



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ANÁLISIS DE CONGLOMERADOS PARA DETERMINAR LA FOCALIZACIÓN DEL ÁMBITO DE INFLUENCIA DEL PROYECTO DE INVERSIÓN DE MERCAGRO

Línea de investigación:

Estadística y bioestadística

Trabajo de suficiencia profesional para optar el título profesional de Licenciada en

Estadística

Autora:

Tejada Bernal, Jedy Salomé

Asesor:

Ruiz Arias, Raúl Alberto

ORCID: 0000-0002-8877-6158

Jurado:

Luján Campos, Yrma

Estrada Cantero, Jeanette Nazaria

Bustillos Borja, Rubén Hernán

Lima - Perú

2024



ANÁLISIS DE CONGLOMERADOS PARA DETERMINAR LA FOCALIZACIÓN DEL AMBITO DE INFLUENCIA DEL PROYECTO DE INVERSIÓN DE MERCAGRO

INFORME DE ORIGINALIDAD

24%

INDICE DE SIMILITUD

21%

FUENTES DE INTERNET

11%

PUBLICACIONES

10%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	2%
2	qdoc.tips Fuente de Internet	2%
3	gredos.usal.es Fuente de Internet	1%
4	cdn.www.gob.pe Fuente de Internet	1%
5	www.fao.org Fuente de Internet	1%
6	Submitted to Universidad Cesar Vallejo Trabajo del estudiante	1%
7	1library.co Fuente de Internet	1%
8	www.agrorural.gob.pe Fuente de Internet	1%



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

**ANÁLISIS DE CONGLOMERADOS PARA DETERMINAR LA
FOCALIZACIÓN DEL ÁMBITO DE INFLUENCIA DEL PROYECTO DE
INVERSIÓN DE MERCAGRO**

**Línea de investigación:
Estadística y bioestadística**

Trabajo de suficiencia profesional para optar el Título Profesional de Licenciada en
Estadística

Autora:
Tejada Bernal, Jedy Salomé

Asesor:
Ruiz Arias, Raúl Alberto
(ORCID: 0000-0002-8877-6158)

Jurado:
Luján Campos, Yrma
Estrada Cantero, Jeanette Nazaria
Bustillos Borja, Rubén Hernán

Lima-Perú

2024

Dedicatoria

El presente trabajo de investigación es dedicado principalmente a mi esposo y a mis hijos, por ser mi mayor motivación y su paciencia en apoyarme para culminar la elaboración del presente trabajo.

Agradecimiento

Agradezco a mis padres por ser mis mejores ejemplos de la vida, a mi esposo por su apoyo incondicional y a mis amigas por su constante seguimiento y confianza para que culmine esta etapa. También agradezco el apoyo de mi asesor y mis profesores por guiarme en la culminación de este trabajo.

ÍNDICE

RESUMEN	8
ABSTRACT	9
I. INTRODUCCIÓN	10
1.1. Trayectoria del Autor	11
1.2. Descripción de la institución	13
1.3. Organigrama de la institución	14
1.4. Área y funciones desempeñadas	15
II. DESCRIPCIÓN DE UNA ACTIVIDAD ESPECÍFICA	17
2.1 Alcances generales del Proyecto de Inversión MERCAGRO	17
2.1.1 Definición del problema central y medios	17
2.1.2 Población Objetivo	20
2.1.3 Componentes del Proyecto de Inversión MERCAGRO	23
2.2 Enfoque metodológico	23
2.2.1 Métodos de análisis de componentes principales	25
2.2.2 Análisis factorial	25
2.2.3 Métodos de análisis de conglomerados	26
2.2.4 Escalamiento multidimensional	28
2.2.5 Definición de variables a ser incluidas	28
2.2.6 Procesamiento de la información	30
III. APORTES MÁS DESTACABLES A LA INSTITUCIÓN	61
IV. CONCLUSIONES	62
V. RECOMENDACIONES	64
VI. REFERENCIAS	65
VII. ANEXOS	67

ÍNDICE DE TABLAS

Tabla 1 Problema central y población afectada de MERCAGRO	20
Tabla 2 Tipología de la Agricultura Familiar.	21
Tabla 3 Variables e Indicadores para la focalización de MERCAGRO	29
Tabla 4 Valores descriptivos de los indicadores	33
Tabla 5 Resultado de la matriz de correlaciones del estudio	34
Tabla 6 Resultado de la prueba de KMO y Bartlett	35
Tabla 7 Resultado de las Comunalidades	35
Tabla 8 Resultado de la varianza total	36
Tabla 9 Resultado de la matriz de componentes	37
Tabla 10 Resultado de la matriz de componentes rotado	38
Tabla 11 Resultado del historial de conglomerados del Componente 1	47
Tabla 12 Resultado del valor promedio de las variables del Componente 1	50
Tabla 13 Resultado de los distritos identificados dentro de cada clúster del componente	151
Tabla 14 Resultado del historial de conglomerados del Componente 2	53
Tabla 15 Resultado del valor promedio de las variables del Componente 2	56
Tabla 16 Resultado de los distritos identificados dentro de cada clúster del componente	257
Tabla 17 Resultado del agrupamiento de distritos en cada conglomerado y componente	58
Tabla 18 Distribución de los distritos focalizados	60

ÍNDICE DE FIGURAS

Figura 1 Organigrama del Programa de Desarrollo Agrario Rural	15
Figura 2 Árbol de problemas y causas del Proyecto de Inversión MERCAGRO	18
Figura 3 Árbol de medios y fines del Proyecto de Inversión MERCAGRO	19
Figura 4 Distribución de los productores agrarios por departamento	21
Figura 5 Esquema de la población objetivo	22
Figura 6 Esquema del servicio y componentes de MERCAGRO	23
Figura 7 Esquema de métodos multivariantes	24
Figura 8 Gráfico de la clasificación del análisis multivariante según el tipo de variable	25
Figura 9 Gráfico del dendograma	27
Figura 10 Esquema de pasos para el procesamiento de información	30
Figura 11 Distribución de valores perdidos por regiones	30
Figura 12 Comparación sin y con datos imputados de la variable “Autoconsumo”	31
Figura 13 Comparación sin y con datos imputados de la variable “sin capacitación”	32
Figura 14 Comparación sin y con datos imputados de la variable “Riego tecnificado”	32
Figura 15 Resultado del gráfico de sedimentación	37
Figura 16 Resultado del gráfico de componentes rotados	39
Figura 17 Variabilidad de los datos del porcentaje de la población rural	40
Figura 18 Variabilidad de los datos del porcentaje de pobreza monetaria	41
Figura 19 Variabilidad de los datos del IVIA	42
Figura 20 Variabilidad de los datos del PPE	43
Figura 21 Variabilidad de los datos del porcentaje de autoconsumo	44
Figura 22 Variabilidad de los datos del porcentaje de productores sin capacitación	45

Figura 23 Variabilidad de los datos del porcentaje de UA sin riesgo	46
Figura 24 Resultado del gráfico Dendograma del Componente 1	49
Figura 25 Resultado del gráfico de los 4 clústeres dentro de Vulnerabilidad Social	52
Figura 26 Resultado del gráfico Dendograma del Componente 2	55
Figura 27 Resultado del gráfico de los 5 clústeres dentro de la Potencialidad de Demanda	58

RESUMEN

Objetivo: La presente investigación titulada: “Análisis de conglomerados para determinar la focalización del ámbito de influencia del proyecto de inversión de MERCAGRO”, tiene como objetivo elaborar la metodología estadística para la focalización del ámbito de intervención del proyecto de inversión denominado MERCAGRO, a través de un análisis multivariante que permita formar parte de los documentos técnicos que sustentan la identificación del proyecto de inversión en el marco del “invierte.pe” del Ministerio de Economía y Finanzas. **Método:** La metodología aplicada es un análisis multivariante, que permitió a través de un análisis de componentes, determinar dos grupos: Vulnerabilidad Social y Potencialidad de demanda, donde se tomaron en cuenta las variables Pobreza monetaria, Índice de vulnerabilidad a la inseguridad alimentaria, Población rural, Potencial productivo efectivo, Producción destinada al autoconsumo, Productores agrarios de la agricultura familiar sin capacitación y Unidades agropecuarias sin riego. **Resultados:** Con el análisis clúster, se logró categorizar a los distritos en cada componente, es así que, dentro del componente de Vulnerabilidad Social, los distritos fueron categorizados en 4 clúster: Muy Alta, Alta, Baja y Muy Baja y para el componente Potencialidad de demanda, los distritos fueron categorizados en 5 clúster: Muy Alta, Alta, Medio, Baja y Muy Baja. **Conclusiones:** Para la focalización se determinó que, los distritos que tienen Vulnerabilidad Social dentro de la categoría de “Muy Alta” y “Alta” y aquellos distritos que tienen Potencialidad de demanda dentro de las categorías de “Muy Alta”, “Alta” o “Medio” son seleccionados para formar parte del territorio a intervenir del proyecto MERCAGRO.

Palabras clave: Mercagro, agricultura familiar, focalización, clúster, productor agrario.

ABSTRACT

Objective: The present investigation with the title “Cluster analysis to determine the targeting of the ambit of influence of the MERCAGRO investment project”, has the objective is to develop the statistical methodology for targeting the ambit of intervention of the investment project called MERCAGRO through a multivariate analysis that allows it to be part of the technical documents that support the identification of the investment project within the framework of the “invierte.pe” of the Ministry of Economy and Finance. **Method:** The methodology applied is a multivariate analysis, which allowed, through a component analysis, to determine two groups: Social Vulnerability and Demand Potential, where the variables Monetary poverty, Vulnerability index to food insecurity, Rural population, Effective productive potential, Production intended for self-consumption, agricultural producers of family agriculture without training and agricultural units without irrigation. **Results:** With the cluster analysis, it was possible to categorize the districts into each component, within the Social Vulnerability component, the districts were categorized into 4 clusters: Very High, High, Low and Very Low and for the Demand Potential component, the districts were categorized into 5 clusters: Very High, High, Medium, Low and Very Low. **Conclusions:** For targeting, it was determined that the districts that have Social Vulnerability within the category of “Very High” and “High” and those districts that have demand potential within the categories of “Very High”, “High” or “Medium” are selected to be part of the territory to intervene in the MERCAGRO project.

Keywords: Mercagro, family agriculture, targeting, cluster, agricultural producer.

I. INTRODUCCIÓN

El Programa de Desarrollo Productivo Agrario Rural (AGRO RURAL) es la unidad ejecutora del Ministerio de Desarrollo Agrario y Riego (MIDAGRI), tiene como finalidad promover el desarrollo agrario rural a través del financiamiento de inversiones en zonas rurales en el ámbito agrario; asimismo, es la responsable de articular las acciones relacionadas con la promoción del desarrollo agrario rural, en el marco de los lineamiento de políticas establecidos por el Ministerio de Desarrollo Agrario y Riego (MIDAGRI).

Los proyectos de inversión pública son intervenciones temporales, que son financiadas total o parcial con financiamiento del tesoro público destinada a la formación de capital fijo, humano, natural, institucional y/o intelectual. Su objetivo es cerrar brechas prioritarias, en ese sentido, debe estar vinculada a una meta establecida en el Programa Multianual de Inversiones del MIDAGRI.

El Sistema Nacional de Programación Multianual y Gestión de Inversiones, conocido como INVIERTE:PE, a cargo de la Dirección General de Inversión Pública del Ministerio de Economía y Finanzas, busca orientar el uso de los recursos públicos destinados a la inversión para la efectiva prestación de servicios y la provisión de la infraestructura necesaria para el desarrollo del país, con un enfoque territorial.

El proyecto de inversión de MERCAGRO, se encuentra en proceso de formulación en el marco del INVIERTE:PE, este proyecto tiene por objeto la “Mejora del acceso de los productores de la agricultura familiar a los servicios de apoyo al desarrollo productivo en el acceso a mercados de forma rentable, sostenible e inclusiva” y busca mejorar el acceso a mercados rentables, sostenibles e inclusivos para los productores dedicados a la agricultura familiar de subsistencia e

intermedia en las zonas de la sierra y selva del Perú a través de su inserción en cadenas de valor inclusivas.

El presente trabajo tiene como objeto brindar la metodología técnica estadística para la focalización del territorio del proyecto de inversión denominado MERCAGRO. En el proceso de evaluación se utilizaron seis variables, entre las cuales, se encuentra la Pobreza monetaria, Índice de vulnerabilidad a la inseguridad alimentaria, Población rural, Potencial productivo efectivo, Producción destinada al autoconsumo, Productores agrarios de la agricultura familiar sin capacitación y Unidades agropecuarias sin riego, donde se desarrolló el análisis de componentes principales, encontrándose dos componentes denominados “Vulnerabilidad Social” y “Potencialidad de demanda”, luego se utilizó el análisis clúster para agrupar los distritos semejantes dentro de cada componente que permitió la selección para la focalización del proyecto de inversión en mención.

El resultado final del análisis determina la selección de 359 distritos focalizados en 21 regiones del país que cumplen con las características asociadas al objetivo del proyecto de inversión.

1.1. Trayectoria del Autor

La autora del presente trabajo de investigación es bachiller en la carrera profesional de Estadística, egresada de la maestría en Gerencia Social con mención en Gerencia de Programas y Proyectos de Desarrollo cursado en la Pontificia Universidad Católica del Perú, con diploma de post grado en gerencia social, cursos culminados de gestión de procesos y formulación de indicadores y gestión de riesgos.

Entre los principales cargos desempeñados, se detalla a continuación:

Instituto Nacional de Estadística e Informática, analista estadística en el desarrollo y ejecución de la Encuesta Post Censal y la Encuesta Nacional de Hogares – ENAHO en la Dirección de Muestreo y Marcos Muestrales.

Programa Nacional de Apoyo Directo a los más Pobres – JUNTOS, analista de operaciones, realizando actividades de seguimiento de los indicadores de gestión que corresponden al cumplimiento de corresponsabilidades de los usuarios para la subvención económica que otorga el Programa Juntos.

Despacho Presidencial, participante del equipo técnico que realizaba el seguimiento al cumplimiento de los compromisos de Estado, asumidos por las diferentes carteras del Sector Ejecutivo y los Gobiernos Regionales, en el Consejo de Ministros Descentralizado.

Ministerio de Desarrollo e Inclusión Social, desempeñándose como analista de procesos, así como, el control de calidad de la información recibida en la ficha socioeconómica del Sistema de Focalización de Hogares – SISFOH, para la obtención de la clasificación socioeconómica del hogar; además, el seguimiento al cierre de brechas del indicador de cobertura de la población sin clasificación socioeconómica a nivel distrital.

Ministerio de Educación, especialista de monitoreo, responsable del recojo de la información y los indicadores para el Programa de Acompañamiento Pedagógico Poli docente.

Programa Integral Nacional para el Bienestar Familiar – INABIF, Coordinadora de la Asistencia Económica, responsable del proceso de gestión de la asistencia económica que se otorga a niñas, niños y adolescentes en situación de orfandad que provienen de un hogar en situación de pobreza, promoviendo la Ley 31405, Ley que promueve la protección y desarrollo integral de las niñas, niños y adolescentes que se encuentren en situación de orfandad.

Programa Nacional de entrega de la pensión no contributiva a personas con discapacidad severa en situación de pobreza – CONTIGO, desempeñándose como especialista de afiliaciones, realizando acciones y tareas en cumplimiento del proceso de afiliación a la pensión no contributiva a favor de las personas con discapacidad severa, luego, además, asumió el cargo de confianza como Jefa de la Unidad de Operaciones, dirigiendo y gestionando el cumplimiento del proceso misional denominado “Entrega de la pensión no contributiva”, que comprende los procesos de afiliaciones, autorizaciones y transferencia monetaria.

Programa de Desarrollo Agrario Rural – AGRO RURAL, consultora de la asistencia técnica en gestión de la información integrando el equipo técnico del Proyecto de Inversión MERCAGRO.

1.2. Descripción de la institución

El Programa de Desarrollo Agrario Rural (AGRO RURAL), tiene por finalidad la promoción del desarrollo agrario en el ámbito rural, a través de la ejecución de proyectos de inversión pública en territorios con un precario desarrollo económico, estos proyectos son aprobados en conformidad a la normativa vigente; asimismo, el Programa es responsable de articular las acciones y actividades que se encuentran en relación a la promoción del desarrollo agrario rural, en función a los lineamientos de política establecidos por el Ministerio de Desarrollo Agrario y Riego. Este año tiene cuenta con un presupuesto total de S/372.4 millones de Soles.

AGRO RURAL, es la Unidad Ejecutora con código 011-1296 adscrita al Ministerio de Desarrollo Agrario y Riego (MIDAGRI) y depende funcionalmente del Viceministerio de Desarrollo de Agricultura Familiar e Infraestructura Agraria y Riego.

La misión y visión, de acuerdo a su manual de operaciones, se muestra a continuación

1.2.1. Misión

Diseñar, promover y gestionar modelos de desarrollo agrario rural que faciliten la articulación de las inversiones público-privadas y que contribuyan a la reducción de la pobreza y a la inclusión de las familias rurales.

1.2.2. *Visión*

Familias rurales del Perú mejoran su calidad de vida mediante la ejecución de planes y políticas de desarrollo rural sostenibles concertadas con los Gobiernos Regionales, locales y otros actores sociales.

1.3. Organigrama de la institución

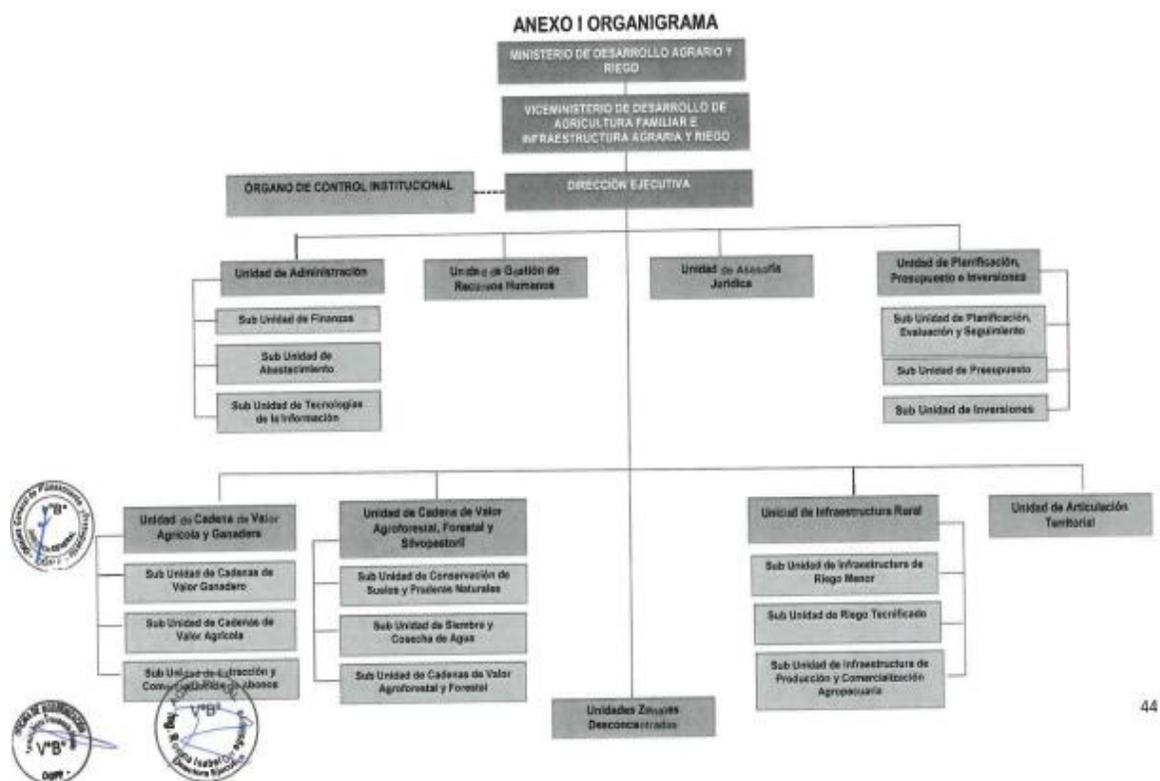
La estructura orgánica de AGRO RURAL, consta de una (01) Dirección Ejecutiva, un (01) órgano de control institucional, dos (02) unidades de asesoramiento, (02) unidades de apoyo, cuatro (04) unidades de línea y veintiuno (21) unidades zonales descentralizadas, ubicadas en 21 departamentos a nivel nacional.

La Unidad de Planificación, Presupuesto e Inversiones es la unidad de asesoramiento, dependiente de la Dirección Ejecutiva, se encarga de asesorar, conducir y coordinar los procesos vinculados con Presupuesto Público, Planeamiento Estratégico, Programación Multianual y Gestión de Inversiones y Modernización de la Gestión Pública; así como, la cooperación técnica, seguimiento y evaluación de la gestión y resultados del Programa. La conforman tres (03) Subunidades:

- Sub Unidad de Planificación, Evaluación y Seguimiento
- Sub Unidad de Presupuesto
- Sub Unidad de Inversiones

Figura 1

Organigrama del Programa de Desarrollo Agrario Rural



Fuente: Programa de Desarrollo Agrario Rural – AGRO RURAL

Nota: El organigrama ha sido aprobado con Resolución Ministerial N°0137-2021-MIDAGRI

1.4. Área y funciones desempeñadas

Respecto al vínculo laboral con AGRO RURAL, la autora trabaja en la Unidad de Planificación, Presupuesto e Inversiones, formando parte integrante del equipo técnico de formulación del proyecto de inversión denominado “Mejoramiento de los servicios de comercialización de productos agropecuarios para su expendio en mercados de consumo – MERCAGRO”.

Las funciones realizadas en este periodo como especialista en la gestión de la información son las siguientes:

- Revisar y desarrollar la formulación del modelo de focalización para la formulación del Proyecto de Inversión de MERCAGRO.
- Revisar la formulación del ítem de aspectos demográficos y económicos del capítulo de Diagnóstico de la población afectada de la formulación del Proyecto de Inversión de MERCAGRO.
- Elaborar los instrumentos de recojo de información a la población afectada para la formulación del proyecto de inversión MERCAGRO.
- Registro, procesamiento y análisis de la información recogida en campo para la formulación del proyecto de inversión MERCAGRO.

II. DESCRIPCIÓN DE UNA ACTIVIDAD ESPECÍFICA

2.1 Alcances generales del Proyecto de Inversión MERCAGRO

2.1.1 *Definición del problema central y medios*

Se define como “Inadecuado acceso de los productores de la agricultura familiar a los servicios de apoyo al desarrollo productivo en el acceso a mercados de forma rentable, sostenible e inclusiva”. El problema identificado surge debido a:

- i. Inadecuados e insuficientes medios o espacios para la comercialización.
- ii. Bajos volumen, calidad, inocuidad y valor agregado de los productos.
- iii. Baja capacidad organizativa, de comercialización y empresarial de los productores.
- iv. Limitada articulación comercial de los productores con empresas privadas en cadenas de valor.
- v. Baja provisión de servicios de apoyo a la producción y comercialización (tecnológicos y financieros)¹ y escasa información para la toma de decisiones.

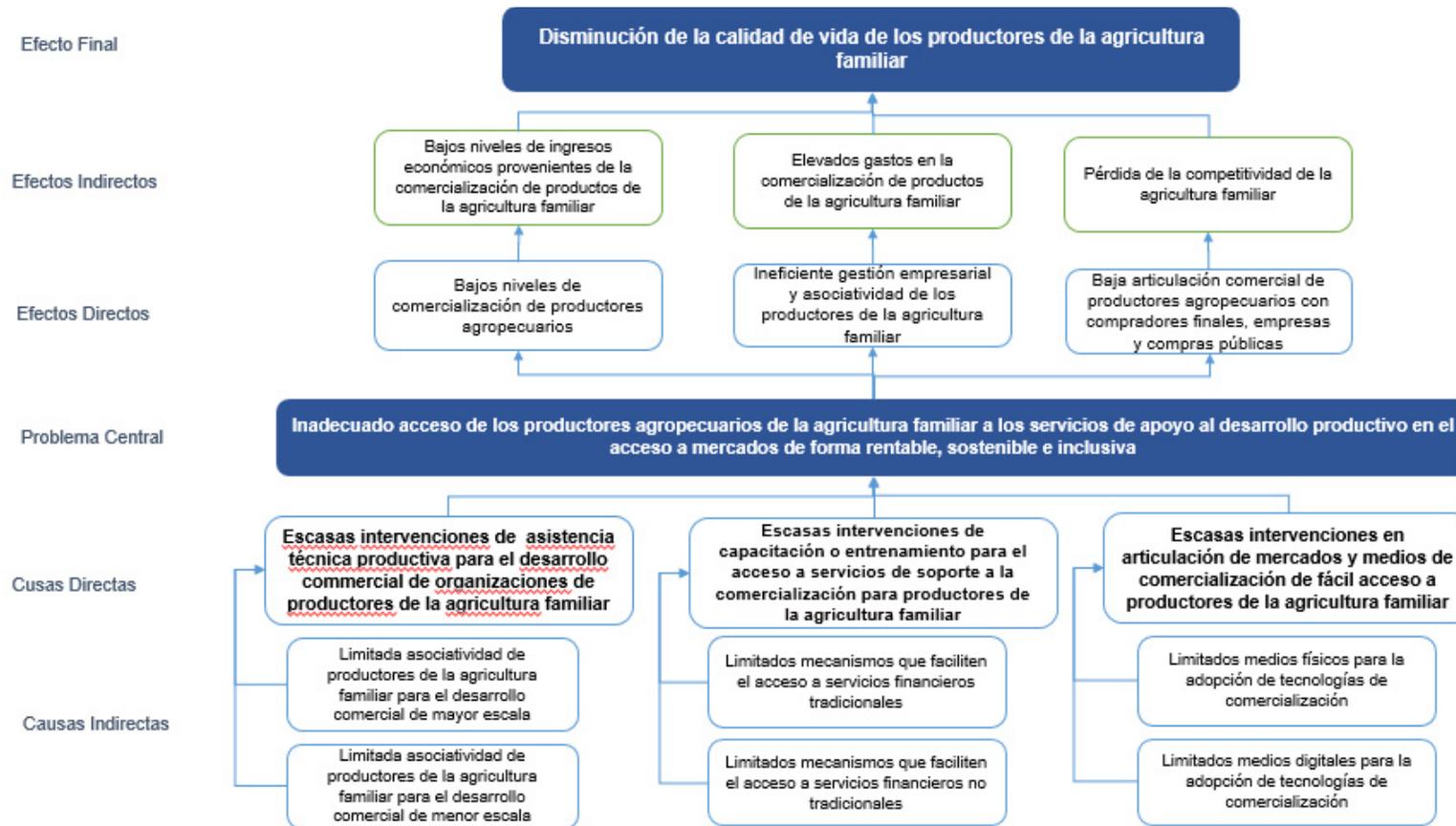
Esta situación, que se explica a través de un proceso de exclusión histórico del medio rural, se traduce en bajos niveles de tenencia de activos y de acceso a servicios básicos, infraestructura y capital humano (FAO, Unión Europea, y CIRAD 2022; Fort y Paredes 2015), lo que conlleva a un incremento en los costos de producción, limitado poder de negociación, limitada escala de producción y escasa capacidad de articulación con los actores de la cadena de valor (FAO, Unión Europea, y CIRAD 2022).

El árbol de problemas, causas y efectos, y el árbol de medios, se presentan a continuación:

¹ Agüero et al. 2020; FAO, Unión Europea y CIRAD 2022; Banco Mundial 2017.

Figura 2

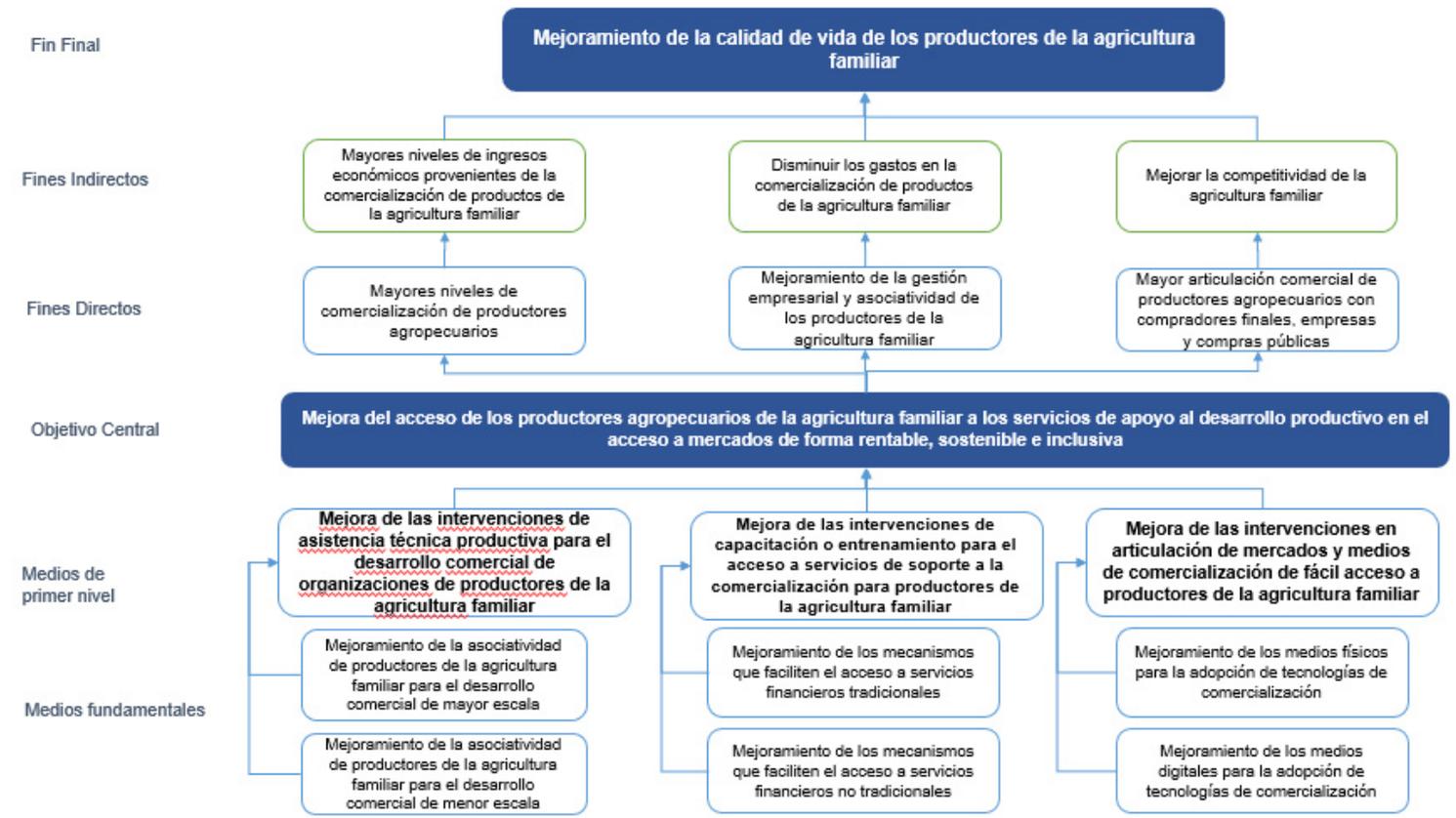
Árbol de problemas y causas del Proyecto de Inversión MERCAGRO



Nota: El árbol de medios y fines se muestran en el capítulo 2 denominado Identificación, que forma parte del Anexo 07: “Contenido mínimo del estudio de pre inversión a nivel perfil para proyectos de inversión”

Figura 3

Árbol de medios y fines del Proyecto de Inversión MERCAGRO



Nota: El árbol de medios y fines se muestran en el capítulo 2 denominado Identificación, que forma parte del Anexo 07: “Contenido mínimo del estudio de pre inversión a nivel perfil para proyectos de inversión”

De acuerdo a lo señalado en la Figura 3, el Proyecto MERCAGRO busca mejorar el acceso a mercados rentables, sostenibles e inclusivos para los productores dedicados a la agricultura familiar de subsistencia (PAFS) y productores agrarios de la agricultura familiar intermedia de nivel 1 (PAFI I) en las zonas de la Sierra y la Selva del Perú a través de su inserción en cadenas de valor inclusivas.

2.1.2 Población Objetivo

El problema central y la población afectada identificada en el Proyecto, se muestra a continuación:

Tabla 1

Problema central y población afectada de MERCAGRO

Problema central	Población afectada
Inadecuado acceso de los productores de la agricultura familiar a los servicios de apoyo al desarrollo productivo en el acceso a mercados de forma rentable, sostenible e inclusiva.	Productores agrarios tipificados como Agricultura Familiar de Subsistencia No Crítica y Agricultura Familiar Intermedia que destinan su producción al mercado en 83 distritos en 8 departamentos del Perú.

Nota: La definición del problema y la identificación de la población afectada se detalla en el capítulo 2 denominado Identificación, que forma parte del Anexo 07: “Contenido mínimo del estudio de pre inversión a nivel perfil para proyectos de inversión”

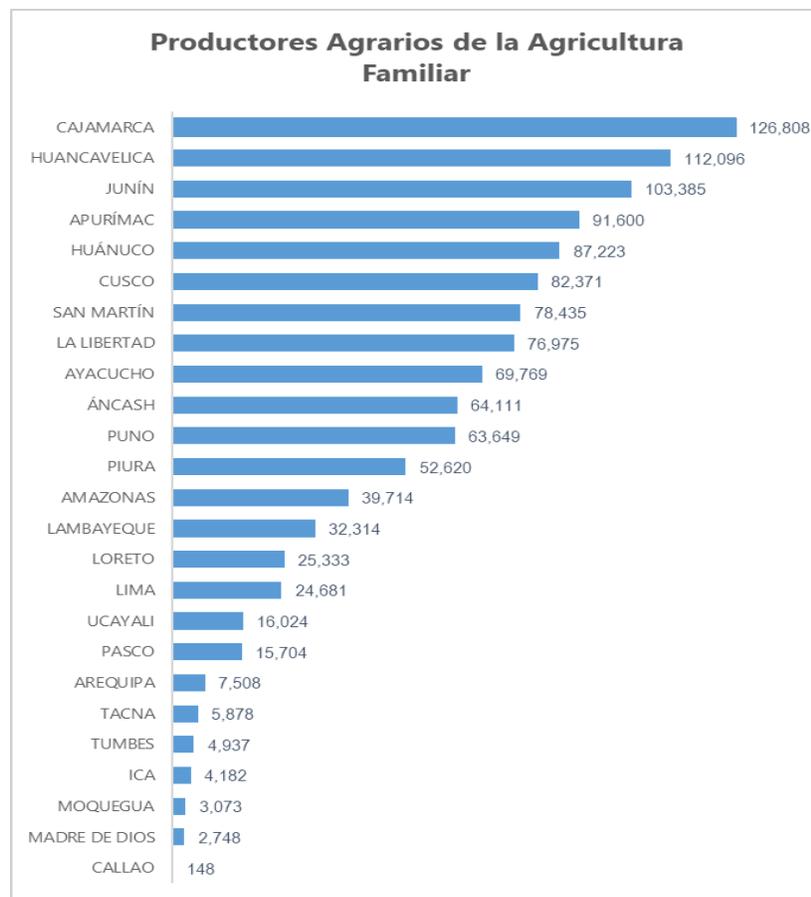
Según el padrón de productores agrarios realizado por el Ministerio de Desarrollo Agrario y Riego, al 31 de diciembre de 2023, se tiene identificado a 1,191,286 productores agrarios a nivel nacional, que pertenecen a la agricultura familiar, según el detalle siguiente:

Tabla 2.*Tipología de la Agricultura Familiar*

Tipología de la Agricultura Familiar	Productores agrarios	%
AF de subsistencia	851,766	71.5%
AF Intermedia	297,574	25.0%
AF Consolidada	41,946	3.5%
Total	1,191,286	100.0%

Fuente: Padrón de Productores Agrarios del MIDAGRI con fecha de corte 31 de diciembre de 2023

Los productores agrarios a nivel departamental, se muestran a continuación:

Figura 4*Distribución de los productores agrarios por departamento*

Fuente: Padrón de Productores Agrarios del MIDAGRI, 31 de diciembre de 2023

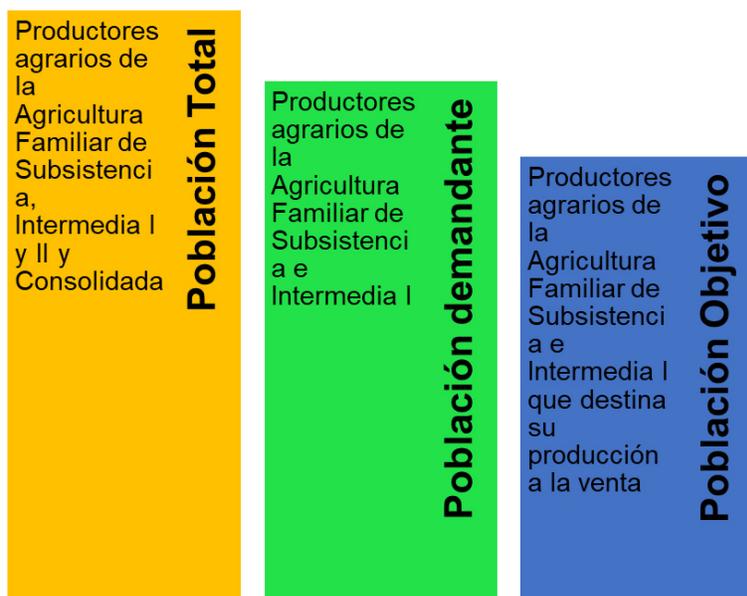
La población total o de referencia ha sido definida como los productores agrarios de la agricultura familiar de tipo Subsistencia, Agricultura Familiar Intermedia I, Agricultura Familiar II y Agricultura Familiar Consolidada.

La población demandante, está referido a los productores agrarios de la agricultura familiar que se encuentran dentro de la tipología de la Agricultura Familiar de Subsistencia y Agricultura Familiar Intermedia I.

Para determinar la población objetivo, está referido al productor agrario de la Agricultura Familiar de Subsistencia y Agricultura Familiar Intermedia I que destina su producción a la venta, tal como se define en el siguiente esquema:

Figura 5

Esquema de la población objetivo



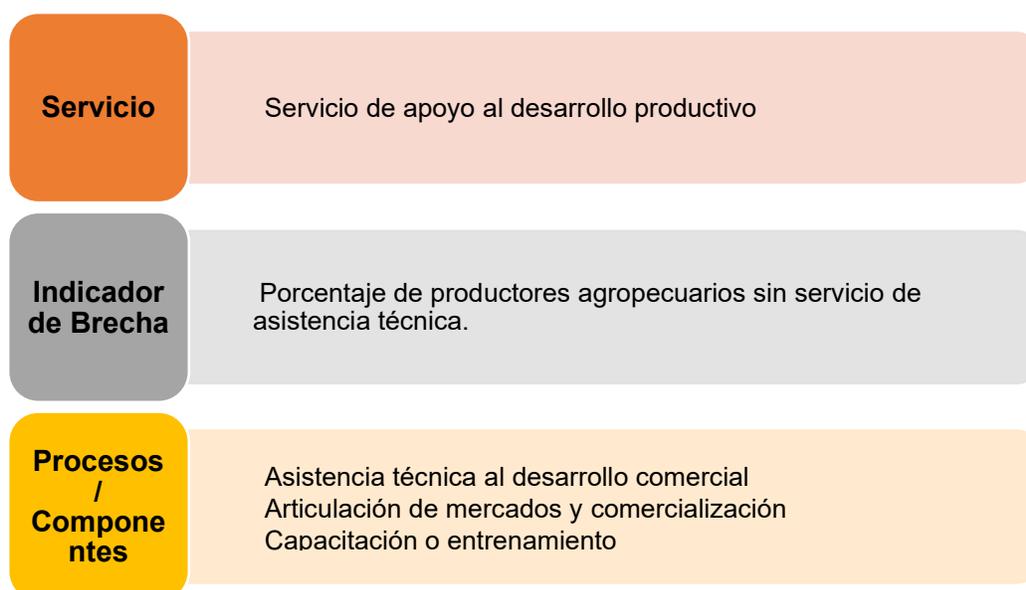
Nota: La definición y el esquema de la población objetivo se detalla en el capítulo 2 denominado Identificación, que forma parte del Anexo 07: “Contenido mínimo del estudio de pre inversión a nivel perfil para proyectos de inversión”

2.1.3 Componentes del Proyecto de Inversión MERCAGRO

Los productores agropecuarios demandan una serie de servicios del Sector Agrario y Riego. Uno de estos servicios es el que se denomina Servicio de Apoyo al Desarrollo Productivo Agrario, que comprende: asistencia técnica, capacitación o entrenamiento, asesoramiento empresarial, implementación de equipos de uso común, entre otros. Para este proyecto, el servicio se determina de la siguiente manera:

Figura 6

Esquema del servicio y componentes de MERCAGRO



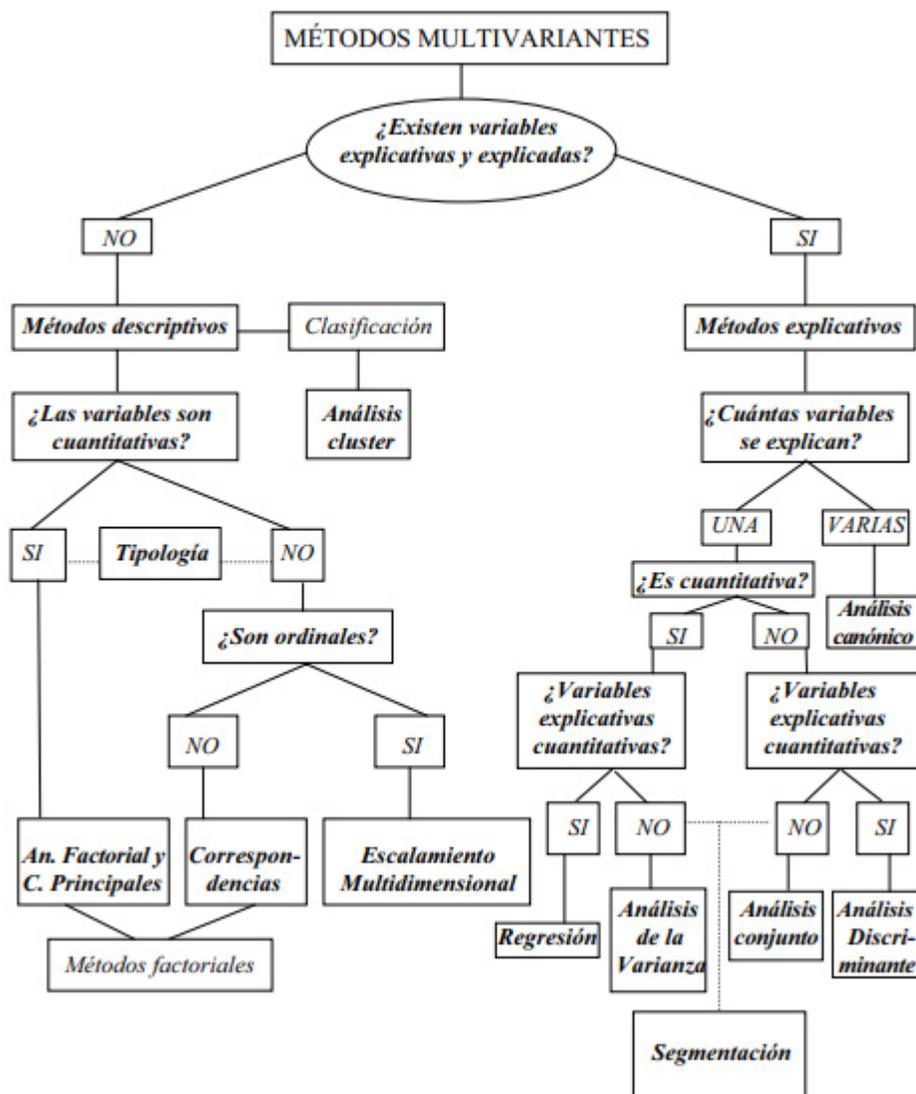
Nota: El servicio, la brecha y los componentes del proyecto son definidos en el capítulo 2 denominado Identificación, que forma parte del Anexo 07: “Contenido mínimo del estudio de pre inversión a nivel perfil para proyectos de inversión”

2.2 Enfoque metodológico

En el libro Técnicas de Análisis Multivariante de Datos, el autor Pérez (2008) clasifica las “técnicas de análisis estadístico en relación al tipo de variable el objetivo principal del tratamiento en conjunto” (p.20), tal como se detalla a continuación:

Figura 7

Esquema de métodos multivariantes

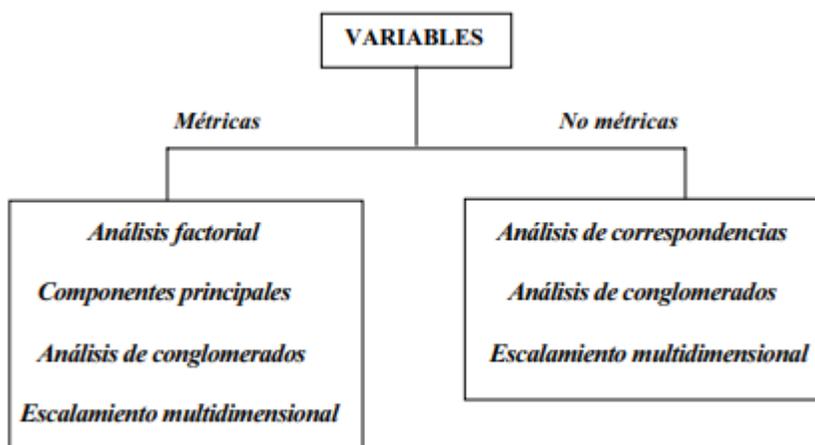


Nota. El gráfico pertenece al libro Técnicas de Análisis Multivariante de Datos (Perez, 2008)

La clasificación del análisis multivariante discrimina entre la existencia de variables explicativas y explicadas; el autor Pérez (2008) muestra la clasificación en función de la naturaleza métrica o no métrica de las variables.

Figura 8

Gráfico de la clasificación del análisis multivariante según el tipo de variable



Nota. El gráfico pertenece al libro Técnicas de Análisis Multivariante de Datos (Perez, 2008, p. 28)

2.2.1 Métodos de análisis de componentes principales

El análisis de componentes principales es una técnica multivariante que busca reducir la dimensión de una tabla de datos muy grande por la gran cantidad de variables que contiene y reducirlas a n componentes que lleguen a explicar significativamente la variabilidad de las variables iniciales.

Es importante verificar que para el análisis de componentes principales las variables tienen que ser cuantitativas.

2.2.2 Análisis factorial

El análisis factorial, al igual que el análisis de componentes principales, son técnicas multivariantes que buscan reducir la dimensión de una tabla muy grande por la gran cantidad de variables y reducirlas a unas cuantas que expliquen significativamente la variabilidad de las variables originales.

La diferencia entre el análisis de componentes principales y el análisis factorial se basa principalmente en el cálculo de sus varianzas, siendo el análisis factorial, el que toma en cuenta dos tipos de varianzas, la varianza común que es la varianza de una variable compartida con otra y la varianza única que es la varianza de una única variable; por otro lado, el análisis de componentes principales no hace ni un análisis de varianzas sólo distingue la varianza total; otra diferencia matemática refiere que el análisis factorial busca un nuevo conjunto de variables en menor número que represente lo que es común a esas variables y el análisis de componentes principales busca combinaciones lineales que expliquen la mayor parte de la variación total.

2.2.3 Métodos de análisis de conglomerados

El análisis clúster es una técnica estadística multivariante que agrupa los casos similares u homogéneos en conjuntos llamados conglomerados y cada conjunto debe ser más heterogéneo entre sí (distancias), existen variadas formas de medir estas distancias para asignar los casos a distintos grupos.

Este análisis inicia con la estimación de similitud entre los casos a través de la correlación con las distintas variables que dispone, en este caso, el proyecto de inversión.

a. Estimación de distancias y similitudes. La proximidad expresa la semejanza que existe entre los casos. Si el valor de la distancia es alto, significa que se encuentra alejado del otro caso; en cambio, el valor de la similitud significa cuánto más próximo se encuentra al caso considerado.

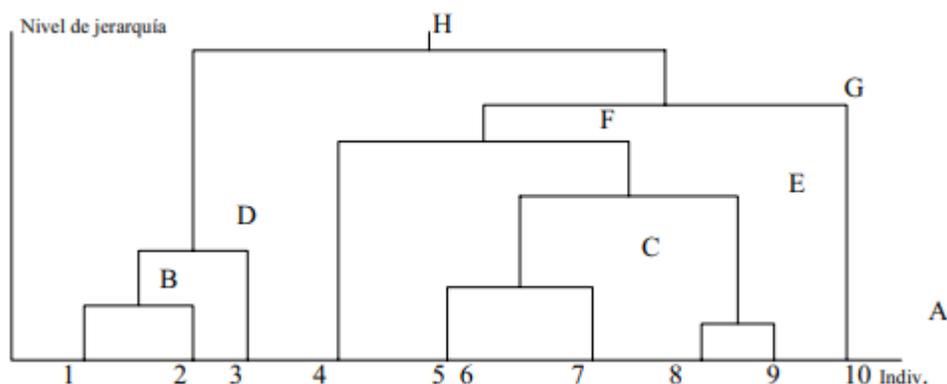
La correlación de Pearson y los coeficientes de Spearman y de Kendall son índices de similitud.

b. Clúster jerárquico: DENDOGRAMA. Muestra los grupos iniciales que se estudian y agrupaciones sucesivas entre ellos de forma que gradualmente se vayan integrando en clúster

que a su vez se unirán entre sí en un nivel superior formando grupos más grandes que al final contendrán todos los casos analizados. La representación gráfica de esta formación de grupos, es denominado dendograma.

Figura 9

Gráfico del dendograma



Nota. El gráfico pertenece al libro *Técnicas de Análisis Multivariante de Datos* (Perez, 2008, p. 427)

Para la agrupación de estos clústeres se utilizan algunos métodos, como el método del promedio, método del centroide, método de la mediana, el método de war y el método de enlace por densidad; a continuación, se describe cada uno de ellos:

Método del promedio entre grupos o Enlace promedio. Considera la distancia media entre cada clúster, utiliza la media ponderada, de tal forma que entre todos los casos en el clúster resultante sea lo más pequeña posible.

Método del centroide o Enlace centroide. Considera como distancia la media aritmética de las variables de los casos que componen los clústeres. Si existiese casos atípicos, es el método más robusto.

Método de la mediana. Considera como distancia la mediana de las variables de los casos que componen los clústeres.

Método de Ward o Enlace por mínima varianza. La distancia entre dos grupos es el resultado del menor incremento de varianza residual global. Este método utiliza más información sobre los casos de los grupos que otros métodos. Es uno de los que ha demostrado mayor eficacia en estudios de simulación.

Método del Enlace por densidad. Utiliza como distancia entre el caso A y el caso B, la media de los inversos de ambas densidades, de tal forma, que se ponderan más próximos dos casos, cada uno de ellos con k casos muy cercanos, que otros dos con k casos lejanos.

2.2.4 Escalamiento multidimensional

Es una técnica de análisis multivariante que representa la distancia entre los casos en una escala euclídeana. Este procedimiento puede utilizarse para tratar datos de poca similitud que proceden de múltiples fuentes de información.

Finalmente, tomando en cuenta las definiciones de los métodos multivariantes descritos y con la finalidad de mostrar los conocimientos adquiridos durante la experiencia laboral de la autora del presente trabajo de investigación, a continuación, se detalla la aplicación del análisis clúster para obtener la focalización de los distritos que formarán el ámbito de influencia del proyecto de inversión denominado “Mejoramiento de los servicios de comercialización de productos agropecuarios para su expendio en mercados de consumo – MERCAGRO”.

2.2.5 Definición de variables a ser incluidas

La selección de variables se encuentra relacionada al objetivo del proyecto de inversión, tomando en cuenta la existencia de información actualizada, debido a su influencia que van a ejercer sobre el resultado, para el análisis estadístico, se debe tener en cuenta los siguientes:

- La estandarización de los valores de las variables al momento de la aplicación del método estadístico de componentes y conglomerados.

- Las variables correlacionadas para evitar que algunas sean combinaciones lineales de otras.
- Evitar el exceso de variables que puedan dificultar el estudio.

Para el presente análisis, se determinaron las siguientes variables e indicadores:

Tabla 3

Variables e Indicadores para la focalización de MERCAGRO

Nº	Variabes	Indicador	Fuente de Información	Valores
1	Población Rural	Porcentaje de población rural en el distrito	INEI	0 a 100%
2	Pobreza Monetaria	Porcentaje de pobreza total en el distrito	INEI (2008)	0 a 100%
3	Vulnerabilidad a la Inseguridad Alimentaria	Índice de la vulnerabilidad a la inseguridad alimentaria	MIDIS (2018)	0 a 1
4	Potencialidad de Producción Efectiva	Potencial Productivo Efectivo	MIDAGRI	1, 2, 3 y 4
5	Producción destinada al Autoconsumo	Porcentaje de producción destinada al autoconsumo	MIDAGRI	0 a 100%
6	Productor sin acceso a capacitación o asistencia técnica	Porcentaje de productores que no recibieron capacitación y/o asistencia técnica	MIDAGRI	0 a 100%
7	Riego Tecnificado	Porcentaje de unidades agropecuarias sin riego tecnificado	MIDAGRI	0 a 100%

Elaboración propia.

2.2.6 Procesamiento de la información

Para el procesamiento de la información, se han seguido los siguientes pasos:

Figura 10

Esquema de pasos para el procesamiento de información

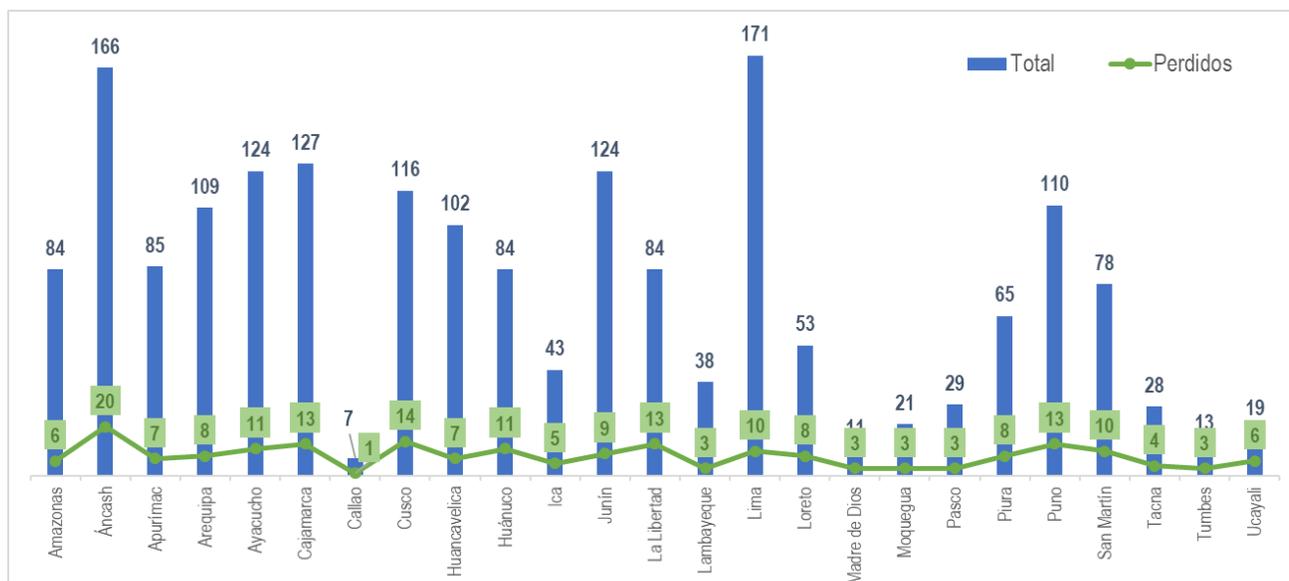


Elaboración propia.

2.2.6.1 Imputación de datos faltantes. En el presente estudio, se ha identificado que en 3 variables existen valores perdidos, variable 1: “Producción destinada al Autoconsumo”, variable 2: “Productor sin acceso a capacitación” y variable 3: “Riego Tecnificado”, corresponden a 200 distritos y al realizar el análisis por región, se identifica que en la región de Ancash tiene la mayor cantidad de valores perdidos.

Figura 11

Distribución de valores perdidos por regiones

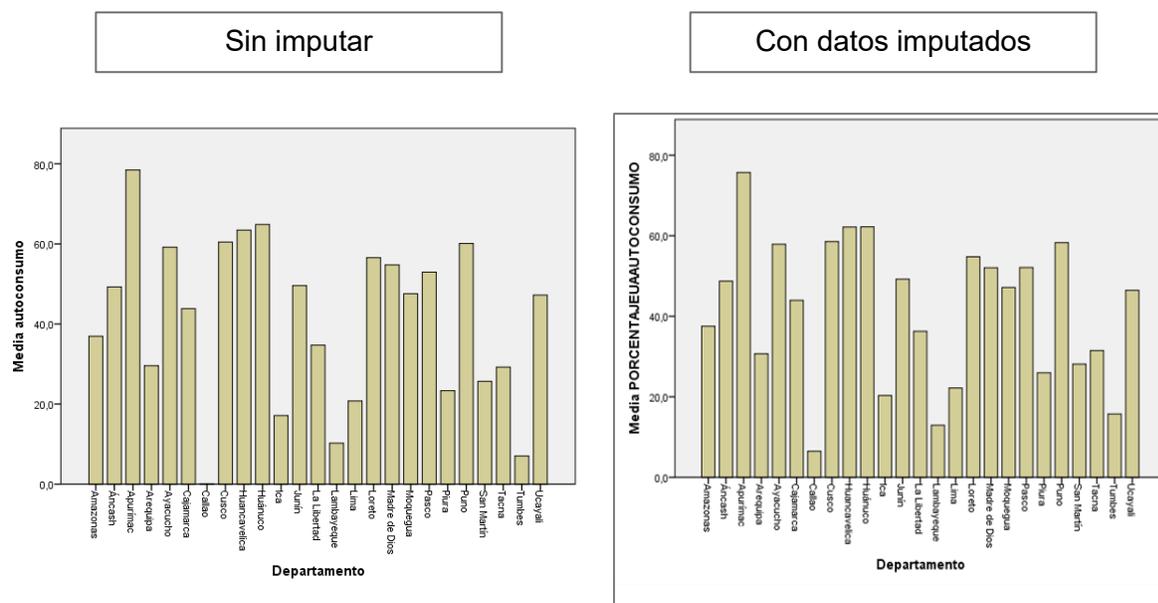


Elaboración propia.

Es importante verificar que los valores imputados no sesguen la distribución de las variables, por ello, se decide imputar el valor perdido con el resultado del promedio a nivel provincia, por cada variable. A continuación, se muestra la distribución de cada variable antes y después del proceso de imputación.

Figura 12

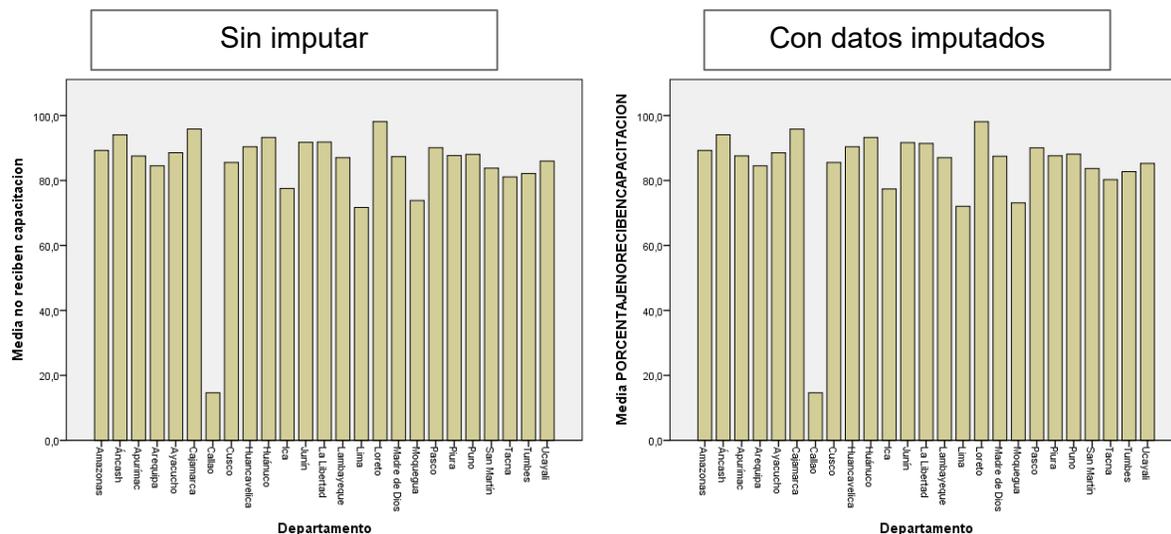
Comparación sin y con datos imputados de la variable “Autoconsumo”



Elaboración propia.

Figura 13

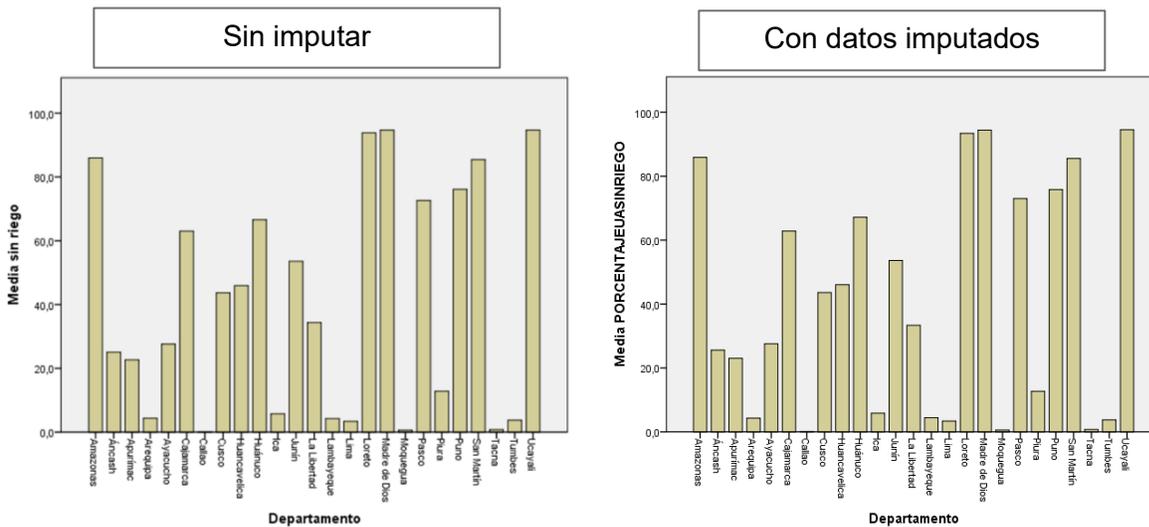
Comparación sin y con datos imputados de la variable “sin capacitación”



Elaboración propia.

Figura 14

Comparación sin y con datos imputados de la variable “Riego tecnificado”



Elaboración propia.

Como se puede apreciar en las Figuras 11, 12 y 13 no se evidencia algún sesgo luego del proceso de imputación; por tanto, se procede a ejecutar el análisis de componentes principales.

2.2.6.2 Análisis factorial. Como paso previo al análisis multivariante, se requiere conocer algunos indicadores estadísticos como la media, desviación estándar, mínimo y máximo valor en cada variable de estudio, para ello, se utiliza el software estadístico SPSS ejecutando el análisis descriptivo, seleccionando los estadísticos mencionados anteriormente, siendo los resultados los siguientes:

Tabla 4

Valores descriptivos de los indicadores

Indicadores	N	Mínim o	Máxim o	Media	Desviación estándar
PORCENTAJE POBLACION RURAL	1 891	0.00	97.58	45.67	30.11
POBREZA MONETARIA	1 891	0.06	81.26	34.12	16.89
ÍNDICE DE LA VULNERABILIDAD A LA INSEGURIDAD ALIMENTARIA	1 891	0.04	0.84	0.51	0.18
POTENCIAL PRODUCTIVO EFECTIVO	1 891	0.00	4.00	2.04	1.01
PORCENTAJE DE AUTOCONSUMO	1 891	0.00	100.00	16.23	19.05
PORCENTAJE NO RECIBEN CAPACITACION	1 891	0.00	100.00	87.34	15.49
PORCENTAJE SIN RIEGO	1 891	0.00	100.00	40.70	38.53

Elaboración propia.

Según los resultados mostrados en la Tabla 4, se puede identificar que los datos ascienden a 1 891 para cada variable, asimismo, este conjunto de datos presenta alta dispersión en todas las variables.

En los 1 891 distritos, en promedio el 45.67% de la población es rural, el 34.12% se encuentra en situación de pobreza, el índice de vulnerabilidad a la inseguridad alimentaria en promedio es 0.51, el potencial productivo efectivo en promedio es 2, en promedio el 16.23% de la producción se destina al autoconsumo, el 87.34% de los productores agrícolas no recibieron capacitación y/o asistencia técnica y en promedio el 40.7% de las unidades agropecuarias se encuentran sin riego.

A continuación, se calcula la matriz de correlaciones, la cual presenta factores comunes, es decir que se encuentran muy correlacionadas entre sí, siendo el valor del determinante 0.06; es decir, que las variables utilizadas están linealmente relacionadas, lo que significa que el análisis factorial es una técnica pertinente.

Tabla 5

Resultado de la matriz de correlaciones del estudio

	POB_RURAL	POBREZA	IVIA	PPE	AUTOCONSUMO	PAF_SINCAPA	UA_SINRIEGO	
Correlación	POB_RURAL	1,000	,666	,788	,082	,188	,319	,359
	POBREZA	,666	1,000	,818	,155	,326	,404	,364
	IVIA	,788	,818	1,000	,183	,331	,435	,421
	PPE	,082	,155	,183	1,000	,166	,220	,270
	AUTOCONSUMO	,188	,326	,331	,166	1,000	,179	,071
	PAF_SINCAPA	,319	,404	,435	,220	,179	1,000	,240
	UA_SINRIEGO	,359	,364	,421	,270	,071	,240	1,000
Sig. (unilateral)	POB_RURAL		,000	,000	,000	,000	,000	,000
	POBREZA	,000		,000	,000	,000	,000	,000
	IVIA	,000	,000		,000	,000	,000	,000
	PPE	,000	,000	,000		,000	,000	,000
	AUTOCONSUMO	,000	,000	,000	,000		,000	,001
	PAF_SINCAPA	,000	,000	,000	,000	,000		,000
	UA_SINRIEGO	,000	,000	,000	,000	,001	,000	

a. Determinante = ,062

Elaboración propia.

Para comprobar que las correlaciones entre las variables son distintas a 0, se ha utilizado la prueba de esfericidad de Bartlett, la misma que ha obtenido un p-valor = 0.0000, con dicho

resultado, se rechaza la hipótesis nula, concluyendo que existe correlación entre las variables; por otro lado, también se ha realizado la prueba de Kaiser-Meyer-Olkin (KMO) que contrasta si las correlaciones parciales entre las variables son pequeñas, siendo su valor 0.787, lo que lleva a considerar que el análisis factorial si es pertinente.

Tabla 6

Resultado de la prueba de KMO y Bartlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,787
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	5252,710
	gl	21
	Sig.	,000

Elaboración propia.

Para el método de extracción, el resultado de las comunalidades, explican que los factores comunes explican la variabilidad de una variable.

Tabla 7

Resultado de las Comunalidades

	Inicial	Extracción
POB_RURAL	1,000	,769
POBREZA	1,000	,796
IVIA	1,000	,887
PPE	1,000	,838
AUTOCONSUMO	1,000	,197
PAF_SINCAPA	1,000	,383
UA_SINRIEGO	1,000	,406

Método de extracción: análisis de componentes principales.

Elaboración propia.

Cómo las pruebas antes mencionadas indican que el análisis factorial es pertinente, se calcula la matriz varianza en el SPSS, que determina la cantidad de factores a retenerse, en este caso los dos primeros factores recogen el 61.1% de la varianza de todas las variables existentes.

Tabla 8

Resultado de la varianza total

Componente	Auto valores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	3,216	45,947	45,947	3,216	45,947	45,947	2,828	40,397	40,397
2	1,061	15,158	61,105	1,061	15,158	61,105	1,450	20,709	61,105
3	,945	13,500	74,605						
4	,743	10,609	85,214						
5	,580	8,288	93,501						
6	,313	4,466	97,968						
7	,142	2,032	100,000						

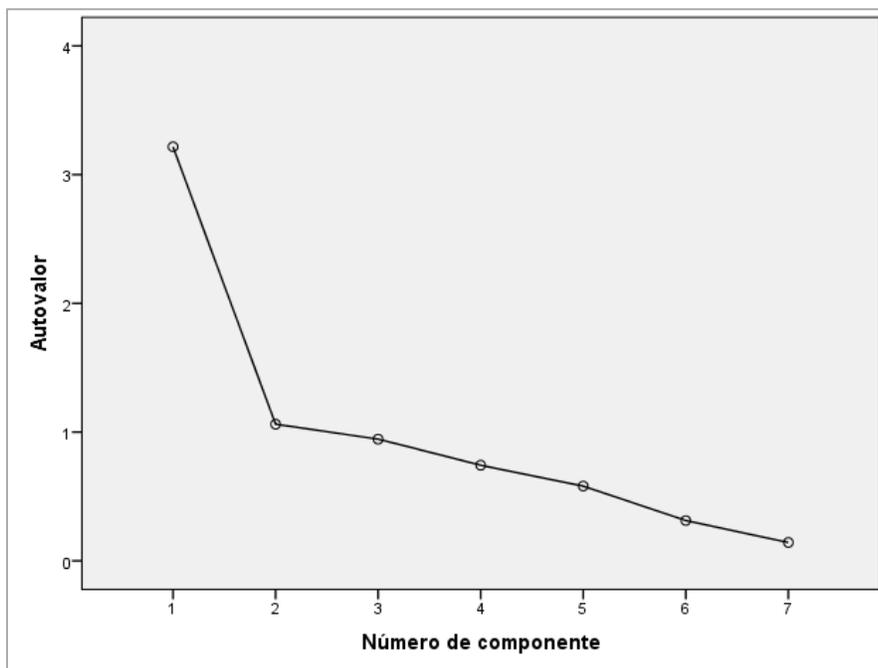
Método de extracción: análisis de componentes principales.

Elaboración propia.

Asimismo, el gráfico de sedimentación, muestra de claramente la ruptura en 2 factores:

Figura 15

Resultado del gráfico de sedimentación



Elaboración propia.

La matriz de componentes, indica las variables asociadas a cada componente, tal como se muestra a continuación:

Tabla 9

Resultado de la matriz de componentes

	Componente	
	1	2
POB_RURAL	,813	-,329
POBREZA	,869	-,201
IVIA	,922	-,191
PPE	,326	,856
AUTOCONSUMO	,425	,131
PAF_SINCAPA	,584	,203
UA_SINRIEGO	,566	,292

Método de extracción: análisis de componentes principales.

a. 2 componentes extraídos.

Elaboración propia.

Para un mejor cálculo, se rotan las variables de cada componente y se puede apreciar que claramente se visualiza los variables que aportan a cada componente, tal como se muestra a continuación:

Tabla 10

Resultado de la matriz de componentes rotado

	Componente	
	1	2
POB_RURAL	,876	,047
POBREZA	,872	,187
IVIA	,916	,218
PPE	-,068	,913
AUTOCONSUMO	,329	,299
PAF_SINCAPA	,443	,432
UA_SINRIEGO	,389	,505

Método de extracción: análisis de componentes principales.

Método de rotación: Varimax con normalización Kaiser.

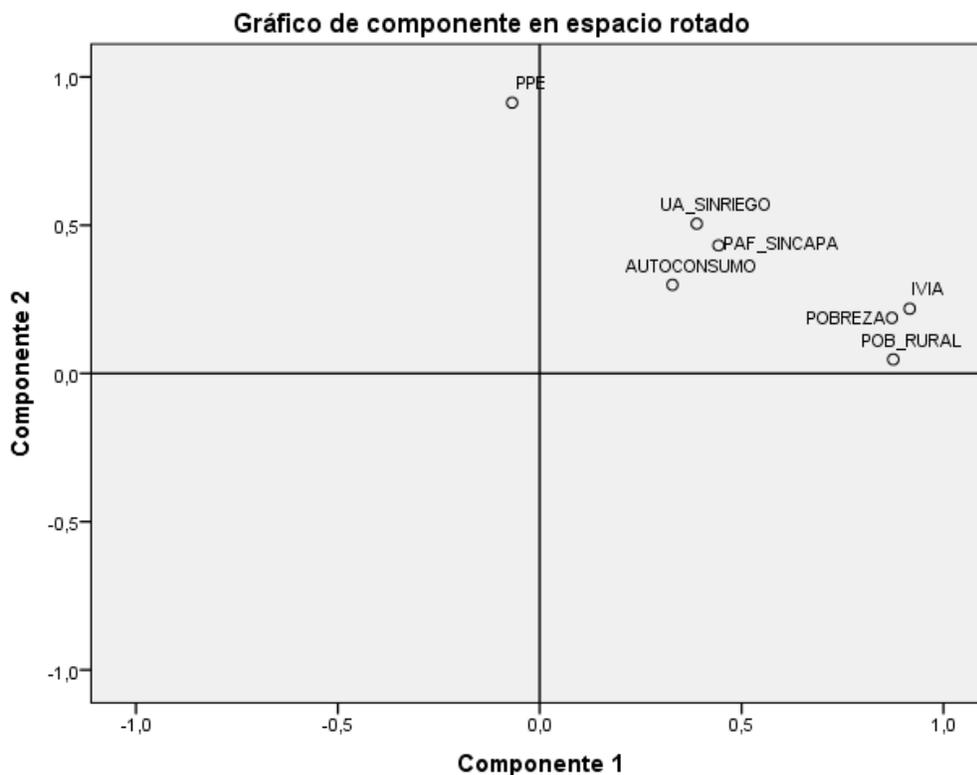
a. La rotación ha convergido en 3 iteraciones.

Elaboración propia.

A continuación, se presenta el gráfico bidimensional para los 2 primeros factores:

Figura 16

Resultado del gráfico de componentes rotados



Elaboración propia.

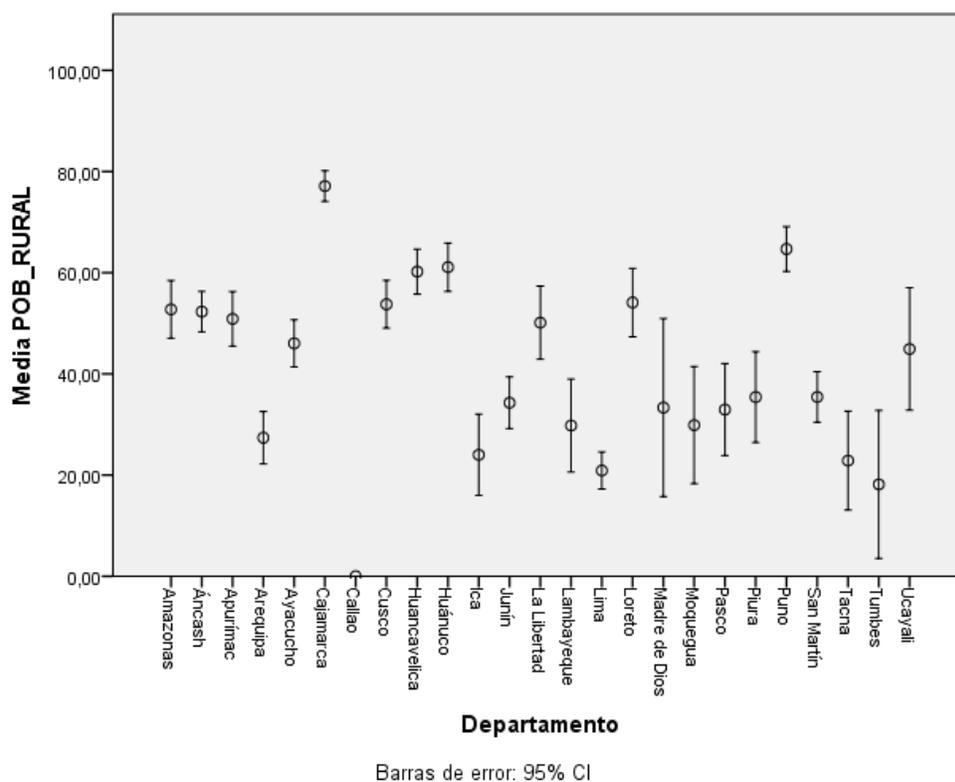
De acuerdo a los resultados obtenidos en la Tabla 10 y la Figura 15, los dos componentes extraídos desde el análisis multivariante, son definidos de la siguiente manera:

- **Componente 1: Vulnerabilidad Social.** La vulnerabilidad social del distrito es determinada por 3 variables del estudio:

- ✓ **Población Rural:** es una medida que determina la proporción de la población que vive en zonas rurales, siendo las regiones de Cajamarca y Puno las que presentan mayor promedio de población en zonas rurales.

Figura 17

Variabilidad de los datos del porcentaje de la población rural

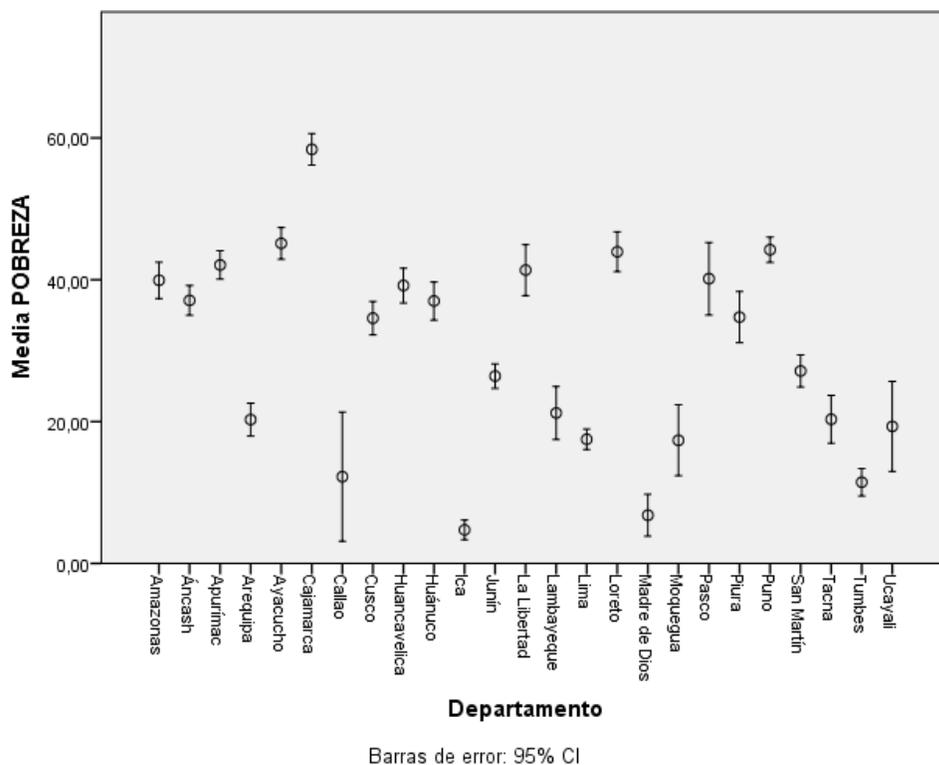


Elaboración propia.

✓ **Pobreza Monetaria:** la pobreza monetaria distrital refleja el porcentaje de la población que se encuentra por debajo de la línea de pobreza.

Figura 18

Variabilidad de los datos del porcentaje de pobreza monetaria

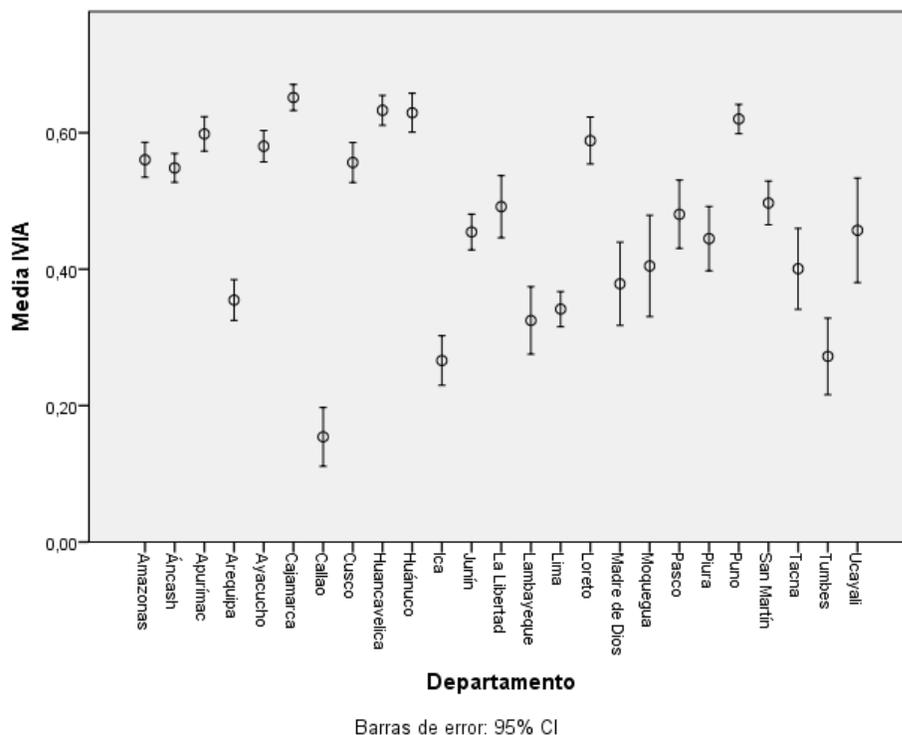


Elaboración propia.

✓ **Vulnerabilidad a la inseguridad alimentaria:** es una medida que refleja la vulnerabilidad de la población a la inseguridad alimentaria, donde los valores cercanos a 0 indica menos vulnerabilidad y valores cercanos a 1 indica mayor vulnerabilidad. En este caso, las regiones Cajamarca, Huancavelica, Huánuco y Apurímac son las más vulnerables.

Figura 19

Variabilidad de los datos del IVIA



Elaboración propia.

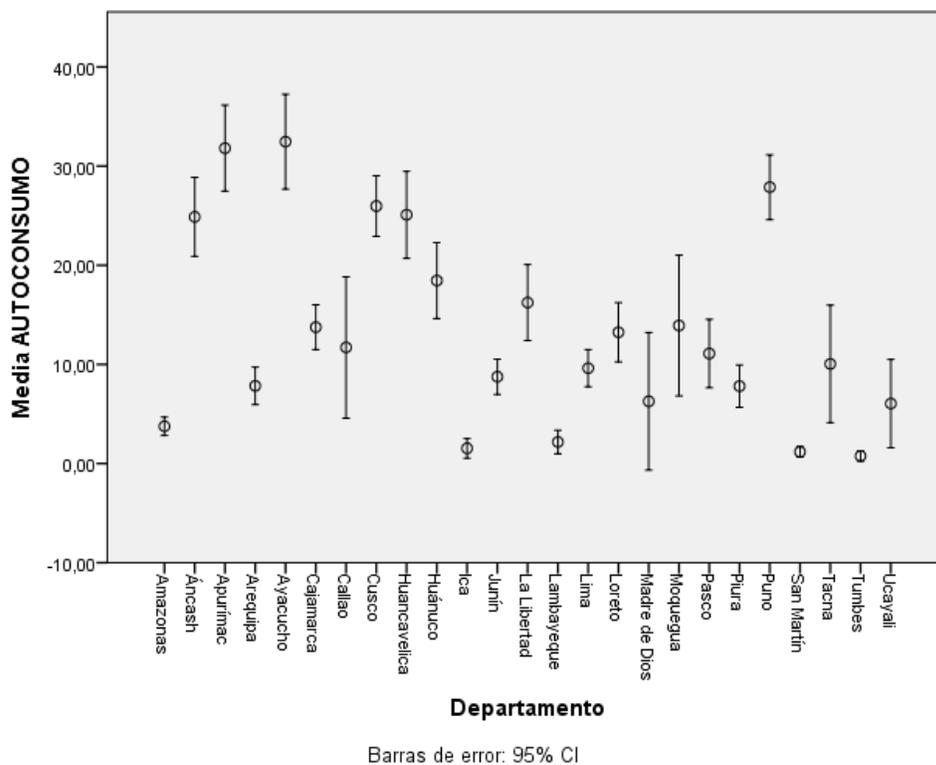
Nota. IVIA es el índice de la vulnerabilidad a la inseguridad alimentaria.

- **Componente 2: Potencialidad de demanda.** La vulnerabilidad social del distrito es determinada por 4 variables del estudio:

- ✓ **Potencial Productivo Efectivo:** es una medida que determina el grado en que el territorio se vincula a sistemas financieros, logísticos, dinámica económica, a través de corredores y rutas óptimas con bastante tránsito relacionando al potencial de sus recursos, como la cantidad de hectáreas, cobertura vegetal y uso de tecnologías de irrigación. El valor 1 muestra que tiene un alto potencial, el valor 2 indica potencial medio, el valor 3 tiene bajo potencial y el 4 indica que el territorio no tiene potencial.

Figura 21

Variabilidad de los datos del porcentaje de autoconsumo

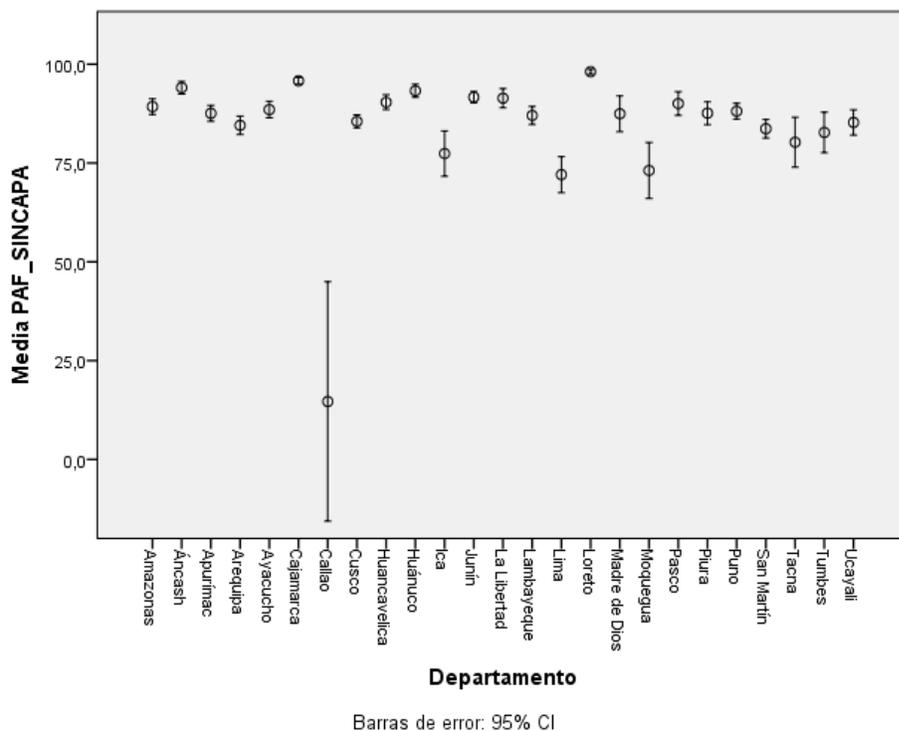


Elaboración propia.

✓ **Productor sin acceso a capacitación:** Este indicador muestra el porcentaje de productores agrarios que no han recibido el servicio de capacitación, presentando un mayor valor en las regiones de Ancash, Loreto y Cajamarca.

Figura 22

Variabilidad de los datos del porcentaje de productores sin capacitación

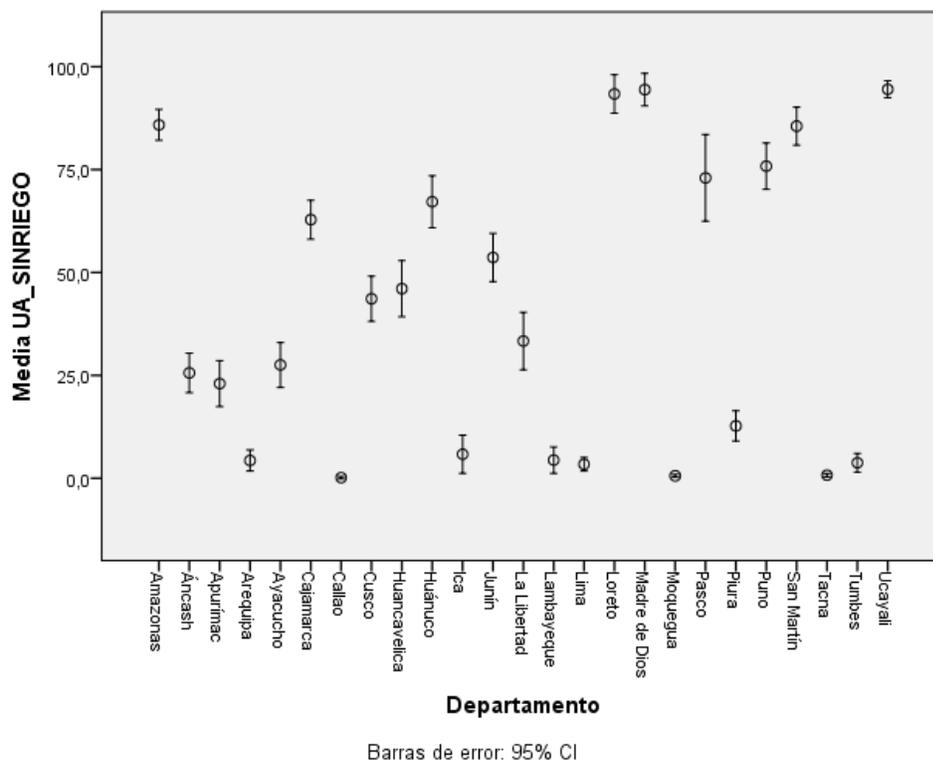


Elaboración propia.

✓ **Unidad agropecuaria sin riego:** Muestra el porcentaje de unidades agropecuarias que no tienen tecnología de riego, siendo Madre de Dios, Moquegua y Ucayali, las regiones que tienen un mayor porcentaje de unidades agropecuarias sin riego tecnificado.

Figura 23

Variabilidad de los datos del porcentaje de UA sin riego



Elaboración propia.

2.2.6.3 Análisis Clúster. En esta etapa, ya se encuentra definido los dos componentes del estudio, por ello, se prosigue con el análisis de conglomerados. A continuación, se procede a analizar el modelo de agrupamiento para cada componente identificado en el análisis factorial

- **Componente 1: Vulnerabilidad Social.** Utilizando el análisis clúster jerárquico, el resultado del historial de conglomeración muestra los outputs que detallan el análisis del clúster, la primera columna se enumeran las etapas del análisis (1891 casos), la segunda y tercera columna (Clúster 1 y Clúster 2) indican los conglomerados unidos en cada etapa y la cuarta columna (Coeficientes) la distancia en la que se produce el agrupamiento, la quinta y sexta columna (Clúster

1 y Clúster 2) indican la etapa en la que el conglomerado aparece por primera vez y finalmente la última columna indica en qué etapa volverá a ser utilizado el agrupamiento.

Tabla 11

Resultado del historial de conglomerados del Componente 1

Etapa	Clúster combinado		Coeficientes	Primera aparición del clúster de etapa		Etapa siguiente
	Clúster 1	Clúster 2		Clúster 1	Clúster 2	
1	1025	1040	,000	0	0	168
2	1436	1440	,000	0	0	875
3	704	787	,000	0	0	1062
4	1632	1636	,000	0	0	121
5	7	475	,001	0	0	969
6	32	747	,001	0	0	606
7	54	1654	,001	0	0	733
8	615	640	,001	0	0	297
9	699	1334	,002	0	0	566
10	373	379	,002	0	0	294
11	1554	1556	,002	0	0	790
12	364	1309	,002	0	0	253
13	1307	1313	,003	0	0	46
14	444	1573	,003	0	0	1041
15	233	898	,003	0	0	972
16	918	1470	,003	0	0	621
17	412	518	,004	0	0	1367
18	1182	1547	,004	0	0	59
19	74	1671	,004	0	0	834
20	710	1549	,005	0	0	182
21	848	1567	,005	0	0	1350
22	1869	1870	,005	0	0	336
23	338	428	,006	0	0	102
24	198	1612	,006	0	0	814
25	1306	1324	,007	0	0	93
26	526	669	,007	0	0	408
27	177	1655	,008	0	0	801
28	589	616	,008	0	0	405
29	1153	1160	,008	0	0	306
30	1103	1628	,009	0	0	360
31	742	924	,009	0	0	983
32	1173	1317	,010	0	0	629
33	1473	1680	,010	0	0	330
34	370	1282	,011	0	0	748
...						

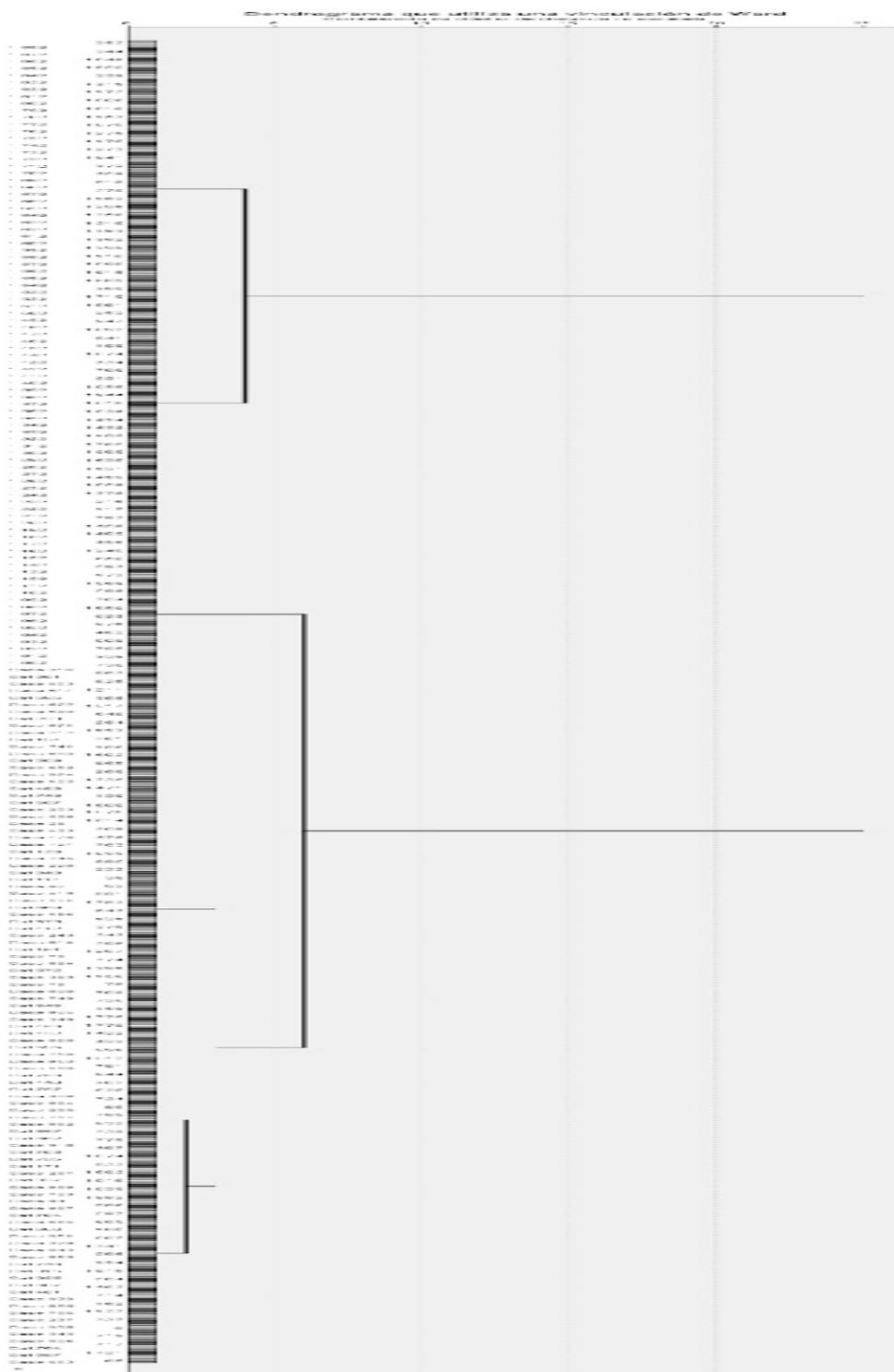
Elaboración propia.

Para determinar el número de conglomerados, se analiza el gráfico del dendograma, que muestra los posibles agrupamientos indicando las distancias entre los distritos, es importante

recordar que el objetivo es agrupar los distritos considerando sus características de vulnerabilidad social; por tanto, tener un solo grupo no aporta información, sabiendo que a menor distancia los conglomerados son más homogéneos, se elige la distancia un poco menor a 5, donde se puede visualizar que los distritos son agrupados en 4 conglomerados.

Figura 24

Resultado del gráfico Dendograma del Componente 1



Elaboración propia.

A continuación, se calcula el valor promedio de las variables por cada conglomerado.

Tabla 12

Resultado del valor promedio de las variables del Componente 1

Ward Method	POB_RURAL	POBREZA	IVIA
1	11,2371	11,9724	0,2275
2	57,6762	35,6129	0,5767
3	16,5766	28,9404	0,4278
4	79,2145	56,4136	0,7046
Total	45,6701	34,1208	0,5090

Elaboración propia.

De acuerdo a los resultados de la tabla 12, los clústeres se denominan de la siguiente forma:

Clúster 1: Muy Baja

La vulnerabilidad social en los distritos del clúster 1 es muy baja, dado que en promedio el 11.97% de la población que viven en estos distritos se encuentran en situación de pobreza, el 11.24% de su población es rural y el índice promedio de la vulnerabilidad a la inseguridad alimentaria es de 0.23.

Clúster 2: Alta

La vulnerabilidad social en los distritos del clúster 2 es alta, dado que en promedio el 35.61% de la población que viven en estos distritos se encuentran en situación de pobreza, el 57.68% de su población es rural y el índice promedio de la vulnerabilidad a la inseguridad alimentaria es de 0.58.

Clúster 3: Baja

La vulnerabilidad social en los distritos del clúster 3 es baja, dado que en promedio el 28.94% de la población que viven en estos distritos se encuentran en situación de pobreza, el 16.58% de su población es rural y el índice promedio de la vulnerabilidad a la inseguridad alimentaria es de 0.42.

Clúster 4: Muy Alta

La vulnerabilidad social en los distritos del clúster 4 es muy alta, dado que en promedio el 56.41% de la población que viven en estos distritos se encuentran en situación de pobreza, el 79.21% de su población es rural y el índice promedio de la vulnerabilidad a la inseguridad alimentaria es de 0.70.

Por otro lado, se presenta a continuación, la cantidad de distritos distribuidos en cada categoría.

Tabla 13

Resultado de los distritos identificados dentro de cada clúster del componente 1

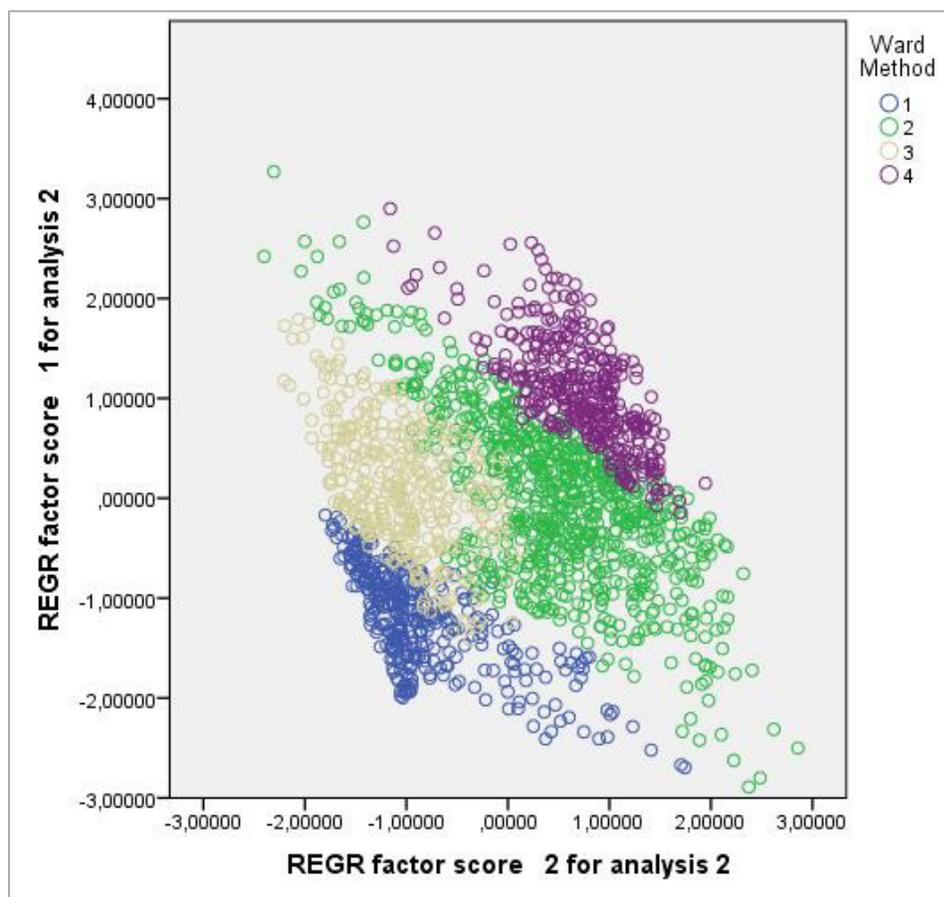
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	359	19,0	19,0	19,0
	2	800	42,3	42,3	61,3
	3	348	18,4	18,4	79,7
	4	384	20,3	20,3	100,0
	Total	1891	100,0	100,0	

Elaboración propia.

El siguiente gráfico, muestra la dispersión y el comportamiento de los distritos en cada clúster.

Figura 25

Resultado del gráfico de los 4 clústeres dentro de Vulnerabilidad Social



Elaboración propia.

- **Componente 2: Potencial de demanda.** Utilizando el análisis clúster jerárquico, el resultado del historial de conglomeración muestra los outputs que detallan el análisis del clúster, la primera columna se enumeran las etapas del análisis (1891 casos), la segunda y tercera columna (Clúster 1 y Clúster 2) indican los conglomerados unidos en cada etapa y la cuarta columna (Coeficientes) la distancia en la que se produce el agrupamiento, la quinta y sexta columna (Clúster 1 y Clúster 2) indican la etapa en la que el conglomerado aparece por primera vez y finalmente la última columna indica en qué etapa volverá a ser utilizado el agrupamiento.

Tabla 14*Resultado del historial de conglomerados del Componente 2*

Etapa	Clúster combinado		Coeficientes	Primera aparición del clúster de etapa		Etapa siguiente
	Clúster 1	Clúster 2		Clúster 1	Clúster 2	
1	1333	1334	,000	0	0	2
2	697	1333	,000	0	1	4
3	1331	1332	,000	0	0	4
4	697	1331	,000	2	3	6
5	1327	1329	,000	0	0	6
6	697	1327	,000	4	5	8
7	1323	1324	,000	0	0	8
8	697	1323	,000	6	7	10
9	1315	1321	,000	0	0	10
10	697	1315	,000	8	9	12
11	1313	1314	,000	0	0	12
12	697	1313	,000	10	11	15
13	1305	1309	,000	0	0	16
14	1306	1308	,000	0	0	15
15	697	1306	,000	12	14	18
16	1298	1305	,000	0	13	1106
17	1297	1300	,000	0	0	18
18	697	1297	,000	15	17	21
19	782	783	,000	0	0	944
20	700	702	,000	0	0	21
21	697	700	,000	18	20	23
22	698	699	,000	0	0	23
23	697	698	,000	21	22	1769
24	497	498	,000	0	0	926
25	82	653	,000	0	0	806
26	1191	1340	,000	0	0	59
27	1190	1262	,000	0	0	133
28	1189	1368	,000	0	0	90
29	1188	1411	,000	0	0	83
30	1255	1357	,000	0	0	39
31	377	1338	,000	0	0	52
32	1423	1832	,000	0	0	114
33	363	1025	,000	0	0	96
34	1766	1778	,000	0	0	511
...						

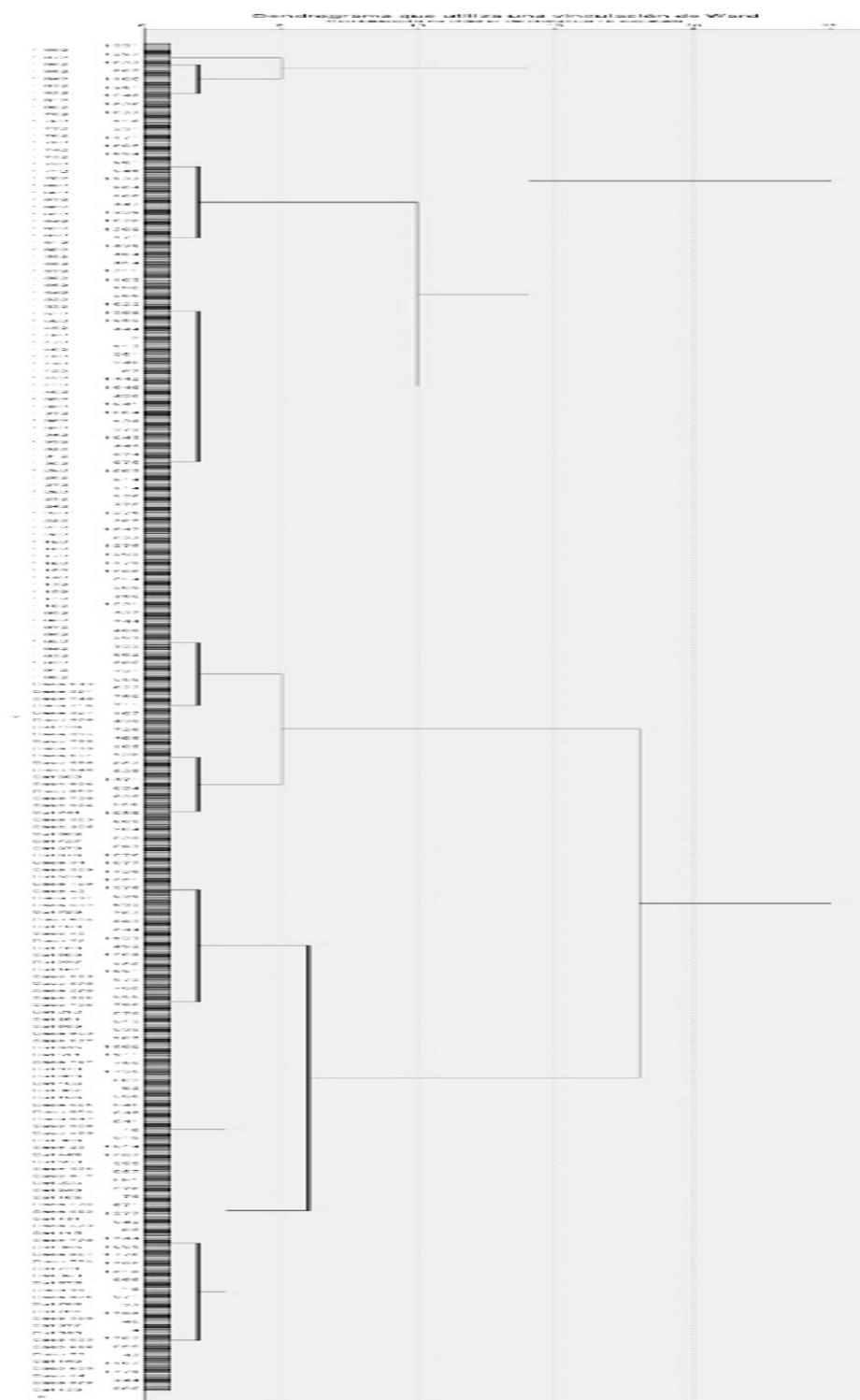
Elaboración propia.

Para determinar el número de conglomerados, se analiza el gráfico del dendograma, que muestra los posibles agrupamientos indicando las distancias entre los distritos, es importante recordar que el objetivo es agrupar los distritos considerando sus características de potencialidad de demanda; por tanto, tener un solo grupo no aporta información, sabiendo que a menor distancia

los conglomerados son más homogéneos, se elige la distancia igual a 10, donde se puede visualizar que los distritos son agrupados en 5 conglomerados.

Figura 26

Resultado del gráfico Dendograma del Componente 2



Elaboración propia.

A continuación, se calcula el valor promedio de las variables por cada conglomerado.

Tabla 15

Resultado del valor promedio de las variables del Componente 2

Ward Method	PPE	AUTOCONSUMO	PAF_SINCAPA	UA_SINRIEGO
1	2,260	11,4712	90,948	77,345
2	1,236	5,8607	89,393	6,579
3	3,359	9,7163	88,680	11,544
4	2,225	50,4104	91,699	40,589
5	1,288	8,5995	44,244	3,484
Total	2,044	16,2228	87,337	40,700

Elaboración propia.

De acuerdo a los resultados de la tabla 15, los clústeres se denominan de la siguiente forma:

Clúster 1: Baja

La potencialidad de demanda en los distritos del clúster 1 es baja, dado que en promedio el 11.47% de la producción se destina al autoconsumo, el 90.95% de los productores agrarios de la agricultura familiar no ha recibido capacitación, el 77.35% de las unidades agropecuarias se encuentra sin riego y el valor del potencial productivo efectivo de 2.3.

Clúster 2: Muy Alta

La potencialidad de demanda en los distritos del clúster 2 es muy alta, dado que en promedio el 5.86% de la producción se destina al autoconsumo, el 89.39% de los productores agrarios de la agricultura familiar no ha recibido capacitación, el 6.58% de las unidades agropecuarias se encuentra sin riego y el valor del potencial productivo efectivo de 1.2.

Clúster 3: Medio

La potencialidad de demanda en los distritos del clúster 3 es medio, dado que en promedio el 9.72% de la producción se destina al autoconsumo, el 88.68% de los productores agrarios de la agricultura familiar no ha recibido capacitación, el 11.54% de las unidades agropecuarias se encuentra sin riego y el valor del potencial productivo efectivo de 3.4.

Clúster 4: Muy Baja

La potencialidad de demanda en los distritos del clúster 4 es muy baja, dado que en promedio el 50.41% de la producción se destina al autoconsumo, el 91.70% de los productores agrarios de la agricultura familiar no ha recibido capacitación, el 40.59% de las unidades agropecuarias se encuentra sin riego y el valor del potencial productivo efectivo de 2.2.

Clúster 5: Alta

La potencialidad de demanda en los distritos del clúster 5 es alta, dado que en promedio el 8.60% de la producción se destina al autoconsumo, el 44.24% de los productores agrarios de la agricultura familiar no ha recibido capacitación, el 3.49% de las unidades agropecuarias se encuentra sin riego y el valor del potencial productivo efectivo de 1.3.

Por otro lado, se presenta a continuación, la cantidad de distritos distribuidos en cada categoría.

Tabla 16

Resultado de los distritos identificados dentro de cada clúster del componente 2

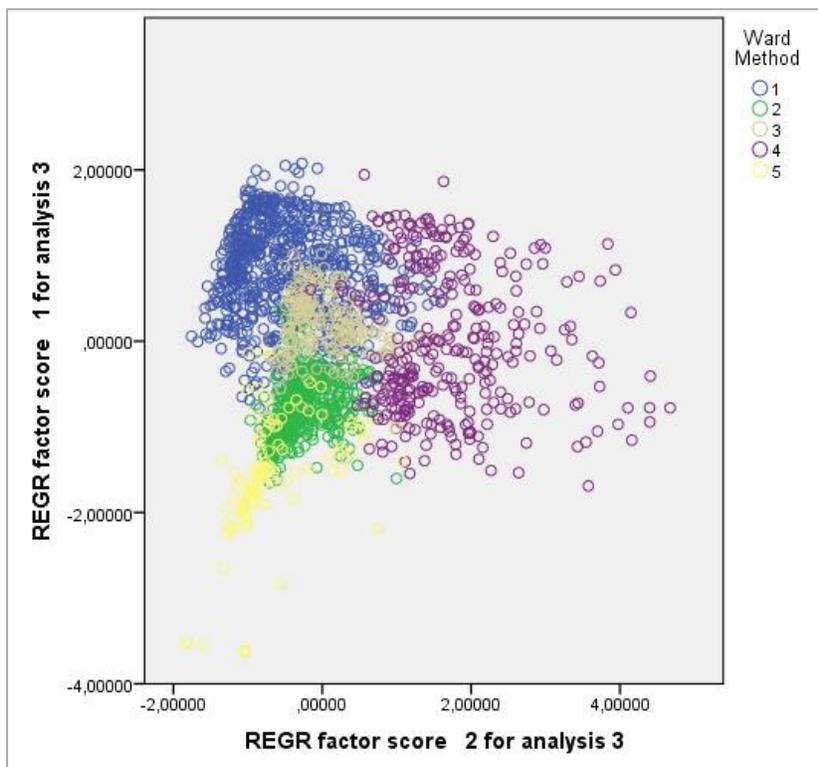
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido 1	749	39,6	39,6	39,6
2	491	26,0	26,0	65,6
3	206	10,9	10,9	76,5
4	320	16,9	16,9	93,4
5	125	6,6	6,6	100,0
Total	1891	100,0	100,0	

Elaboración propia.

El siguiente gráfico, muestra la dispersión y el comportamiento de los distritos en cada clúster.

Figura 27

Resultado del gráfico de los 5 clústeres dentro de la Potencialidad de Demanda



Elaboración propia.

2.2.6.4 Selección de distritos. Luego del análisis clúster en los dos componentes, la distribución de los distritos es la siguiente:

Tabla 17

Resultado del agrupamiento de distritos en cada conglomerado y componente

Vulnerabilidad Social	Potencialidad de demanda				
	Muy Baja	Baja	Medio	Alta	Muy Alta
Muy Baja	4	58	77	76	144
Baja	60	122	30	20	116
Alta	177	346	73	25	179
Muy Alta	79	223	26	4	52

Elaboración propia.

Para el proyecto se decide focalizar aquellos distritos que se encuentran con una potencialidad de demanda “Medio”, “Alta” y “Muy alta” y con una vulnerabilidad social “Alta” y “Muy alta”.

De la tabla 17, se identifica 359 distritos, con las siguientes características:

- Selección 1: 73 distritos tienen potencialidad de demanda medio y vulnerabilidad social alta.
- Selección 2: 25 distritos tienen potencialidad de demanda alta y vulnerabilidad social alta.
- Selección 3: 179 distritos tienen una potencialidad de demanda muy alta y vulnerabilidad social alta.
- Selección 4: 26 distritos tienen potencialidad de demanda medio y vulnerabilidad social muy alta.
- Selección 5: 4 distritos tienen potencialidad de demanda alta y vulnerabilidad social muy alta.
- Selección 6: 52 distritos tienen potencialidad de demanda muy alta y vulnerabilidad social muy alta.

La distribución de los distritos a nivel nacional, es el siguiente:

Tabla 18

Distribución de los distritos focalizados

Departamento	Distritos seleccionados	Porcentaje
Total	359	100%

Amazonas	3	1%
Áncash	67	19%
Apurímac	25	7%
Arequipa	35	10%
Ayacucho	29	8%
Cajamarca	16	4%
Cusco	16	4%
Huancavelica	28	8%
Huánuco	11	3%
Ica	3	1%
Junín	20	6%
La Libertad	12	3%
Lambayeque	13	4%
Lima	38	11%
Loreto	2	1%
Moquegua	6	2%
Piura	17	5%
Puno	10	3%
San Martín	2	1%
Tacna	5	1%
Tumbes	1	0%

Elaboración propia.

III. APORTES MÁS DESTACABLES A LA INSTITUCIÓN

- Se han generado sintaxis en el software SPSS que permite replicar el análisis de los datos con el análisis de conglomerados.
- Aporta el análisis técnico para formar parte del diagnóstico del proyecto de inversión en el marco del Invierte.pe, insertando un anexo metodológico en el documento.
- El análisis de conglomerados permite focalizar los distritos con sustento estadístico
- Con la focalización identificada, se puede proceder a la priorización del ámbito de influencia del proyecto de inversión, con los criterios establecidos por la unidad formuladora.

IV. CONCLUSIONES

- 4.1 El proyecto de inversión denominado MERCAGRO tiene como objetivo mejorar el acceso de los productores agropecuarios de la agricultura familiar a los servicios de apoyo al desarrollo productivo en el acceso a mercados de forma rentable, sostenible e inclusiva
- 4.2 Para la formulación de un proyecto de inversión, en el marco del Invierte.pe del Ministerio de Economía y Finanzas, se requiere elaborar en una primera etapa el Diagnóstico, donde se define el territorio que es el ámbito geográfico en el que se ubica la población afectada, en este caso, los productores agrarios de la agricultura familiar de subsistencia e intermedia I.
- 4.3 Como resultado del análisis multivariante a los 1 891 distritos a nivel nacional, se logró identificar dos componentes principales, el primer componente denominado “Vulnerabilidad Social”, clasificado en 4 categorías: Muy alta, Alta, Baja y Muy baja; y, el componente denominado “Potencialidad de demanda”, clasificado en 5 categorías: Muy alta, Alta, Medio, Baja y Muy baja.
- 4.4 Para el caso del primer componente, se ha identificado que el 42% de los distritos se encuentra en vulnerabilidad social alta, el 20% de los distritos se encuentran en muy alta vulnerabilidad social, el 19% se encuentra en vulnerabilidad muy baja y el 18% de los distritos se encuentran en vulnerabilidad social baja.
- 4.5 Para el caso del segundo componente, se ha identificado que el 40% de los distritos se encuentran con potencialidad de demanda baja, el 26% de los distritos se encuentran con potencialidad de demanda muy alta, el 17% tienen potencialidad de demanda muy baja, el

11% tienen potencialidad de demanda en un nivel medio y el 7% de los distritos se encuentran con potencialidad de demanda alta.

- 4.6 Teniendo en cuenta lo anteriormente señalado, la selección de los distritos se determina con base a los resultados de las clasificaciones en cada componente, por ello, se establece que aquellos distritos que cumplan con encontrarse con vulnerabilidad social alta y muy alta y con potencialidad de demanda de nivel medio, alta y muy alta.
- 4.7 Por tanto, se concluye que, luego del análisis multivariante realizado a los 1 891 distritos de las 25 regiones, tomando en cuenta las variables de pobreza, población rural, índice de la vulnerabilidad a la inseguridad alimentaria, producción destinada al autoconsumo, productores agrarios sin capacitación y unidades agropecuarias sin riego, se focaliza 359 distritos distribuidos en 21 regiones.

V. RECOMENDACIONES

- 5.1 Se recomienda utilizar el presente análisis estadístico para sustentar la focalización del proyecto de inversión, para que se pueda establecer a partir de este estudio, la priorización de atención.
- 5.2 Utilizar métodos estadísticos para determinar la intervención de los proyectos o programas del Sector Público, en beneficio de la población vulnerable.
- 5.3 El presente análisis, permite dimensionar el monto presupuestal del proyecto de inversión, en función a la existencia de los corredores económicos y cadenas de productos en los distritos focalizados.

VI. REFERENCIAS

- Castaño, E. (2006). *Introducción Al Análisis De Datos Multivariados En Ciencias Sociales*.
Facultad de Ciencias, Universidad Nacional de Colombia, Medellín.
https://www.inec.gov.pa/IASI/docs/announcements/documentos/MemoriasCursillos/4%20Casta%C3%B1o_An%C3%A1lisis%20de%20datos%20multivariados.pdf
- Cuadras, C. (2014). *Nuevos métodos de análisis multivariante*. CMC Editions
https://gc.scalahed.com/recursos/files/r161r/w24899w/Semana5/METODOS_S5.pdf
- Cochran, W. (1998). *Técnicas de Muestreo. (13a reimpresión)*. México: CECSA.
- De la Fuente, S. (2011). *Análisis Conglomerados*. Universidad Autónoma de Madrid
http://www.estadistica.net/Master-Econometria/Análisis_Cluster.pdf
- Díaz, L. y Morales, M. (2012). *Análisis Estadístico de Datos Multivariados*. Bogota-Colombia.
- Instituto Nacional de Estadística e Informática (Mayo de 2024). *Evolución de la Pobreza Monetaria, 2014-2023*.
<https://www.gob.pe/institucion/inei/informes-publicaciones/5558423-peru-evolucion-de-la-pobreza-monetaria-2014-2023>
- Ministerio de Desarrollo Agrario y Riego (2021). *Política Nacional Agraria*.
<https://cdn.www.gob.pe/uploads/document/file/2071814/DECRETO%20SUPREMO%2017-2021-MIDAGRI.pdf>
- Ministerio de Desarrollo e Inclusión Social (2018). *Vulnerabilidad a la inseguridad alimentaria por departamento, provincia y distrito*.
<https://evidencia.midis.gob.pe/infor-final-ivia-2018/>
- Ministerio de Economía y Finanzas (2019). *Contenido mínimo del Estudio de Preinversión a nivel perfil para Proyectos de Inversión*.

https://www.mef.gob.pe/contenidos/inv_publica/anexos/anexo7_directiva001_2019EF63_01.pdf

Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill.

Perez C. (2004). *Técnicas del Análisis Multivariante de Datos. Aplicaciones con SPSS*.

Madrid España. Editorial Pearson Education S.A.

Programa de Desarrollo Productivo Agrario Rural (2021). *Manual de Operaciones*.

<https://www.gob.pe/institucion/midagri/normas-legales/1931145-0137-2021-midagri>

Ramirez, R., Guevara, F (2017). *Análisis Multivariante: Teoría y práctica de las principales técnicas*. Ediciones Holguín S.A.

VII. ANEXOS

ANEXO A. Sintaxis para el análisis de componentes y clúster

Para hallar los componentes

FACTOR

/VARIABLES POB_RURAL POBREZA IVIA

PPE AUTOCONSUMO PAF_SINCAPA

UA_SINRIEGO

/MISSING LISTWISE

/ANALYSIS POB_RURAL POBREZA IVIA

PPE AUTOCONSUMO PAF_SINCAPA

UA_SINRIEGO

/PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO EXTRACTION

ROTATION FSCORE

/PLOT EIGEN ROTATION

/CRITERIA MINEIGEN(1) ITERATE(25)

/EXTRACTION PC

/CRITERIA ITERATE(25)

/ROTATION VARIMAX

/SAVE REG(ALL)

/METHOD=CORRELATION.

*****inicio de clúster jerárquico en COMPONENTE 1: rural, pobreza, seguridad
alimentaria*****

DATASET DECLARE D0.3034726667610079.

PROXIMITIES POB_RURAL POBREZA IVIA

/MATRIX OUT(D0.3034726667610079)

/VIEW=CASE

/MEASURE=SEUCLID

/PRINT NONE

/STANDARDIZE=VARIABLE Z.

CLUSTER

/MATRIX IN(D0.3034726667610079)

/METHOD WARD

/PRINT SCHEDULE

/PLOT DENDROGRAM VICICLE.

Dataset Close D0.3034726667610079.

se decide 4 clústeres

DATASET DECLARE D0.34982090830188894.

PROXIMITIES POB_RURAL POBREZA IVIA

/MATRIX OUT(D0.34982090830188894)

/VIEW=CASE

/MEASURE=SEUCLID

/PRINT NONE

/STANDARDIZE=VARIABLE Z.

CLUSTER

/MATRIX IN(D0.34982090830188894)

/METHOD WARD

/PRINT CLUSTER(4)

/PLOT VICICLE

/SAVE CLUSTER(4).

Dataset Close D0.34982090830188894.

compara medias

MEANS TABLES=POB_RURAL POBREZA IVIA BY CLU4_1

/CELLS=MEAN.

*****extraer 2 factores para el gráfico

FACTOR

/VARIABLES POB_RURAL POBREZA IVIA

/MISSING LISTWISE

/ANALYSIS POB_RURAL POBREZA IVIA

/PRINT UNIVARIATE INITIAL KMO EXTRACTION ROTATION

/PLOT EIGEN ROTATION

/CRITERIA FACTORS(2) ITERATE(25)

```

/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.

```

*****grafico de factores y metodo ward

* Generador de gráficos.

GGRAPH

```

/GRAPHDATASET NAME="graphdataset" VARIABLES=FAC2_2 FAC1_2 CLU4_1
MISSING=LISTWISE REPORTMISSING=NO
/GRAPHSPEC SOURCE=INLINE.

```

BEGIN GPL

```

SOURCE: s=userSource(id("graphdataset"))
DATA: FAC2_2=col(source(s), name("FAC2_2"))
DATA: FAC1_2=col(source(s), name("FAC1_2"))
DATA: CLU4_1=col(source(s), name("CLU4_1"), unit.category())
GUIDE: axis(dim(1), label("REGR factor score 2 for analysis 2"))
GUIDE: axis(dim(2), label("REGR factor score 1 for analysis 2"))
GUIDE: legend(aesthetic(aesthetic.color.exterior), label("Ward Method
" "))
ELEMENT: point(position(FAC2_2*FAC1_2), color.exterior(CLU4_1))
END GPL.

```

*****inicio de cluster jerarquico en COMPONENTE 2 ppe, sin riego, sin capacitacion,
autoconsumo**

DATASET DECLARE D0.8581347483611834.

PROXIMITIES PPE AUTOCONSUMO PAF_SINCAPA

UA_SINRIEGO

/MATRIX OUT(D0.8581347483611834)

/VIEW=CASE

/MEASURE=SEUCLID

/PRINT NONE

/STANDARDIZE=VARIABLE Z.

CLUSTER

/MATRIX IN(D0.8581347483611834)

/METHOD WARD

/PRINT SCHEDULE

/PLOT DENDROGRAM VICICLE.

Dataset Close D0.8581347483611834.

se decide 5 cluster**

DATASET DECLARE D0.05107437426101691.

PROXIMITIES PPE AUTOCONSUMO PAF_SINCAPA

UA_SINRIEGO

/MATRIX OUT(D0.05107437426101691)

```
/VIEW=CASE  
/MEASURE=SEUCLID  
/PRINT NONE  
/STANDARDIZE=VARIABLE Z.  
CLUSTER  
/MATRIX IN(D0.05107437426101691)  
/METHOD WARD  
/PRINT SCHEDULE CLUSTER(5)  
/PLOT DENDROGRAM VICICLE  
/SAVE CLUSTER(5).
```

Dataset Close D0.05107437426101691.

compara medias

```
MEANS TABLES=PPE AUTOCONSUMO PAF_SINCAPA  
    UA_SINRIEGO BY CLU5_2  
/CELLS=MEAN.
```

*****se extraen 2 factores para el grafico

```
FACTOR  
/VARIABLES PPE AUTOCONSUMO PAF_SINCAPA  
    UA_SINRIEGO  
/MISSING LISTWISE
```

```

/ANALYSIS PPE AUTOCONSUMO PAF_SINCAPA
  UA_SINRIEGO
/PRINT UNIVARIATE INITIAL KMO EXTRACTION ROTATION
/PLOT EIGEN ROTATION
/CRITERIA FACTORS(2) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.

```

*****grafico de factores y metodo ward

* Generador de gráficos.

GGRAPH

```

/GRAPHDATASET  NAME="graphdataset"  VARIABLES=FAC2_3  FAC1_3  CLU5_2
MISSING=LISTWISE REPORTMISSING=NO
/GRAPHSPEC SOURCE=INLINE.

```

BEGIN GPL

```

SOURCE: s=userSource(id("graphdataset"))
DATA: FAC2_3=col(source(s), name("FAC2_3"))
DATA: FAC1_3=col(source(s), name("FAC1_3"))
DATA: CLU5_2=col(source(s), name("CLU5_2"), unit.category())
GUIDE: axis(dim(1), label("REGR factor score 2 for analysis 3"))

```

```

GUIDE: axis(dim(2), label("REGR factor score 1 for analysis 3"))
GUIDE: legend(aesthetic(aesthetic.color.exterior), label("Ward Method
" "))
ELEMENT: point(position(FAC2_3*FAC1_3), color.exterior(CLU5_2))
END GPL.

```

*****distribución de los distritos*****

CTABLES

```

/VLABELS VARIABLES=CLU4_1 CLU5_2 DISPLAY=LABEL
/TABLE CLU4_1 [C][COUNT F40.0] BY CLU5_2 [C]
/CATEGORIES VARIABLES=CLU4_1 CLU5_2 ORDER=A KEY=VALUE
EMPTY=EXCLUDE.

```

```

IF ((CLU4_1 = 2 & CLU5_2 = 3) | (CLU4_1 = 2 & CLU5_2 = 5) | (CLU4_1 = 2 & CLU5_2 = 2)
| (CLU4_1 =
4 & CLU5_2 = 3) | (CLU4_1 = 4 & CLU5_2 = 5) | (CLU4_1 = 4 & CLU5_2 = 2))
focalizacion=1.

```

EXECUTE.

```

RECODE focalizacion (SYSMIS=0).

```

EXECUTE.

* Tablas personalizadas.

CTABLES

/VLABELS VARIABLES=Departamento focalizacion DISPLAY=LABEL

/TABLE Departamento [COUNT F40.0] BY focalizacion

/CATEGORIES VARIABLES=Departamento focalizacion ORDER=A KEY=VALUE

EMPTY=EXCLUDE.