

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



**“IDENTIFICACIÓN DE PERFILES DE LOS CENTROS DE
EDUCACIÓN TÉCNICO-PRODUCTIVA PÚBLICOS USANDO
INDICADORES DE CONDICIONES BÁSICAS DE CALIDAD
MEDIANTE CLÚSTER BIETÁPICO”**

**TRABAJO DE SUFICIENCIA PROFESIONAL
PARA OPTAR EL TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO**

CESAR ANTHONY TONCONI CALISAYA

LIMA – PERÚ

2021

La UNALM es titular de los derechos patrimoniales de la presente investigación (Art. 24-
Reglamento de Propiedad Intelectual.

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN**

**“IDENTIFICACIÓN DE PERFILES DE LOS CENTROS DE EDUCACIÓN
TÉCNICO-PRODUCTIVA PÚBLICOS USANDO INDICADORES DE
CONDICIONES BÁSICAS DE CALIDAD MEDIANTE CLÚSTER BIETÁPICO”**

Presentado por:

CESAR ANTHONY TONCONI CALISAYA

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO**

SUSTENTADO Y APROBADO ANTE EL SIGUIENTE JURADO:

Mg. Sc. Celso Gonzáles Chavesta

PRESIDENTE

Dr. Raphael Félix Valencia Chacón

ASESOR

Dr. Jorge Chue Gallardo

MIEMBRO

MS. Luz Jeanet Bullón Camarena

MIEMBRO

LIMA – PERÚ

2021

DEDICATORIA

“Dedicado a mis padres, por su comprensión y
por todo el apoyo mostrado”

Cesar Anthony Tonconi Calisaya

AGRADECIMIENTO

Agradezco a mis padres, hermano y amigos por siempre darme la fuerza para seguir adelante y luchar por mis objetivos.

Una mención especial para al Departamento de Estadística e Informática y su plana docente por brindarme toda la asesoría requerida para la realización del presente trabajo.

ÍNDICE GENERAL

I.	INTRODUCCIÓN.....	1
1.1.	Problemática.....	2
1.2.	Objetivos.....	3
1.2.1.	Objetivo principal.....	3
1.2.2.	Objetivos específicos.....	3
II.	MARCO TEÓRICO.....	5
III.	MARCO METODOLÓGICO.....	9
3.1.	Tipo o Enfoque.....	9
3.2.	Diseño.....	9
3.3.	Población y Muestra.....	9
3.4.	Método de recojo de información.....	9
3.5.	Fases del análisis de la información.....	12
3.5.1.	Análisis descriptivo.....	12
3.5.2.	Limpieza de datos.....	12
3.5.3.	Identificación de los conglomerados.....	12
3.5.4.	Definición de los perfiles.....	12
3.6.	Definiciones importantes.....	12
3.6.1.	Hipótesis principal.....	14
3.6.2.	Hipótesis específicas.....	14
3.7.	Introducción al Método Cluster Bietápico.....	14
3.7.1.	Objetivos del Clúster Bietápico.....	14
3.7.2.	Supuestos que se deben de cumplir para utilizar el análisis de clúster bietápico.....	15
3.7.3.	Método BIRCH.....	16
3.7.4.	Algoritmo del Método de Clúster Bietápico.....	17
3.7.5.	Cálculo para la medida de la distancia en el clúster bietápico.....	19
3.7.6.	Criterios de agrupamiento utilizado por el método de Clúster Bietápico.....	20
IV.	RESULTADOS Y ANÁLISIS.....	23
4.1.	Análisis exploratorio y descriptivo.....	23
4.1.1.	Análisis descriptivo de las variables cuantitativas.....	23
4.1.2.	Análisis exploratorio y descriptivo de variables cualitativas.....	25
4.2.	Limpieza de datos de las variables consideradas en el estudio.....	26
4.2.1.	Limpieza de datos para variables cuantitativas.....	26
4.3.	Obtención de los conglomerados.....	31

4.4.	Propuesta de perfiles de los conglomerados	33
4.4.1.	Perfil del conglomerado 1 en función de las condiciones básicas de calidad analizadas	33
4.4.2.	Perfil del conglomerado 2 en función de las condiciones básicas de calidad analizadas	34
V.	CONCLUSIONES Y RECOMENDACIONES.....	36
5.1.	Conclusiones	36
8.2.	Recomendaciones.....	39
VI.	REFERENCIAS BIBLIOGRÁFICAS	40
VII.	ANEXOS	43

ÍNDICE DE TABLAS

Tabla 1: Variables cualitativas y cuantitativas	11
Tabla 2: Estadísticos descriptivos de variables cuantitativa del conjunto de datos inicial	23
Tabla 3: Estadísticos descriptivos de variables cuantitativa del conjunto de datos inicial	27
Tabla 4: Estadísticos descriptivos de variables cuantitativa luego de la limpieza de datos.....	28
Tabla 5: Puntuaciones de las variables cuantitativas en cada conglomerado	32

ÍNDICE DE FIGURAS

Figura 1: Árbol CF de características	18
Figura 2: Diagrama de cajas de variable cuantitativa: Matrícula 2019	24
Figura 3: Diagrama de cajas de variable cuantitativa: Docentes 2019	24
Figura 4: Gráfico de sector de las variables cualitativas	26
Figura 5: Diagrama de cajas de la Distancia de Mahalanobis con la presencia de outliers multivariados	27
Figura 6: Diagrama de cajas de la Distancia de Mahalanobis sin la presencia de outliers multivariados	28
Figura 7: Histograma de la variable Matrícula 2019 antes de la limpieza de datos	29
Figura 8: Histograma de la variable Matrícula 2019 luego de la limpieza de datos	29
Figura 9: Histograma de la variable Docentes 2019 antes de la limpieza de datos.....	30
Figura 10: Histograma de la variable Docentes 2019 luego de la limpieza de datos	30
Figura 11: Tamaños de los conglomerados	31
Figura 12: Comparación de los conglomerados	32

ÍNDICE DE ANEXOS

Anexo 1: Resumen del modelo.....	43
Anexo 2: Importancia del predictor.....	44
Anexo 3: Comportamiento de las variables en cada conglomerado.....	45

RESUMEN

El aumento continuo de grandes volúmenes de datos y la importancia de la utilización de éstos, junto con la búsqueda de información se han vuelto un gran reto que las grandes entidades públicas hoy en día quieren superar. En la actualidad las entidades públicas conocen la importancia que tiene el almacenamiento, la captura de datos y el beneficio que le puede resultar si se explotan correctamente.

Este estudio presenta una técnica estadística para segmentar e identificar perfiles de los Centros de Educación Técnico-Productiva (CETPRO) públicos. El conjunto de datos está conformado por información de todos los CETPRO públicos a nivel nacional recogida en el 2019. El conjunto de datos inicial estaba compuesto por 704 instituciones educativas; posteriormente, luego de proceder con el análisis exploratorio y la limpieza de datos, correspondiente a la eliminación de datos outliers y faltantes; se trabajó un conjunto de datos compuesto por 684 instituciones. Se aplicó análisis de clúster bietápico, que es una técnica de segmentación que permite trabajar con variables cuantitativas y categóricas.

El resultado de la aplicación de la técnica brindó 2 conglomerados: el primero con 324 CETPRO (47.4%), el segundo conglomerado con 360 (52.6%). Después de esto se procedió a describir los perfiles de cada conglomerado y se identificaron las principales características en función de las variables relacionadas a las 5 condiciones básicas de calidad planteadas.

Palabras clave: análisis de clúster bietápico, perfiles de CETPRO, gestión institucional, académica, infraestructura, previsión de ingresos y gastos.

ABSTRACT

The continuous increase in large volumes of data and the importance of its use with the search for information, have become a great challenge that today large public entities want to overcome. Today, public entities know the importance of storage, data capture and the benefit that can result if they are exploited correctly.

This study shows a statistical technique to segment and identify profiles of the public Productive Technical Education Centers (CETPRO). The dataset consists of information from all public CETPROs at the national level collected in 2019. The initial dataset was composed of 704 educational institutions; later, after proceeding with the exploratory analysis and data cleaning, corresponding to the elimination of outliers and missing data, a data set consisting of 684 institutions was used. Two-stage cluster sampling analysis was applied, which is a segmentation technique that allows working with quantitative and categorical variables.

The result of the application of the technique provided 2 clusters: the first with 324 CETPRO (47.4%), the second with 360 (52.6%). Subsequently, the profile of each cluster was described, and the main characteristics were identified based on the variables related to the 5 basic quality conditions proposed.

Key Words: management, institutional, academic, infrastructure, analysis, cluster, two-stage sampling, profile, CETPRO (PRODUCTIVE TECHNICAL EDUCATION CENTERS) income of and expenditure.

I. INTRODUCCIÓN

La caracterización de los Centro de Educación Técnico-Productiva (CETPRO) públicos a nivel nacional es fundamental para el sector educación; puesto que, permite conocer a profundidad la situación actual en la que dichas instituciones vienen impartiendo oferta educativa tomando en consideración los indicadores relacionados a las condiciones básicas de calidad que se detallan en la normativa vigente “Decreto Legislativo N° 1375 que modifica diversos artículos de la Ley N° 28044, Ley General de Educación, sobre Educación Técnico – Productiva y dicta otras disposiciones” del Ministerio de Educación.

Según el Censo Educativo del 2018 realizado por el Ministerio de Educación, en el Perú existen 1 800 CETPRO, de ellos 808 son públicos y 992 privados, los cuales cuentan con 250,000 estudiantes matriculados.

En ese sentido, las técnicas estadísticas multivariadas nos permiten analizar, a partir de grandes volúmenes de información, relaciones y ensayar distintas maneras de organizar dicha información. El análisis clúster permite caracterizar a los CETPRO mediante la conformación de conglomerados con alto grado de homogeneidad dentro del conglomerado y heterogeneidad entre los mismos.

El análisis de Cluster Bietápico aplicado tendrá como resultado las siguientes mejoras principales: sistematización de información de los CETPRO públicos a nivel nacional organizada en una base de datos en SPSS; asimismo, establecer conglomerados y de esta manera conocer el perfil de estos en función a las variables de condiciones básicas de calidad y obtener evidencia estadística para la realización de acciones destinadas al fortalecimiento de aspectos relacionados a las condiciones básicas de calidad que se identifiquen como deficientes y de esta manera se pueda garantizar que los CETPRO públicos puedan llegar en óptimas condiciones al proceso de licenciamiento institucional.

1.1. Problemática

La “Dirección de Gestión de Instituciones de Educación Técnico-Productiva y Superior Tecnológica y Artística (DIGEST)” es una subdirección (órgano de línea) de la Dirección General de Educación Técnico – Productiva y Superior Tecnológica y Artística que actualmente tiene a su cargo el licenciamiento institucional de los Institutos de Educación Superior Tecnológica (IEST) públicos y privados como Institutos de Educación Superior, este procedimiento forma parte de la “Ley de Institutos y Escuelas de Educación Superior y de la Carrera Pública de sus Docentes (Ley N° 30512)” publicada el 02 de noviembre de 2016; de esta manera se aborda la reforma de educación superior que busca garantizar una oferta educativa de calidad. Posteriormente, mediante Decreto Supremo N° 004-2019 MINEDU, se establece el procedimiento de licenciamiento para los CETPRO públicos y privados; estableciendo como requisito la verificación del cumplimiento de condiciones básicas de calidad; dicho procedimiento en la actualidad no se viene ejecutando por no tener la denominada norma técnica de la ley antes mencionada; que es un documento donde se detallan todos los medios de verificación relacionados a cada condición básica de calidad requerida para alcanzar el licenciamiento.

Hasta el 2018, el Ministerio de Educación no contaba con información de los CETPRO relacionada a las condiciones básicas de calidad planteadas en las normativas antes mencionada; siendo el Censo Educativo que se realiza todos los años la única fuente de información; sin embargo, lo recogido es insuficiente puesto que solo abarca variables relacionadas a información de infraestructura de los CETPRO. En base a la información recogida a través del censo se ha podido identificar deficiencias importantes en aspectos relacionados a infraestructura; tales como, solo el 17% cuenta con un espacio destinado al funcionamiento de una biblioteca física; asimismo, solo el 9% cuenta con laboratorio de ciencias lo cual pone de manifiesto que existe una brecha enorme por cubrir en el aspecto relacionado a espacios destinados a la enseñanza educativa.

Por otro lado, tenemos que solo el 74% de los CETPRO públicos cuenta con local propio y; de este grupo solo el 70% tiene su local inscrito en registros públicos; lo cual, evidencia la necesidad urgente de implementar acciones destinadas a la atención de esta problemática puesto que constituyen requisitos primordiales para alcanzar el licenciamiento institucional. Por último, tenemos que la tenencia de servicios básicos es preocupante; puesto que, existe

un 2% que no cuenta con energía eléctrica que provenga de una red pública, otro 4% que no cuenta con agua potable proveniente de red pública, 5% que no tiene conexión a un servicio de desagüe y por último un 54% que no tiene conexión a internet.

La información estadística que se menciona guarda relación directa con las condiciones básicas de calidad; sin embargo, solo es un pequeño porcentaje del total de variables que se requiere analizar. Es por ello, que actualmente no se tiene una caracterización de los CETPRO públicos a nivel nacional que permita conocer el estado en el cual vienen funcionando tomando como referencia aspectos como infraestructura y equipamiento, propuesta pedagógica, gestión institucional, disponibilidad docente y otros relacionados a las condiciones básicas de calidad; en ese sentido, si no se implementan políticas destinadas al fortalecimiento de la oferta educativa pública a través de un proceso de optimización que busque el cierre de brechas a través de asignación de presupuestos; estas instituciones de educación técnico productiva públicos no estarán en la capacidad de afrontar de manera adecuada un proceso de licenciamiento institucional.

Por lo tanto, ante la limitación que se tiene de información relacionada a las condiciones básicas de calidad es que se realiza un levantamiento de información destinada únicamente a los CETPRO en el 2019, para ello se conformaron equipos compuestos por un docente y un arquitecto; el primero porque es necesario que tenga conocimiento de las características académicas y el segundo porque participa en una evaluación de la infraestructura; finalmente, con la información recogida es que se aplica el denominado Cluster Bietápico.

1.2. Objetivos

1.2.1. Objetivo principal

Identificar conglomerados de los Centros de Educación Técnico-Productiva públicos según los indicadores de condiciones básicas de calidad mediante clúster bietápico

1.2.2. Objetivos específicos

- Establecer las principales características de los CETPRO en relación con la gestión institucional usando indicadores de condiciones básicas de calidad mediante análisis

clúster bietápico.

- Establecer las principales características de los CETPRO en relación con la gestión académica usando indicadores de condiciones básicas de calidad mediante análisis clúster bietápico.
- Establecer las principales características de los CETPRO en relación con la infraestructura y equipamiento usando indicadores de condiciones básicas de calidad mediante análisis clúster bietápico.
- Establecer las principales características de los CETPRO en relación con el personal docente usando indicadores de condiciones básicas de calidad mediante análisis clúster bietápico.
- Establecer las principales características de los CETPRO en relación con la previsión económica y financiera usando indicadores de condiciones básicas de calidad mediante análisis clúster bietápico.

II. MARCO TEÓRICO

Bacher, Wenzig y Volger (2014) indican que la segmentación bietápica es una técnica no jerárquica exploratoria de agrupamiento de un conjunto de datos. Asimismo, tiene como finalidad resolver deficiencias que presentan las técnicas tradicionales; adicionalmente, esta técnica puede procesar datos numéricos y categóricos.

De la Fuente (2011) señala que la aplicación de análisis de clúster bietápico en SPSS es un método de exploración diseñado para determinar agrupaciones de datos o variables; asimismo, cuenta con la ventaja de poder aplicarse teniendo una gran cantidad de información que analizada con otra técnica tradicional no sería factible.

Valero (2011) describe al clúster como un método bastante reciente; en el cual, no es necesario indicar desde el inicio cuantos clústers se desea obtener y permite trabajar con variables de tipo categóricas y numéricas considerando previamente un análisis de consistencia.

Rubio-Hurtado, Vila-Baños y Duarte (2016) indican que el análisis de clúster en dos fases, conocido como bietápico, es una metodología de exploración utilizada para conocer las agrupaciones de una cantidad de información; de ese modo, permite acceder a estadísticos de frecuencia, descriptivos, sectores y demás que se consideren necesarios.

Espigares-Pinazo, Bautista-Vallejo y Duarte (2017) definen que el análisis clúster bietápico es considerado también un método de conformación de conglomerados. Los conglomerados finales se construyen en dos pasos y es de mucha utilidad cuando se cuenta con bastante información disponible.

Bacher et al. (2004) señala que la agrupación en clústeres en dos etapas ofrece la posibilidad de manejar variables continuas y categóricas. Por lo tanto, este modelo proporciona una solución para un caso especial de variables de tipo mixto. Variables cuantitativas con diferentes unidades de escala y nominales. El usuario debe decidir manejar variables ordinales como continuas o categóricas si están presentes.

Romero (2019) indica que La calidad educativa debe ser mostrada por las instituciones educativas en sus ofertas académicas, programas y planes de estudios, así como en una gestión académica e institucional con logros significativos que les permita constituirse en instituciones de primer nivel y elevados grados de eficiencia.

Ministerio de Educación (2015) señala que “los Centro de Educación Técnico-Productivo (CETPRO) públicos son las instituciones donde se imparte la enseñanza técnico productivo; asimismo, los CETPRO pueden implementar acciones de capacitación, actualización y reconversión laboral y fortalecen a la Educación Básica Regular” brindándoles dichos servicios.

Ley N° 30512 (2016) señala que el “Ministerio de Educación es la institución encargada de supervisar el cumplimiento del proceso de licenciamiento, cumplimiento de las condiciones básicas de calidad solicitadas y las políticas nacionales y sectoriales contenidas en la presente ley y su reglamento, en los programas de estudios conducentes al grado de bachiller, título profesional y segunda especialidad”.

Decreto Supremo N° 004 (2019, artículo 106B) especifica que “la creación de un CETPRO público está condicionada a la opinión favorable de la Unidad de Gestión Educativa Local y la opinión presupuestal favorable del Gobierno Regional correspondiente; las cuales deben asegurar el cumplimiento de las Condiciones Básicas de Calidad requeridas”.

Haya de la Torre (2005) señala que los problemas vistos en torno a los desencuentros entre

la oferta de profesionales técnicos y la demanda de empleo en un determinado sector, no sólo son fruto de los problemas del empleo atribuibles al crecimiento poblacional, migración, mayor participación de la mujer y a la baja de la dinámica económica y productiva; adicionalmente, obedecen a diversas falencias provenientes del sistema de educación profesional técnica, siendo la principal la poca adecuación que presenta dicho sistema ante los requerimientos, potenciales y tendencias del sector productivo.

Brawijaya (2011) define el procedimiento de clúster bietápico como exploratorio que ha sido diseñado para identificar conglomerados a partir de información disponible que de otro modo no serían evidentes. El algoritmo empleado por este procedimiento tiene varias características deseables que lo diferencian de las técnicas tradicionales de agrupamiento.

Hinneburg y Keim (1999) señalan la existencia de algoritmos de agrupación en clúster efectivos y eficientes para grandes conjuntos de datos de alta dimensión con alto nivel de ruido que tiene por finalidad determinar características estándar en cada grupo que se conforme.

Littau and Boley (2006) indican que ante la necesidad de agrupar grandes conjuntos de datos surgen muchos enfoques, como el clúster jerárquico, K-medias y otros. El enfoque de agrupamiento consiste en adaptar algoritmos de agrupación para datos pequeños a grandes conjuntos de datos.

Chiu, Fang (2001) menciona que una de las principales dificultades cuando se aplican las técnicas tradicionales de segmentación de datos se presenta cuando tenemos una cantidad de observaciones demasiado grandes para analizar y en consecuencia ser agrupados; es por ello, que surge como alternativa el Método “Bietápico”.

Huang (1998) y Kaufman and Rousseeuw (1990) implementaron medidas de distancia para variables de tipo cualitativa como el de compatibilidad por diferencia utilizado en el método de K-modas (Huang, 1998). Para modelos de variables de tipo numéricas y categóricas se desarrollaron distancias basadas en ponderaciones de la suma de las variables numéricas y categóricas.

Meilă y Heckerman (1998) contribuyeron con la aplicación del concepto probabilístico dado por Banfield y Raftery y se derivó otra medida de distancia para datos con solo atributos categóricos.

Banfield (1993) en un inicio propusieron un modelo que se sustentaba en la medida de distancias y que solo trabaja con variables cuantitativas. Posteriormente, Chiu et al. (2001) mejoraron la propuesta inicial e introdujeron modelos con variables numéricas y categóricas.

Tirlangi, Mohan, Kalyampudi y Krishna (2014) expresan que la agrupación en clúster es el proceso de encontrar grupos de observaciones o individuos de tal manera que los individuos de un grupo sean similares entre sí y diferentes de los individuos que pertenecen a otros grupos. Asimismo, es una técnica usada en muchos contextos y por investigadores de muchas disciplinas; esta refleja su amplio atractivo y utilidad como uno de los pasos en análisis exploratorio de datos.

III. MARCO METODOLÓGICO

3.1. Tipo o Enfoque

El estudio que se llevará a cabo es de tipo explicativo; puesto que, permitirá ahondar en las variables relacionadas a las condiciones básicas de calidad y establecer características diferenciadoras mediante el análisis clúster bietápico.

3.2. Diseño

En relación con el diseño de la presente investigación es de tipo cuantitativo de tipo no-experimental debido a que la información ha sido recolectada de manera directa de los CETPRO sin ninguna intervención previa, del tipo transversal, puesto que se procedió a recoger información de estos a través de visitas en sus instalaciones en un solo momento, obteniendo datos directamente para analizarlos posteriormente.

3.3. Población y Muestra

La presente monografía no considera un tipo de muestreo; puesto que considera a toda la población conformada por los CETPRO públicos a nivel nacional (704 en total).

3.4. Método de recojo de información

Considera la aplicación de un cuestionario en cada uno de los CETPRO públicos a nivel nacional; para lo cual, se reclutará a un docente y un arquitecto para que en forma conjunta recojan información relacionada a condiciones básicas de calidad. El cuestionario en mención será elaborado por el equipo de licenciamiento de la Dirección de Gestión de Instituciones de Educación Técnico-Productiva y Superior Tecnológica y Artística. Asimismo, la información analizada está constituida por las siguientes variables:

La variable condiciones básicas de calidad.

La variable denominada conglomerados.

La variable condiciones básicas de calidad está agrupada en cinco dimensiones:

- a. **Gestión Institucional:** comprende variables relacionadas a la tenencia de documentos de gestión; tales como, tenencia de reglamento interno, proyecto educativo institucional, adicionalmente, la tenencia de servicio de bienestar estudiantil y servicio de atención básica de emergencias dentro del local de la institución.
- b. **Gestión Académica:** comprende información de la cantidad de matriculados, egresados, número de programas de estudio ofertado que al mismo tiempo estos estén actualizados.
- c. **Infraestructura y Equipamiento:** comprende información de tenencia de local, tenencia de servicios básicos, tenencia de biblioteca física y virtual, acceso a internet y material predominante en el local del CETPRO.
- d. **Personal docente:** comprende información del número de docentes por tipo de contrato, por tiempo de dedicación y años de experiencia laboral.
- e. **Previsión económica y financiera:** comprende información relacionada a demostrar disponibilidad de recursos financieros que garanticen el desarrollo sostenible, solvencia, continuidad y calidad en la enseñanza.

La información recolectada cuenta con dos variables cuantitativas y diez variables cualitativas. Estas doce variables miden los aspectos relacionados a las condiciones básicas de calidad divididas en las dimensiones antes mencionadas y fueron proporcionadas bajo el criterio de confidencialidad.

Tabla 1: Variables cualitativas y cuantitativas

Gestión Institucional	
¿El CETPRO tiene Proyecto Educativo Institucional (PEI)?	1. Si 2. No
¿El CETPRO tiene Manual de Perfil de Puestos (MPP)?	1. Si 2. No
¿El CETPRO tiene Reglamento Interno (RI)?	1. Si 2. No
¿El CETPRO cuenta con servicio de atención básica de emergencias?	1. Si 2. No
Gestión académica	
¿Al menos un programa de estudios que oferta el CETPRO se ha adecuado al DS 004-2019-MINEDU?	1. Si 2. No
Matriculados 2019	Cantidad de alumnos matriculados en el 2019
Infraestructura y equipamiento	
EL local donde funciona el CETPRO es:	1. Alquilado 2. Cedido o donado 3. Cedido por convenio. 4. Prestado 5. Propio
El agua que utiliza el local proviene de:	1. Camión o cisterna 2. No tiene agua 3. Pilón de uso público. 4. Pozo 5. Red pública dentro del local 6. Río acequia o manantial
El alumbrado eléctrico que tiene el local proviene de:	1. Generador o motor 2. No tiene alumbrado eléctrico 3. Panel Solar. 4. Red pública
Personal docente	
¿El CETPRO cuenta con un plan de actualización y capacitación docente?	1. Si 2. No
Docentes 2019	Cantidad de docentes laborando en el CETPRO
Previsión económica y financiera	
¿El CETPRO realiza estudios verificables de previsión de ingresos y gastos con los recursos directamente recaudados?	1. Si 2. No

3.5. Fases del análisis de la información

3.5.1. Análisis descriptivo

Se analizará la información recolectada inicialmente para evaluar las características de las variables numéricas y categóricas.

3.5.2. Limpieza de datos

Procedimiento que tiene la intención de eliminar las observaciones inconsistentes; así como, abordar datos ausentes, entre otros. Por otro lado, para las variables de tipo cualitativas se evaluará eliminar aquellas cuyo valor difiere de las alternativas mencionadas en cada pregunta del instrumento de recojo, así también variables con datos vacíos o nulos.

3.5.3. Identificación de los conglomerados

A través del Análisis de Clúster Bietápico se va a identificar los conglomerados.

3.5.4. Definición de los perfiles

Se identificaron perfiles para cada clúster o conglomerado encontrado.

3.6. Definiciones importantes

Educación técnico productiva: es una “forma de educación orientada a la adquisición y desarrollo de competencias laborales y empresariales en una perspectiva de desarrollo sostenible, competitivo y humano, así como a la promoción de la cultura innovadora que responda a la demanda del sector educativo y a los avances de la tecnología, del desarrollo local, regional y nacional, así como a las necesidades educativas de los estudiantes en sus respectivos entornos” (Decreto Supremo N° 022-2004-ED) .

Condiciones básicas de calidad: son condiciones mínimas para ofertar servicio educativo en los CETPRO. Es necesario su cumplimiento para alcanzar el licenciamiento institucional y contempla las siguientes dimensiones:

- **Gestión institucional:** el grado de eficiencia de la institución educativa para conducir la planificación estratégica, la administración de los procesos académicos, así como, de los sistemas de soporte administrativo y de bienestar estudiantil, con la finalidad de garantizar un servicio de calidad al estudiante.
- **Gestión académica:** está orientada a la capacidad del CETPRO para ofertar programas de estudios pertinentes y que tengan relación con las actividades económicas de la región donde se imparte servicio educativo; asimismo, considera necesario que los programas de estudios se actualicen constantemente.
- **Infraestructura y equipamiento:** la capacidad de la institución educativa para evidenciar la disponibilidad de una adecuada infraestructura física y el equipamiento correspondiente y relacionado a cada uno de los programas ofertados. La infraestructura debe de ser propia y debe de ser de uso exclusivo por el CETPRO.
- **Disponibilidad de personal docente:** la institución educativa debe de garantizar la disponibilidad de personal docente y altamente capacitado, adicionalmente, este personal debe de tener formación relacionada al programa de estudios que se brinde y de esta manera se garantiza estudiantes altamente capacitados.
- **Previsión económica y financiera:** orientada a evidenciar que se realiza la previsión económica, financiera de los recursos directamente recaudados a fin de asegurar la continuidad del servicio y su sostenibilidad.

Hipótesis

3.6.1. Hipótesis principal

Existen características diferenciadas de los Centros de Educación Técnico-Productiva públicos, según las variables relacionadas a las condiciones básicas de calidad.

3.6.2. Hipótesis específicas

- Existen características diferenciadas de los Centros de Educación Técnico-Productiva públicos, según las variables relacionadas a la gestión institucional.
- Existen características diferenciadas de los Centros de Educación Técnico-Productiva públicos, según las variables relacionadas a la gestión académica.
- Existen características diferenciadas de los Centros de Educación Técnico-Productiva públicos, según las variables relacionadas a la infraestructura y equipamiento.
- Existen características diferenciadas de los Centros de Educación Técnico-Productiva públicos, según las variables relacionadas al personal docente.

3.7. Introducción al Método Clúster Bietápico

El clúster bietápico identifica de manera automática el número de clúster o conglomerados; asimismo, difiere de los métodos tradicionales de agrupamiento de información, por otro lado, tenemos que el clúster bietápico tiene como ventaja principal que permite el procesamiento de grandes volúmenes de información y para ello utiliza el denominado árbol de características (CF). Esta ventaja radica principalmente en que optimiza los recursos del procesador del computador para obtener resultados de forma automática en comparación con los métodos tradicionales que no soportan grandes volúmenes de información.

3.7.1. Objetivos del Clúster Bietápico

El clúster bietápico es una técnica de análisis que sirve para identificar agrupaciones naturales dentro de un conjunto de datos; asimismo, el algoritmo o método que emplea tiene una serie de funciones que lo diferencian las técnicas tradicionales:

- Permite trabajar con variables de tipo continuas y categóricas. Por lo tanto, tiene como supuesto que las variables sean independientes; por lo tanto, es óptimo utilizar una distribución normal multinomial de manera conjunta a las variables antes mencionadas.
- Identificación del número de clúster de forma automática. Para ello, utiliza la comparación de los valores de un criterio de selección del modelo empleado para una variedad de soluciones de agrupamiento; por lo tanto, puede seleccionar de manera automática el número óptimo de clúster.

3.7.2. Supuestos que se deben de cumplir para utilizar el análisis de clúster bietápico

Es importante mencionar que el análisis de clúster bietápico tiene las siguientes consideraciones:

1. Debido a que utiliza la distancia de la verosimilitud es necesario que las variables del modelo de clúster sean independientes. Por lo tanto, se debe de realizar las siguientes comprobaciones:
 - Para comprobar que las variables continuas sean independientes se recomienda utilizar correlaciones bivariadas.
 - Para variables categóricas utilizar tablas de contingencia.
 - Para el análisis de relación entre una variable continua y categórica se debe de utilizar el procedimiento de medias.
2. Se debe de utilizar la prueba de Kolmogórov-Smirnov para evaluar que las variables continuas tengan una distribución normal.
3. Para las variables no numéricas se debe de utilizar la prueba de chi-cuadrado y verificar si sigue una distribución multinomial.
4. Finalmente, hay que mencionar que en análisis de clúster bietápico es una técnica robusta cuya aplicación puede omitir que se cumplan los supuestos antes mencionados; lo cual, es una gran ventaja para el análisis de información con variables continuas y categóricas.

3.7.3. Método BIRCH

El análisis de clúster bietápico tiene como referencia base al método BIRCH cuyo principal aporte es que trabaja con grandes cantidades de información; asimismo, emplea un método de agrupamiento basado en distancias (Euclídea y Manhattan) que no requiere que las variables utilizadas sean independientes.

Por otro lado, el método BIRCH introduce el método de árbol de Características del Clúster (CF); que es una forma de representar de manera más resumida a los clústeres creados; por lo tanto, se tiene gran velocidad de procesamiento para grandes cantidades de información.

Este árbol CF tiene los siguientes parámetros

$$CCCC = \{N, LS, SS\}$$

Los parámetros utilizados se definen como:

N: Cantidad de individuos incluidos en un clúster.

LS: Suma de los atributos de los *N* individuos $\sum_{i=1}^N X_i$.

SS: Suma al cuadrado de los *N* individuos $\sum_{i=1}^N X_i^2$ dentro un clúster.

El aporte del método BIRCH presenta una gran deficiencia; puesto que, es un método que solo es aplicable cuando se tiene solo variables continuas o categóricas es decir que no permite trabajar con ambas simultáneamente. Es por ello, que se hizo necesario la realización de mejoras y en ese contexto aparece la investigación de Chiu et al. (2001) para trabajar con variables de tipo numéricas y categóricas para un mismo conjunto de datos y que se conoce como método “Clúster Bietápico”.

3.7.4. Algoritmo del Método de Clúster Bietápico

El método de Clúster Bietápico propuesto por Chiu et al. (2001) se realiza en dos pasos:

3.7.4.1. Pasos del método de Clúster Bietápico

1) Formación del Pre-Clúster

La primera parte del análisis de clúster bietápico consiste en la formación de los denominados pre-clúster o también conocida como fase de pre-agrupamiento.

Para este proceso se utiliza el árbol de Características del Clúster (CF) como:

$$CCCC = \{N, SA, SA^2, N_B\}$$

Los parámetros se definen como:

N : Tamaño del Clúster o conglomerado, consiste en la cantidad de individuos que están incluidos en un clúster concreto.

SA : Suma de los atributos de las variables de tipo continuas que pertenecen al clúster.

SA^2 : Suma al cuadrado de los atributos de las variables de tipo continuas que pertenecen al clúster.

N_B : $\sum(L_k - 1)$, que es la suma de categorías de cada variable categórica.

En resumen, cuando se tiene dos clústers y estos requieren agruparse se da un proceso de fusión de las características asociadas a cada uno para formar un nuevo CF, por lo tanto:

$$CF_{(j,s)} = \{N_j + N_s, SA_j + SA_s, SA_j^2 + SA_s^2, N_{Bj} + N_{Bs}\}$$

Esta nueva representación es de mucha utilidad; puesto que, se tiene a toda la información inicial de forma resumida y se puede calcular las medidas de similitud que se necesitan.

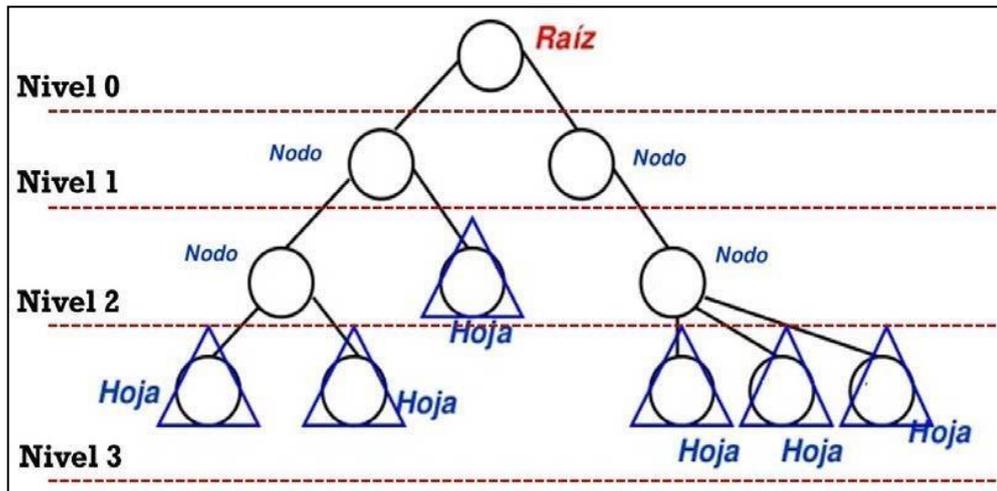


Figura 1: Árbol CF de características

Fuente: Elaboración propia

El árbol de características del Clúster (CF) consiste en la representación de información de forma resumida a través de nodos que se relacionan entre sí por medio de ramas, y éstos al mismo tiempo tienen más nodos de los que salen más ramas. La estructura del árbol comienza con un nodo base que se denomina “Nodo Raíz” que se encuentra en el primer nivel del árbol y es desde este nivel que se van generando de forma ascendente el resto de los niveles que tiene el CF; asimismo, el número de niveles estará determinado por el algoritmo.

Puesto de manera más práctica lo que hace el método es colocar al primer individuo en la raíz del árbol CF y la información de la variable en estudio es colocada en un nodo de hoja, el procedimiento continúa añadiendo al resto de individuos a un nodo existente o caso contrario se forma uno nuevo, para evaluar este escenario se toma en consideración el grado de similitud o disimilitud que tiene el individuo añadido con los nodos ya existentes y utilizando las medidas de distancia previamente definidas.

2) Agrupamiento del Árbol de Características

Como ya se tiene el CF elaborado en la primera parte el paso siguiente es agrupar los nodos utilizando un algoritmo aglomerativo, este algoritmo tiene la característica de proporcionar una variedad de soluciones. Por lo tanto, para seleccionar la mejor alternativa se utilizan comparaciones a través del Criterio Bayesiano de Schwarz (BIC)

o el Criterio de Información de Akaike (AIC).

3.7.5. Cálculo para la medida de la distancia en el clúster bietápico

Para efectos del presente estudio que contempla la utilización de variables continuas y categóricas se tiene a la medida de verosimilitud como medida de distancia.

3.7.5.1. Distancia de Máxima Verosimilitud

Se utiliza esta distancia porque permite trabajar con variables de tipo continuas y categóricas. La definición de la distancia del logaritmo de máxima verosimilitud entre dos clústers es la siguiente:

$$d(i, s) = \varepsilon_i + \varepsilon_s + \varepsilon_{(i,s)}$$

Por lo tanto, una disminución en el logaritmo de la función de máxima verosimilitud al juntarse dos clúster es tomada como la medida de distancia entre los grupos.

Siendo sus componentes:

$$\begin{aligned} \varepsilon_i &= -n_i \left(\sum_{j=1}^p \frac{1}{2} \log(\delta_{ij}^2 + \delta_j^2) \right) - \sum_{j=1}^q \sum_{l=1}^{m_j} \pi_{ijl} \log(\pi_{ijl}) \\ \varepsilon_s &= -n_s \left(\sum_{j=1}^p \frac{1}{2} \log(\delta_{sj}^2 + \delta_j^2) \right) - \sum_{j=1}^q \sum_{l=1}^{m_j} \pi_{sjl} \log(\pi_{sjl}) \\ \varepsilon_{(i,s)} &= -n_{(i,s)} \left(\sum_{j=1}^p \frac{1}{2} \log(\delta_{(i,s)j}^2 + \delta_j^2) \right) - \sum_{j=1}^q \sum_{l=1}^{m_j} \pi_{(i,s)jl} \log(\pi_{(i,s)jl}) \end{aligned}$$

En donde ε_v nos evidencia la varianza que hay dentro del clúster.

El parámetro ε_v , está conformado por dos partes:

- La primera, definida por “ $-n_v \left(\sum_{j=1}^p \frac{1}{2} \log(\delta_{vj}^2 + \delta_j^2) \right)$ ” es el grado de dispersión que

tienen las variables continuas x_j dentro del clúster definido por “ v ”.

- La segunda, definida por “ $n_v(\sum_{j=1}^p \sum_{l=1}^{m_j} \pi_{vjl} \log(\pi_{vjl}))$ ” y mide el grado de dispersión que tienen las variables categóricas dentro del clúster.

Finalmente, la regla de agrupamiento se da de la siguiente manera: los clústers que tengan menor distancia entre ellos se juntan para conformar un nuevo clúster.

3.7.6. Criterios de agrupamiento utilizado por el método de Clúster Bietápico

Los criterios de agrupamiento son el Bayesiano y el criterio de Akaike. Sin embargo, el criterio que se utilizó en el presente trabajo es el Bayesiano que permite tener de forma automática el número óptimo de clúster.

3.7.6.1. Criterio de SCHWARTZ (BIC)

El criterio está representado de la siguiente manera:

$$BIC = (-2 * \ln L(\theta) + \ln(n) * k)$$

La regla de decisión para seleccionar el mejor modelo está dada por lo siguiente: el modelo con el valor BIC más bajo es considerado como la mejor opción para explicar toda la información con el mínimo número de parámetros.

3.7.7. Importancia del Predictor de Agrupamiento

La importancia que tiene cada variable en la formación de un clúster o conglomerado, tanto dentro como entre clúster está definido por:

$$VI_i = \frac{-\log_{10}(p_valor_i)}{\max_{j \in \Omega} (-\log_{10}(p_valor_j))}$$

Donde Ω representa la matriz de datos, el p-valor se calcula tal como se describe a continuación, y si el $(p - valor)_i = 0$ entonces $(p - valor)_i$ es el valor mínimo.

a. Entre Clúster

VARIABLES CATEGÓRICAS: El p-valor se basa en una chi-cuadrado

$$p_{valor} = Prob(x_d^2 > x^2)$$

$$x^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - \hat{N}_{ij})^2}{\hat{N}_{ij}}$$

$$\hat{N}_{ij} = \frac{N_i \cdot N_j}{N}$$

Grados de libertad: $(I - 1, J - 1)$:

$N=0$ La importancia será indefinida o desconocida

$N_i = 0$ I se calcula como $(I = I - 1)$

$N_j = 0$ J se calcula como $(J = J - 1)$

Si $J \leq 1$ o $I \leq 1$ La importancia será indefinida o desconocida

VARIABLES CONTINUAS: El p-valor se basa en una F de Snedecor:

$$p_{valor} = Prob \{F(J - 1, N - J) > F\}$$

$$F = \frac{\sum_{j=1}^J (x_j - \bar{x})^2 / (J - 1)}{\sum_{j=1}^J (N_j - 1) s_j^2 / (N - J)}$$

Grados de libertad: $(J - 1, N - J)$:

$N=0$ La importancia será indefinida o desconocida

$N_j = 0$ J se calcula como $(J = J - 1)$

Si $J \leq 1$ o $N \leq J$ La importancia será indefinida o desconocida

b. Dentro del Clúster

Para variables categóricas el p-valor se determina mediante la Chi-cuadrado.

$$x^2 = \sum_{i=1}^I \frac{(N_{ij} - N_j p_i)^2}{N_j p_i}$$

Grados de libertad: $d = (I - 1)$:

$N_j = 0$ la importancia será indefinida o desconocida

$N_j = 0$ I se calcula como ($I = I - 1$)

Si $I \leq 1$ La importancia será indefinida

Para variables de tipo continuas el p-valor está basado mediante una T-student:

$$t = \frac{(x_j - \bar{x})}{s_j \sqrt{N_j}}$$

Sabiendo que los grados de libertad son $d = (N_j - 1)$

Si $N_j \leq 1$ o $s_j = 0$ la importancia será indefinida o desconocida

Si el denominador de la fórmula para la t-Student es cero, entonces el p-valor se tomará como igual a 1, y se calculará como:

$$p_valor = 1 - Prob\{|T(d)| \leq |t|\}$$

IV. RESULTADOS Y ANÁLISIS

4.1. Análisis exploratorio y descriptivo

4.1.1. Análisis descriptivo de las variables cuantitativas

A continuación, se muestra los resultados obtenidos para el conjunto inicial que incluye dos variables cuantitativas denominadas: cantidad de matriculados en el CETPRO el 2019 y cantidad de docentes en cada CETPRO en el mismo año.

Tabla 2: Estadísticos descriptivos de variables matriculados y cantidad de docentes

	N	Mínimo	Máximo	Media	Desv. Desviación
Matrícula 2019	704	8	2637	271.12	339.299
Docentes 2019	704	1	100	6.98	7.942

Fuente: Elaboración propia

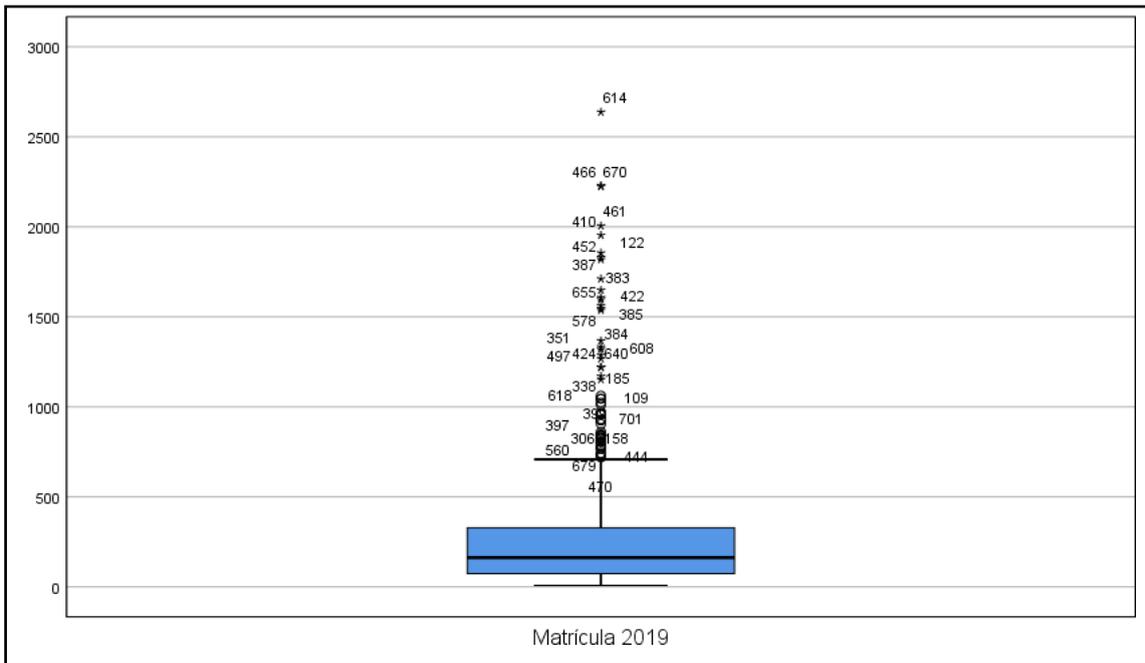


Figura 2: Diagrama de cajas de variable cuantitativa: Matrícula 2019

Fuente: Elaboración propia

El gráfico de cajas de la figura 2 evidencia que los datos pertenecientes a la variable Matrícula 2019 presentan outliers, esto se evidencia en que se tiene como valor máximo de matriculados un total de 2637 y un valor mínimo de 8.

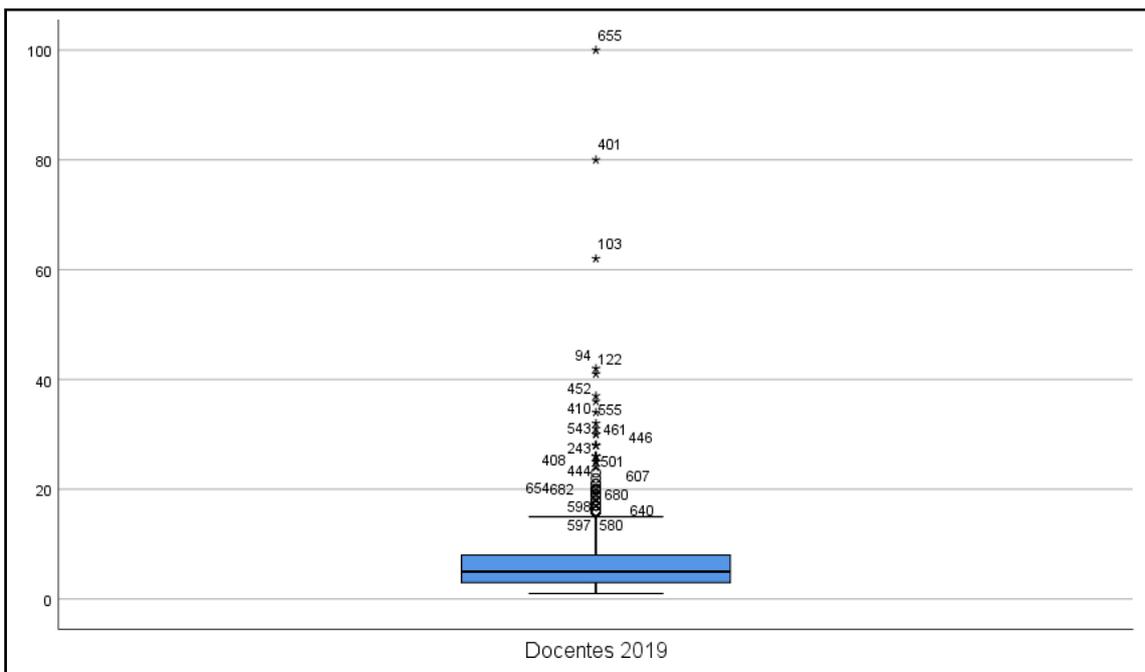
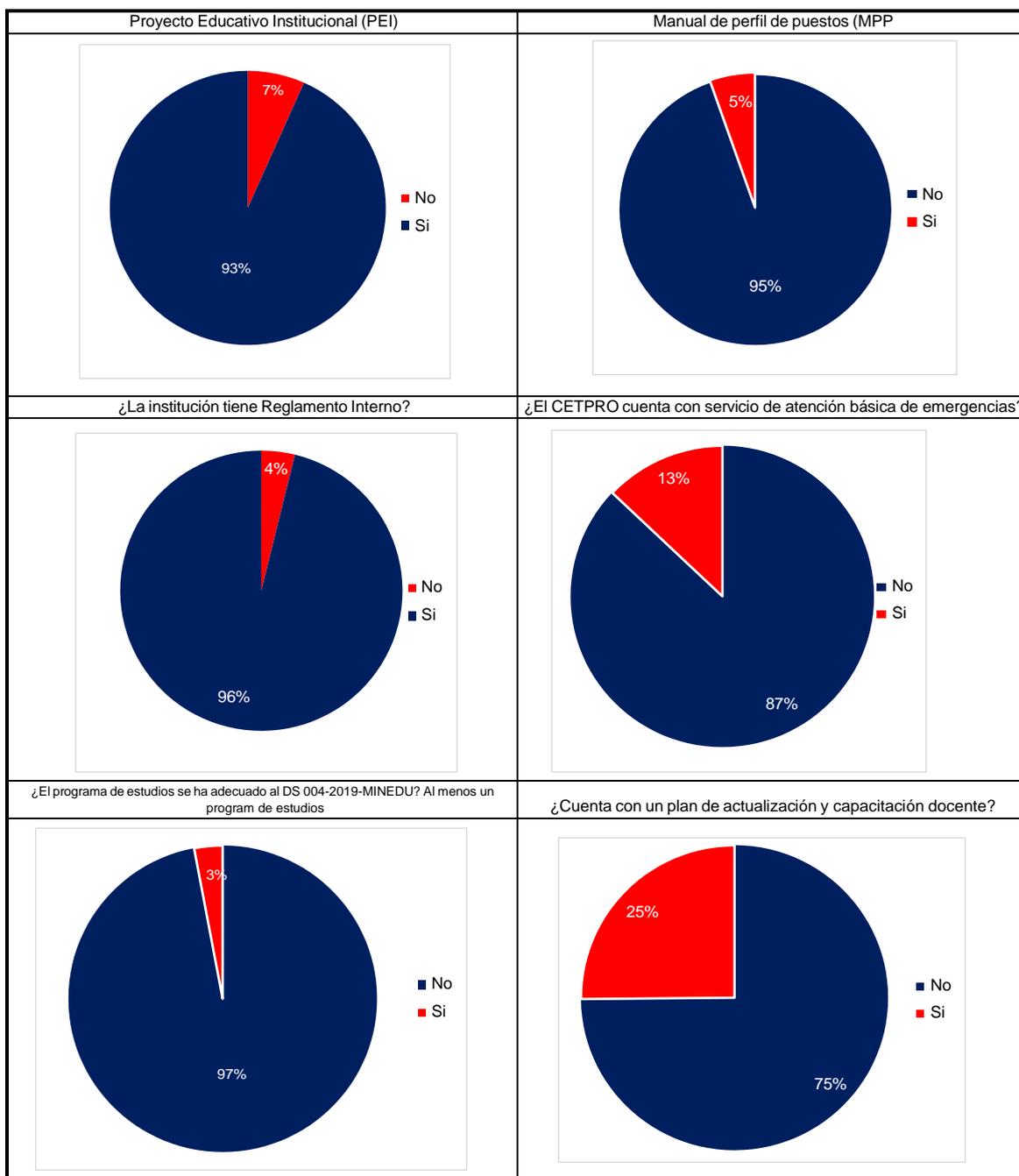


Figura 3: Diagrama de cajas de variable cuantitativa: Docentes 2019

Fuente: Elaboración propia

De igual forma que lo antes mencionado, la variable Docentes 2019 presenta outliers; puesto que, presenta un valor máximo de 100 y un valor mínimo de 1.

4.1.2. Análisis exploratorio y descriptivo de variables cualitativas



«continuación»

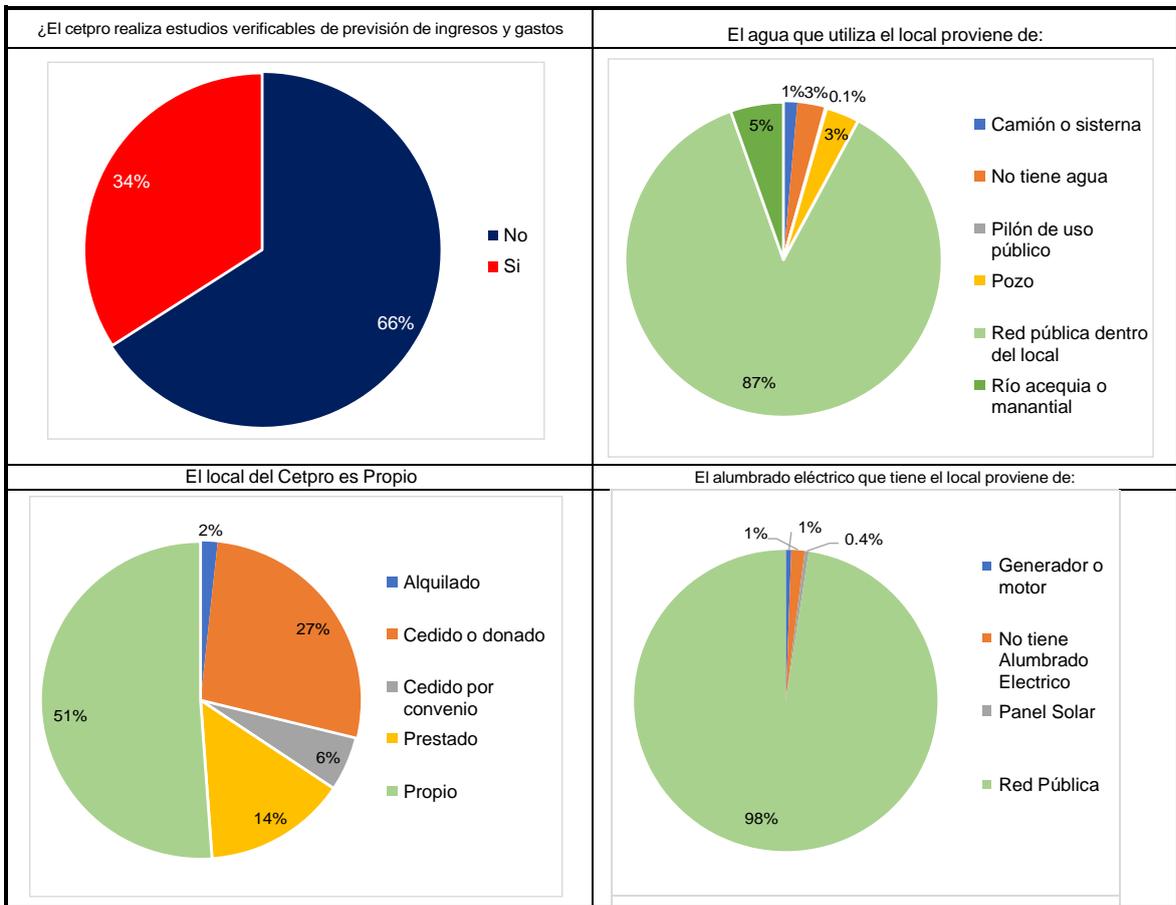


Figura 4: Gráfico de sector de las variables cualitativas

Fuente: Elaboración propia

Respecto al análisis de las variables cualitativas, no se identificó la presencia de datos ausentes; asimismo, no se identificó variables etiquetadas con la opción no responde o la categoría de otros.

4.2. Limpieza de datos de las variables consideradas en el estudio

4.2.1. Limpieza de datos para variables cuantitativas

Para la limpieza de datos para las variables cuantitativas (Matrícula 2019 y Docentes 2019) se identificó outliers multivariantes utilizando la distancia de Mahalanobis. En ese sentido, se obtuvo un valor mínimo de 0.00001 y un valor máximo de 153.42564.

Tabla 3: Estadísticos descriptivos de variables cuantitativa del conjunto de datos inicial

	N	Mínimo	Máximo	Media	Desv. Desviación
Distancia de Mahalanobis	704	0.00001	153.42564	1.9971591	8.80769329

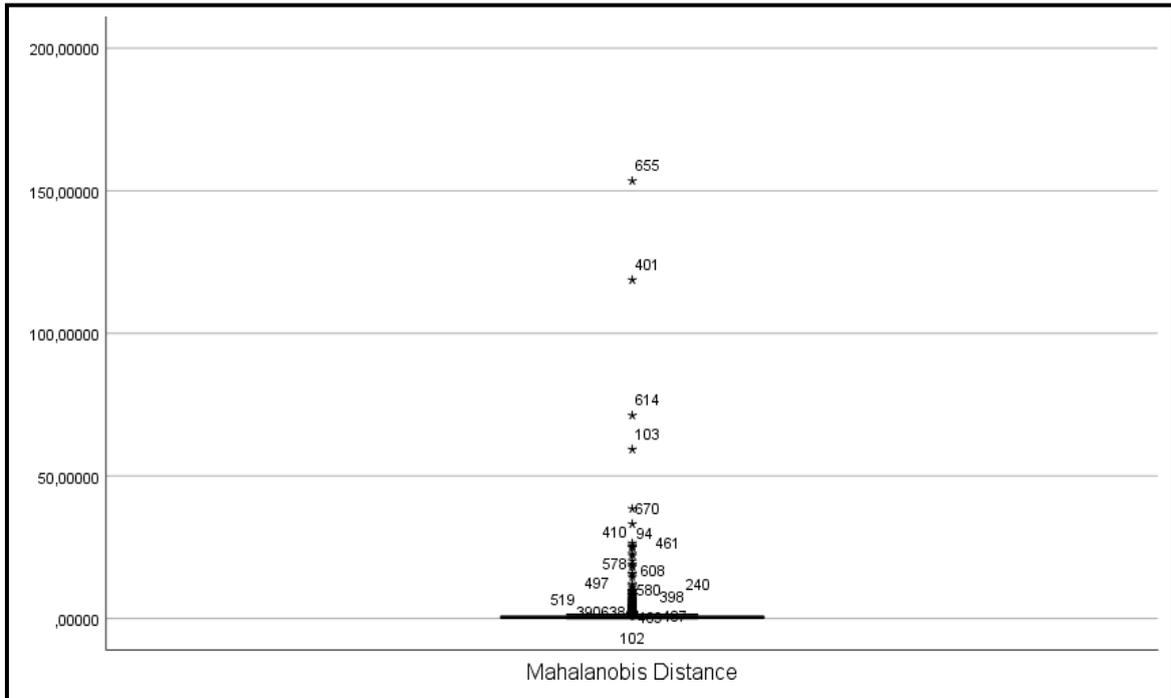


Figura 5: Diagrama de cajas de la Distancia de Mahalanobis con la presencia de outliers multivariados

Fuente: Elaboración propia.

El diagrama de caja de la figura 5 evidencia la presencia de outliers; es por ello, que para identificarlos se utilizó el cálculo de la probabilidad superior de la distribución Chi-cuadrada con 2 grados de libertad (porque se analizaron dos variables). Por lo tanto, los valores extremos multivariados son aquellos cuya probabilidad sea inferior a 0.001.

Luego de retirar los valores extremos (outliers) multivariados se obtuvo el siguiente diagrama de cajas

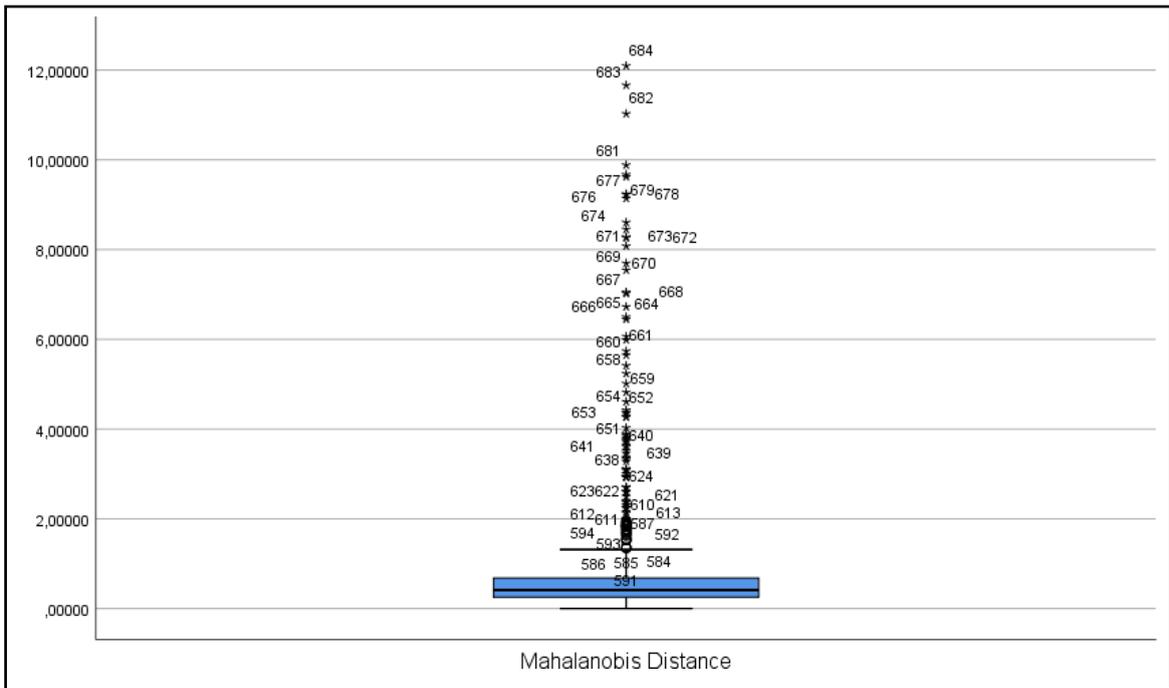


Figura 6: Diagrama de cajas de la Distancia de Mahalanobis sin la presencia de outliers multivariados

Fuente: Elaboración propia.

Como resultado de la limpieza de datos (solo para las variables cuantitativas) se eliminaron un total de 20 registros. Por lo tanto, el nuevo conjunto de datos final para el análisis de clúster bietápico es de 684 registros.

Tabla 4: Estadísticos descriptivos de variables cuantitativa sin outliers

	N	Mínimo	Máximo	Media	Desv. Desviación
Matrícula 2019	684	8	1333	233.63	240.586
Docentes 2019	684	1	31	6.20	5.315

Fuente: Elaboración propia

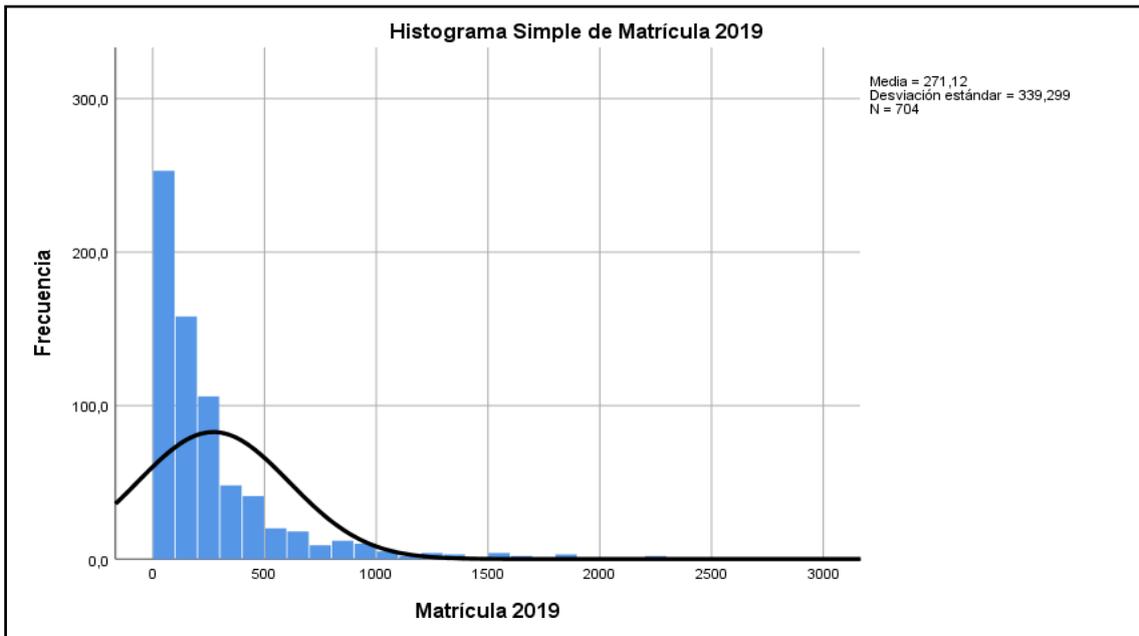


Figura 7: Histograma de la variable Matrícula 2019 antes de la limpieza de datos

Fuente: Elaboración propia

La variable Matrícula 2019 presenta una distribución asimétrica hacia la derecha y sus valores más frecuentes se encuentran en un rango de 0 a 250 matriculados.

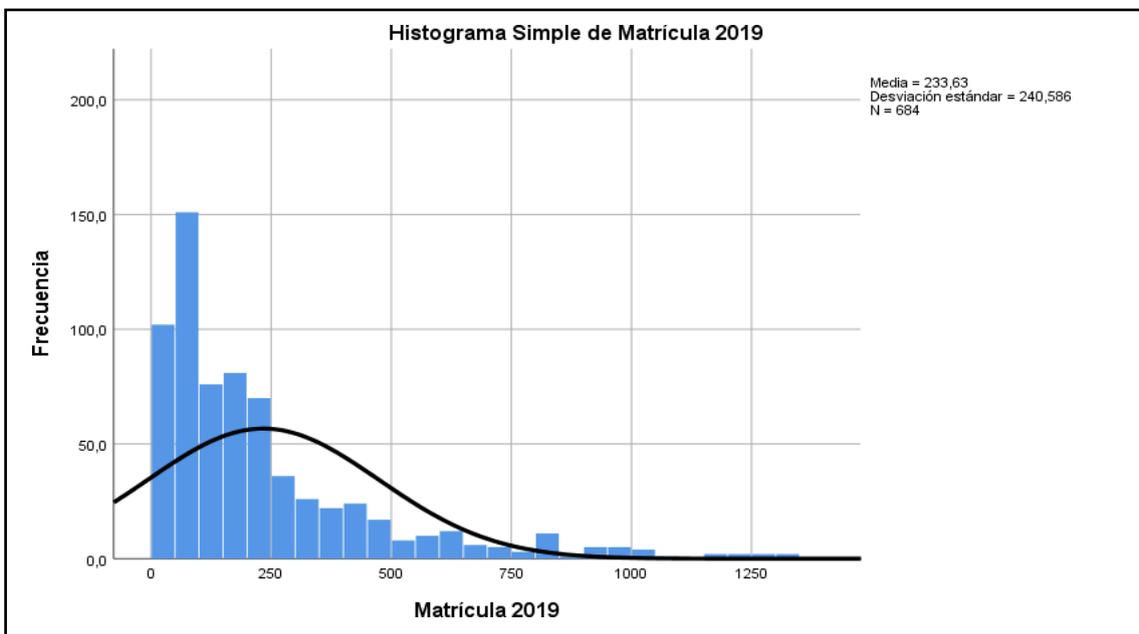


Figura 8: Histograma de la variable Matrícula 2019 luego de la limpieza de datos

Fuente: Elaboración propia

Como resultado se observa que la asimetría presentada anteriormente disminuye.

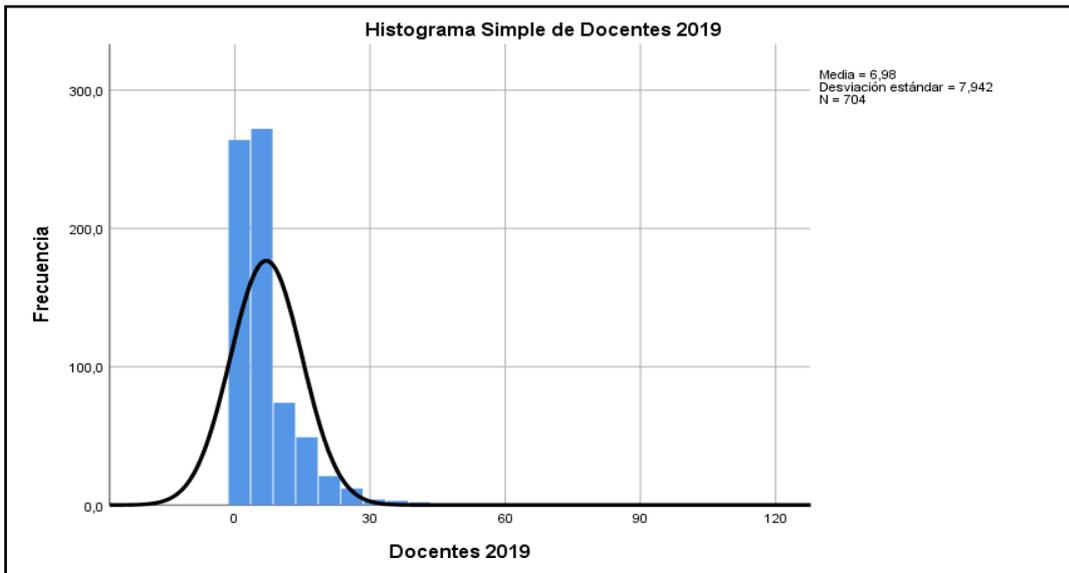


Figura 9: Histograma de la variable Docentes 2019 antes de la limpieza de datos

Fuente: Elaboración propia

El conjunto de datos para la variable número de docentes en el 2019 presenta una distribución asimétrica hacia la derecha.

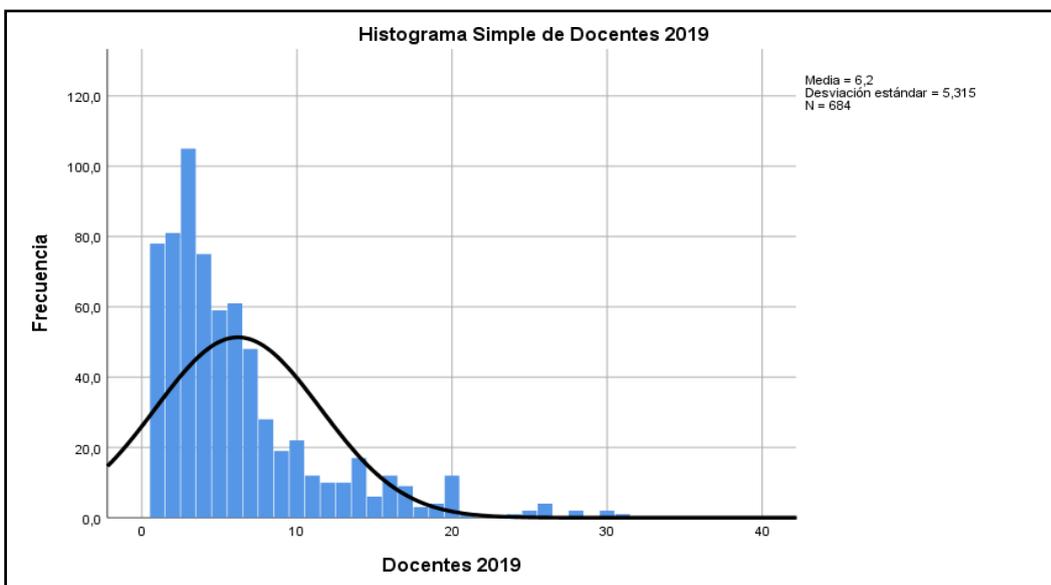


Figura 10: Histograma de la variable Docentes 2019 luego de la limpieza de datos

Fuente: Elaboración propia

Como resultado de la limpieza de datos se observa que la asimetría presentada anteriormente disminuye.

4.3. Obtención de los conglomerados

Con los 684 registros finales, se aplicó el análisis de clúster bietápico; asimismo, la medida de distancia utilizada fue log-verosimilitud (para variables cuantitativas y categóricas) y el criterio de agrupamiento utilizado fue el BIC (criterio bayesiano) obteniendo los siguientes resultados:

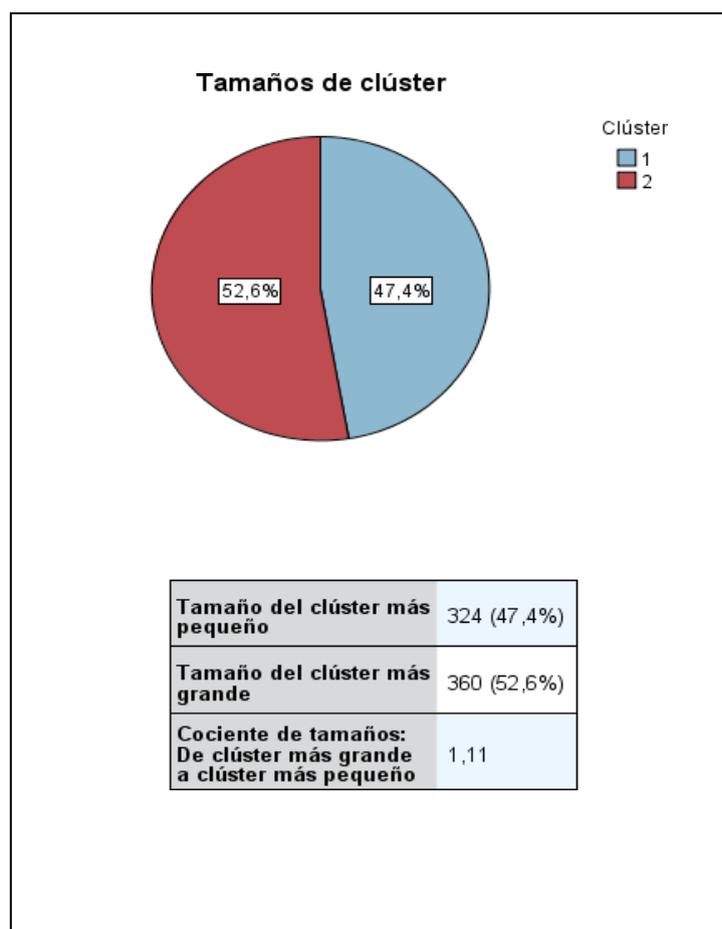


Figura 11: Tamaños de los conglomerados

Los conglomerados obtenidos se distribuyen de la siguiente manera: el primero conformado por un total de 324 (47.4%) y el segundo por 360 registros (52.6%).

Tabla 5: Estadísticos de resumen de las variables cuantitativas en cada conglomerado

Número de conglomerados en dos fases	Matrícula 2019	Docentes 2019
	Media	Media
1	331.60	8.57
2	145.45	4.06

De los resultados obtenidos para las variables cuantitativas se puede apreciar que el primer conglomerado presenta una media mayor que la segunda para la variable cantidad de matriculados en el 2019. Por otro lado, analizando la variable cantidad de docentes vemos que el primer conglomerado presenta una mayor media que el segundo conglomerado.

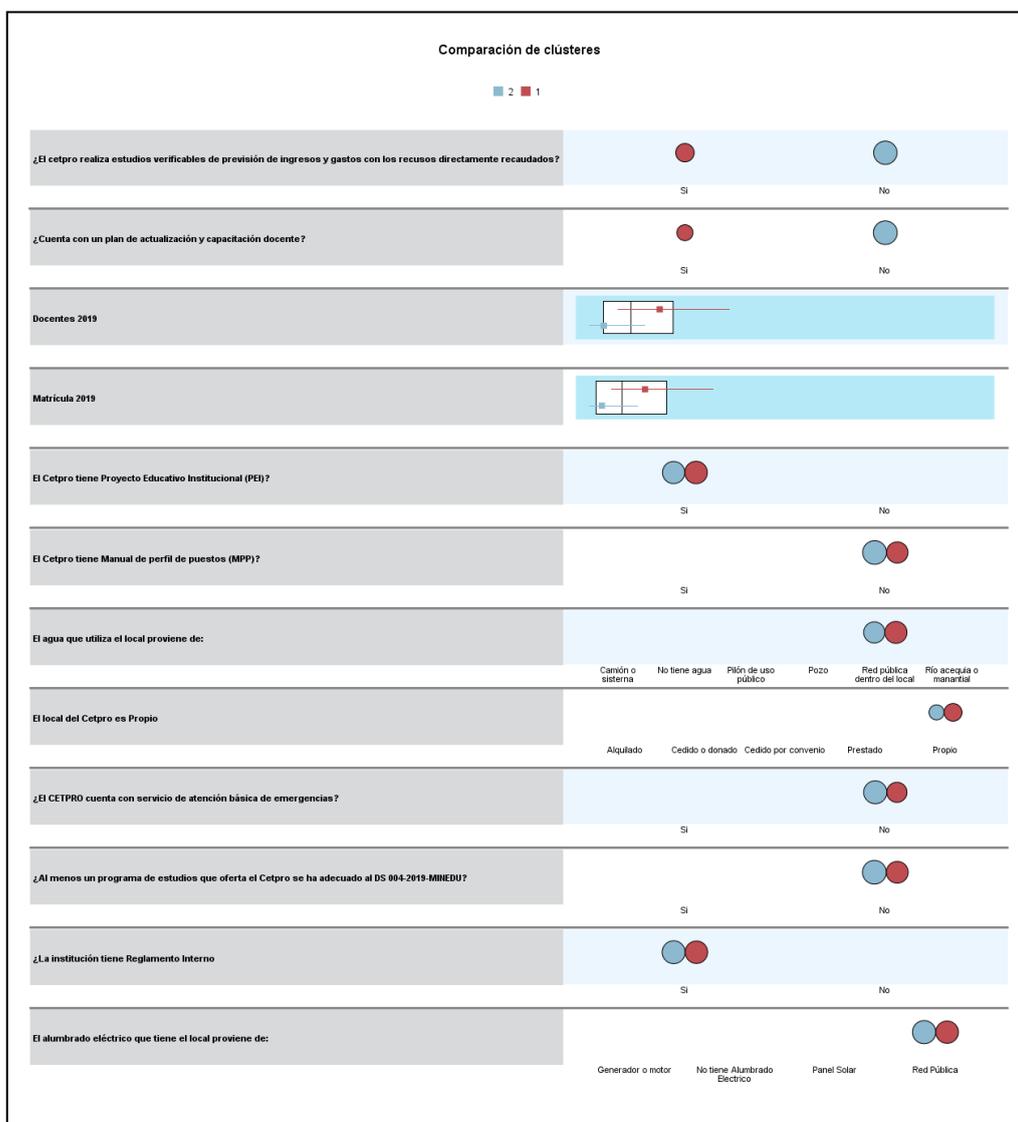


Figura 12: Comparación de los conglomerados

De los resultados obtenidos tenemos que, el primer conglomerado está conformado por un 66.4% de CETPRO que realizan estudios verificables de ingresos y gastos y evidencian tener recursos directamente recaudados, un 50.6% de los mismos cuenta con un plan de actualización y capacitación docente, un 99.7% cuenta con un Proyecto Educativo Institucional (PEI), un 89.5% no cuenta con un Manual de Perfil de Puestos (MPP), un 95.1% cuenta con servicio de agua potable proveniente de red pública, un 66.1% cuenta con local propio, un 79.6% no cuenta con un servicio de atención básica de emergencias, un 94.1% no ha cumplido con adecuar al menos uno de sus programas ofertados al D.S. 004-2019-MINEDU, un 99.4% cuenta con Reglamento Interno (RI) y finalmente un 99.7% cuenta con alumbrado eléctrico proveniente de red pública.

Por otro lado, el segundo conglomerado está conformado por un 95.8% de CETPRO que no realizan estudios verificables de ingresos y gastos y evidencian tener recursos directamente recaudados, un 99.4% de los mismos que no cuenta con un plan de actualización y capacitación docente, un 87.2% cuenta con un Proyecto Educativo Institucional (PEI), el 100% no cuenta con un Manual de Perfil de Puestos (MPP), un 79.2% cuenta con servicio de agua potable proveniente de red pública, un 40.6% cuenta con local propio, un 94.2% no cuenta con un servicio de atención básica de emergencias, un 99.7% no ha cumplido con adecuar al menos uno de sus programas ofertados al D.S. 004-2019-MINEDU, un 93.1% cuenta con Reglamento Interno (RI) y finalmente un 95.6% cuenta con alumbrado eléctrico proveniente de red pública.

4.4. Propuesta de perfiles de los conglomerados

4.4.1. Perfil del conglomerado 1 en función de las condiciones básicas de calidad analizadas

Analizando la Gestión Institucional, tenemos CETPRO que cuentan con un Proyecto Educativo Institucional (PEI), poseen Reglamento Interno (RI); sin embargo, adolecen de un Manual de Perfil de Puestos (MPP) y no cuentan con servicio de atención básica de emergencias (Tópico dentro de la institución).

En relación con la Gestión Académica, se evidencia que predomina el incumplimiento de la adecuación de los programas de estudio de los CETPRO a la normativa emitida por el Ministerio de Educación. Por otro lado, el conglomerado en mención contiene a las instituciones con mayor demanda de matriculados.

En cuanto a la Infraestructura, tenemos a los CETPRO que cuenta con local propio lo cual pone de manifiesto su capacidad para darle continuidad al servicio educativo en el tiempo, asimismo, el local cuenta con servicio de agua y alumbrado eléctrico proveniente de red pública.

Del personal docente, tenemos a instituciones que demuestran su preocupación por contar con un plan de actualización y capacitación docentes que fortalezca la capacidad para brindar servicio educativo, asimismo, cuentan con mayor disponibilidad de personal docentes.

Por último, en relación con la previsión de ingresos y gastos; son instituciones que demuestran tener ingresos propios denominados como directamente recaudados; asimismo, ponen de manifiesto la buena gestión de estos a través de estudios verificables de previsión de ingresos y gastos.

4.4.2. Perfil del conglomerado 2 en función de las condiciones básicas de calidad analizadas

Analizando la Gestión Institucional, tenemos CETPRO que cuentan con un Proyecto Educativo Institucional (PEI), poseen Reglamento Interno (RI); sin embargo, adolecen de un Manual de Perfil de Puestos (MPP) y no cuentan con servicio de atención básica de emergencias (Tópico dentro de la institución). Si bien existe bastante coincidencia respecto al conglomerado uno, se aprecia que la cantidad de CETPRO que presentan estas características difieren en cantidades.

En relación con la Gestión Académica, se evidencia que predomina el incumplimiento de la adecuación de los programas de estudio de los CETPRO a la normativa emitida por el Ministerio de Educación. Por otro lado, el conglomerado en mención contiene a las instituciones con menor demanda de matriculados.

En cuanto a la Infraestructura, vemos que lo conforman CETPRO que cuentan con local propio lo cual pone de manifiesto su capacidad para darle continuidad al servicio educativo en el tiempo, asimismo, el local cuenta con servicio de agua y alumbrado eléctrico proveniente de red pública.

Del personal docente, tenemos a instituciones que ponen de manifiesto su poca capacidad y preocupación por contar con un plan de actualización y capacitación docente que fortalezca la capacidad para brindar servicio educativo, asimismo, cuentan con menor disponibilidad de docentes en comparación que el primer conglomerado

Por último, en relación con la previsión de ingresos y gastos son instituciones que no cuentan con ingresos propios denominados como directamente recaudados; esto se pone en evidencia porque no cuentan con estudios verificables de previsión de ingresos y gastos.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

A partir de los resultados mostrados en la sección de resultados se tiene las siguientes conclusiones:

1. El análisis de Clúster Bietápico dio como resultado la identificación de dos conglomerados de los Centros de Educación Técnico-Productiva con variables diferenciadoras entre cada grupo identificado.
2. En relación con las principales características asociadas a la gestión institucional, se tiene lo siguiente para los dos conglomerados identificados: a) la tenencia de Proyecto Educativo Institucional (PEI), b) tenencia de Reglamento Interno), c) ausencia de un Manual de Perfil de Puestos, d) ausencia de servicio básico de atención de emergencias. Por lo tanto, podemos afirmar que no existen características diferenciadoras entre conglomerados según las variables relacionadas a la gestión institucional.
3. En relación con las principales características asociadas a la gestión académica, se tiene que para ambos conglomerados predomina como variable principal el incumplimiento de la adecuación de los programas de estudio de los Centros de Educación Técnico-Productiva a la normativa emitida por el Minedu “DS 004-2019-Minedu”; asimismo, se ha podido establecer que el conglomerado uno agrupa a los CETPRO con mayor demanda de matriculados en comparación que el conglomerado dos con una menor demanda. Por lo tanto, podemos afirmar que existe al menos una característica diferenciadora en los CETPRO según las variables relacionadas a la gestión académica.

4. Asimismo, en relación con las principales características asociadas a la infraestructura y equipamiento se tiene que para ambos conglomerados predomina: a) la tenencia de local propio, b) tenencia de servicio de agua potable proveniente de red pública dentro del local, c) tenencia de servicio de alumbrado eléctrico proveniente de red pública. Por lo tanto, podemos afirmar que no existen características diferenciadas entre conglomerados según las variables relacionadas a infraestructura y equipamiento.

5. En cuanto a las principales características asociadas al personal docente, tenemos que el conglomerado uno tiene a la variable asociada a la tenencia de un plan de actualización y capacitación docente como herramienta para fortalecer del servicio educativo brindado; caso contrario, en el conglomerado dos en el que predomina la ausencia de un plan de actualización y capacitación docente. Asimismo, los CETPRO del conglomerado uno posee mayor disponibilidad de personal docente en comparación al conglomerado dos. Por lo tanto, podemos afirmar que existen características diferenciadas entre los CETPRO del conglomerado uno y dos.

6. En cuanto a las principales características asociadas a la previsión económica y financiera, tenemos que el conglomerado uno tiene como principal variable la realización de estudios verificables de previsión de ingresos y gastos con los recursos directamente recaudados; caso contrario, en el conglomerado dos en el que predomina la ausencia de estudios de previsión de ingresos y gastos. Por lo tanto, podemos afirmar que existen características diferenciadas entre los CETPRO del conglomerado uno y dos.

7. Se identificaron dos grandes grupos de CETPRO con un determinado perfil para cada uno. En ese sentido, tenemos que las variables: Estudios verificables de previsión de ingresos y gastos con los recursos directamente recaudados y la tenencia de un plan de actualización y capacitación docente, son las principales para diferenciar a un

conglomerado respecto al otros. Por lo tanto, como producto de dicho análisis se tiene un primer conglomerado conformado por 324 (47.4%) instituciones y un segundo conformado por 360 (52.6%).

8. Para cada uno de los conglomerados se establecieron sus perfiles respectivos en función de las variables analizadas, producto de ello, se ha podido identificar CETPRO con mejores condiciones, teniendo para ello el conglomerado 1 que evidencia contar con instituciones con recursos generados por gestión propia,

8.2. Recomendaciones

1. Incluir variables adicionales que se relacionen con las condiciones básicas de calidad que ayuden a un mejor perfilamiento de los conglomerados.
2. Comparar la técnica con conglomerados propuesta con otras técnicas de agrupamiento como K-medias, segmentación de Kohonen entre otras para poder seleccionar el mejor agrupamiento.
3. Recoger los resultados mostrados que están centrados en las condiciones básicas de calidad de los CETPRO públicos para implementar acciones y estrategias direccionadas al fortalecimiento de cada una de las deficiencias identificadas y mencionadas en la conclusión número cinco.
4. De igual forma, se recomienda tomar en cuenta los resultados mostrados como línea base para la implementación de condiciones básicas de calidad que se ajusten a la realidad de los CETPRO públicos; los cuales, han evidenciado presentar deficiencias considerables en los cinco aspectos analizados.

VI. REFERENCIAS BIBLIOGRÁFICAS

Bacher, J.; Wenzig, K.; Volger, M. (2004). *SPSS TwoStep Cluster – A First Evaluation*. Recuperado de https://www.ssoar.info/ssoar/bitstream/handle/document/32715/ssoar-2004-bacher_et_al-SPSS_TwoStep_Cluster_-_a.pdf?sequence=1

Bacher, J.; Wenzig, K.; Vogler, M. (2014). *SPSS TwoStep Cluster – a first evaluation*. Recuperado de https://www.researchgate.net/publication/215665884_SPSS_TwoStep_Cluster_-_a_first_evaluation/link/542af606cf27e39fa917a92/download

Brawijaya. (2011). *Two Step Cluster Analysis*. Recuperado de https://www.academia.edu/4064153/Estudio_Situacion_Educacion_Superior_Tecnologica_Peru

Chiu, T.; Fang, D.; Chen, J.; Wang, Y.; Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In International conference on Knowledge discovery and data mining (pp. 263-268). San Francisco, USA.

De la Fuente, S. (2011). *Análisis de Conglomerados*. Recuperado de https://www.researchgate.net/publication/215665884_SPSS_TwoStep_Cluster_-_a_first_evaluation/link/542af606cf27e39fa917a92/download.

Decreto Supremo N° 004 de 2019 [con fuerza de ley]. “Decreto Supremo que modifica el Reglamento de la Ley N° 28044, Ley General de Educación, aprobado por Decreto Supremo N° 011-2012-ED, y lo adecúa a lo dispuesto en el Decreto Legislativo N° 1375 que modifica diversos artículos de la Ley N° 28044, sobre educación técnico-productiva y dicta otras disposiciones”. 12 de marzo de 2019. D.O. No. 21.

- Espigares-Pinazo, M., Bautista-Vallejo, J. y Duarte, M. (2017). *Uso del Método Bietápico en el estudio de los procesos de enseñanza y aprendizaje musical a través de Moodle*. Recuperado de <https://core.ac.uk/reader/236361319>
- Haya de la Torre, R. (2005). *Estudio sobre la situación de la Educación Superior Tecnológica en el Perú*. Recuperado de https://www.academia.edu/4064153/Estudio_Situacion_Educacion_Superior_Tecnologica_Peru
- Hinneburg, A.; Keim, D. (1999). *Clustering Techniques for Large Data Sets*. Recuperado de https://www.academia.edu/4064153/Estudio_Situacion_Educacion_Superior_Tecnologica_Peru
- Huang, Z. (1998). Extensions to the kmeans algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery* 2,3, Kluwer Academic Publisher, no 3, 283-304, 1384-5810.
- Ley N° 30512 (2016). Ley de Institutos y Escuelas de Educación Superior y de la Carrera Pública de sus Docentes. 02 de noviembre del 2016. D.O. No. 603245.
- Littau, D.; Boley, D. (2006). *Clustering Very Large Data Sets with Principal Direction Divisive Partitioning*. Recuperado de https://www.academia.edu/4064153/Estudio_Situacion_Educacion_Superior_Tecnologica_Peru
- Meilă, M.; Heckerman, D. (febrero de 1998). Una comparación experimental de varios métodos de agrupación e inicialización. Informe técnico MSR-TR-98-06, Microsoft Research, Redmond, WA.
- Ministerio de Educación (2015). *Situación de la Educación Superior Tecnológica y Técnico Productiva hacia una Política de Calidad*. Recuperado de [http://www2.congreso.gob.pe/sicr/cendocbib/con5_uibd.nsf/DAFD29C47494BD7005258312006FA34D/\\$FILE/SITUACION_DE_LA_EDUCACION_SUPERIOR_TE](http://www2.congreso.gob.pe/sicr/cendocbib/con5_uibd.nsf/DAFD29C47494BD7005258312006FA34D/$FILE/SITUACION_DE_LA_EDUCACION_SUPERIOR_TE)

CNO.pdf

Romero, A. (2019). *Condiciones Básicas de Calidad y su relación con la optimización de la oferta educativa en un Instituto de Educación Superior Tecnológica* (Tesis de Maestría). Lima, Perú: Universidad Enrique Guzmán y Valle.

Santos, C. (2015). *Two-Step Clúster en SPSS y técnicas relacionadas* (Trabajo para optar el Máster en Análisis avanzado de Datos Multivariantes). Salamanca, España: Universidad de Salamanca.

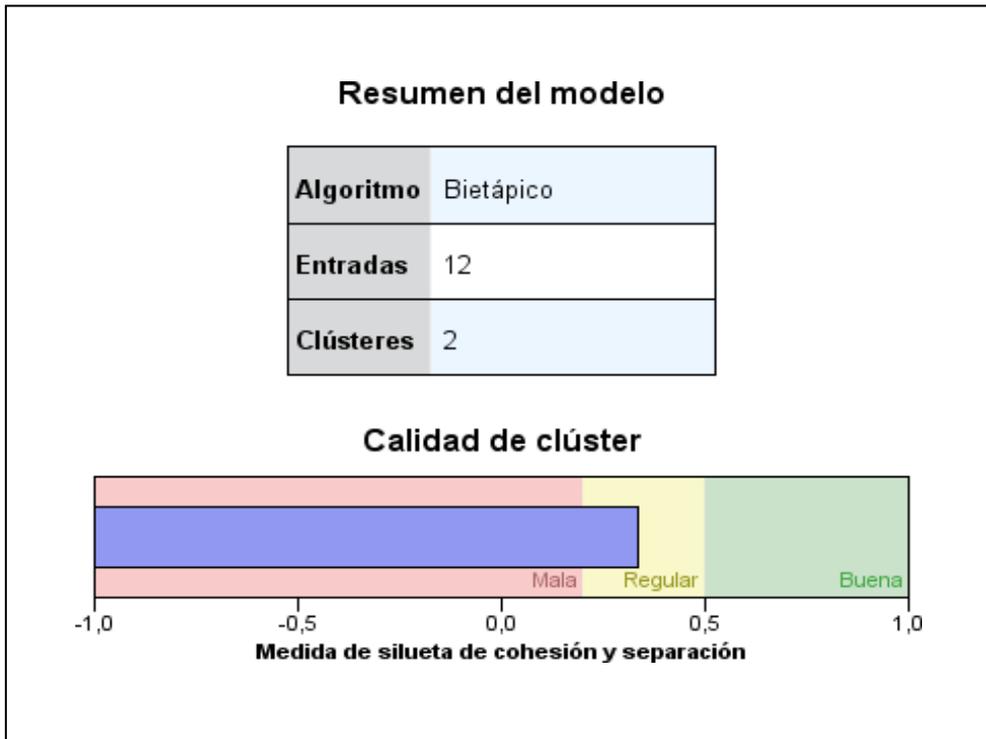
Rubio-Hurtado, M.; Vila-Baños, R. y Duarte, M. (2016). *El análisis de conglomerados bietápico o en dos fases con SPSS*. Recuperado de <https://core.ac.uk/reader/83829700>.

Tirlangi, R.; Mohan, C.; Kalyampudi, L.; Krishna G. (2014). *The Clustering in Large Databases using Clustering Huge Data Sets (CLHDS) Algorithm*. Recuperado de <http://www.ijsrp.org/research-paper-0614/ijsrp-p30111.pdf>

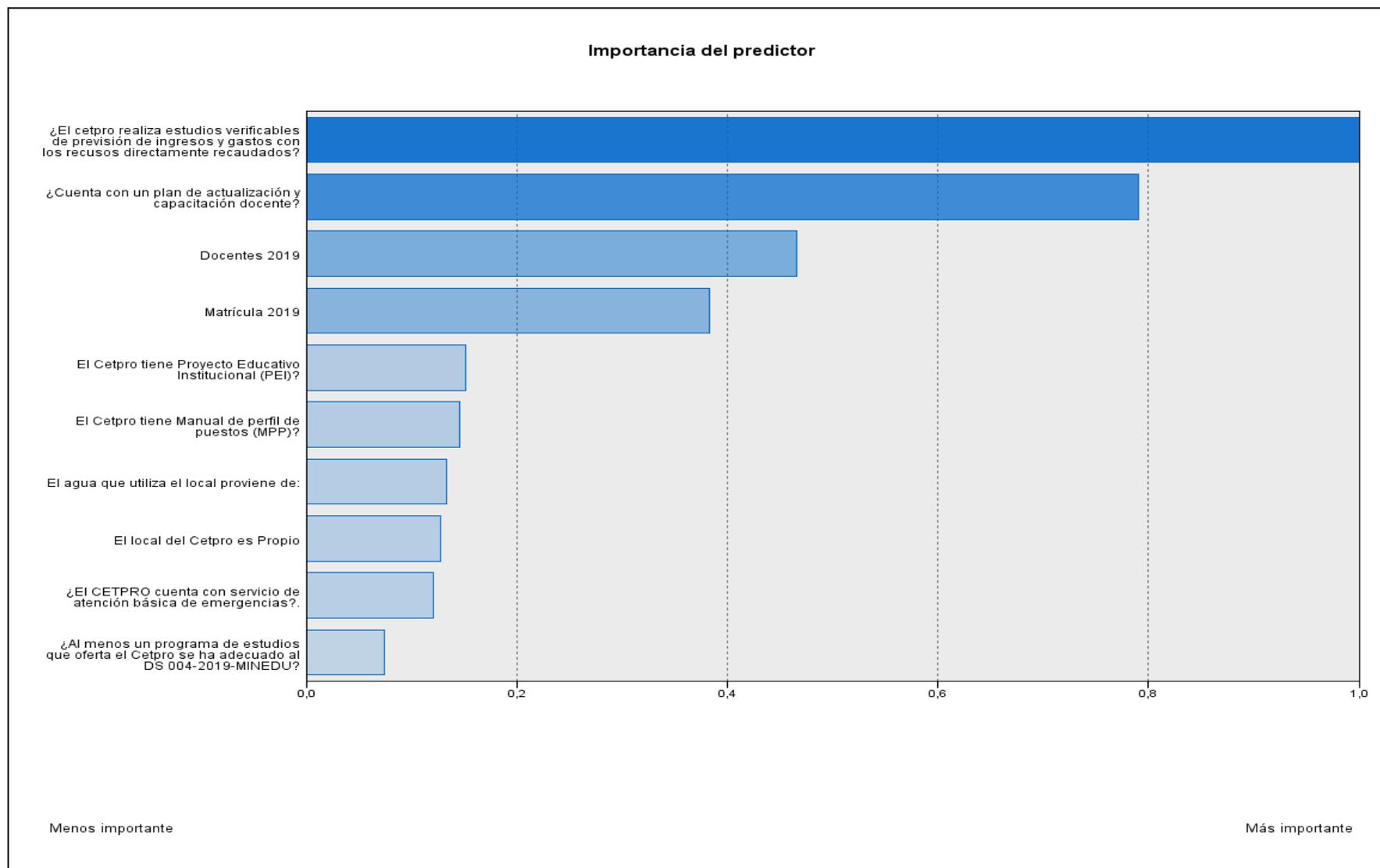
Valero, P. (2011). *Análisis Clúster*. Recuperado de <https://www.uv.es/secretasfp/pdf/2011/docspssav.pdf>

VII. ANEXOS

Anexo 1: Resumen del modelo



Anexo 2: Importancia del predictor



Anexo 3: Comportamiento de las variables en cada conglomerado

