
Tacna	2.97	1.77	1.04	1.27	3.08	1.67	2.55	2.57	2.69	2.59	1.69	2.64	2.97
Tumbes	3.21	3.05	3.29	2.18	2.91	1.17	2.69	1.43	2.81	1.80	1.69	2.06	2.51
Ucayali	2.74	2.81	3.05	2.64	2.96	3.38	3.61	2.48	2.98	3.04	2.33	2.81	3.10

Anexo 20: Índice de desempeño general en las regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	2.72	2.75	2.87	3.18	3.00	3.02	3.29	3.06	2.80	2.97	3.09	3.45	3.51
Ancash	3.55	3.81	4.05	4.40	3.84	4.45	4.69	4.37	3.92	4.45	4.29	3.96	4.22
Apurímac	3.42	3.34	3.26	3.68	3.60	3.34	3.89	3.59	3.96	3.48	4.06	3.64	3.43
Arequipa	4.88	5.10	4.86	5.47	4.63	4.95	4.60	4.88	5.30	5.20	4.93	5.19	5.10
Ayacucho	3.46	3.54	2.92	3.52	2.96	3.25	3.29	3.19	3.50	3.42	3.15	3.47	3.40
Cajamarca	3.14	3.70	2.87	3.91	3.25	3.52	3.43	3.34	3.76	3.30	3.70	3.45	3.21
Cusco	3.04	2.81	3.33	4.25	3.59	3.96	3.01	4.00	3.91	3.52	3.67	3.46	3.63
Huancavelica	2.75	2.93	2.84	3.19	3.11	3.31	3.25	2.87	2.85	3.65	3.71	2.68	2.97
Huánuco	3.25	3.19	2.64	3.55	3.51	3.06	3.32	3.01	3.28	3.35	3.27	3.18	2.91
Ica	4.61	4.74	4.92	5.08	4.61	4.75	5.17	4.75	5.10	4.97	5.19	4.86	5.06
Junín	3.89	3.81	3.55	4.01	3.88	4.27	4.00	3.75	4.04	3.74	3.94	3.60	3.72
La Libertad	4.38	4.40	3.80	4.73	4.40	4.62	4.46	4.27	4.82	4.23	4.18	4.29	4.90
Lambayeque	4.31	4.34	4.25	4.86	4.08	4.59	4.70	4.31	4.74	4.72	4.20	5.12	4.81
Lima	6.24	6.06	5.34	4.98	5.78	5.76	5.69	5.49	5.53	5.44	5.45	5.83	5.69
Loreto	2.60	2.88	2.53	2.66	2.47	2.39	2.39	2.23	2.82	2.44	2.89	2.75	2.19
Madre de Dios	3.86	3.54	3.41	4.13	4.19	4.26	3.97	4.46	3.98	4.17	3.61	3.98	4.02
Moquegua	3.72	4.97	4.70	5.07	4.65	4.67	4.88	4.99	5.31	4.56	5.01	5.15	4.91
Pasco	2.93	2.73	3.54	3.35	2.90	3.26	2.97	2.75	3.08	3.17	3.59	2.78	2.92
Piura	3.72	3.65	3.08	3.93	3.57	3.97	3.62	3.52	3.86	3.52	2.92	4.01	3.61
Puno	3.46	3.51	2.79	3.61	3.33	2.84	3.08	3.17	3.04	2.54	2.88	2.98	3.45
San Martín	2.98	3.21	2.87	3.44	3.24	3.46	3.64	3.20	2.82	3.10	2.78	3.36	3.46
Tacna	4.73	4.83	4.53	5.13	4.78	4.81	4.80	4.98	5.25	4.98	5.06	4.99	5.35
Tumbes	3.65	3.66	3.63	4.05	3.82	3.92	4.03	4.15	4.52	4.09	3.51	3.84	4.15
Ucayali	2.65	2.95	2.26	3.06	2.32	3.40	2.86	2.48	3.05	2.86	3.15	3.23	2.78

Anexo 21: Índice de eficiencia insumos de salud por regiones en el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.37	0.46	0.53	0.59	0.54	0.35	0.48	0.52	0.39	0.49	0.20	0.20	0.52
Ancash	0.16	0.29	0.21	0.26	0.29	0.13	0.22	0.24	0.24	0.27	0.24	0.23	0.24
Apurímac	0.40	0.34	0.33	0.35	0.32	0.31	0.28	0.35	0.31	0.34	0.29	0.30	0.31
Arequipa	0.15	0.17	0.18	0.19	0.22	0.22	0.21	0.23	0.19	0.19	0.18	0.18	0.19
Ayacucho	0.29	0.37	0.27	0.25	0.33	0.23	0.25	0.16	0.17	0.20	0.22	0.20	0.24
Cajamarca	0.23	0.33	0.19	0.20	0.26	0.14	0.17	0.21	0.19	0.21	0.19	0.19	0.20
Cusco	0.20	0.19	0.19	0.21	0.23	0.18	0.13	0.20	0.18	0.20	0.21	0.18	0.19
Huancavelica	0.56	0.40	0.43	0.42	0.41	0.29	0.31	0.41	0.33	0.36	0.36	0.41	0.34
Huánuco	0.33	0.36	0.31	0.34	0.33	0.21	0.21	0.29	0.26	0.30	0.31	0.29	0.27
Ica	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.29	0.29	1.00	1.00	1.00	1.00
Junín	0.19	0.18	0.22	0.24	0.22	0.19	0.19	0.14	0.17	0.19	0.16	0.18	0.17
La Libertad	0.17	0.18	0.19	0.20	0.20	0.17	0.17	0.18	0.18	0.19	0.18	0.17	0.18
Lambayeque	0.23	0.31	0.24	0.24	0.24	0.30	0.28	0.26	0.26	0.27	0.25	0.26	0.27
Lima	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Loreto	0.16	0.27	0.24	0.27	0.27	0.17	0.23	0.24	0.20	0.21	0.19	0.21	0.21
Madre de Dios	1.00	1.00	0.94	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Moquegua	1.00	1.00	0.49	1.00	1.00	1.00	1.00	0.51	0.66	1.00	1.00	1.00	1.00
Pasco	0.70	0.90	0.69	1.00	0.69	0.50	0.72	0.96	0.63	0.53	0.43	0.79	0.50
Piura	0.12	0.20	0.22	0.24	0.26	0.16	0.19	0.20	0.20	0.22	0.19	0.18	0.18
Puno	0.22	0.25	0.21	0.21	0.23	0.17	0.19	0.26	0.22	0.26	0.22	0.21	0.21
San Martín	0.31	0.35	0.35	0.39	0.23	0.27	0.24	0.22	0.18	0.23	0.28	0.32	0.30
Tacna	1.00	0.49	0.49	0.50	0.41	0.48	0.53	0.62	0.64	1.00	1.00	0.45	0.44
Tumbes	1.00	1.00	1.00	0.49	0.61	0.47	0.71	1.00	0.70	1.00	1.00	1.00	1.00
Ucayali	0.48	0.55	0.45	0.48	0.49	0.36	0.48	0.50	0.39	0.31	0.30	0.35	0.29

Anexo 22: Índice de eficiencia resultados de salud por regiones en el periodo 2007- 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.49	0.48	0.49	0.48	0.47	0.47	0.46	0.45	0.45	0.43	0.43	0.43	0.42
Ancash	0.63	0.64	0.64	0.64	0.64	0.64	0.63	0.63	0.63	0.62	0.63	0.62	0.62
Apurímac	0.44	0.45	0.45	0.44	0.43	0.42	0.41	0.40	0.40	0.39	0.39	0.38	0.38
Arequipa	0.83	0.84	0.84	0.84	0.85	0.86	0.86	0.87	0.87	0.87	0.87	0.86	0.86
Ayacucho	0.40	0.40	0.40	0.41	0.40	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
Cajamarca	0.58	0.57	0.58	0.57	0.56	0.56	0.56	0.55	0.54	0.53	0.53	0.51	0.52
Cusco	0.35	0.36	0.36	0.36	0.36	0.37	0.36	0.37	0.37	0.37	0.37	0.38	0.38
Huancavelica	0.39	0.39	0.40	0.39	0.38	0.37	0.36	0.36	0.36	0.35	0.35	0.35	0.34
Huánuco	0.46	0.47	0.47	0.47	0.47	0.47	0.47	0.48	0.48	0.47	0.48	0.47	0.47
Ica	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Junín	0.58	0.58	0.58	0.58	0.57	0.56	0.56	0.55	0.54	0.53	0.52	0.52	0.51
La Libertad	0.80	0.80	0.80	0.81	0.82	0.83	0.84	0.84	0.84	0.84	0.84	0.84	0.84
Lambayeque	0.83	1.00	0.84	0.84	0.85	0.85	0.86	0.86	0.86	0.86	0.87	0.86	0.85
Lima	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Loreto	0.41	0.41	0.42	0.41	0.41	0.41	0.41	0.42	0.41	0.41	0.41	0.41	0.40
Madre de Dios	1.00	1.00	0.45	1.00	0.46	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Moquegua	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pasco	0.47	0.47	0.46	1.00	0.45	0.44	0.43	0.42	0.41	0.39	0.39	0.38	0.37
Piura	0.58	0.58	0.58	0.58	0.58	0.58	0.57	0.56	0.56	0.55	0.55	0.54	0.54
Puno	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.14	0.15	0.15	0.15
San Martín	0.43	0.43	0.43	0.43	0.41	0.41	0.40	0.40	0.39	0.39	0.38	0.38	0.37
Tacna	1.00	0.70	0.71	0.71	0.72	0.73	0.74	1.00	0.75	1.00	0.75	0.75	0.75
Tumbes	1.00	1.00	1.00	0.65	0.65	0.66	0.65	1.00	1.00	0.64	1.00	1.00	1.00
Ucayali	0.35	0.35	0.36	0.36	0.36	0.36	0.35	0.36	0.36	0.35	0.36	0.36	0.36

Anexo 23: Índice de eficiencia insumos de educación por regiones en el periodo 2007-2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.03	0.03	0.02	0.19	0.14	0.16	0.24	0.16	0.11	0.14	0.17	0.16	0.17
Ancash	0.07	0.07	0.07	0.43	0.54	0.39	0.76	0.43	0.60	0.56	0.60	0.52	0.54
Apurímac	0.02	0.02	0.02	0.11	0.07	0.12	0.12	0.11	0.14	0.25	0.41	0.31	0.32
Arequipa	1.00	1.00	0.11	1.00	1.00	0.71	0.88	0.76	0.62	1.00	1.00	0.88	0.89
Ayacucho	0.03	0.03	0.03	0.18	0.13	0.17	0.22	0.15	0.16	0.17	0.21	0.18	0.21
Cajamarca	0.04	0.04	0.04	0.28	0.29	0.26	0.34	0.23	0.27	0.22	0.24	0.20	0.23
Cusco	0.06	0.05	0.07	0.40	0.38	0.39	0.50	0.42	0.35	0.49	0.59	0.43	0.46
Huancavelica	0.03	0.03	0.03	0.14	0.13	0.16	0.18	0.13	0.16	0.14	0.17	0.16	0.17
Huánuco	0.04	0.03	0.03	0.18	0.13	0.14	0.18	0.15	0.22	0.18	0.23	0.19	0.23
Ica	0.10	0.11	0.11	0.60	0.32	0.65	0.66	0.69	0.74	0.81	1.00	0.80	0.88
Junín	0.05	0.06	0.05	0.35	0.27	0.37	0.39	0.37	0.44	0.38	0.50	0.42	0.44
La Libertad	0.08	0.08	0.08	0.51	0.40	0.45	0.61	0.41	0.50	0.44	0.54	0.49	0.55
Lambayeque	0.06	0.06	0.07	0.40	0.27	0.41	0.52	0.42	0.43	0.43	0.50	0.50	0.51
Lima	1.00	1.00	1.00	1.00	1.00	1.00	0.56	1.00	1.00	0.75	1.00	1.00	1.00
Loreto	0.04	0.04	0.04	0.29	0.23	0.26	0.36	0.27	0.33	0.21	0.25	0.23	0.25
Madre de Dios	0.09	0.08	0.08	0.58	0.30	0.38	0.41	0.25	0.32	0.42	0.38	0.26	0.31
Moquegua	0.16	0.17	0.15	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00
Pasco	0.11	0.07	0.09	0.39	0.39	0.36	0.65	0.39	0.44	0.45	0.54	0.42	0.46
Piura	0.07	0.07	0.07	0.46	0.47	0.45	0.64	0.43	0.42	0.43	0.45	0.43	0.46
Puno	0.03	0.03	0.03	0.21	0.19	0.19	0.29	0.18	0.23	0.20	0.25	0.22	0.23
San Martín	0.03	0.03	0.03	0.24	0.20	0.23	0.21	0.21	0.21	0.22	0.25	0.22	0.22
Tacna	0.08	0.06	1.00	0.43	0.43	0.46	0.56	0.56	0.58	0.61	0.68	0.70	0.85
Tumbes	0.03	0.04	0.04	0.30	0.14	0.20	0.33	0.22	0.28	0.28	0.30	0.28	0.32
Ucayali	0.04	0.04	0.04	0.23	0.20	0.23	0.34	0.28	0.28	0.25	0.29	0.24	0.27

Anexo 24: Índice de eficiencia resultados de educación por regiones en el periodo 2007-2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.44	0.45	0.54	0.58	0.52	0.52	0.58	0.56	0.51	0.55	0.57	0.59	0.62
Ancash	0.57	0.63	0.76	0.80	0.66	0.77	0.82	0.80	0.71	0.82	0.79	0.68	0.74
Apurímac	0.55	0.55	0.61	0.67	0.62	0.58	0.68	0.65	0.72	0.64	0.75	0.62	0.60
Arequipa	1.00	1.00	1.00	1.00	0.80	1.00	0.81	0.89	0.96	0.96	0.90	1.00	1.00
Ayacucho	0.56	0.58	0.55	0.64	0.51	0.56	0.58	0.58	0.63	0.63	0.58	0.60	0.60
Cajamarca	0.50	0.61	0.54	0.71	0.56	0.61	0.60	0.61	0.68	0.61	0.68	0.59	0.56
Cusco	0.49	0.46	0.62	0.78	0.62	0.69	0.53	0.73	0.71	0.65	0.67	0.59	0.64
Huancavelica	0.44	0.48	0.53	0.58	0.54	0.57	0.57	0.52	0.51	0.67	0.68	0.46	0.52
Huánuco	0.52	0.53	0.49	0.65	0.61	0.53	0.58	0.55	0.59	0.62	0.60	0.54	0.51
Ica	0.74	0.78	1.00	1.00	0.80	1.00	0.91	0.87	0.92	0.91	1.00	0.83	0.89
Junín	0.62	0.63	0.66	0.73	0.67	0.74	0.70	0.68	0.73	0.69	0.72	0.62	0.65
La Libertad	0.70	0.73	0.71	0.86	0.76	0.80	0.78	0.78	0.87	0.78	0.77	0.74	0.86
Lambayeque	0.69	0.72	0.80	0.89	0.71	0.80	0.82	0.79	0.86	0.87	0.77	0.88	0.85
Lima	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Loreto	0.42	0.48	0.47	0.49	0.43	0.41	0.42	0.41	0.51	0.45	0.53	0.47	0.38
Madre de Dios	0.62	0.58	0.64	0.75	0.73	0.74	0.70	0.81	0.72	0.77	0.66	0.68	0.71
Moquegua	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pasco	1.00	0.45	0.66	0.61	0.50	0.57	0.52	0.50	0.56	0.58	0.66	0.48	0.51
Piura	0.60	0.60	0.58	0.72	0.62	0.69	0.64	0.64	0.70	0.65	0.53	0.69	0.63
Puno	0.56	0.58	0.52	0.66	0.58	0.49	0.54	0.58	0.55	0.47	0.53	0.51	0.61
San Martín	0.48	0.53	0.54	0.63	0.56	0.60	0.64	0.58	0.51	0.57	0.51	0.58	0.61
Tacna	0.76	0.80	0.85	0.94	1.00	0.83	0.84	0.91	0.95	0.92	0.93	0.86	1.00
Tumbes	0.58	0.60	0.68	0.74	0.66	0.68	0.71	0.76	0.82	0.75	0.64	0.66	0.73
Ucayali	0.42	0.49	0.42	0.56	0.40	0.59	0.50	0.45	0.55	0.53	0.58	0.55	0.49

Anexo 25: Índice de eficiencia insumos general por regiones en el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.22	0.22	0.18	0.22	0.17	0.26	0.23	0.23	0.16	0.21	0.17	0.10	0.20
Ancash	0.74	0.66	0.55	0.51	0.48	0.62	0.65	0.60	0.60	0.74	0.68	0.65	0.54
Apurímac	0.28	0.22	0.18	0.16	0.14	0.18	0.17	0.18	0.14	0.36	0.42	0.35	0.40
Arequipa	1.00	1.00	1.00	1.00	0.66	1.00	0.69	0.75	0.62	0.89	0.87	1.00	1.00
Ayacucho	0.26	0.28	0.26	0.28	0.18	0.20	0.21	0.21	0.16	0.23	0.23	0.23	0.26
Cajamarca	0.49	0.49	0.47	0.44	0.30	0.35	0.34	0.35	0.27	0.32	0.32	0.32	0.33
Cusco	0.54	0.41	0.37	0.33	0.39	0.39	0.38	0.43	0.35	0.47	0.52	0.41	0.50
Huancavelica	0.33	0.30	0.28	0.28	0.24	0.25	0.24	0.22	0.16	0.19	0.20	0.27	0.23
Huánuco	0.37	0.30	0.28	0.32	0.25	0.28	0.28	0.28	0.22	0.26	0.28	0.24	0.25
Ica	0.72	0.86	1.00	1.00	0.68	1.00	0.89	0.93	0.74	0.93	1.00	0.70	0.72
Junín	0.56	0.53	0.43	0.50	0.39	0.52	0.44	0.54	0.44	0.49	0.51	0.43	0.49
La Libertad	0.76	0.68	0.64	0.67	0.56	0.70	0.61	0.66	0.50	0.57	0.61	0.34	0.65
Lambayeque	0.51	0.48	0.46	0.51	0.39	0.53	0.52	0.57	0.43	0.52	0.47	0.29	0.43
Lima	1.00	1.00	1.00	0.45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Loreto	0.48	0.48	0.50	0.51	0.33	0.44	0.41	0.44	0.33	0.34	0.33	0.28	0.35
Madre de Dios	0.64	0.60	0.22	0.44	0.42	0.38	0.34	0.36	0.32	0.43	0.38	0.19	0.17
Moquegua	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pasco	1.00	0.67	0.57	0.64	0.53	0.63	0.50	0.66	0.44	0.52	0.56	0.45	0.44
Piura	0.62	0.62	0.60	0.64	0.55	0.62	0.56	0.60	0.42	0.53	0.49	0.28	0.44
Puno	0.37	0.34	0.29	0.31	0.24	0.32	0.29	0.29	0.23	0.26	0.28	0.21	0.23
San Martín	0.30	0.29	0.31	0.34	0.28	0.32	0.29	0.32	0.21	0.27	0.28	0.22	0.27
Tacna	0.53	0.46	0.40	0.51	1.00	0.58	0.53	0.59	0.58	0.51	0.51	0.54	1.00
Tumbes	0.31	0.39	0.35	0.40	0.30	0.35	0.37	0.39	0.28	0.36	0.37	0.19	0.29
Ucayali	0.41	0.37	0.35	0.38	0.35	0.42	0.36	0.35	0.28	0.28	0.29	0.30	0.37

Anexo 26: Índice de eficiencia resultados general por regiones para el periodo 2007 - 2019

Año	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Departamentos													
Amazonas	0.44	0.45	0.54	0.58	0.52	0.52	0.58	0.56	0.51	0.55	0.57	0.59	0.62
Ancash	0.57	0.63	0.76	0.80	0.66	0.77	0.82	0.80	0.71	0.82	0.79	0.68	0.74
Apurímac	0.55	0.55	0.61	0.67	0.62	0.58	0.68	0.65	0.72	0.64	0.75	0.62	0.60
Arequipa	1.00	1.00	1.00	1.00	0.80	1.00	0.81	0.89	0.96	0.96	0.90	1.00	1.00
Ayacucho	0.56	0.58	0.55	0.64	0.51	0.56	0.58	0.58	0.63	0.63	0.58	0.60	0.60
Cajamarca	0.50	0.61	0.54	0.71	0.56	0.61	0.60	0.61	0.68	0.61	0.68	0.59	0.56
Cusco	0.49	0.46	0.62	0.78	0.62	0.69	0.53	0.73	0.71	0.65	0.67	0.59	0.64
Huancavelica	0.44	0.48	0.53	0.58	0.54	0.57	0.57	0.52	0.51	0.67	0.68	0.46	0.52
Huánuco	0.52	0.53	0.49	0.65	0.61	0.53	0.58	0.55	0.59	0.62	0.60	0.54	0.51
Ica	0.74	0.78	1.00	1.00	0.80	1.00	0.91	0.87	0.92	0.91	1.00	0.83	0.89
Junín	0.62	0.63	0.66	0.73	0.67	0.74	0.70	0.68	0.73	0.69	0.72	0.62	0.65
La Libertad	0.70	0.73	0.71	0.86	0.76	0.80	0.78	0.78	0.87	0.78	0.77	0.74	0.86
Lambayeque	0.69	0.72	0.80	0.89	0.71	0.80	0.82	0.79	0.86	0.87	0.77	0.88	0.85
Lima	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Loreto	0.42	0.48	0.47	0.49	0.43	0.41	0.42	0.41	0.51	0.45	0.53	0.47	0.38
Madre de Dios	0.62	0.58	0.64	0.75	0.73	0.74	0.70	0.81	0.72	0.77	0.66	0.68	0.71
Moquegua	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pasco	1.00	0.45	0.66	0.61	0.50	0.57	0.52	0.50	0.56	0.58	0.66	0.48	0.51
Piura	0.60	0.60	0.58	0.72	0.62	0.69	0.64	0.64	0.70	0.65	0.53	0.69	0.63
Puno	0.56	0.58	0.52	0.66	0.58	0.49	0.54	0.58	0.55	0.47	0.53	0.51	0.61
San Martín	0.48	0.53	0.54	0.63	0.56	0.60	0.64	0.58	0.51	0.57	0.51	0.58	0.61
Tacna	0.76	0.80	0.85	0.94	1.00	0.83	0.84	0.91	0.95	0.92	0.93	0.86	1.00
Tumbes	0.58	0.60	0.68	0.74	0.66	0.68	0.71	0.76	0.82	0.75	0.64	0.66	0.73
Ucayali	0.42	0.49	0.42	0.56	0.40	0.59	0.50	0.45	0.55	0.53	0.58	0.55	0.49

Anexo 27: Funciones y Procedimientos en R

```
# CARGANDO PAQUETES
library(proxy)
library(dtwclust)
library(cluster)
library(purrr)
library(carData)
library(ggplot2)

# COPIANDO LOS DATOS EN BRUTO
data1<-read.delim("clipboard")

# TRANSFORMANDO LOS DATOS EN LISTAS
data1<-tselect(data1[1:24,1:13],simplify="T")

data1

# RENOMBRANDO
nodeD <- list(`1` = "Amazonas",
              `2` = "Ancash",
              `3` = "Apur?mac",
              `4` = "Arequipa",
              `5` = "Ayacucho",
              `6` = "Cajamarca",
              `7` = "Cusco",
              `8` = "Huancavelica",
              `9` = "Hu?nuco",
              `10` = "Ica",
              `11` = "Jun?n",
              `12` = "La Libertad",
              `13` = "Lambayeque",
              `14` = "Lima",
              `15` = "Loreto",
```

```

`16` = "Madre de Dios",
`17` = "Moquegua",
`18` = "Pasco",
`19` = "Piura",
`20` = "Puno",
`21` = "San Mart?n",
`22` = "Tacna",
`23` = "Tumbes",
`24` = "Ucayali")

```

```

rename_items <- function(item){
  if (is.list(item)){
    item <- lapply(item, rename_items)
    names(item) <- unname(nodeD[names(item)])
  }
  item
}
datal2 <- rename_items(datal)
head(datal2)

#=====
# APLICACION A SERIES DE TIEMPO DE CLUSTERS JERARQUICO
#=====

# PAQUETES NECESARIOS PARA PROCESAR LOS DATOS.

library(dtwclust)
library(cluster)
library(tidyverse)
library(gghighlight)

# PARA VER MEDIDAS DE DISTANCIA
summary(pr_DB, "long")

```

```

# SE VALIDO 2 a 10 CLUSTERS

# SE APLICARA EL ENLACE WARD; PRIMERO EL EUCLIDEANO Y LUEGO EL
DTW.

#method = ward.D2

hc_dist<-tsclust(data12, type = "h", k=2L:20L, distance = "DTW2",
                control =hierarchical_control(method = "complete"))
names(hc_dist) <- paste0("k_", 2L:20L)
indicadores<-sapply(hc_dist, cvi, type = "internal")

# Criterios
max(indicadores[1,]) #SILUETA: 2
max(indicadores[2,]) # FUNCION SCORE:
max(indicadores[3,]) #INDICE CALINSKI-HARABASZ:2
min(indicadores[4,]) #INDICE DAVIES-BOULDIN:
min(indicadores[5,]) #INDICE DAVIES-BOULDIN MODIFICADO:
max(indicadores[6,]) # INDICE DUNN:
min(indicadores[7,]) # INDICE COP:

# CORRIENDO EL CLUSTER CORRECTO
hc_dist<-tsclust(data12, type = "h", k=2L, distance = "DTW2",
                control =hierarchical_control(method = "complete"))

# GRAFICO EN LINEA CON SU PROTOTIPO.
par(mfrow = c(1, 1))
plot(hc_dist, type="sc")
plot(hc_dist, type = "centroids", clus = 1L:2L)
plot(hc_dist, family="Arial")

```

```

rect.hclust(hc_dist, k=2, border="cadetblue") # OK

# CORTANDO EL ARBOL
clustered_data <- cutree(hc_dist, k=2)
table(clustered_data)
clustered_data_tidy <- as.data.frame(as.table(clustered_data)) %>% glimpse()
View(clustered_data_tidy)
table(clustered_data_tidy$Freq)
##### PARA INSUMOS #####
data1<-read.delim("clipboard")
# TRANSFORMANDO LOS DATOS EN LISTAS
data1<-tlist(data1[1:24,1:13],simplify="T")
data1
# RENOMBRANDO
nodeD <- list(`1` = "Amazonas",
              `2` = "Ancash",
              `3` = "Apurimac",
              `4` = "Arequipa",
              `5` = "Ayacucho",
              `6` = "Cajamarca",
              `7` = "Cusco",
              `8` = "Huancavelica",
              `9` = "Huanuco",
              `10` = "Ica",
              `11` = "Junin",
              `12` = "La Libertad",
              `13` = "Lambayeque",
              `14` = "Lima",
              `15` = "Loreto",
              `16` = "Madre de Dios",

```

```
`17` = "Moquegua",  
`18` = "Pasco",  
`19` = "Piura",  
`20` = "Puno",  
`21` = "San Martin",  
`22` = "Tacna",  
`23` = "Tumbes",  
`24` = "Ucayali")
```

```
rename_items <- function(item){  
  if (is.list(item)){  
    item <- lapply(item, rename_items)  
    names(item) <- unname(nodeD[names(item)])  
  }  
  item  
}
```

```
datal2 <- rename_items(datal)  
head(datal2)  
View(datal2)
```

```
#=====
```

```
# APLICACION A SERIES DE TIEMPO DE CLUSTERS JERARQUICO
```

```
#=====
```

```
# PAQUETES NECESARIOS PARA PROCESAR LOS DATOS.
```

```
library(dtwclust)  
library(cluster)  
library(tidyverse)  
library(gghighlight)
```

```

# PARA VER MEDIDAS DE DISTANCIA
summary(pr_DB, "long")

# SE VALIDO 2 a 10 CLUSTERS
# SE APLICARA EL ENLACE WARD; PRIMERO EL EUCLIDEANO Y LUEGO EL
DTW.
#method = ward.D2

hc_dist<-tsclust(data12, type = "h", k=2L:20L, distance = "DTW2",
                control =hierarchical_control(method = "complete"))
names(hc_dist) <- paste0("k_", 2L:20L)
indicadores<-sapply(hc_dist, cvi, type = "internal")
# Criterios
max(indicadores[1,]) #SILUETA: 2
max(indicadores[2,]) # FUNCION SCORE:
max(indicadores[3,]) #INDICE CALINSKI-HARABASZ:2
min(indicadores[4,]) #INDICE DAVIES-BOULDIN:
min(indicadores[5,]) #INDICE DAVIES-BOULDIN MODIFICADO:
max(indicadores[6,]) # INDICE DUNN:
min(indicadores[7,]) # INDICE COP:

# CORRIENDO EL CLUSTER CORRECTO
hc_dist<-tsclust(data12, type = "h", k=3L, distance = "DTW2",
                control =hierarchical_control(method = "complete"))
# PROTOTIPO ES DTW.
# armar como hacer el grafico de la data frame.
# GRAFICO EN LINEA CON SU PROTOTIPO.
par(mfrow = c(1, 1))
plot(hc_dist, type="sc")
names(hc_dist)

```

```
str(hc_dist)
hc_dist@centroids
hc_dist@cluster
hc_dist@datalist
plot(hc_dist, type = "centroids", clus = 1L:3L)
plot(hc_dist, family="Arial")
rect.hclust(hc_dist, k=3, border="cadetblue") # OK

# CORTANDO EL ARBOL
clustered_data <- cutree(hc_dist, k=3)
table(clustered_data)
clustered_data_tidy <- as.data.frame(as.table(clustered_data)) %>% glimpse()
View(clustered_data_tidy)
table(clustered_data_tidy$Freq)

#=====
```