



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA



**SEGMENTACIÓN DE ALPACAS HUACAYA DE LA CALIDAD DE
FIBRA EN EL DISTRITO DE CUYOCUYO - SANDIA, PUNO - 2016**

TESIS

PRESENTADA POR:

Bach. OLINDA YESENIA PACHA PHOCCO

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2018



DEDICATORIA

Dedico este trabajo principalmente a Dios quién supo guiarme por el buen camino, darme fuerzas para seguir adelante y no desmayar en los problemas que se presentaban, enseñándome a encarar las adversidades sin perder nunca la dignidad ni desfallecer en el intento. A mis queridos padres Marina y Guillermo que, con su apoyo incondicional, amor y confianza permitieron que logre culminar mi carrera profesional.

*A Senaida y Kenedin, mis hermanos queridos
Gracias por darme esas fuerzas para cumplir
con cada uno de mis objetivos.*

A mis amigas Maju y Juana, que forman parte de mi familia y son como mis hermanas gracias por brindarme su amistad, comprensión y siempre motivándome a continuar con mi formación profesional.

Olinda Yesenia



AGRADECIMIENTOS

- ❖ *A Dios por haberme dado la vida, y permitido llegar hasta este punto y haberme dado salud para lograr mis objetivos, además de su infinita bondad y amor, me ha ayudado a enfrentar todas las dificultades que se han presentado en el trayecto de la vida.*
- ❖ *A mis padres por ser el pilar fundamental y haberme apoyado, pese a las adversidades que se presentaron.*
- ❖ *A la Universidad Nacional del Altiplano-Puno, a los Docentes de La Facultad de Ingeniería Estadística e Informática por haber compartido sus valiosos conocimientos hicieron que pueda crecer día a día como profesional.*
- ❖ *De manera especial a mi director de tesis Dr. Edgar Eloy Carpio Vargas, quien ha guiado con su paciencia, experiencia, conocimiento y motivación para el logro de esta investigación.*
- ❖ *A mis jurados: D.Sc. Alejandro Apaza Tarqui, M.Sc. Samuel Donato Pérez Quispe, M.Sc. Leonid Alemán Gonzales que aportaron a través de sus observaciones para la conclusión de esta investigación.*

Olinda Yesenia



ÍNDICE GENERAL

DEDICATORIA

AGRADECIMIENTOS

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

ÍNDICE DE ACRÓNIMOS

RESUMEN 10

ABSTRACT..... 11

CAPÍTULO I

INTRODUCCIÓN

1.1 FORMULACIÓN DEL PROBLEMA 13

1.1.3 Problema General 13

1.2 OBJETIVOS DE LA INVESTIGACIÓN..... 14

1.2.1 Objetivo General 14

1.2.2 Objetivos Específicos 14

1.3 HIPÓTESIS DE LA INVESTIGACIÓN..... 14

1.3.1 Hipótesis General 14

1.3.2 Hipótesis Específicos 14

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES..... 15

2.1.1 A Nivel Internacional 15

2.1.2 A Nivel Nacional..... 16

2.1.3 A Nivel Local 18

2.2 MARCO TEÓRICO 20



2.2.1	Análisis Exploratorio de Datos.....	20
2.2.2	Diagrama de Cajas (Box Plot).....	21
2.2.3	Análisis de Componentes Principales	23
2.2.4	Análisis Clúster	25
2.2.5	Análisis Clúster Jerárquico.....	26
2.2.6	Método de Distancia Euclídea Al Cuadrado	29
2.2.7	Análisis Clúster a nivel de Individuos.....	29
2.2.8	Dendograma	30
2.2.9	Análisis Clúster K-Means	32
2.3	MARCO CONCEPTUAL.....	37
2.3.1	Importancia Socioeconómica de la Fibra de la Alpaca	37
2.3.2	Camélidos Sudamericanos	38
2.3.3	Acopio de la Fibra de Alpaca	38
2.3.4	Identificación de Alpaca.....	39
2.3.5	Alpaca.....	39
2.3.6	Raza Huacaya (Alpaca)	41
2.3.7	Características Tecnológicas de la Fibra de Alpaca.....	41
2.3.8	Método del Ofda.....	47
CAPÍTULO III		
MATERIALES Y MÉTODOS		
3.1	MÉTODO DE INVESTIGACIÓN.....	48
3.2	TÉCNICAS E INSTRUMENTOS PARA LA INVESTIGACIÓN	48
3.3	POBLACIÓN Y MUESTRA DEL ESTUDIO	48
3.4	IDENTIFICACIÓN DE VARIABLES:.....	49
3.5	METODOLOGÍA DE LA INVESTIGACIÓN.....	49
3.5.1	Análisis Descriptivo y Exploratorio	49
3.5.2	Prueba de la Normalidad	49



3.5.3 Obtención del Clúster Jerárquico y No Jerárquico.....	50
3.5.4 Comparación de Medias y Prueba de Varianza.....	51
3.6 CARACTERIZACIÓN DEL AREA DE INVESTIGACIÓN	52
3.6.1 Ámbito De Estudio	52
CAPÍTULO IV	
RESULTADOS Y DISCUSIÓN	
4.1 RESULTADOS OBTENIDOS ANÁLISIS EXPLORATORIO.....	53
4.2 DETECCIÓN DE VALORES ATÍPICOS	54
4.3 ESTANDARIZACIÓN DE DATOS.....	54
4.4 CONTRASTACIÓN DE LA HIPÓTESIS GENERAL	54
4.5 SEGMENTACIÓN CON EL ANÁLISIS CLÚSTER JERÁRQUICO .	58
4.6 FORMACIÓN DEL DENDOGRAMA.....	60
4.7 SEGMENTACIÓN MEDIANTE ANÁLISIS CLÚSTER K-MEAN.....	62
4.8 FORMACIÓN DE CLÚSTER CON K-MEANS.....	67
4.9 COMPARACIÓN DE MEDIAS (ACNJ)	68
4.10 VALIDACIÓN DEL ANÁLISIS CLÚSTER NO JERÁRQUICO	69
V. CONCLUSIONES.....	72
VI. RECOMENDACIONES	74
VII. REFERENCIAS BIBLIOGRAFICAS.....	75
ANEXOS.....	79

Área: Estadística

Tema: Análisis Multivariado

Fecha de sustentación: 07/09/2018



ÍNDICE DE FIGURAS

Figura 1. Diagrama de Cajas	21
Figura 2. Método Jerárquico Aglomerativo.....	29
Figura 3. Representación de un Dendograma.....	31
Figura 4. Dendograma a Nivel de Individuos Similitud por Tamaño	32
Figura 5. Algoritmo del Diagrama del K-medias	34
Figura 6. Convergencia Adecuada del Algoritmo Centros Clústeres.....	36
Figura 7. Convergencia adecuada del algoritmo de 7 clústeres.....	36
Figura 8. Convergencia Inadecuada del Algoritmo K- media.....	37
Figura 9. Población Mundial de Alpacas Dirección General de Negocios Agrarios	40
Figura 10. Población Mundial 2013 de la Dirección General de Negocios Agrarios ...	40
Figura 11. Interacción Entre el Tejido, los Terminales de la Fibra y la Piel.....	44
Figura 12. Representación Bidimensional de la Forma de Una Fibra de Lana	46
Figura 13. Equipo del Análisis de las Características de la Fibra (OFDA 2000).....	47
Figura 14. Esquema General del Análisis Clúster	51
Figura 15. Mapa de Localización	52
Figura 16. Histograma, Diagrama de Caja y Bigote.....	57
Figura 17. Gráfico del Comportamiento de los Individuos (Alpacas).....	58
Figura 18. Dendograma Formado en 3 grupos en función de la Calidad de Fibra.....	61
Figura 19. Salida de Individuos en Cada Etapa.....	67
Figura 20. Diagrama de Dispersión para Cada Caso de Clúster.....	71



ÍNDICE DE TABLAS

Tabla 1 Identificación de Variables	49
Tabla 2 Estadística Descriptiva de los Datos	53
Tabla 3 Prueba de Normalidad	55
Tabla 4 Historial de Conglomeración- Promedio Entre Grupos.....	59
Tabla 5 Centros de Clusters Iniciales.....	63
Tabla 6 Centros de Clusters Finales.....	64
Tabla 7 Distancia Entre Los Centro de los Clusters Finales	64
Tabla 8 Cuadro de la Varianza del Análisis Cluster NO Jerárquico.....	66
Tabla 9 Número de Casos en Cada Cluster	66
Tabla 10 Comparación de Medias con el Método de Enlace Promedios	68
Tabla 11 Análisis de la Varianza de los 3 Clusters.....	70



ÍNDICE DE ACRÓNIMOS

AC	Análisis Clúster
ACJ	Análisis Clúster Jerárquico
ACNJ	Análisis Clúster No Jerárquico
AKM	Análisis K-means
DEC	Distancia Euclídea al cuadrado
AED	Análisis Exploratorio de Datos.
AF	Análisis Factorial
ACP	Análisis de Componentes Principales
CT	Características Tecnológicas
DF	Diámetro de Fibra
CV	Coefficiente de Variabilidad
IC	Índice de Curvatura
LM	Longitud de Mecha
FG	Fibra gruesa
FC	Factor de Confortabilidad
OFDA (ADFO)	Optical Fibre Diameter Analyser (Analizador de Diámetro de Fibra Óptica)
GRP	Gobierno Regional Puno
PECSA	Proyecto Especial Camélidos Sudamericanos
MINAGRI	Ministerio de Agricultura y Riego



RESUMEN

El presente estudio de investigación se ejecutó en la comunidad de Puna Ayllu del distrito de Cuyocuyo, Provincia Sandia del Departamento de Puno, cuyo objetivo es Segmentar las alpacas Huacaya de la calidad de fibra (características físicas de lana). El diseño de investigación es no experimental y los datos se analizaron con el análisis clúster jerárquico y no jerárquico, esta metodología permitirá diseñar distintas estrategias de promocionar similitudes y disimilitudes entre individuos, Los criterios utilizados para la segmentación de los 208 individuos fueron con el método de enlace de promedios y la medida de distancia Euclídea segmentados en tres grupos, El **C1**: conformado por 53 alpacas Huacaya los individuos, EL DIÁMETRO DE FIBRA hace un mayor aporte con una media mayor de 25,6551 μm , al del C2 y C3 con una desviación de 1,37719 y la FIBRA GRUESA con una media de 10,7377 μm , El **C2** conformado por 88 alpacas Huacaya los individuos que más se agrupan según las características en común que tienen es el INDICE DE CURVATURA hace un mayor aporte en un promedio de 48,5848, y **EL C3** conformado por 67 alpacas Huacaya el COEFICIENTE DE VARIABILIDAD que hace un mayor aporte en un promedio de 24,8857 % con una desviación de 2,57157.

Palabras Clave: Análisis Clúster, Distancia Euclídea, Dendograma, Características físicas de lana.



ABSTRACT

The present research study was carried out in the Puna Ayllu community of the Cuyocuyo district, Sandia Province of the Department of Puno, whose objective is to segment the Huacaya alpacas of the fiber quality (physical characteristics of wool). The research design is non-experimental and the data were analyzed with hierarchical and non-hierarchical cluster analysis, this methodology will allow the design of different strategies to promote similarities and dissimilarities between individuals. The criteria used for the segmentation of the 208 individuals were with the method of Averages link and the Euclidean distance measure segmented into three groups, C1: made up of 53 Huacaya alpacas individuals, THE FIBER DIAMETER makes a greater contribution with a mean greater than 25.6551 μm , than that of C2 and C3 with a deviation of 1.37719 and the THICK FIBER with an average of 10.7377 μm , The C2 made up of 88 Huacaya alpacas the individuals that are most grouped according to the characteristics in common they have is the CURVATURE INDEX makes a greater contribution on an average of 48.5848, and EL C3 made up of 67 Huacaya alpacas, the COEFFICIENT OF VARIABILITY that makes a greater contribution by an average of 24.8857% with a deviation of 2.57157.

Key Words: Cluster Analysis, Euclidean distance, Dendrogram, wool physical characteristics.



CAPÍTULO I

INTRODUCCIÓN

La población conformada de alpacas en el Perú es de 3 millones 592 mil 249 y el 89.7% se encuentra principalmente en las zonas alto andinas: Puno, Cusco, Arequipa, Huancavelica y Apurímac (INEI, 2012).

El Perú es el principal productor de fibra de alpaca en el mundo pues alberga el 87% de la población mundial (CENAGRO, 2012) La misma que se comercializa a través de rescatistas que representa el 70%, productores de hilados artesanales 10%, agentes comerciales 17% y el resto es destinado al autoconsumo del productor (Ministerio de Agricultura, 2003)

Una de las características más importantes del vellón de la alpaca es el diámetro de fibra que constituye el criterio principal para determinar el precio (Canaza, 2009), con el aprovechamiento racional de los pastizales alto andinos para la producción de fibra como ventaja productiva y el uso de la mejora genética de los camélidos es uno de los retos que tiene el país como el medio más efectivo de lucha contra la pobreza, que afecta a numerosas familias que viven de la crianza de esta especie ganadera.

La alpaca es considerada la fuente de recurso principal para el poblador alto andino, cuya crianza constituye el principal sustento socioeconómico; debido a que de esta actividad se obtiene la producción de la fibra la cual se destina el 90% al mercado exterior; no obstante que su proceso de producción es incipiente en la innovación tecnológica, por no planificar las actividades en relación al objetivo que es la producción de la fibra fina, actualmente se observa engrosamiento del diámetro de fibra al no practicar la selección por finura y no implementar registros productivos, ni mucho menos en determinar diámetro de fibra, peso vellón, longitud de mecha, peso al nacimiento, etc., por esta razón el productor no tiene márgenes que le permitan las condiciones mínimas



de vida convirtiéndose cada vez más dependiente del mercado especialmente cuando los precios de la fibra se encuentran en sus niveles más bajos (Díaz , 2014).

En la actualidad en el Distrito Cuyocuyo, comunidad Puna Ayllu, las características tecnológicas de la fibra de alpaca Huacaya no se conocen respecto a finura al hilado, Diámetro de Fibra, Longitud de Mecha, Índice de Curvatura, Fibra Gruesa y Factor de Confort. Estas variables son importantes desde el punto de vista de la industria textil. Con los resultados del presente estudio de investigación se aportará sobre mantener la calidad de fibra con las con el método del Análisis clúster.

El desarrollo de la presente investigación empezó con la formulación del problema y bajo el siguiente objetivo Segmentar las alpacas huacaya de la calidad de fibra, seguidamente en el capítulo II, presenta la revisión de la literatura, donde se expone los antecedentes que sustentan el trabajo de la investigación, en el capítulo III, presenta los materiales y métodos, donde se da conocer la población y muestra, la metodología de investigación para poder alcanzar los objetivos propuestos.

1.1 FORMULACIÓN DEL PROBLEMA

Segmentar las alpacas Huacaya de la calidad de fibra nos permitirá determinar la estructura correcta de la segmentación de las alpacas Huacaya, analizar y comparar el comportamiento de las variables en función a las características tecnológicas.

1.1.3 Problema General

¿Cuáles son los grupos segmentados de alpacas Huacaya de la calidad de fibra en el Distrito Cuyocuyo, Provincia Sandía?



1.2 OBJETIVOS DE LA INVESTIGACIÓN

1.2.1 Objetivo General

Segmentar las alpacas Huacaya de la calidad de fibra en el Distrito de Cuyocuyo, Provincia de Sandia del Departamento de Puno.

1.2.2 Objetivos Específicos

- Seleccionar y analizar la formación de grupos homogéneos conformado por alpacas huacaya de la calidad de fibra del Distrito de Cuyocuyo, Provincia Sandia.
- Elegir los criterios de agrupación basado en la distancia existente que cuantifique el grado de similitud de cada grupo homogéneo conformado por alpacas Huacaya de la calidad de fibra del Distrito de Cuyocuyo, Provincia Sandia.

1.3 HIPÓTESIS DE LA INVESTIGACIÓN

1.3.1 Hipótesis General

Al realizar un análisis de conglomerados de las alpacas Huacaya se obtendrá grupos segmentados con mayores similitudes que formen grupos definidos para mantener la calidad de fibra.

1.3.2 Hipótesis Específicos

- Los grupos segmentados de las alpacas huacaya de la calidad de fibra identifican la homogeneidad y afinidad entre ellos del Distrito de Cuyocuyo, Provincia Sandia.
- El criterio elegido basado en la distancia existente si cuantifica el grado de similitud de cada grupo homogéneo conformado por alpacas huacaya de la calidad de fibra del Distrito de Cuyocuyo, Provincia Sandia.



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES

Habiendo realizado la búsqueda de información acerca de antecedentes respecto a nuestra investigación se encontró lo siguiente:

2.1.1 A Nivel Internacional

Ferreira et al.,(2010), refiere la importancia del turismo en la economía así mismo ser potencia en crecimiento sustentable equilibrado de los países Portugal, España y Francia. El principal objetivo del presente es ilustrar las ventajas asociadas a la aplicación de técnicas multivariadas, EL ANALISIS CLUSTER en dos fases, realizándose una segmentación a posteriori (post hoc) basada en las preferencias de los turistas españoles. Los resultados obtenidos permitieron identificar 5 clústeres: los Culture Seekers Oriented by Low Prices. Turistas que prefieren destinos con elevada oferta cultural y precios accesibles; los Sun and Night Lovers, referente a los turistas que prefieren destinos de sol, con elevada oferta de diversión nocturna; y los Night Lovers oriented by Low Prices, que son turistas que buscan destinos con elevada oferta de diversión nocturna y precios accesibles.

Candia, (2014), refiere que esta investigación tiene como propósito hallar segmentos o conglomerados de usuarios en el centro deportivo Energy Fitness Club-Chile, perfilados entorno a variables motivacionales, demográficas y del propio uso del servicio. Utilizo la metodología de segmentación, a través del Análisis Clúster de SPSS (jerárquico a través del método Ward de varianza mínima), Concluye que de un total de 275 usuarios seleccionados de las principales sucursales Energy: Los condes, Cerrillos, La Florida. Se identificaron 5 clústeres: a) no competitivos: para ellos entrenar es una



entretención que les permite conocer a nuevas personas, b) Orientados al logro: persiguen un objetivo o una meta, por lo que se les suele ver concentrados en su rutina, c) en sus vidas, salud y estado físico, d) enérgicos: son principalmente hombres, que en algunos casos son deportistas e) orientados al ocio: entrenan porque son usuarios que usualmente los ves distraídos realizando las actividades dentro de un gimnasio.

Juárez et al., (2016), expresa la falta de información sobre la calidad de la leche y la heterogeneidad en los sistemas de doble propósito del trópico mexicano, la capacitación no se destina a los grupos que la requieren. El objetivo de este estudio fue caracterizar la leche de estos sistemas, compararla con las normas de referencia y relacionarla con las practicas del manejo, además de proponerse la metodología de identificar los grupos que requieren acciones de intervención. Se analizó en 192 muestras de leche cruda en siete localidades. Por ello se aplicó un Análisis Clúster para agrupar las muestras con base en el contenido de grasa, solidos no grasos, crioscopia y acidez titulable. El 65 % de las muestras presentaron valores fuera de norma (sólidos, densidad y crioscopía, sugiriendo una adulteración por adición de agua). El uso de oxitocina, tipo de ordeño, raza y suplementación afectaron ($P < 0.05$) las características de la leche. Mediante el análisis clúster se identificaron cinco grupos de calidad.

Los grupos conformados de excelente y buena calidad presentaron valores dentro de las normas de referencia en todos los parámetros evaluados. Los grupos de deficiente, mala y muy mala calidad (47 %) presentaron valores elevados de crioscopía y bajo contenido de sólidos. Con la metodología propuesta se espera facilitar la intervención enfocada a resolver problemas específicos, evitando el dispendio de recursos.

2.1.2 A Nivel Nacional

Tang et al.,(2016), señala que el estudio presenta una técnica estadística para segmentar clientes de una tienda de retail en el Perú; se trabajó un conjunto de datos



compuestos por 4980 clientes. Donde se aplicó análisis de conglomerados bietapico, que es una técnica de segmentación que permite trabajar con variables cuantitativas y categóricas. Se concluye que el resultado de la aplicación de la técnica brindó 3 conglomerados: el primero con 1817 clientes (36.5%), el segundo conglomerado con 1390 clientes (27.9%) y el tercer conglomerado con 1773 clientes (35.6%). Después de esto se procedió a describir los perfiles de cada conglomerado y se propusieron estrategias de marketing, con el objetivo de fidelizar al cliente.

Lajo & Vizcarra, (2016), da a conocer el problema que existe en las empresas en el Perú especialmente en el área de marketing, esta investigación indica el uso de Análisis de Conglomerados Jerárquico y no jerárquico, y los Árboles de clasificación CHAID exhaustivo en el tema de clasificación clientes que usan smartphones. Concluye que el Análisis de Conglomerados Jerárquico permitió encontrar las posibles cantidades de segmentos óptimos para poder trabajar en los siguientes pasos. El Análisis de Conglomerados No Jerárquico ayudó a validar los resultados obtenidos previamente y permitió determinar una cantidad adecuada de tres segmentos llamados TRADICIONALES (personas que les importa poco tener lo último en tecnología pero buscan que su teléfono móvil funcione como debe ser y que les permita realizar llamadas y mensajes), MODERNOS (ellos valoran que su equipo luzca moderno y sea de marca reconocida por las características físicas exteriores de sus teléfonos móviles) y TECNOLÓGICOS (buscan no solamente de acorde a sus necesidades sino también aprovechan sus funcionalidades y conocen bastante sobre tecnología, buscan la calidad y seguridad que una marca les puede brindar).

Gutiérrez et al., (2015), enfatiza en la presente investigación como objetivo segmentar las causas de la morbilidad materna del hospital San Bartolomé, aplicando el



modelo de clustering basado en el algoritmo de k- means de SQL server, la metodología usada fue Crisp-DM, finalmente los resultados del algoritmo mostraron el modelo de minería de datos presentado por diez grupos segmentados de acuerdo a la mayor similitud que presenta entre ellos y las relaciones que se dan entre cada grupo.

2.1.3 A Nivel Local

Justo, (2015), basado en el análisis económico se realiza a través del concepto de clúster turístico, referido al conjunto de empresas e instituciones que desarrolladas a partir de un recurso turístico base, en este caso de Luquina Chico-Chucuito, dan forma al producto global que caracteriza el destino. Su utilidad radica en la posibilidad de analizar los eslabonamientos productivos que conforman la cadena de valor y las relaciones que vinculan a las empresas y a las instituciones y la comunidad emprendedora. El objetivo de la investigación es determinar los efectos de la clusterización para la optimización del Turismo Rural Vivencial de la Comunidad de Luquina Chico. La metodología aplicada para la investigación es cuantitativa a nivel descriptivo. Cuyos resultados se muestra que los impactos de la clusterizacion en los aspectos socio cultural, económico y ambiental en Luquina Chico se debe al soporte técnico que hicieron las entidades cooperante como Swiss Contac Suiza, ONG CEDESO, DIRCETUR, Municipalidad Distrital de Chucuito, OGD Puno por el lado de la gestión pública y por el lado de la gestión privada se tiene a las Agencias de Viaje y Operadores de Turismo como Edgar Adventure, All ways Travel, Explorandes y Amaru Tours, son las que más proveen turistas resultado de ello tienen una buenísima oferta turística de los servicios de los emprendedores rurales.

Chura, (2013), enfatiza la caracterización molecular y de distancias genéticas en variedades nativas y sus parientes silvestres de quinua, región Puno. Se concluye que los resultados de la investigación señalan que el dendograma obtenido del análisis de las 13 muestras de variedades nativas con un coeficiente de similitud de 0,65 forman dos



clústeres. El primer clúster estuvo conformado por: las variedades: Chullpi, Choclito, Wariponcho, Toledo, Rosa Frutilla, Huallata, Negra Collona, Antahuara, Airampo y Pucakello y el segundo clúster fue -constituido por pasankalla, Salcedo INIA y Kancolla. Mientras que con un coeficiente de similitud de 0,77 se llegan a formar 5 clúster, el primero está formado por 5 variedades nativas, el segundo clúster está formado por tres variedades, el tercer clúster por una variedad, el cuarto clúster por solo una variedad, y el quinto clúster está constituido por 5 variedades nativas. Esto indica que existe similitudes genéticas entre ellas. Analizado los parientes silvestres con un coeficiente de similitud de 0,78 se forman 5 grupos: el primero integrado por un solo pariente silvestre, el segundo grupo está formado por 4 parientes, el tercer grupo está conformado por un pariente silvestre, el cuarto grupo está formado por tres y el quinto grupo está formado por un solo pariente silvestre. Por otro lado, el análisis combinado de 13 variedades nativas y 10 parientes silvestres, muestra un dendograma a un coeficiente de similitud de 0,64 en el que se forman tres clústeres, el clúster (a) formado por 10 cultivares, el clúster (b) constituido por 7 cultivares y el clúster (c) constituido por 6 cultivares. Por lo que podemos asumir que muchas variedades nativas, poseen características genéticas similares en los tres clústeres (a, b y c). Es así que el pariente silvestre HAY1 (Ayara - Huataraqui) se encuentra dentro del clúster (a), indicando que posee características genéticas coincidentes con las variedades nativas: Chullpi, Choclito, Wariponcho, Toledo, Rosa Frutilla, Huallata, Negra Collona, Antahuara y Airampo, por tanto, se deduce que es el ancestro -de estas variedades -nativas. Palabras claves: Quinoa, AFLP, sistema, genética molecular, variedad, silvestres.

Apaza Tito, (2014), refiere una aplicación para la segmentación de imágenes basada en conglomerados, denominado Algoritmos Genéticos K-medias (AGKM) en la región Puno. Esta aplicación fue propuesta debido al deficiente método de selección del

valor de inicialización del algoritmo K-medias al tomar un número de conglomerados inicial de forma aleatoria o por cálculo de la observación visual, esto puede influir en el desempeño del algoritmo, haciendo que tenga una separación inadecuada o demore más tiempo en la búsqueda del número de conglomerados. Se implementó usando la metodología de los Algoritmos Genéticos (AGs) y K-medias, el primero tiene la finalidad de encontrar un número de conglomerados existentes en la imagen y el segundo realiza el proceso de separación. La métrica usada para evaluar la eficiencia de este algoritmo es el valor de la entropía en las imágenes, los resultados obtenidos son sometidos a una prueba estadística que nos indica que existe una ligera mejoría. Se concluye que el AGKM ofrece una ligera mejoría con respecto al algoritmo K-medias tradicional en la segmentación de imágenes satelitales.

2.2 MARCO TEÓRICO

2.2.1 Análisis Exploratorio De Datos

Es un conjunto de técnicas estadísticas, que tiene como objetivo describir los datos, organizándolos y presentándolos de forma que pueda entenderse fácilmente su estructura. (Vargas, 1995)

La finalidad del Análisis Exploratorio de Datos es examinar los datos previamente a la aplicación de cualquier técnica estadística. Nos recomiendan que en la realización de un Análisis Exploratorio de Datos se considere las siguientes etapas:

- a) Preparar los datos para hacerlos accesibles a cualquier técnica estadística.
- b) Realizar un examen gráfico de la naturaleza de las variables individuales que serán analizadas y un análisis descriptivo numérico que permita cuantificar algunos aspectos gráficos de los datos.

- c) Realizar un examen gráfico de las relaciones entre variables analizadas y un análisis descriptivo numérico que cuantifique el grado de interrelación existente entre ellas.
- d) Evaluar, si fuera necesario, algunos supuestos básicos subyacentes a muchas técnicas estadísticas como, por ejemplo, la normalidad, linealidad y homocedasticidad.
- e) Identificar los posibles datos discordantes (outliers) y evaluar el impacto potencial que puedan ejercer en análisis estadísticos posteriores.
- f) Evaluar, si fuera necesario, el impacto potencial que pueden tener los datos ausentes (missing) sobre la representatividad de los datos analizados.

2.2.2 Diagrama De Cajas (Box Plot)

Es una técnica gráfica que nos permite determinar: los valores extremos (outliers), valores atípicos o datos discordantes, la asimetría de la distribución que corresponde al 50% central de los datos y la variabilidad del 50% central de dos o más conjuntos de datos. (Miranda & Salinas, 2012)

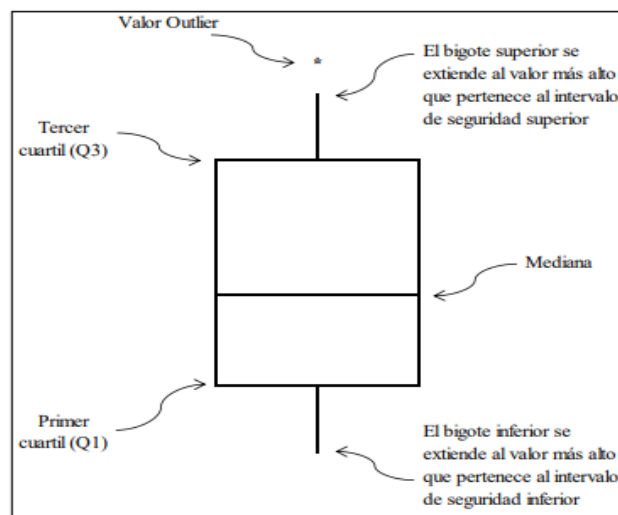


Figura 1. Diagrama de Cajas

Fuente: Adaptado de Diagrama de Cajas por Miranda & Salinas , 2012



La caja contiene información del 50% de los datos ya que la línea inferior de la caja es el primer cuartil ($Q1=P25$), y la línea superior de la caja es el tercer cuartil ($Q3=P75$). La línea que divide a la caja es la mediana y su posición dentro de ella nos brindará información sobre la asimetría de la distribución del 50 % central de datos. Los bigotes son las líneas que se extienden de la parte inferior y superior de la caja hasta la observación más alta o más bajas dentro de los intervalos de seguridad. Estos intervalos son zonas donde teóricamente la variabilidad de los datos que forman parte de ella es similar a la que corresponde al 50% central de datos.

Intervalo de Seguridad Inferior (ISI): $Q1-1.5(Q3-Q1)$

Intervalo de Seguridad Superior (ISS): $Q3+1.5(Q3-Q1)$

Los outliers son observaciones fuera de los límites inferiores y/o superiores.(Miranda & Salinas, 2012)

2.2.2.1 Detección De Outliers Multivariantes

según el autor indica que la estadística que se usa para identificar outliers multivariantes es la distancia de Mahalanobis, dicha distancia se utiliza como la medida entre el centroide de cada caso al conjunto de los datos, donde el centroide es la intersección de las medias de todas las variables en el espacio multidimensional. (Muñoz, 2013)

Para datos multivariantes distribuidos normalmente, los valores de la distancia de Mahalanobis tienen aproximadamente una distribución Chi-cuadrado con tantos grados de libertad como variables consideradas. En consecuencia, aquellas observaciones con una distancia de Mahalanobis grande se indican como valores outliers. (Muñoz, 2013)



$$MSD_i = \sqrt{(x_i - \bar{x})^T - S_n^{-1}(x_i - \bar{x})}$$

Dónde:

- MSD_i es la distancia mahalanobis
- El súper índice T, denota la matriz transpuesta
- (\bar{x}) expresa la media del vector muestral.
- S_n es la matriz de covarianza muestral:

$$S_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n) (x_i - \bar{x}_n)^T$$

Los efectos de enmascaramiento y empantanamiento juegan un rol importante en la adecuación de la distancia de Mahalanobis como criterio para la detección de valores outliers. Es decir, los efectos de enmascaramiento podrían disminuir la distancia de Mahalanobis de un valor outliers. Esto puede ocurrir, por ejemplo, cuando un pequeño grupo de outliers atrae a X_n e infla S_n hacia su dirección. Por otra parte, los efectos de empantanamiento podrían aumentar la distancia de Mahalanobis de las observaciones que no son outliers. Por ejemplo, cuando un pequeño grupo de valores outliers atrae X_n e infla S_n lejos del patrón de la mayoría de las observaciones. (Muñoz Garcia, 2013).

2.2.3 Análisis De Componentes Principales

El ACP trata de encontrar grupos homogéneos de variables a partir de un conjunto numeroso. Consiste en buscar el número mínimo de factores capaces de explicar la máxima información contenida en el conjunto de variables originales, y aplicable solo para variables cuantitativas. Y las observaciones deben de ser independientes entre ellas.



a. Test Esfericidad De Barlett

El test pone a prueba el H_0 mediante X^2 que dice que la matriz de correlaciones empírica R es igual a la matriz identidad I (cuya diagonal principal son unos y el resto de los coeficientes son 0):

$H_0: |R| = 1$ (Hay incorrelación)

$H_a: |R| \neq 1$ (Hay correlación)

Si se confirma el H_0 significa que las variables no están correlacionadas, por lo tanto, la nube de puntos en el espacio formaría una esfera (esfericidad). Si el p-valor es > 0.05 se rechaza el H_0 , esto nos ayudara a tomar una decisión en base a la significatividad estadística. (Rodríguez, 2012)

b. Medida De Adecuación Muestral (KMO Káiser-Meyer-Olkin)

Introducida por Bartlett en 1937, es una modificación del test de Neyman y Pearson para “corregir el sesgo”; esta prueba es la que se utiliza con más frecuencia para probar la homogeneidad de las varianzas. En esta prueba los n_i en cada tratamiento no necesitan ser iguales; sin embargo, se recomienda que los n_i no sean menores que 3 y muchos de los n_i deben ser mayores que 5, según el autor (Rodríguez, 2012)

- Es un índice, no un estadístico.
- Evalúa si las correlaciones parciales entre las variables son suficientemente pequeñas.
- Varía entre 0 y 1
- Valores pequeños indican que el ACP puede no ser una buena idea
- $KMO < 0,5$: no debe utilizarse con los datos suministrados.

$$KMO = \frac{\sum_{i \neq j} r^2_{ij}}{\sum_{i \neq j} r^2_{ij} + \sum_{i \neq j} r^2_{ij} \cdot m}$$

Dónde:

R_{ij} = coeficiente correlación X_i, X_j

$R_{ij,m}$ = coeficiente correlación X_i, X_j eliminado el efecto de las restantes m variables.

2.2.4 Análisis Clúster

Conjunto de técnicas que se utilizan para clasificar los objetos o casos en grupos relativamente homogéneos llamados conglomerados (clústeres), los objetos en cada grupo tienden a ser similares entre sí (alta homogeneidad interna, dentro del clúster) y diferentes a los objetos de los otros grupos (alta heterogeneidad externa, entre clúster) con respecto al criterio de selección predeterminado. El AC tiene como propósito esencial, agrupar aquellos que reúnan idénticas características, es decir, se convierte así en una técnica de análisis exploratorio, diseñada para revelar las agrupaciones naturales dentro de una colección de datos. Este análisis no hace ninguna distinción entre variables dependientes y variables independientes, sino que calcula las relaciones interdependientes de todo el conjunto de variables enfatiza el autor (De La Fuente Crespo, 2015)

Como dice (McDaniel & Gates, 2005), señalan que: “El término Análisis de Conglomerados generalmente se refiere a los procedimientos estadísticos utilizados para identificar objetos o personas que son similares en lo que concierne a ciertas variables o mediciones.

Según el autor Perez, (2004) Proporciona la siguiente clasificación de los métodos de análisis de conglomerados basada en los algoritmos de agrupación de individuos:

- **Métodos Aglomerativos-Divisivos:** el método aglomerativo es aquel que considera tantos grupos como elementos y sucesivamente va agrupando los



dos grupos más similares, hasta llegar a una clasificación determinada; mientras que el método divisivo parte de un solo grupo formado por todos los individuos, y en cada etapa posterior va apartando individuos de los grupos establecido anteriormente, formando nuevos grupos.

- **Métodos Jerárquicos-No Jerárquicos:** el método jerárquico consiste en una secuencia de $g+1$ clúster: G_0, \dots, G_g en la que G_0 es la participación disjunta de todos los elementos y G_g es el conjunto partición. Progresivamente, el número de partes de cada una de las particiones disminuye, lo que hace que estas sean cada vez más amplias y menos homogéneas. Por el contrario, en el método no jerárquico se forman grupos homogéneos sin establecer relaciones de orden o jerárquicas entre dichos grupos. En este método se conoce de antemano el número de conglomerados que se desea formar.

2.2.5 Análisis Clúster Jerárquico

Existen dos formas de generar los conglomerados de manera jerárquica, estos métodos como aglomerativos y divisivos. (Métodos jerárquicos aglomerativos): con el cual empieza con los individuos separados, luego se van agrupando los más similares y al final todos los subgrupos se unen en un único segmento. (Métodos jerárquicos divisivos): se realiza de manera contraria a los métodos aglomerativos. Se parte de un solo grupo que contiene a todas las observaciones y se va dividiendo según qué tan grande sean las diferencias.

Una vez determinadas las variables y los objetos a clasificar el siguiente paso consisten en establecer una medida de proximidad o de distancia entre ellos que cuantifique el grado de similitud entre cada par de objetos. (Matriz de proximidad): A veces la data es representada directamente en términos de proximidad (diferencia o afinidad) entre pares de objetos. Estas pueden ser tanto similitudes o disimilitudes

(diferencia o falta de afinidad). Por ejemplo, en experimentos de ciencias sociales, se pide a los participantes juzgar en cuanto difieren ciertos objetos, unos de otros. Las disimilitudes pueden ser entonces calculadas, promediando sobre el conjunto de tales juicios. Este tipo de data puede ser representado por una matriz D , $N \times N$, donde N es el número de objetos, y cada elemento $d_{ii'}$ registra la proximidad entre el objeto i y el i' . Esta matriz es entonces entregada como entrada al algoritmo de clustering. La mayoría de los algoritmos presume una matriz de disimilitudes con entradas no negativas y elementos de la diagonal nulos: $d_{ii} = 0$, $i = 1, 2, \dots, N$. Si la data original es coleccionada como similitudes, puede usarse una función monótona decreciente apropiada para convertirlas a disimilitudes. Además, la mayoría de los algoritmos asumen matrices simétricas de disimilitud, por lo que, si la matriz original D no es simétrica, debe ser sustituida por $(D + D^T) / 2$. Las disimilitudes subjetivamente juzgadas son, raras veces, las distancias en el estricto sentido, dado que el triángulo de desigualdad $d_{ii'} \leq d_{ik} + d_{i'k}$, para todo $k \in \{1, \dots, N\}$, no se mantiene. Por ello, algunos algoritmos que asumen distancias no pueden usarse con ese tipo de data. Al formar grupos homogéneos, el investigador puede conseguir los siguientes objetivos:

- Descripción de una taxonomía, permitiendo obtener una clasificación de los individuos que a su vez puede ser comparada con una tipología (clasificación basada en la teoría).
- Simplificación de los datos, ya que las observaciones pueden agruparse para análisis posteriores, es decir, permite ver las observaciones como miembros de un conglomerado y perfiladas por sus características generales.
- Validación, La decisión sobre el número óptimo de clústeres es subjetiva, se suelen representar los distintos pasos del algoritmo y la distancia a la que produce la fusión. La obtención del clúster Se trata, fundamentalmente, de



resolver el siguiente problema: Dado un conjunto de n individuos caracterizados por la información de P variables;

X_j , ($j = 1, 2, \dots, p$), nos planteamos clasificarlos de manera que los individuos pertenecientes a un grupo (clúster) (y siempre con respecto a la información disponible de las variables) sean lo más similares posibles entre sí y los distintos grupos sean entre ellos tan disimilares como sea posible, en la regresión logística.

- **Selección Del Número Óptimo Del Clústeres**

Uno de los problemas que encontramos a la hora de aplicar estos métodos de clustering (k-means) es la selección del número de clústeres. No existe un criterio objetivo ni ampliamente valido para la elección de un número óptimo de los clústeres; pero tenemos en cuenta, que una mala elección de los mismos puede dar lugar a realizar agrupaciones de datos muy heterogéneos (pocos clústeres); o datos, que siendo muy similares unos a otros los agrupemos en clústeres diferentes (muchos clústeres), según el autor (Pérez, 2004)

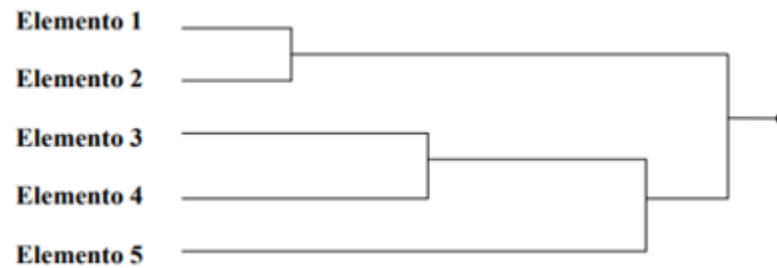


Figura 2. Método Jerárquico Aglomerativo

Fuente: Adaptado por “ Análisis Clúster” , (p, 8), por De La Fuente Crespo, 2015

2.2.6 Método De Distancia Euclídea Al Cuadrado

La medida euclídea es la distancia según una “línea recta” entre dos conglomerados. Solo se puede utilizar cuando todas las variables son continuas, enfatiza el autor (Perez, 2004).

La expresión de la distancia euclídea al cuadrado.

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Donde:

- $d_E(P_1, P_2)$: son las distancias del P_1 y P_2
- $(x_2 - x_1)^2$: representan en el punto del eje de las abscisas
- $(y_2 - y_1)^2$: representas los valores en el punto del eje de las ordenadas

2.2.7 Análisis Clúster a nivel de Individuos

Según el autor (De la Fuente, 2008), El punto de partida para este análisis sea X una muestra de m individuos sobre los que se miden p variables.

Donde: x es un conjunto de valores numéricos que se pueden ordenar en una matriz:

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & x_{22} & \vdots \\ x_{m1} & \dots & x_{mp} \end{pmatrix}$$

x_{11} : valor que presente el primer individuo en la primera variable

x_{ij} : valor que presente el individuo i -esimo en la variable j -esima.

Cada columna contiene los valores que toman todos los individuos para cada variable que se estudia.

- Distancia métrica: una función $d: U \times U \rightarrow \mathbb{R}$ se llama distancia métrica si $\forall x, y, z \in U$ se verifica.

$$D(x, x) \geq 0$$

$$D(x, y) = 0 \Leftrightarrow x = y$$

$$D(x, y) = d(y, x)$$

$$D(x, z) \leq d(x, y) + d(y, z)$$

- Similaridad: una función $s: U \times U \rightarrow \mathbb{R}$ se llama similaridad si para todo $x, y \in U$ se verifica.

$$s(x, x) \leq S_0$$

$$s(x, y) = S_0$$

$$s(x, y) = S_0(y, x)$$

Donde: $S_0 \equiv$ numero real finito arbitrario

- Promedio de grupo distancia entre clústeres

$$A \text{ y } B \text{ son los clústeres donde } d(A, B) = \frac{1}{n_A \times n_B} \sum_{i \in A, j \in B} d(i, j)$$

2.2.8 Dendograma

Es un gráfico usado en el procedimiento que permite visualizar el proceso de agrupamiento del clúster en los distintos pasos, formando un diagrama de árbol, el

dendograma es una valiosa herramienta visual que puede ayudar a decidir el número de grupos que podrían representar mejor que la estructura de los datos teniendo en cuenta la forma en que se van anidando los clústeres y la medida de similitud a la cual hacen. Cortando transversalmente, a una distancia determinada, las ramas del gráfico se obtiene una partición según el autor (Perez, 2004).

Según el autor (De la Fuente, 2008), indica que un Dendograma nos ayuda para la decisión en el corte horizontal que se le asigne en la figura 2 claramente se puede visualizar que se tiene 2 cluster denominados A y C.

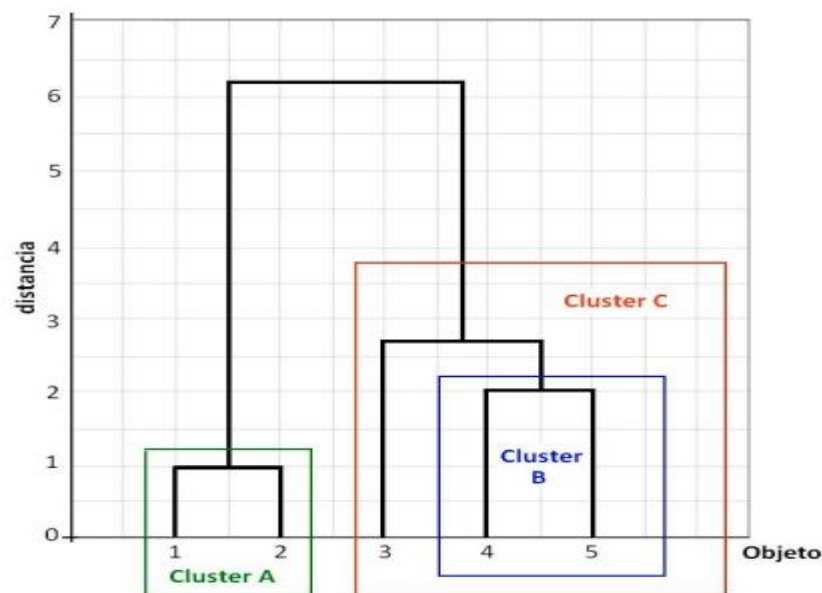


Figura 3. Representación de un Dendograma

Fuente: Adaptado de " Analisis de Conglomerados ", (p-12) por De la Fuente, 2008

El establecimiento del número de grupos no existe una norma fija para establecer cuantos grupos pueden considerarse, el Dendograma puede servir de ayuda visual para determinar dicho número, la similitud entre dos objetos viene dada por la “por la altura” del nodo común más cercano, tal y como se aprecia en la **figura 4** la vaca es mas grande en tamaño y así sucesivamente.

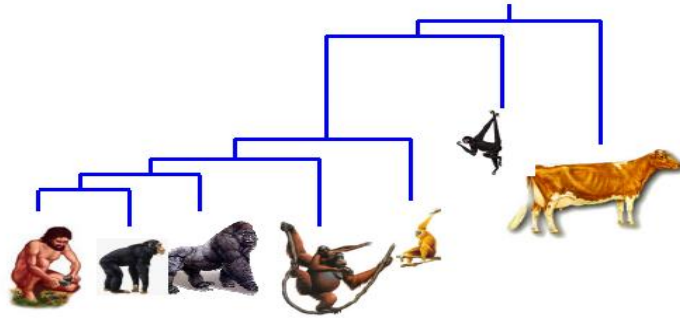


Figura 4. Dendrograma a Nivel de Individuos Similitud por Tamaño

Fuente: Adaptado de “Dendrogram para data Mining” (P.9), por Wang, Keogh, lee, 2009,

University of California

2.2.9 Análisis Clúster K-Means

El análisis de conglomerados de k-means es especialmente útil cuando se dispone de un gran número de casos. Existe la posibilidad de utilizar la técnica de manera exploratoria, clasificando los casos e iterando para encontrar la ubicación de los centroides, o solo como técnica de clasificación, agrupando los casos a partir de los centroides conocidos suministrados por el usuario. Cuando se utiliza como técnica exploratoria, es habitual que se desconozca el número adecuado de conglomerados, por lo que es conveniente repetir el análisis con distintos números de conglomerados y comparar las soluciones obtenidas. En estos casos también puede utilizarse el método de análisis clúster jerárquico con una sub muestra de casos, según el autor (Perez, 2004)

2.2.9.1 Algoritmo de K-medias

El autor menciona (Apaza Tito, 2014), sea del conjunto de datos con n instancias y sea C_1, C_2, \dots, C_k los k grupos disjuntos de D. entonces la función error es definida como:

$$E = \sum_{i=1}^k \sum_{x \in C_i} d(x, u(C_i))$$



Donde $\mu (C_i)$ es el centroide del conglomerado C_i y $d (x, \mu (C_i))$ denota la distancia entre x y $\mu (C_i)$. Aquí definimos la distancia euclidiana $d (x, y_i)$ como medida de distancia estándar entre dos puntos x, y .

$$x = (x_1, x_2, \dots, x_d)$$

$$y = (y_1, y_2, \dots, y_d)$$

$$d (x, y) = \sqrt{\sum_{j=1}^d (x_j - y_j)^2}$$

El algoritmo tiene un procedimiento iterativo. por tanto, primero se atribuye un vector de centroides inicial arbitrario, en la segunda etapa clasifica cada punto para el siguiente conglomerado, en la tercera etapa nuevos centroides son calculados con base en todos los puntos en un conglomerado, finalmente la segunda y tercera etapa son repetidas hasta que el cambio en la iteración es pequeño. El cambio puede ser definido de varias maneras, sea por medio de la medición de distancias entre el centroide y los otros puntos del grupo, o por el porcentual de puntos que cambiaron de grupos entre las iteraciones indicado por el autor (Lucchese & Mitra, 1999 citado por Apaza Tito, 2014).

El pseudocódigo del algoritmo k- medias es presentado a continuación.

Algoritmo: Algoritmo k- medias

Entrada: Distribución de los n puntos de D entre los k conglomerados

Salida: Distribución de los n puntos de D entre los k conglomerados

Sea C_i es el i – ésimo conglomerado.

$$C_1, C_2, \dots, C_k = \text{particion inicial de } D$$

Repita

$d_{i,j} = \text{distancia entre el caso } i \text{ y el conglomerado } j$

Para todo $1 \leq j \leq k$, $n_i = \text{argmin} \{d_i: \forall i, j\}$

Atribuir el caso i al conglomerado n_i ,

Recalcular los centroides

Fin

Ningún centroide cambie de posición

Retorna su salida

La figura 5, presenta la secuencia del algoritmo k- medias.

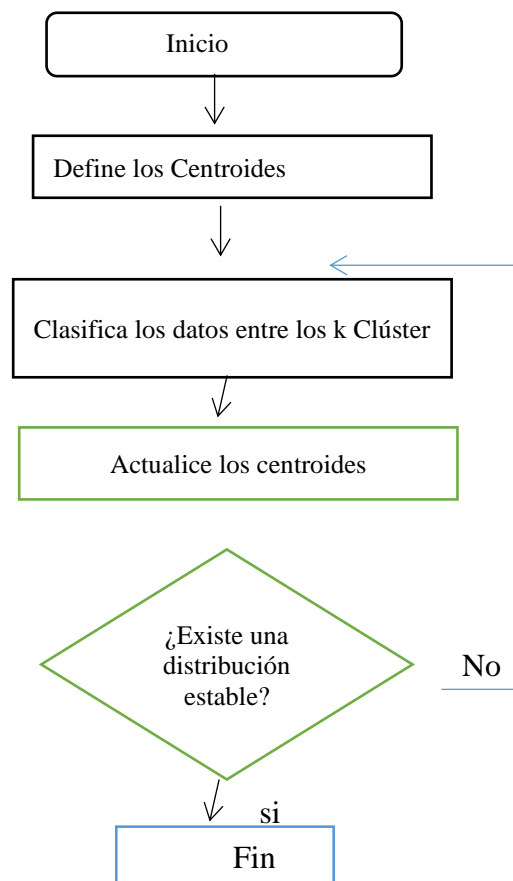


Figura 5. Algoritmo del Diagrama del K-medias

Fuente: Elaboración Propia



Las características claves del K-medias, las que lo hacen eficiente vienen a convertirse en su principal problema:

- La distancia Euclidiana se usa como una métrica y la varianza es usada como una medida de la dispersión de los grupos.
- El número de grupos k es un parámetro de entrada: una elección inapropiada puede acarrear malos resultados. Por eso es muy importante cuando corremos el K-medias tener en cuenta la importancia de determinar el número de grupos para un conjunto de datos.

Una limitación clave del K-medias es su modelo de agrupamiento, el concepto se basa en grupos esféricos que son separables de una forma en que el valor de la media converge hacia el centro del grupo. Se espera que los grupos tengan igual tamaño, por lo que la asignación al grupo más cercano es la asignación correcta.

Por ejemplo, si se tiene un conjunto de datos visiblemente agrupables en 5 conglomerados como se puede observar en la Figura 6.

Caso 1: utilizando el algoritmo K-medias con un valor de inicialización de $k=5$, el resultado es 5 conglomerados como el esperando, este sería el mejor caso donde el algoritmo tiene un buen desempeño. Pero no siempre es así, como se muestra a seguir.

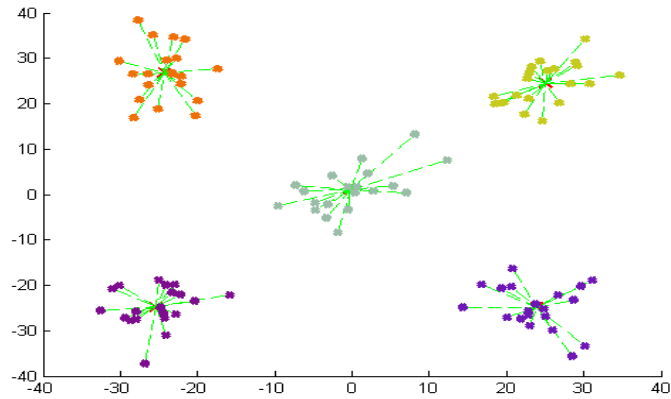


Figura 6. Convergencia Adecuada del Algoritmo Centros Clústeres

Fuente: Adaptado de Tesis, (p.90), por Apaza, Melitón

Caso 2: Dado el conjunto de datos de la figura anterior, donde a simple vista es agrupable en 5 grupos, ejecutando el algoritmo con $k=7$, los cinco grupos visibles no fueron agrupados adecuadamente, esto podría ser debido a que el número de conglomerados iniciales fue $k=7$, esto es superior al número de conglomerados existentes en la figura, el algoritmo genera grupos donde no era posible observar en la Figura 7.

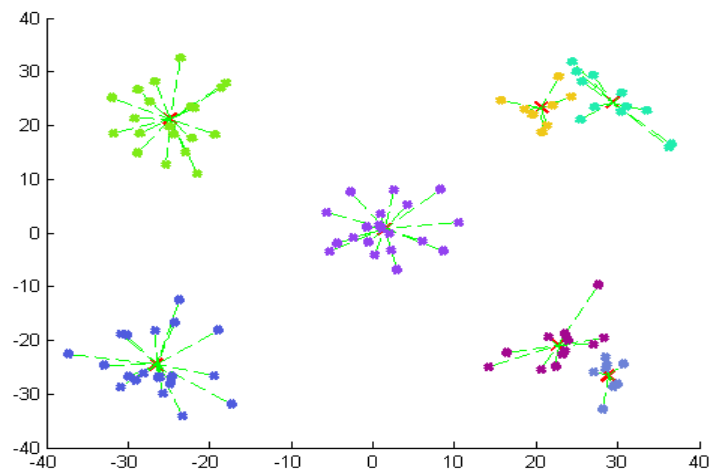


Figura 7. Convergencia adecuada del algoritmo de 7 clústeres

Fuente: Adaptado de Tesis, (p.90), por Apaza, Melitón

Caso 3: Para el mismo ejemplo de 5 grupos visibles, esta vez el algoritmo se encuentra ante una situación donde podemos observar un comportamiento inadecuado debido a que se inicializó con $k=4$, que es un número inferior a la cantidad de conglomerados existentes como se observa en la Figura 8.

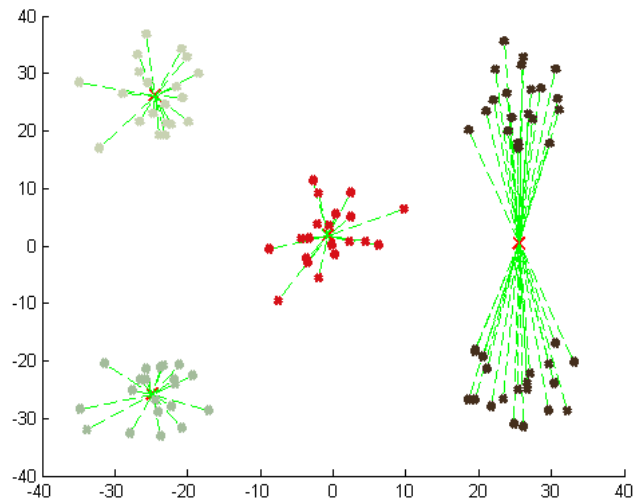


Figura 8. *Convergencia Inadecuada del Algoritmo K- media*

Fuente: Adaptado de Tesis, (p.90), por Apaza, Melitón

2.3 MARCO CONCEPTUAL

2.3.1 Importancia Socioeconómica De La Fibra De La Alpaca

El autor menciona Cadwell *et al.*, (2001); citado por (Díaz, 2014), menciona la principal característica productiva y económica de la alpaca es su fibra, que actualmente se considera como una fibra exótica y sus características textiles de calidad hacen que tenga un precio mayor frente a la lana de ovino en el mercado mundial. La fibra de los camélidos es denominada como una fibra especial al igual que otras fibras animales como la de Cashemire, Mohair y Yak. Entre las características de la fibra de los camélidos sudamericanos existen algunas similitudes, que en cierto sentido podrían establecer competencias entre ellas, desde el punto de vista del uso textil, el estado peruano considera a la fibra de alpaca como producto bandero; sin embargo, existen grandes



dificultades para impulsar la producción definida de fibras con características deseadas por la industria textil, por lo que se debe iniciar procesos de organización y fortalecimiento de capacidades en los productores alpaqueros.

2.3.2 Camélidos Sudamericanos

Los camélidos (Camelidae) son una familia de mamíferos artiodáctilos del suborden tilópodos. El concepto de Camélidos Sudamericanos, incluye a dos grupos de especies. Un primer grupo de especies domésticas, la Llama (*Lama glama*) y la Alpaca (*Lama pacos*), y un segundo grupo de especies silvestres como el Guanaco (*Lama guanicoe*) y la Vicuña (*Vicugna*). La crianza de las especies domésticas, es de alta importancia económica en zonas alto andinas, principalmente en Perú y Bolivia, y en menor grado en Argentina, Ecuador y Chile.

2.3.3 Acopio De La Fibra De Alpaca

En el acopio, la materia prima se compra en vellones de fibra de alpaca. Un vellón equivale a la fibra producida por un animal esquilado. La lana sucia comprada sin clasificar es inspeccionada para ser separada de acuerdo a su procedencia como a su calidad menciona el autor (Antúnez et al., 1996); citado por (Díaz, 2014) .

Para la clasificación por grupos de calidades de la fibra de alpaca se deberá tener en cuenta los criterios siguientes: * Por la finura, seleccionado manual y visual de acuerdo al micronaje de la fibra, por personal calificado. * Por la longitud, seleccionado manual y visual por el largo de la fibra, pudiéndose obtener fibra larga o corta, por personal calificado. * Por el color, seleccionado manual y visual de las diferentes tonalidades de los colores básicos naturales, por personal calificado.



2.3.4 Identificación De Alpaca

Las alpacas se identifican de distinta manera por ejemplo con una numeración llevan un registro, para así tener un control de estos; llamados también número de arete.

2.3.5 Alpaca

La alpaca (del quechua allpaqa, paqu) (*Vicugna pacos*) es una especie doméstica de mamífero artiodáctilo proveniente familia Camélida. Evolutivamente está emparentada con la vicuña, aunque en las poblaciones actuales hay una fuerte introgresión genética de la llama (*Lama glama*). La alpaca representa al miembro más importante de los camélidos sudamericanos en cuanto se refiere a la producción de fibra (Dueñas, 2015).

La industria textil menciona a la fibra de alpaca como una fibra especial, están clasificadas como artículos de lujo (Wang, Liu, & Wang, 2004). Su población mundial se estima en unos 3,7 millones (FAO, 2005) y el 80% de ellas (aprox. 3 millones) se encuentran principalmente en las zonas alto andinas de Perú, de los que alrededor del 86% son alpacas de color blanco (INIA, 2014).

Perú es el primer productor de fibra de alpaca en el mundo, al contar con el 87% de la población mundial, seguido de Bolivia con el 9.5% (CENAGRO, 2012), el Perú cuenta con 3'685,516 cabezas, de las cuales, y el 99% se encuentra en poder de personas naturales (pequeños criadores) y el 1% restante en diversas formas organizativas (medianas y grandes empresas, cooperativas, asociaciones, comunidades campesinas, etc.).

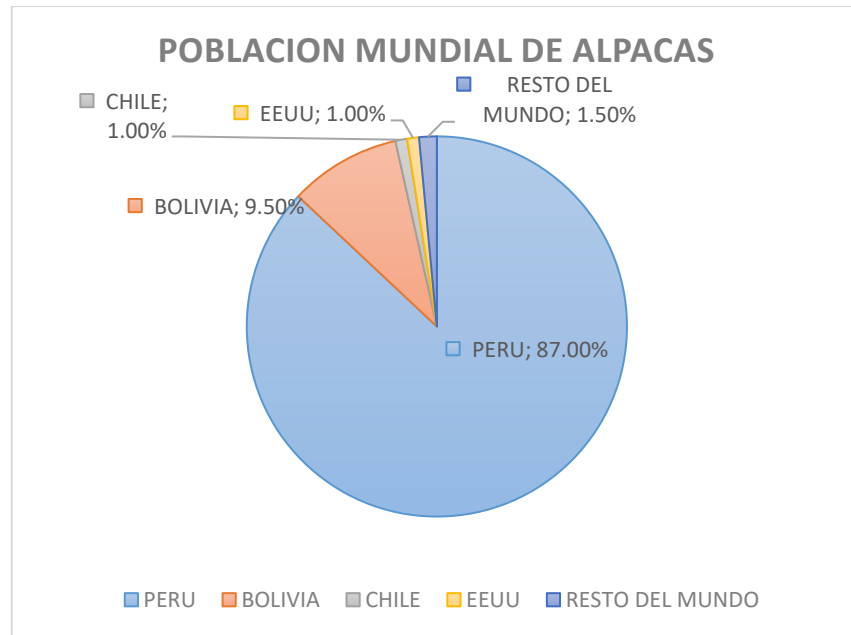


Figura 9. Población Mundial de Alpacas Dirección General de Negocios Agrarios

Fuente: Ministerio de Agricultura y Riego (Minagri) – 2015

La población de alpacas se encuentra distribuida en 17 departamentos, siendo Puno y Cusco los que concentran la mayor producción con 1,880,395 y 552,509 respectivamente. Le siguen Arequipa (446,010); Huancavelica (269,650); Ayacucho (209,947) entre otras Regiones.

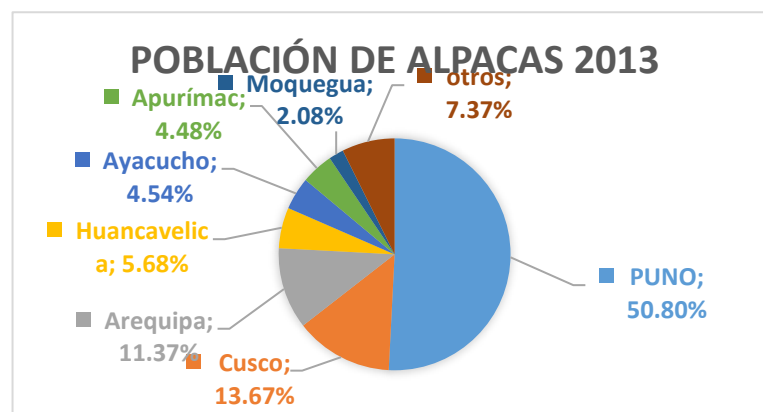


Figura 10. Población Mundial 2013 de la Dirección General de Negocios Agrarios

Fuente: Ministerio de Agricultura y Riego (Minagri) - 2015



Existen dos razas de alpacas según el (INIA, 2014):

Huacaya: que se caracteriza principalmente por ser compacta, de fibra suave y con presencia de ondulaciones, asimismo el vellón manifiesta una apariencia esponjosa, lo que le da una apariencia más voluminosa.

Suri: la cual presenta fibras de gran longitud que se organizan en rizos que caen por los costados del cuerpo, agrupadas en mechadas rizadas, que crecen paralelas al cuerpo.

2.3.6 Raza Huacaya (Alpaca)

Es la raza de mayor difusión en el país, representa el 85% del total de alpacas, siendo sus características zootécnicas las siguientes: Es un animal de buen desarrollo corporal, con fibra que crece perpendicularmente al cuerpo, de cabeza relativamente pequeña, orejas de forma triangular, ollares amplios y pigmentados, boca con bellos muy móviles también pigmentados, con copete bien formado y cara limpia, cuello largo y fuerte. El tamaño aceptable es de 80cm. a la cruz; el vellón debe cubrir todo el cuerpo incluyendo las extremidades hasta las cañas, la línea superior del animal es ligeramente convexa, que continúa hasta la cola, con extremidades fuertes y de buen aplomo, lo que en conjunto le da una armoniosa apariencia general al animal detalla el autor (Crispin, 2014).

2.3.7 Características Tecnológicas De La Fibra De Alpaca

Muchas son las características que tienen influencia en la transformación de la fibra en tejidos u otros terminales de uso, haciendo que el producto tenga una mayor rentabilidad y sea de preferencia para el consumidor. Las principales características son:



Diámetro de la Fibra (DF), Coeficiente de Variación (CV), Índice de Curvatura (IC), Longitud de mecha (LM), Fibra Gruesa (FG) y Factor de confortabilidad (FC).

2.3.1.1 Diámetro De La Fibra (Df)

Como lo dice Gillespie & Flanders (2010); citado por (Gil, 2017, p.13), Se refiere al diámetro que existe cuando la fibra se corta transversalmente Se mide en micrones (micras), lo que equivale a una milésima parte de un milímetro (Cottle, 2010; Poppi y McLennan, 2010; Rowe, 2010; citado por Gil, 2017,p.13) El diámetro de fibra es ampliamente reconocido como la característica más importante de la fibra (Lee *et al.*, 2001; Edriss *et al.*, 2007; Kelly *et al.*, 2007; Rowe, 2010; citado por Gil, 2017, p.13). En consecuencia, las fibras más finas pueden ser transformadas en hilos de tal manera que se adecuen para la confección de una gran variedad de productos textiles (Warn *et al.*, 2006; Rowe, 2010; citado por (Gil, 2017, p.13). Con las fibras más finas se pueden confeccionar tejidos lujosos con peso ligero (Cottle, 2010; citado por Gil, 2017, p.13). Basada en la norma técnica (NTP. 231. 300:2004).

Las fibras con mayor diámetro son las que se encuentran en la espalda, pecho y muslo; así mismo el diámetro se incrementa en la dirección dorso-ventral y Fleecelmente es bien claro que las fibras de menor diámetro se encuentran en la línea media superior del animal, es decir en la cruz, lomo y grupa (Bustinza, 2001).

Los diámetros en todas las fibras naturales están sujetos a variaciones, la misma que depende de las características genéticas y del medio ambiente de donde provienen. De tal manera, existen muchos factores que condicionan la variación del diámetro, tales como la raza, la edad, el sexo y el color del vellón (Calle Escobar, 1982). Las variaciones en el diámetro son causadas también por cambios fisiológicos en el animal debido a la nutrición, gestación, lactación, destete o enfermedades, así como por factores tales como



la edad, sexo, raza, temperatura, radiación solar fotoperiodo, estrés, época del año, época de empadre, época de esquila, sanidad y otros factores característicos del medio ambiente alto andino (Solis, 1991).

2.3.1.2 Coeficiente De Variación (Cv)

El coeficiente de variación del diámetro de la fibra (CVDF) es una medida de amplitud relativa del diámetro de la fibra alrededor de la media dentro de un vellón. Es una variación de medida estandarizada en función al diámetro de la fibra. Un vellón con CVDF más bajo, indica una mayor uniformidad de los diámetros de las fibras individuales dentro del vellón (McLennan y Lewer, 2005; citado por Gil Quispe, 2017). Matemáticamente se expresa como el cociente entre la desviación estándar y el promedio multiplicado por 100, por lo tanto, su magnitud está expresado en porcentaje (Greeff, 2006; citado por Gil Quispe, 2017).

2.3.1.3 Fibra Gruesa (Fg)

Es la característica textil que se utiliza para describir la prioridad espacial de una masa de fibras que pueden ser aceptables entre 0,99 micras-14,56 micras describe el autor (Fish, 1999; citado por Peña et al., 2015)

2.3.1.4 Factor De Confortabilidad (Fc)

El factor de confort se define como el porcentaje de las fibras menores a 30 micras y se conoce también como factor de comodidad (McCull, 2004; Mueller, 2007; citado por Peña et al., 2015). En contraste con el factor de confort es el factor de picazón, que describe el porcentaje de fibras con diámetros mayores a 30 micras. Las prendas confeccionadas con fibras finas son altamente confortables en cambio prendas confeccionadas con fibras mayores a 30 micras causan la sensación de picazón debido a que los extremos de la fibra que sobresalen desde la superficie de los hilos son relativamente gruesas, sin embargo, si estos hilos fueran más delgados serían más

flexibles y existiría menor probabilidad de que provoquen picazón en la piel (Sacchero, 2008; McColl, 2004; Mueller, 2007).

En contraste con el factor de confort es el factor de picazón, que describe el porcentaje de fibras con diámetros mayores a 30 micras. Contrariamente, el porcentaje de fibras mayores a 30 micrones se conoce como el factor de picazón (FP). Por tanto, la industria textil de prendas prefiere vellones con un FC igual o mayor a 95% con un FP igual o menor a 5%. Estos dos parámetros valoran los intercambios de sensaciones que genera entre el cuerpo humano y la prenda de fibra ante las respuestas fisiológicas y sensoriales de las personas (Sacchero, 2008).

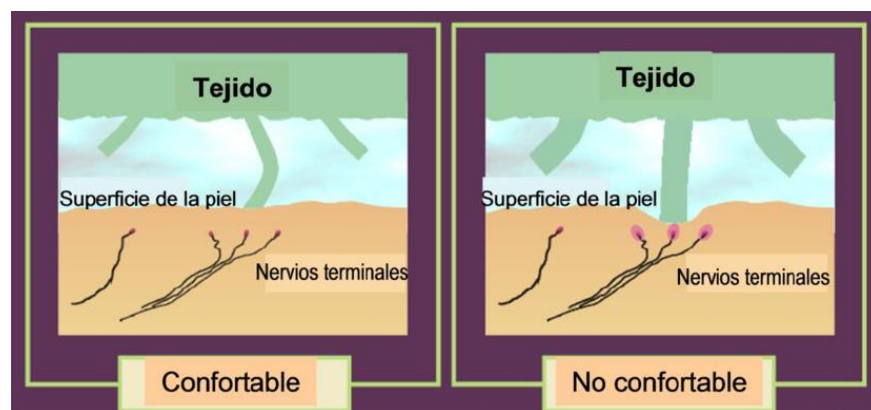


Figura 11. Interacción Entre El Tejido, Los Terminales De La Fibra Y La Piel

Fuente: Tomado de (Garnsworthy R.K., Gully R.L., Kandiah R.P., Kenins P., Mayfield R.J. y Westerman, 1988)

2.3.1.5 Índice De Curvatura (Ic)

Hace referencia el autor (Quispe Peña et al., 2015), al realizar una apreciación visual de las mechas de fibra, las ondulaciones o el aspecto ondulado es evidente (Rogers, 2006). Tradicionalmente, la frecuencia de rizo se utilizó como un marcador indirecto del diámetro de fibra durante la venta de lotes de ovinos, (Cottle, 2010; Hatcher y Atkins, 2000). Sin embargo, en las últimas décadas, el rizo está siendo evaluado en términos de



curvatura de fibra, que describe la frecuencia de rizos que existe en la fibra (McGregor, 2004) o como el número de rizos por unidad de longitud (Hatcher y Atkins, 2000).

La importancia del rizado de la fibra siempre ha abierto debate en la industria textil de la lana. Generalmente se cree que lanas con bajo rizado son mejores que lanas con alto rizado.

- **Definición del rizo.** - La primera puede ser simplemente descrita como el grado de alineamiento del rizo, de modo que lanas donde el rizo de la fibra no se encuentra bien alineado tienen definiciones pobres
- **Frecuencia del rizo.** - se define como el número de longitudes de ondas curvadas por centímetro.

Ambas características conjuntamente con el color de la grasa, longitud de mecha, la suciedad y el desgaste representan el “estilo de lana”, el cual es considerado muy importante para determinar el rendimiento al procesamiento, prácticas de comercialización y la calidad de los productos de lana Fleecer. El rizado de la lana, expresada como curvatura de fibras, se puede medir utilizando los instrumentos comerciales como la OFDA (Wang et al., 2004).

La curvatura de las fibras puede ser en tres dimensiones, debido a que las fibras se encuentran flexionadas y torcidas a lo largo de su longitud. Sin embargo, debido a que la mayor parte de la curvatura ocurre en un plano y teniendo la flexión la mayor contribución, la forma de la fibra puede ser representada en una forma de onda bidimensional (Fish et al., 1999).

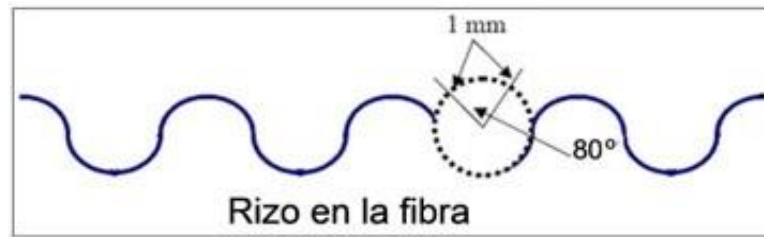


Figura 12. Representación Bidimensional De La Forma De Una Fibra De Lana

Fuente: adaptado por una tesis (p. 25), Fish et al. (1999).

El IC en alpacas ha sido estudiado en Perú por Siguayro, (2009), quien encuentra valores entre 47.66 y 54.01 %/mm en alpacas, mientras que Quispe, (2007) encuentra una media de 38.8 %/mm. Así también, el IC está bien documentado en países como Australia, Nueva Zelanda y Estados Unidos principalmente, basta referir a los resultados de Liu et al. (2004), Wang et al. (2004), Lupton et al. (2006), McGregor (2006) quienes encontraron valores de 28.0, 32.0, 32.5, 32.2 y 27.8 %/mm, respectivamente. Al parecer, la fibra de alpaca Suri tiene menor curvatura que la Huacaya 15 a 35 contra 25 a 60 %/mm respectivamente (Holt, 2006),

En alpacas Huacaya la edad no afecta la curvatura, pero en Suris la curvatura de la fibra fue el doble de los registrados para edades de dos o más años (McGregor, 2006)

2.3.1.6 Longitud De Mecha (Lm)

Al igual que el diámetro es una característica muy importante que determinan las propiedades manufactureras del material textil, siendo recomendable de 8 a 10 cm de crecimiento para la esquila. Hace referencia el autor (Flores Gutiérrez - 2009; citado por Quispe Peña et al., 2015)

Es la longitud de fibra crecida en un periodo determinado, generalmente de un año lo que sería el intervalo entre las esquilas (Consiste en cortar toda la fibra del animal a máquina o con tijeras especiales.).

Norma Técnica Peruana (NTP 231.309) Indica determinar la longitud de mecha en milímetros (mm) y determinar la finura de fibra en micras (μ).

2.3.8 Método Del Ofda

Esencialmente, el OFDA 2000 (Optical Fibre Diameter Analyser), es un microscopio automático que aumenta y captura imágenes de la fibra individual y adjunta esta imagen a una cámara de vídeo (Van Schie, H.F.M., Marler, J.W., Barry, 1990). Las fibras se cortan en snippets de 2 mm de longitud y son extendidas sobre un portaobjetos de vidrio, luego son identificadas y medidas individualmente. El método al igual que el Laserscan produce también un histograma de la distribución del diámetro de fibra, indicando el diámetro de fibra promedio, desviación estándar y coeficiente de variación. Adicionalmente, el equipo proporciona mediciones de largo de mecha, finura al hilado, curvatura media y factor de confort. Los resultados son fiables y reproducibles, debido al gran número de mediciones realizadas (Grishanov, S.A., Harwood, R.J., 2006).



Figura 13. Equipo Del Análisis De Las Características De La Fibra (OFDA 2000)

Fuente: Adaptado de OFDA 2000, por Western, Australia



CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 MÉTODO DE INVESTIGACIÓN

En este capítulo se explicará brevemente los procedimientos y /o métodos que se utilizaron para la solución del problema planteado, explicando cómo se realizó la investigación y la técnica estadística a utilizar

Tipo de investigación: Según, (Sampieri, 2014) el presente trabajo de investigación es no experimental, transversal del tipo descriptivo ya que tiene como propósito encontrar segmentos de las alpacas huacaya de la calidad de fibra (características físicas de la lana).

3.2 TÉCNICAS E INSTRUMENTOS PARA LA INVESTIGACIÓN

Se obtuvo los datos del Gobierno Regional de Puno con el proyecto PECSA (Proyecto Especial de Camélidos Sudamericanos), lo cual nos permite seleccionar nuestras variables de Estudio de un total de población de 208 alpacas huacayas.

Estos datos fueron analizados y obtenidos con el instrumento OFDA en función a la calidad de fibra (características físicas de la lana) para su posterior estudio.

3.3 POBLACIÓN Y MUESTRA DEL ESTUDIO

a) Población

Está conformada por la totalidad de 208 Alpacas Huacaya de la comunidad Puna Ayllu del distrito de Cuyocuyo.

b) Muestra

La muestra fue determinada por toda la población, es decir lo conforman 208 registros de Alpacas Huacaya de cuatro productores.



3.4 IDENTIFICACIÓN DE VARIABLES:

- Variables seleccionadas:

Tabla 1

Identificación de Variables

VARIABLE	INDICADOR	ÍNDICE
DIAMETRO DE FIBRA	Micras	Entre 23 μ m- a 29 μ m
COEFICIENTE DE VARIABILIDAD	Porcentaje	Entre: 9,04% - 50,22%
LONGITUD DE MECHA	Centímetros	De 8 a 10 cm
INDICE_CURVATURA	Grados/mm	Entre: 17,62°/mm - 87,29°/mm
FIBRA GRUESA	Micras	Entre: 0,99 μ m - 14,56 μ m
FACTOR DE CONFORMIDAD	Porcentaje	Entre: 29,79% - 100%

Fuente: Elaboración Propia. Gobierno Regional Puno PECSA, en base a los resultados del software Spss 19.

3.5 METODOLOGÍA DE LA INVESTIGACIÓN

El procedimiento y análisis de datos:

3.5.1 Análisis Descriptivo y Exploratorio

El primer paso a realizar es un análisis exploratorio y descriptivo de las variables cuantitativas y cualitativas que permitió observar el comportamiento de dichas variables.

3.5.2 Prueba de la Normalidad

Se aplicó el método de Kolmogorov –Smirnov por lo que nuestra población es mayor a 50.

3.5.3 Obtención del Clúster Jerárquico y No Jerárquico

- Clúster Jerárquico

Se determina mediante las distancias para el corte adecuado de la conformación de conglomerados (Dendograma).

- Clúster No Jerárquico (K -Medias)

En este método, la medida de distancia o de similaridad entre los casos se calcula utilizando la distancia euclídea. Es muy importante el tipo de escala de las variables para realizar el análisis de conglomerados de k-medias (estandarizadas).

- Algoritmo de k medias, se detalla los centroides. Ver figura 4.

Algoritmo: k- medias

Entrada: Distribución de los n puntos de D entre los k conglomerados

Salida: Distribución de los n puntos de D entre los k conglomerados

Sea C_i es el i – ésimo conglomerado.

$C_1, C_2, \dots, C_k = \text{particion inicial de } D$

Repita

$d_{i,j} = \text{distancia entre el caso } i \text{ y el conglomerado } j$

Para todo $1 \leq j \leq k, n_i = \text{argmin} \{d_i: \forall i, j\}$

Atribuir el caso i al conglomerado n_i ,

Recalcular los centroides

Fin

Ningún Centroide Cambia De Posición Y Retorna Su Salida.

3.5.4 Comparación de Medias y Prueba de Varianza.

Para la validación de los conglomerados conformados se realiza comparación de medias, para determinar que los individuos se aglomeren mediante las características físicas de la calidad de fibra, para así no perder la calidad de fibra de la lana en cada individuo.

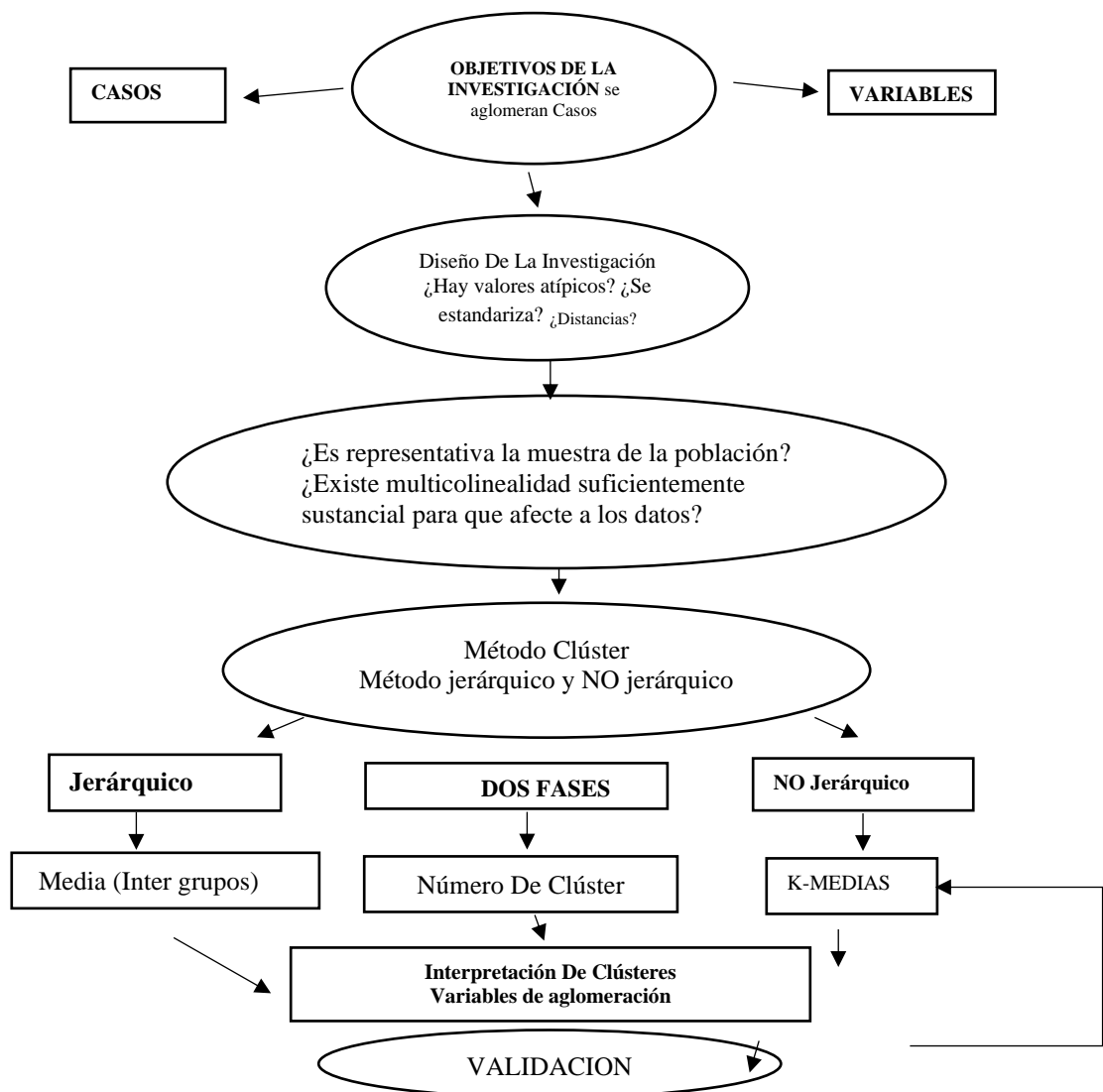


Figura 14. Esquema General del Análisis Clúster

Fuente: Adaptado de Análisis Multivariados por (Perez Lopez, 2004)

3.6 CARACTERIZACIÓN DEL AREA DE INVESTIGACIÓN

3.6.1 Ámbito De Estudio

El trabajo de investigación se realizó en el en la comunidad de Puna Ayllu del distrito de Cuyocuyo, Provincia Sandia del Departamento de Puno, ubicado a una altitud de 3401 m.s.n.m., situada a 200 Km. de distancia de la ciudad de Juliaca, y está entre los paralelos 14° 28' 00" de latitud sur y 69° 32' 06" de longitud oeste del Meridiano de Greenwich.

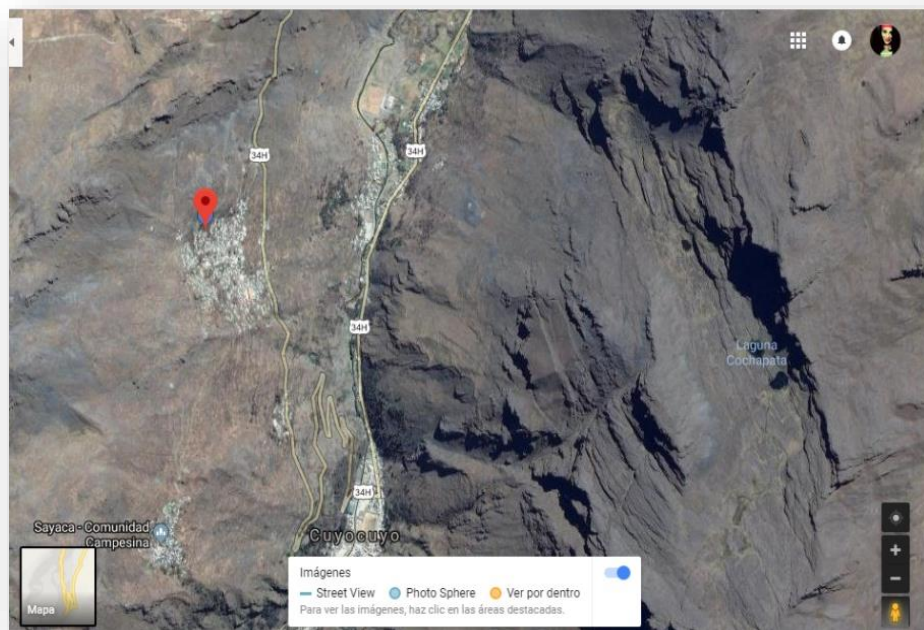


Figura 15. Mapa de Localización

Fuente: Google Earth

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 RESULTADOS OBTENIDOS ANÁLISIS EXPLORATORIO

Para desarrollar el análisis clúster jerárquico y no jerárquico, en primer lugar, analizamos los estadísticos descriptivos. Se presenta a continuación los resultados obtenidos con el conjunto de datos inicial.

Tabla 2

Estadística Descriptiva de los Datos

	DIAMETRO DE FIBRA (μm)	COEF DE VARIABILIDAD (%)	LONGITUD DE MECHA (cm)	ÍNDICE DE CURVATURA (°/mm)	FIBRA GRUESA (μm)	FACTOR DE CONFORMIDAD (%)	
Número de casos válidos	208	208	208	208	208	208	
Perdidos	0	0	0	0	0	0	
Media	1,87	22,5886	21,9747	10,3514	45,2070	9,4007	92,3704
Desviación estándar	1,108	2,57707	3,42052	2,85902	7,37613	1,43540	5,46845
Mínimo	0	16,04	15,64	4,00	26,30	5,73	78,38
Máximo	3	28,40	30,40	17,00	63,86	13,10	100,00

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software Spss 19

La Tabla 2 muestra los estadísticos de las características de fibra de alpaca de un total de 208 alpacas encontramos el Diámetro de Fibra con una media de 22,5886 μm ,entre un mínimo de 16,04 y un máximo de 28,40 μm, El Coeficiente de Variabilidad de la fibra de alpaca se obtuvo en promedio 21,9747 % entre un mínimo de 15,64 % y un máximo de 30,40% , la Longitud de Mecha de la fibra de alpaca con un promedio de 10,3514 cm entre un mínimo de 4,00 cm y un máximo de 17,00 cm, el Índice de Curvatura con un promedio de 45,2070 °/mm entre un mínimo de 26,30 °/mm y un máximo de 63,86 °/mm , la Fibra Gruesa se encuentra en un promedio de 9,4007 (μm) entre un mínimo de

5,73 (μm) y un máximo de 13,10 (μm), finalmente el Factor de confortabilidad con una media al 92,3707 % entre un mínimo al 78,38 % y un máximo al 100,00 %.

4.2 DETECCIÓN DE VALORES ATÍPICOS

El análisis clúster es muy sensible a la presencia de objetos muy diferentes del resto (valores atípicos).

4.3 ESTANDARIZACIÓN DE DATOS

El orden de las similitudes puede cambiar bastante con sólo un cambio de escala de una de las variables por lo que sólo se trabaja con valores estandarizados. Ver Figura 16 Y ANEXO A

4.4 CONTRASTACIÓN DE LA HIPÓTESIS GENERAL

- PRUEBA DE NORMALIDAD

Se utiliza la prueba de normalidad de Kolmogorov Smirnov, debido a que la población es mayor a 50.

- PLANTEAMIENTO DE LA HIPÓTESIS

H_0 : Los datos tienden a una distribución normal

H_1 : Los datos no tienden a una distribución normal

A un nivel de confianza del 95% y un margen de error del 5% los resultados se muestran en la siguiente tabla.

Decisión

Si: $p - \text{valor} \Rightarrow 0,05$ se acepta la **H_0**

$p - \text{valor} < 0,05$ se acepta la **H_1**

Tabla 3

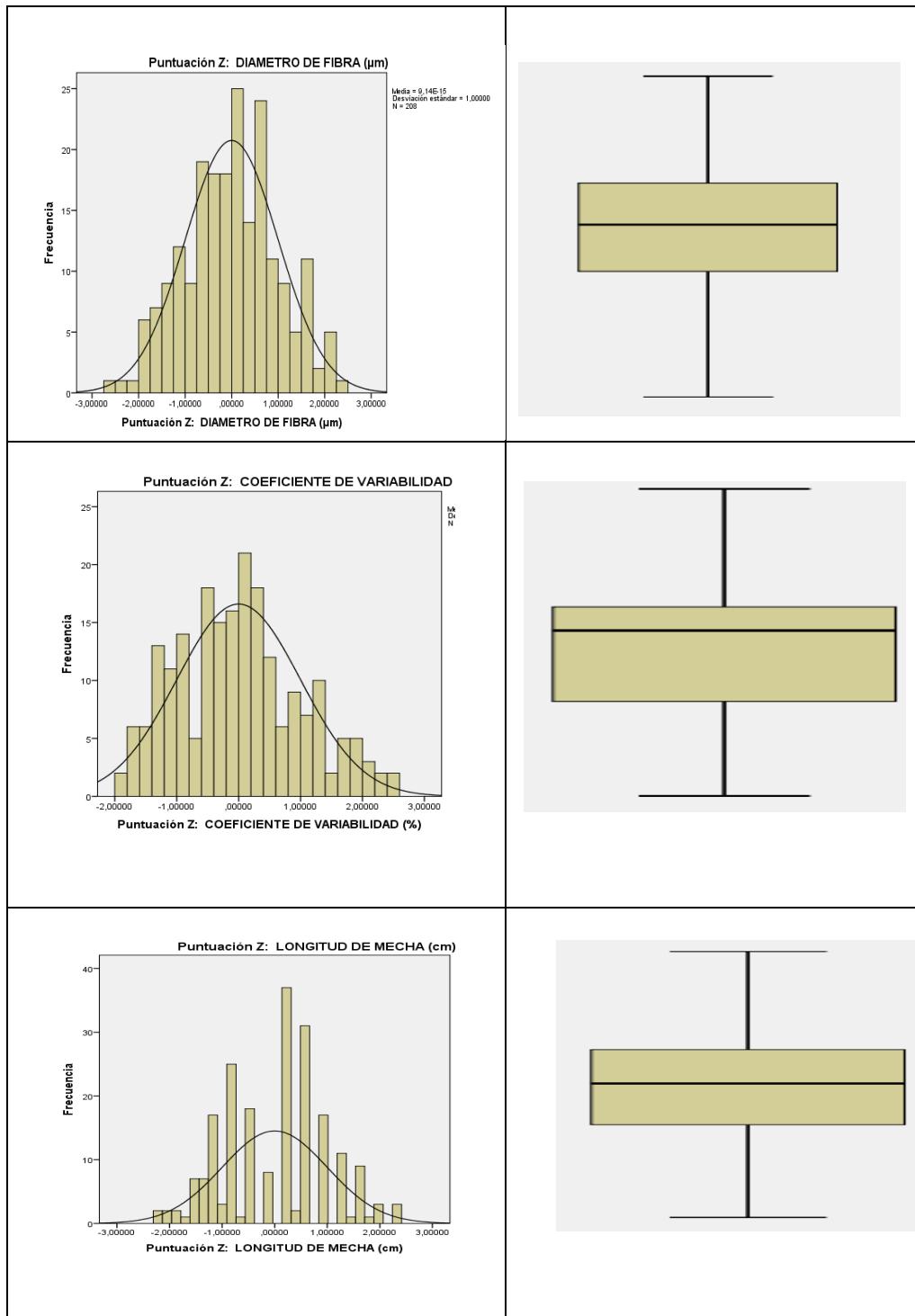
Prueba de Normalidad

		Z	Z	Z	Z	Z	Z
		DIAM_	COEF_	LONG_	IND_C	FIB_GR	FACT_
		FIBRA	VAR	MECH	URV	UESA	CONFO
		A					R
N		5	5	5	5	5	5
		3	3	3	3	3	3
Paráme	Media	1	,	-	-	,	-
tros normales'	^b	,1899	2380	,2912	,9276	9315	1,4127
	Desviación	,	,	,	,	,	,
	estándar	53440	84510	89040	66659	82780	57944
Máxim	Absoluta	,	,	,	,	,	,
as diferencias		095	053	147	084	084	071
extremas	Positivo	,	,	,	,	,	,
		095	052	147	084	064	071
	Negativo	-	-	-	-	-	-
		,083	,053	,116	,084	,084	,067
	Estadístico de	,	,	,	,	,	,
	prueba	095	053	147	084	084	071
	Sig. asintótica	,	,	,	,	,	,
	(bilateral)	200 ^{c,d}	200 ^{c, d}	006 ^c	200 ^{c, d}	200 ^{c,d}	200 ^{c, d}

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

En la Tabla 3 se muestran los resultados de la prueba de normalidad Kolmogorov - Smirnov para la variable DF con p-valor 0,200>0,05, CF con p-valor 0,200>0,05, LM con p-valor 0,06>0,05, IC con p-valor 0,200>0,05, FG con p-valor 0,200>0,05 y el FC con p-valor 0,200>0,05, en tal sentido aceptamos la hipótesis nula, los datos tienden a una distribución normal.



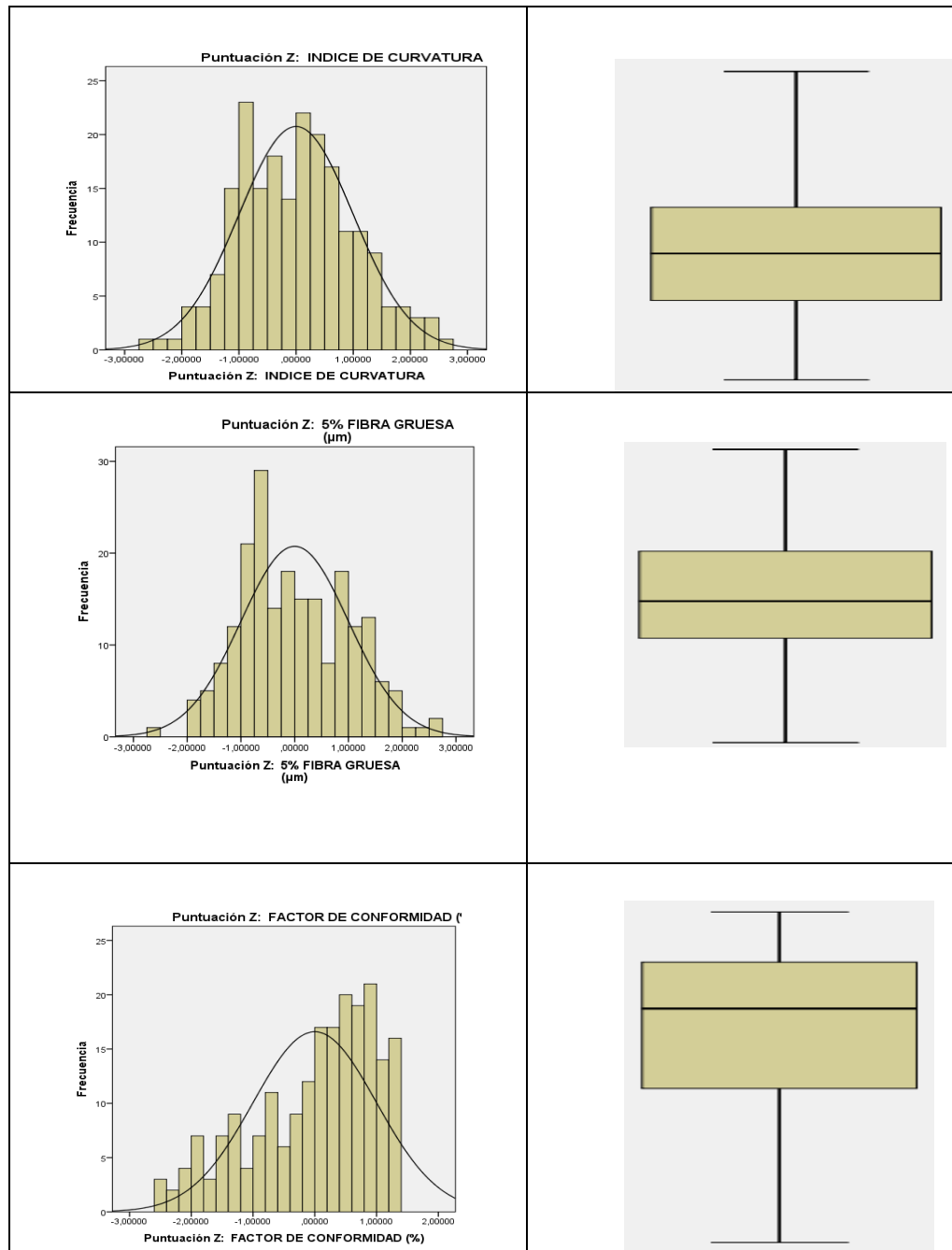


Figura 16. Histograma, Diagrama de Caja y Bigote

Fuente: Elaboración Propia cuadros obtenidos en imagen del software Spss 19

De la Figura 16, los Datos ya estandarizados de las variables DF, CV, LM, IC, FG, FC no presentan *outliers* con respecto a la segunda columna (Diagrama de caja y bigotes). Para la primera columna (Histogramas) el conjunto de datos para las 6 variables se observa que hay simetría por lo tanto tiende a una distribución normal.

4.5 SEGMENTACIÓN CON EL ANÁLISIS CLÚSTER JERÁRQUICO

La Caracterización de las alpacas de la comunidad Puna Ayllu del distrito de Cuyocuyo. Mediante la clasificación jerárquica con el método enlace promedio entre grupos. Obtenemos lo siguiente:

- A Nivel de Individuos

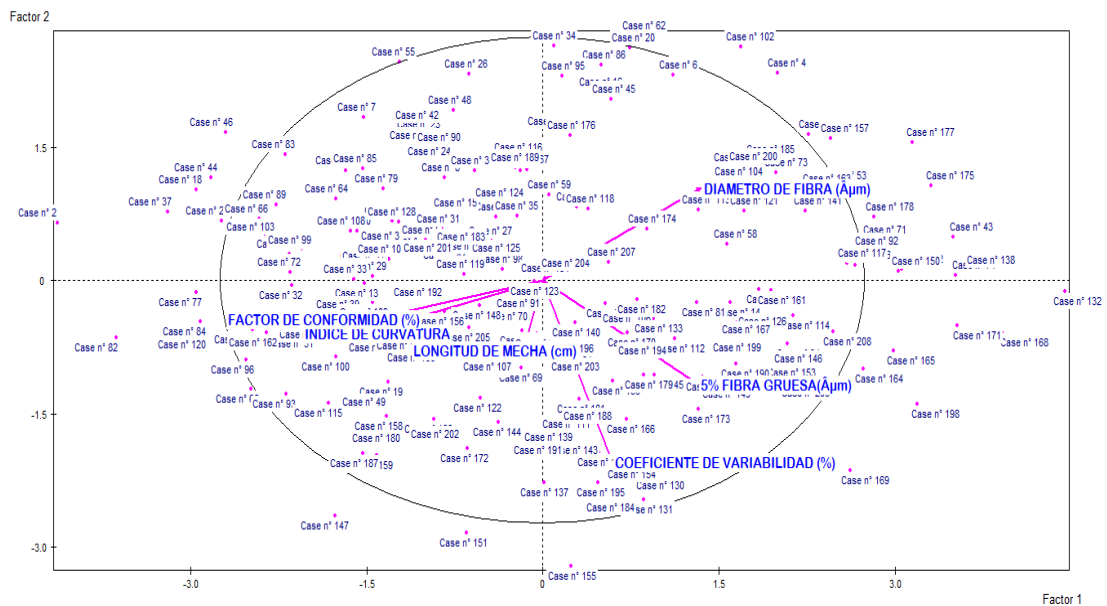


Figura 17. Gráfico del Comportamiento de los Individuos (Alpacas)

Fuente: salida de resultados en el software SPAD N.

- Enlace promedio entre grupos, se combinan los grupos buscando que la distancia promedio dentro de cada clúster sea la menor posible. Ver tabla 4.



Tabla 4

Historial de Conglomeracion- Promedio Entre Grupos

Etapa	Clúster combinado		Coeficient e	Primera aparición del clúster de etapa		Etapa siguiente
	Clúster 1	Clúster 2		Clúster 1	Clúster 2	
		1		2	,09	
	34	04	9	0	0	48
	4	6	,15	0	0	10
	1	0	3	0	0	82
	9	9	,17	0	0	82
	1	8	0	0	0	87
	2	4	,18	0	0	87
	3	8	0	0	0	59
	1	1	,18	0	0	59
	49	60	8	0	0	19
	3	4	,19	0	0	19
	1	7	0	0	0	109
	1	2	,19	0	0	109
	96	03	3	0	0	72
	6	9	,19	0	0	72
	5	6	7	0	0	43
	8	8	,22	0	0	43
	0	8	7	0	0	26
	4	7	,22	2	0	26
0	1	0	9	2	0	26

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Sps 19

En la **Tabla 4** se muestra las 10 primeras etapas del enlace promedios entre grupos. En la primera etapa se unen las alpacas 134 y 204 porque son las que tienen la distancia más pequeña (0,099). El grupo creado por 134 y 204 aparece de nuevo en la



etapa 48 para unirse con la etapa 44 (quinta columna), Así uniéndose los individuos (alpacas) 123 y 125 sucesivamente se van formando los grupos. **Ver Anexo C**

4.6 FORMACIÓN DEL DENDOGRAMA

- Se obtuvo utilizando la técnica de análisis de conglomerados jerárquicos. Mediante el método de enlace promedio entre grupos.
- La lectura se realiza de izquierda a derecha, el objetivo es agrupar los individuos en base a sus características físicas de la fibra de alpaca para una mejor interpretación revisar en el Anexo Historial de Conglomeración donde se ve la conformación de las 208 alpacas.

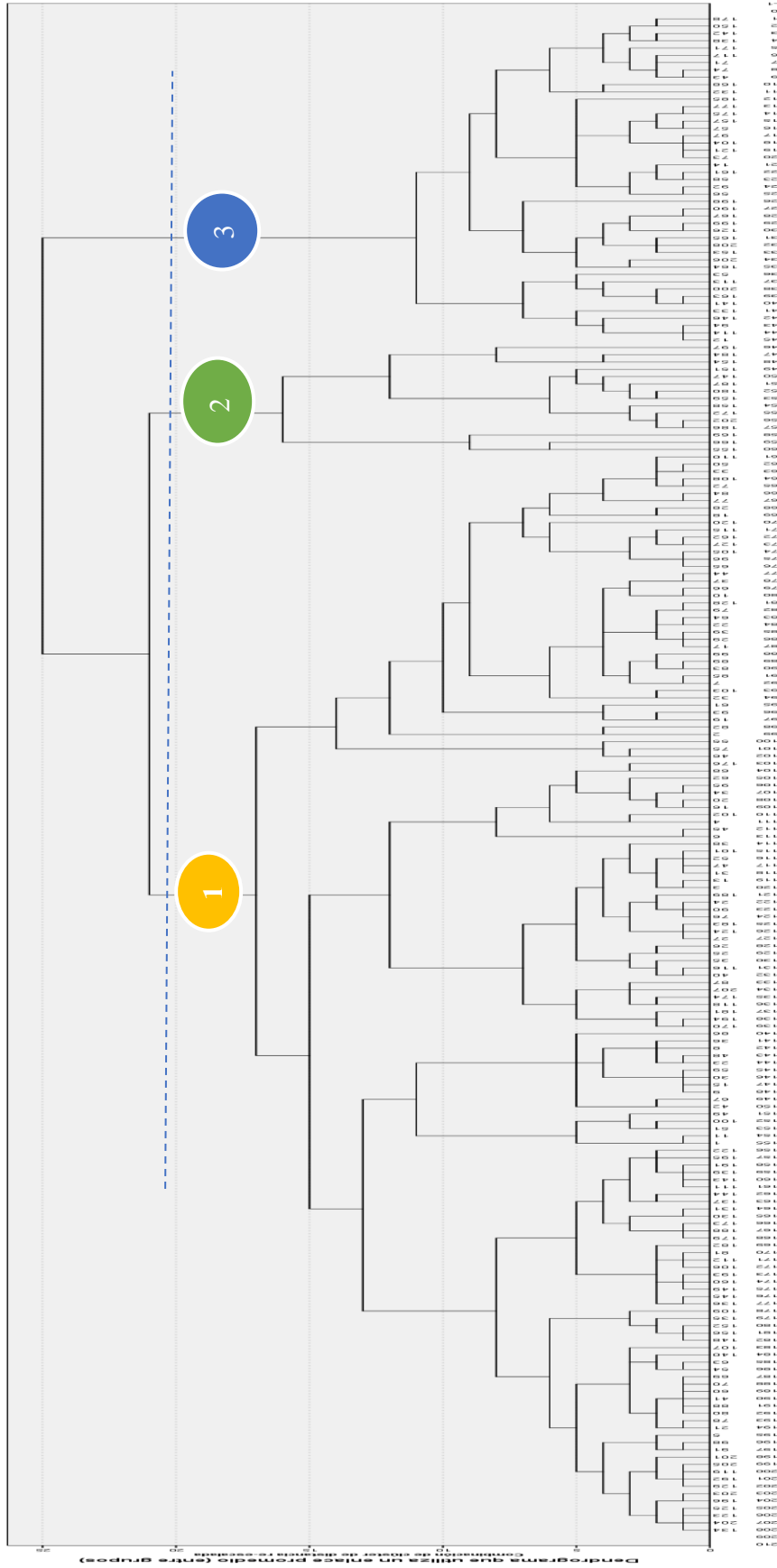


Figura 18. Dendrograma Formado en 3 grupos en función de la Calidad de Fibra

Fuente: Gobierno Regional Puno PECSA Elaboración en base a los resultados del software Spss 19



En la FIGURA 18 nos muestra el dendograma formado en 3 grupos en función a la calidad de fibra de cada individuo (alpaca huacaya), los individuos se conectan mediante enlaces cuya posición en el diagrama está determinada por el nivel de similitud entre cada individuo, es por ello que se realiza el corte horizontal para su división de clústeres.

4.7 SEGMENTACIÓN MEDIANTE ANÁLISIS CLÚSTER K-MEAN

Entre el método no jerárquico, se utiliza el K-medias sin especificar los centros de los conglomerados.

Con centros desconocidos, el método K-medias comienza con una división del conjunto de los individuos en (x) grupos configurados al azar y posteriormente busca mejorar esta primera clasificación reasignando los elementos al centroide del clúster más cercano, tratando de reducir la distancia media entre cada elemento de un grupo y su centroide. El proceso de funcionamiento de este método es el siguiente:

- Se comienza con una partición inicial de los individuos en un específico número de agrupamientos, para calcular posteriormente el centroide de cada uno. Esta partición inicial comienza con los casos más alejados entre sí.
- El siguiente paso trata de reasignar cada caso al agrupamiento más cercano, aquel cuya distancia al centro de gravedad del conglomerado sea menor. No hay que olvidar que en el método de K-medias, al formar parte de los métodos de reasignación, un individuo asignado a un conglomerado en una determinada iteración puede ser reasignado a otro caso en una iteración posterior.
- Calcula los nuevos centroides de los conglomerados cada vez que se incorpora un nuevo caso.



- Repite alternativamente el segundo y el tercer paso hasta que ninguna reasignación de un caso a un nuevo clúster permita reducir más la distancia entre los individuos dentro de cada agrupamiento, ni aumentar la distancia entre los distintos clústeres.(Hernández, 2016)

Tabla 5
Centros de Clusters Iniciales

		Clúster		
		1	2	3
Puntuación Z:				
DIÁMETRO DE FIBRA (μm)		2,02222	-	-
			1,88532	,50003
Puntuación Z:				
COEFICIENTE DE VARIABILIDAD (%)		,85523	-	
			1,59761	2,14159
Puntuación Z:				
LONGITUD DE MECHA (cm)		,22685	1,27616	-
			1,17223	
Puntuación Z: INDICE				
DE CURVATURA		-	1,91876	1,42257
	2,56326			
Puntuación Z: 5%				
FIBRA GRUESA (μm)		2,22884	-	
			2,55729	1,81084
Puntuación Z:				
FACTOR DE CONFORMIDAD (%)		-	1,36228	-
	2,05184			,01288

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

Tabla 6

Centros de Clusters Finales

	Clúster		
	1	2	3
Puntuación Z: DIAMETRO DE FIBRA (μm)	1,18991	-,39239	-,42589
Puntuación Z: COEFICIENTE DE VARIABILIDAD (%)	,23798	-,79128	,85104
Puntuación Z: LONGITUD DE MECHA (cm)	-,29121	-,14041	,41478
Puntuación Z: INDICE DE CURVATURA	-,92765	,45794	,13234
Puntuación Z: 5% FIBRA GRUESA (μm)	,93146	-,85630	,38786
Puntuación Z: FACTOR DE CONFORMIDAD (%)	-1,41274	,64591	,26918

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

Tabla 7

Distancia Entre Los Centro de los Clusters Finales

Clúster	1	2	3
1		3,597	2,781
2	3,597		2,191
3	2,781	2,191	

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

En la Tabla 7 muestra la distancia entre los centroides de los clústeres finales.

Si comparamos estos resultados con los de la Tabla 5 y la Tabla 6 observaremos que los centros de clúster cambian al inicio hay distancia mayor y finalmente se acortan en los tres grupos.

Los grupos se han formado deliberadamente de acuerdo con la distancia entre ellos en el espacio multidimensional, la condición para la aleatoriedad de las observaciones en los diferentes grupos no se halla, los resultados del análisis de dispersión son puramente descriptivos. Es decir, no podemos usar el nivel relevante (columna de la



tabla 8 ANOVA- Análisis de Dispersión de los resultados de clasificación) para verificar la hipótesis de las principales variables. Sin embargo, las diferencias entre las F- ratios (columna F en la Tabla (ANOVA) hace posible extraer conclusiones generales sobre el comportamiento de las variables en la formación del clúster.

En la Tabla 8 tenemos los resultados de dispersión. Nos indica que todas las variables (DF, CV, LM, IC, FG, FC) son significativas al P-valor < 0.05 , a si mismo la cuarta columna muestra la media cuadrática DENTRO de cada grupo; y la media cuadrática (variabilidad) ENTRE grupos aparece en la segunda columna.

Así mismo la ratio entre ambas medias se presenta en la sexta columna, de forma que los altos valores des estadístico F, indican que la variabilidad entre los 3 grupos es mucho mayor que la variabilidad dentro de cada grupo ($50,372/0,518= 97,184$) indicando que dentro de cada grupo son homogéneos y heterogéneos entre cada grupo.

Tabla 8

Cuadro de la Varianza del Analisis Cluster No Jerarquico

	Clúster		Error		F	Sig.
	Media cuadrática	Gl	Media cuadrática	Gl		
Puntuación Z: DIAMETRO DE FIBRA (μm)	50,372	2	,518	205	97,184	,000
Puntuación Z: COEFICIENTE DE VARIABILIDAD (%)	53,314	2	,490	205	108,887	,000
Puntuación Z: LONGITUD DE MECHA (cm)	8,878	2	,923	205	9,617	,000
Puntuación Z: INDICE DE CURVATURA	32,618	2	,692	205	47,167	,000
Puntuación Z: FIBRA GRUESA (μm)	60,294	2	,422	205	143,039	,000
Puntuación Z: FACTOR DE CONFORMIDAD (%)	73,674	2	,291	205	253,186	,000

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

Tabla 9

Numero de Casos en Cada Cluster

Clúster	1	53,000
	2	88,000
	3	67,000
Válidos		208,000
Perdidos		,000

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados

en el software Spss 19

La Tabla 9 presenta los datos del número de unidades que hay en cada clúster, así como el número total y los elementos perdidos (si hay alguno).

El primer clúster está conformado por 53 alpacas, el segundo clúster con 88 alpacas, tercer clúster conformado por 67 alpacas.

4.8 FORMACIÓN DE CLÚSTER CON K-MEANS

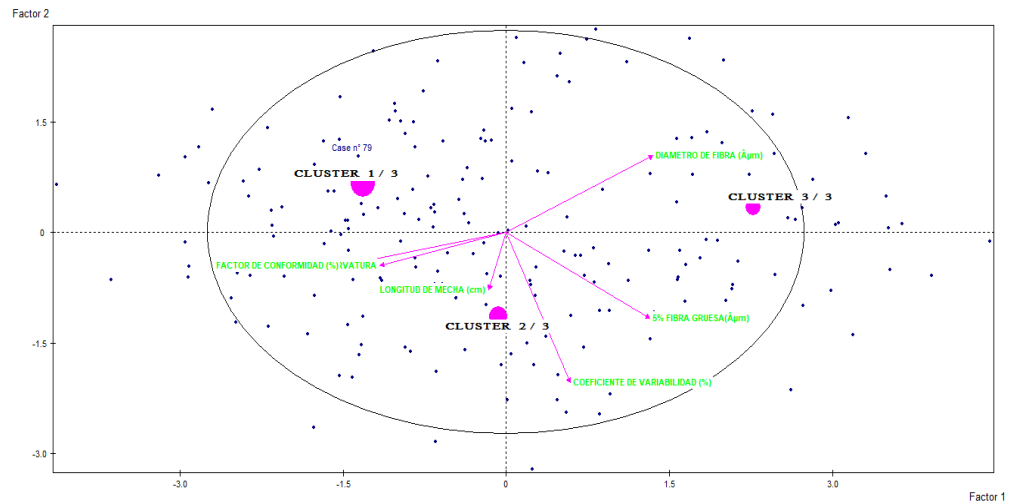


Figura 19. Salida de Individuos en Cada Etapa

Fuente: *Programa Spad N versión 4.5.*

Interpretación, en la figura 18 nos muestra la salida de cada individuo formados en los 3 clúster.

4.9 COMPARACIÓN DE MEDIAS (ACNJ)

Tabla 10

Comparacion De Medias Con El Metodo De Enlace Promedios

Número de caso de clúster		COEFICIENTE DE				INDICE DE CURVATURA	FACTOR DE CONFORMIDAD
		DIAMETRO DE FIBRA	VARIABILIDAD	LONGITUD DE MECHA	FIBRA GRUESAD		
1	Mean	25,	22,	9,5	38,	10,	84,
		6551	7887	189	3645	7377	6449
	N	53	53	53	53	53	53
	Std. Deviation	1,3	2,8	2,5	4,9	1,1	3,1
2	Mean	21,	19,	9,9	48,	8,1	95,
		5774	2681	500	5848	716	9026
	N	88	88	88	88	88	88
	Std. Deviation	2,1	1,8	3,0	7,2	,73	2,9
3	Mean	21,	24,	11,	46,	9,9	93,
		4910	8857	5373	1831	575	8424
	N	67	67	67	67	67	67
	Std. Deviation	1,7	2,5	2,4	5,3	,93	2,7
Total	Mean	22,	21,	10,	45,	9,4	92,
		5886	9747	3514	2070	007	3704
	N	20	20	20	20	20	20
	Std. Deviation	8	8	8	8	8	8
	Std. Deviation	2,5	3,4	2,8	7,3	1,4	5,4
		7707	2052	5902	7613	3540	6845

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

Según el Dendograma (ver Figura 18) del total de 208 individuos (alpacas) como se muestra se toma la decisión de segmentar en tres conglomerados según la distancia mínima de la calidad de la fibra (características físicas de la lana) que más en común tengan entre ellos, por ello se hace una comparación de medias entre los tres clústeres a continuación se detalla la aportación de cada variable en cada clúster:

EL CLÚSTER 1 (color amarillo) los individuos que más se agrupan según las características en común que tienen. La tabla 19 muestra que EL DIÁMETRO DE FIBRA con una media mayor de 25,6551 μm , al del clúster 2 (color verde) y clúster 3 (color azul) con una desviación de 1,37719 y la FIBRA GRUESA con una media de 10,7377 μm con una desviación de 1,18823.

EL CLUSTER 2 (color verde) los individuos que más se agrupan según las características en común que tienen. La tabla 19 muestra que el INDICE DE CURVATURA hace un mayor aporte en un promedio de 48,5848 con una desviación de 7,28357 y el FACTOR DE CONFORMIDAD con un promedio de 95,9026 % y una desviación de 2,96882.

EL CLUSTER 3 (color azul) los individuos que más se agrupan según las características en común que tienen. La tabla 19 muestra que el COEFICIENTE DE VARIABILIDAD hace un mayor aporte en un promedio de 24,8857 % con una desviación de 2,57157 y la LONGITUD DE MECHA hace un mayor aporte en un promedio de 11,5373 cm con una desviación de 2,400314.

4.10 VALIDACIÓN DEL ANÁLISIS CLÚSTER NO JERÁRQUICO

Para validar la segmentación en grupos homogéneos de las alpacas Huacaya según las características físicas de la calidad de la fibra, usando el método de distancia euclídea y el método de enlace de promedios, se plantea la siguiente hipótesis:

H_0 : $\mu_{C1}=\mu_{C2}=\mu_{C3}$ Las medias del clúster son iguales

H_1 : $\mu_{C1}\neq\mu_{C2}\neq\mu_{C3}$ No todas las medias de los clusters son iguales

En la Tabla 11 Análisis de Varianza se puede apreciar para las 6 variables respecto a los 3 clústeres segmentados usando la distancia euclídea y el método de enlace promedios, el p-valor nos permitirá obtener los resultados de contraste,



utilizamos el nivel de error al 5 % ($\alpha=0.05$), se aprecia en la tabla de los 3 clústeres segmentados que el p valor de las 6 variables involucradas en la segmentación a nivel individuos (alpacas Huacaya) son menores al nivel de significancia ($\alpha=0.05$), en conclusión se acepta la Hipótesis alterna no todas las medias de los clústeres son iguales evidencia la heterogeneidad entre sí y la homogeneidad entre ellos.

Tabla 11

Analisis de la Varianza de los 3 Clusters

		ANOVA				
		Sum of Squares	df	Mean Square	F	Sig.
DIÁMETRO DE FIBRA * Número de caso de clúster	Bet (Combined)	669,0	2	334,	97,	,
	ween Groups	74		537	184	000
	Within Groups	705,6	20	3,44		
	Total	75	5	2		
		1374,	20			
		749	7			
COEFICIENTE DE VARIABILIDAD * Número de caso de clúster	Bet (Combined)	1247,	2	623,	108	,
	ween Groups	533		766	,887	000
	Within Groups	1174,	20	5,72		
	Total	353	5	9		
		2421,	20			
		886	7			
LONGITUD DE MECHA * Número de caso de clúster	Bet (Combined)	145,1	2	72,5	9,6	,
	ween Groups	42		71	17	000
	Within Groups	1546,	20	7,54		
	Total	878	5	6		
		1692,	20			
		020	7			
ÍNDICE DE CURVATURA * Número de caso de clúster	Bet (Combined)	3549,	2	1774	47,	,
	ween Groups	292		,646	167	000
	Within Groups	7713,	20	37,6		
	Total	023	5	25		
		1126	20			
		2,315	7			
FIBRA GRUESA * Número de caso de clúster	Bet (Combined)	248,4	2	124,	143	,
	ween Groups	55		228	,039	000
	Within Groups	178,0	20	,868		
	Total	40	5			

	Total	426,4	20		
		95	7		
FACTOR DE	Bet (Combined)	4406,	2203	253	,
CONFORMI	ween Groups	285	,142	,186	000
DAD	* Within Groups	1783,	20	8,70	
Número de		840	5	2	
caso de clúster	Total	6190,	20		
		125	7		

Fuente: Gobierno Regional Puno, PECSA, elaboración en base a los resultados en el software

Spss 19

Diagrama de dispersión de REGR factor score 2 for analysis 1 por REGR factor score 1 for analysis 1 por Número de caso de clúster

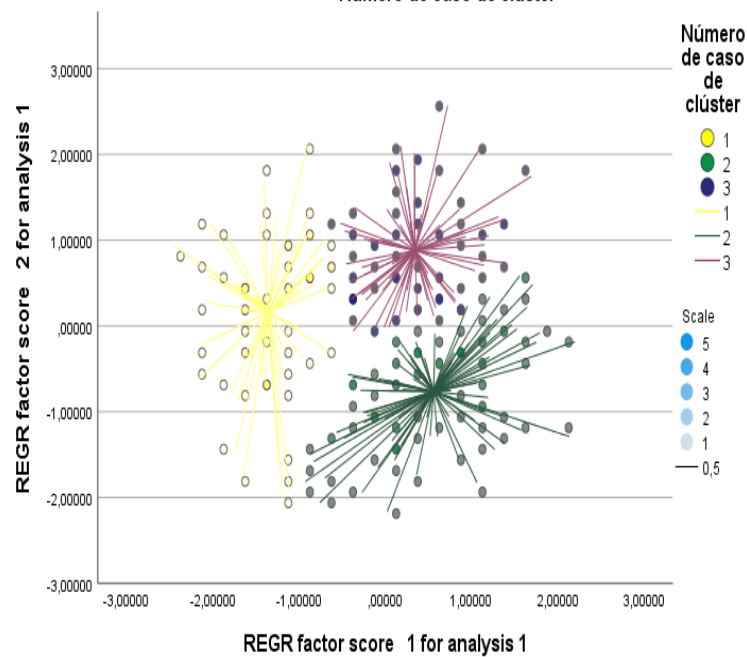


Figura 20. Diagrama De Dispersión Para Cada Caso De Clúster.

Fuente: Resultados Obtenidos En El Software Spss 23

Finalmente, en la Figura 20, nos muestra la asociación de los individuos(alpacas) en cada clúster segmentado.



V. CONCLUSIONES

Las conclusiones según lo desarrollado en el capítulo resultados son los siguientes:

PRIMERO: Las variables consideradas en estudio son diámetro de fibra, coeficiente de variación, longitud de mecha, índice de curvatura, fibra gruesa y factor de conformidad (Características Físicas De La Calidad De La Fibra De La Lana) se eligieron estas variables cuantitativas datos cedidos por el PECSA.

SEGUNDO: Para la determinación de la segmentación primero se analizó el Análisis Clúster Jerárquico para la elección de grupos mediante la obtención del DENDOGRAMA los resultados sugieren que existe una estructura correcta de segmentación en grupos homogéneos entre sí y heterogéneos entre ellos; posteriormente se realiza el Análisis Clúster No Jerárquico (método K-means) para la segmentación a nivel de individuos (alpacas Huacaya) que poseen características similares según las variables (Características Físicas De La Calidad De La Fibra De La Lana), y la aportación de dichas variables en cada grupo esto con la finalidad de seguir manteniendo la calidad de fibra, finalmente se especifica la segmentación de alpacas Huacaya en 3 Clústeres.

TERCERO: Se encontraron tres grandes grupos de Alpacas con diferentes características de medida cada una homogéneos entre sí y heterogéneos entre ellos. Los criterios utilizados para la segmentación de los 208 individuos (Alpacas huacaya) fueron con el método de enlace de promedios y la medida de distancia Euclídea como medida de proximidad y para la validación correspondiente se realiza el análisis de comparación de medias donde El **CLÚSTER 1** conformado por 53 alpacas Huacaya los individuos que más



se agrupan según las características en común que tienen. EL DIÁMETRO DE FIBRA con una media mayor de 25,6551 μm , al del clúster 2 (color verde) y clúster 3 (color azul) con una desviación de 1,37719 y la FIBRA GRUESA con una media de 10,7377 μm con una desviación de 1,18823, **EL CLÚSTER 2** conformado por 88 alpacas Huacaya los individuos que más se agrupan según las características en común que tienen es el INDICE DE CURVATURA hace un mayor aporte en un promedio de 48,5848 con una desviación de 7,28357 y el FACTOR DE CONFORMIDAD con un promedio de 95,9026 % y una desviación de 2,96882, y **EL CLÚSTER 3** conformado por 67 alpacas Huacaya los individuos que más se agrupan según las características en común que tienen es el COEFICIENTE DE VARIABILIDAD que hace un mayor aporte en un promedio de 24,8857 % con una desviación de 2,57157 y la LONGITUD DE MECHA hace un mayor aporte en un promedio de 11,5373 cm con una desviación de 2,400314; Para cada grupo segmentado encontrado se estableció un perfil de acuerdo a sus características más resaltantes, así mismo la validación de la tabla ANOVA evidencia en los resultados la heterogeneidad entre sí y la homogeneidad entre ellos.



VI. RECOMENDACIONES

Según los resultados obtenidos en la siguiente investigación:

PRIMERO: Se recomienda el uso de otros métodos de clasificación que permiten obtener agrupamientos más precisos entre los que se encuentra el análisis de conglomerados bietapico o en dos fases el cual permite detectar y describir subgrupos de sujetos o variables homogéneas en función de los valores observados dentro de un conjunto aparentemente heterogéneo.

SEGUNDO: Se recomienda a las autoridades del gobierno local, instituciones públicas y privadas, ONG, empresa, y sociedad civil deben comprometerse a asegurar y promover programas para el mejoramiento de producción genética de camélidos. Para que de esta manera los productores de la zona tengan un mayor conocimiento sobre la importancia de la fibra de alpaca para un posterior estudio de un análisis de mercado.

TERCERO: Se recomienda incluir las técnicas de clasificación en los programas de estudio en la ingeniería industrial, debido a su importancia y amplio campo de aplicaciones.



VII. REFERENCIAS BIBLIOGRAFICAS

- Apaza Tito, M. (2014). Implementacion de Algoritmos Geneticos Para La Segmentacion de Imagenes Satelitales por Conglomerados de la Region Puno-2013. *Tesis*, 139. http://repositorio.unap.edu.pe/bitstream/handle/UNAP/4875/Apaza_Tito_Meliton.pdf?sequence=1&isAllowed=y
- Bustanza, A. (2001). *La alpaca, conocimiento del gran potencial andino*. Puno. Univ. Nacional del Altiplano.
- Calle Escobar, R. (1982). Produccion y mejoramiento de la produccion de la alpaca. In *Fondo del Libro Banco Agrario del Peru*.
- Canaza, A. W. (2009). *Evaluación Cualitativa y Cuantitativa de la Fibra de Alpaca Mediante Espectroscopia de Reflectancia en el Infrarrojo Cercano (NIRS)*.
- Candia, M. (2014). *Segmentación en base a motivaciones deportivas : caso Energy Fitness Club*. 95.
- CENAGRO. (2012). *IV Censo Nacional Agropecuario*.
- Chura, E. J. (2013). Caracterizacion Molecular y Determinacion de Distancias Geneticas en variedades Nativas y Parientes Silvestres de Quinoa (*Chenopodium quinoa* Willd.) Mediante el Marcador Molecular AFLP. *Tesis*, 165. <http://repositorio.unap.edu.pe/handle/UNAP/253>
- Crispin, M. (2014). Análisis comparativo de la productividad y distribución de fibra de alpaca entre Huancavelica y Puno. *Pensamiento Crítico*, 11, 033. <https://doi.org/10.15381/pc.v11i0.9007>
- De La Fuente Crespo, L. (2015). Analisis Cluster. *Libro*, 84.
- De la Fuente, S. (2008). Análisis de Conglomerados (Análisis Cluster). *Técnicas de Análisis Multivariante de Datos*, 14.
- Diaz Rosas, J. A. (2014). Principales Características de la Fibra de Alpacas Huacaya y Suri



- del sector Chocoquilla - Carabaya. *Tesis*, 1–56.
http://repositorio.unap.edu.pe/bitstream/handle/UNAP/2053/Diaz_Rozas_Jaime_Alain.pdf?sequence=1&isAllowed=y
- Dueñas, M. (2015). *Modelos de respuesta discreta en R y aplicacion con datos reales*.
- FAO. (2005). *Situación actual de los camélidos sudamericanos en Perú*. 63.
- Ferreira, S., Rial, A., & Varela, J. (2010). Segmentación post hoc del mercado turístico español. Aplicación del análisis cluster en dos etapas. *Estudios y Perspectivas En Turismo*, 19(5), 592–606. <http://www.redalyc.org/articulo.oa?id=88302205>
- Fish, V. E., Mahar, T. J., & Crook, B. J. (1999). *Fibre curvature morphometry and measurement. International Wool Textile Organization*. <https://doi.org/NiceMeeting>
- Garnsworthy R.K., Gully R.L., Kandiah R.P., Kenins P., Mayfield R.J. y Westerman, R. A. (1988). Understanding the Causes of Prickle and Itch from Skin Contact of Fabrics, Australiasian Textiles. *Understanding the Causes of Prickle and Itch from Skin Contact of Fabrics, Australiasian Textiles*, 2629.
- Gil Quispe, R. (2017). Evaluación de las características textiles de la fibra de alpacas Huacaya del Instituto de Investigación y Promoción de Camélidos Sudamericanos, Puno. *Repositorio Institucional Digital de La Universidad Nacional Del Altiplano*, 1–67.
<http://repositorio.unap.edu.pe/handle/UNAP/6394>
http://repositorio.unap.edu.pe/bitstream/handle/UNAP/6394/Gil_Quispe_Rubén.pdf?sequence=1&isAllowed=y
- Grishanov, S.A., Harwood, R.J., B. I. (2006). A method of estimating the single flax fibre fineness using data from the LaserScan system. *Industrial Crops and Products* 23, 273–287.



- Gutiérrez, K. D. L. C., Rivasplata, J. C., & Sahuanga, C. F. (2015). Aplicación Del Modelo De Clusterización Basado En El Algoritmo De K-Means Para La Segmentación De La Morbilidad Materna En El Hospital San Bartolomé De La Ciudad De Lima-2012. *Revista de Investigación Business Intelligence*, 2(1), 15–21. https://revistas.upeu.edu.pe/index.php/ri_bi/article/view/901/869
- Hernández, J. (2016). Clustering basado en el algoritmo K-means para la identificación de grupos de pacientes quirúrgicos. *Articulo*, 8. www.udu.edu.co/congreso/historial/.../descargar_pdf1.php?f...%0A
- Icart. (2012). DIAGRAMA DE CAJAS. *Libro*, 30.
- INEI. (2012). IV Censo Nacional. *Articulo*, 47.
- INIA. (2014). Camélidos. *Camelidos Sudamericanos*.
- Juárez, J. M., Díaz, P., Rodríguez, J., Martínez, C. E., Hernández, B., Ramírez, E., Torruco, J. G., & Herman, E. (2016). Caracterización de la Leche y clasificación de calidad mediante análisis Cluster en sistemas de doble propósito. *Revista Mexicana de Ciencias Pecuarias*, 7(4), 525. <https://doi.org/10.22319/rmcp.v7i4.4280>
- Justo, V. (2015). Clusterización y turismo rural vivencial en la comunidad de Luquina Chico-Chucuito. *Tesis*, 108. <http://repositorio.unap.edu.pe/handle/UNAP/13915>
- Lajo, J. J., & Vizcarra, C. A. (2016). Análisis De Preferencia De Servicios De Telefonía Móvil Para Segmentar a Los Clientes Que Usan Smartphone. *Tesis*, 75. <https://repositorio.lamolina.edu.pe>
- McDaniel, C., & Gates, R. (2005). Investigación de Mercados. *Libro*, 6ta edición, 450.
- Ministerio de Agricultura, 2003. (2003). *Ministerio de Agricultura*.
- Miranda Villagomez, F., & Salinas Flores, J. (2012). Estadística General.UNALM. *Revista*, 60.
- Muñoz Garcia, J. A. (2013). Técnicas para Detección de Outliers Multivariantes. *Revista*,



3(5), 11–25.

Perez Lopez, C. (2004). Técnicas de Análisis Multivariante de Datos Aplicaciones con SPSS. *Libro*, 672.

Quispe, E. (2007). *Algunos aspectos de la fibra y peso vivo de alpacas huacaya de color blanco en la Region de Huancavelica. July 2007.*

Quispe Peña, E., Poma Gutierrez, A., & Purroy Unanua, A. (2015). Características Productivas y Textiles de la Fibra de Alpacas de Raza Huacaya. *Revista*, 121(June), 7–32. <https://doi.org/10.5209/rev>

Rodriguez, R. J. (2012). Analisis De Componentes Principales y Analisis De Factores Comunes. *Articulo, Plan 2008*, 28.

Sacchero, D. (2008). *Biotecnología aplicada en camélidos sudamericanos*. 155.

Sampieri, R. (2014). *METODOLOGIA DE INVESTIGACION* (6ta edicio).

Siguayro, R. (2009). *COMPARACIÓN DE LAS CARACTERÍSTICAS FÍSICAS DE LAS FIBRAS DE LA LLAMA CH'AKU (Lama glama) Y LA ALPACA HUACAYA (Lama pacos) DEL CENTRO EXPERIMENTAL QUIMSACHATA DEL INIA - PUNO.*

Solis, H. (1991). *Tecnologías de lana de lanas y fibras de animales especiales*. Facultad de Ciencias Agropecuarias Universidad Nacional Daniel Alcides Carrión.

Tang, F., Vargas, C., & Porras, J. (2016). Segmentacion de clientes de una tienda de electrodomesticos utilizando el analisis de conglomerados. *Tesis*, 53. <http://repositorio.lamolina.edu.pe/handle/UNALM/2214>

Van Schie, H.F.M., Marler, J.W., Barry, L. J. H. (1990). *Measurement of fibre diameter by image analysis*.

Vargas, A. (1995). Estadística Descriptiva e Inferencial. *Libro*, 576.

Wang, X., Wang, L., Liu, X., & Wang, H. (2004). *Reproduced with the kind permission of the copyright owner Copyright: 2004, Australian Alpaca Association.*



ANEXOS

Anexo A Estadísticos Descriptivos del Conjunto de Datos Estandarizados

	Estadísticos Estandarizados				
	N	M ínimo	M áximo	M edia	Desvi ación estándar
Puntuación Z:	2	-	2	,	1,000
DIAMETRO DE FIBRA (μm)	08	2,54110	,25504	0000000	00000
Puntuación Z:	2	-	2	,	1,000
COEFICIENTE DE VARIABILIDAD (%)	08	1,85196	,46318	0000000	00000
Puntuación Z:	2	-	2	,	1,000
LONGITUD DE MECHA (cm)	08	2,22154	,32547	0000000	00000
Puntuación Z:	2	-	2	,	1,000
INDICE DE CURVATURA	08	2,56326	,52884	0000000	00000
Puntuación Z:	2	-	2	,	1,000
5% FIBRA GRUESA (μm)	08	2,55729	,57718	0000000	00000
Puntuación Z:	2	-	1	,	1,000
FACTOR DE CONFORMIDAD (%)	08	2,55838	,39520	0000000	00000
N válido (por lista)	2 08				

Anexo B Grupos formados de los 3 clúster

Número del caso	Clúster	Distancia
1	2	2,524
2	2	3,229
3	2	1,708
4	1	2,399
5	2	1,551
6	1	2,753
7	2	1,270
8	2	1,579



9	2	2,143
10	2	1,347
11	2	2,181
12	1	1,978
13	2	2,005
14	1	1,066
15	2	1,671
16	2	2,428
17	2	,611
18	2	2,300
19	3	2,491
20	1	2,797
21	3	1,825
22	2	1,197
23	2	1,484
24	2	1,720
25	2	1,640
26	2	2,159
27	2	1,181
28	2	1,644
29	2	,865
30	2	1,676
31	2	1,525
32	2	2,409
33	2	1,761
34	2	2,535
35	2	1,605
36	2	1,784
37	2	2,101
38	2	2,099
39	2	1,294
40	2	2,170
41	3	1,392
42	2	2,932
43	1	1,319
44	2	1,716
45	2	2,848
46	2	2,723
47	2	1,554
48	2	1,667
49	3	2,799
50	2	1,483
51	2	1,970
52	2	1,339
53	1	2,420
54	3	1,813



55	2	2,426
56	1	1,508
57	1	1,499
58	1	1,247
59	2	1,930
60	3	1,303
61	2	2,190
62	1	2,881
63	3	1,734
64	2	1,332
65	2	2,316
66	2	1,803
67	2	2,594
68	2	2,024
69	3	1,382
70	3	1,567
71	1	1,231
72	2	1,963
73	1	1,009
74	1	1,497
75	2	2,332
76	2	1,494
77	2	2,342
78	3	1,610
79	2	,708
80	3	,926
81	1	1,589
82	2	2,876
83	2	1,380
84	2	2,351
85	2	,892
86	2	2,902
87	3	1,696
88	3	1,319
89	2	1,375
90	2	1,288
91	3	1,993
92	1	1,101
93	2	2,767
94	1	1,771
95	2	2,657
96	2	2,056
97	1	1,247
98	2	1,954
99	2	1,016
100	2	2,378



101	2	1,034
102	1	2,519
103	2	1,986
104	1	1,282
105	2	1,500
106	3	1,345
107	3	1,785
108	2	1,925
109	2	2,004
110	2	1,662
111	3	,949
112	3	1,449
113	1	1,955
114	1	1,669
115	3	1,850
116	2	1,848
117	1	,644
118	2	1,727
119	2	1,022
120	2	2,586
121	1	,820
122	3	1,204
123	3	1,236
124	2	1,157
125	2	1,237
126	1	1,539
127	2	1,897
128	2	,618
129	3	1,471
130	3	1,529
131	3	1,691
132	1	2,472
133	3	2,619
134	3	1,229
135	3	1,106
136	3	1,085
137	3	1,231
138	1	1,752
139	3	1,205
140	3	1,134
141	1	1,362
142	1	1,478
143	3	,866
144	3	,720
145	3	1,570
146	1	1,870



147	3	2,535
148	3	1,403
149	3	1,682
150	1	1,296
151	3	2,117
152	3	1,690
153	1	1,437
154	3	1,738
155	3	2,324
156	3	1,385
157	1	1,646
158	3	2,805
159	3	2,035
160	3	1,623
161	1	,970
162	2	1,656
163	1	1,114
164	1	1,716
165	1	1,699
166	3	1,932
167	1	1,756
168	1	2,175
169	1	3,170
170	3	1,820
171	1	1,763
172	3	2,222
173	3	1,470
174	1	1,786
175	1	1,524
176	2	2,130
177	1	1,593
178	1	1,360
179	3	1,035
180	3	2,374
181	3	1,788
182	3	1,743
183	2	1,140
184	3	2,821
185	1	2,193
186	3	1,651
187	3	1,891
188	3	,609
189	2	1,880
190	1	2,039
191	3	,965
192	2	1,334



193	1	1,625
194	3	1,506
195	3	1,651
196	3	,612
197	3	2,700
198	1	2,464
199	1	1,742
200	1	1,526
201	2	,953
202	3	1,622
203	3	,748
204	3	1,338
205	3	,988
206	1	1,616
207	3	2,144
208	1	1,509

Anexo C. Historial de conglomerados

Etapa	Clúster combinado		Coeficientes	Primera aparición del clúster de etapa		Etapa siguiente
	Clúster 1	Clúster 2		Clúster 1	Clúster 2	
1	134	204	,099	0	0	48
2	41	60	,153	0	0	10
3	91	98	,170	0	0	82
4	23	48	,180	0	0	87
5	149	160	,188	0	0	59
6	31	47	,190	0	0	19
7	196	203	,193	0	0	109
8	65	96	,197	0	0	72
9	80	88	,227	0	0	43
10	41	70	,229	2	0	26
11	17	29	,239	0	0	40



12	129	192	,240	0	0	69
13	111	143	,263	0	0	24
14	27	124	,266	0	0	38
15	148	156	,274	0	0	56
16	170	194	,276	0	0	92
17	9	15	,287	0	0	42
18	22	64	,304	0	0	84
19	31	52	,315	6	0	54
20	76	90	,330	0	0	55
21	33	50	,333	0	0	89
22	73	121	,338	0	0	50
23	3	13	,352	0	0	88
24	111	139	,374	13	0	36
25	83	89	,378	0	0	103
26	41	69	,396	10	0	43
27	56	92	,403	0	0	136
28	6	45	,407	0	0	190
29	119	205	,414	0	0	69
30	7	85	,430	0	0	126
31	106	112	,438	0	0	46
32	77	84	,442	0	0	158
33	12	114	,456	0	0	60
34	57	157	,457	0	0	108
35	21	78	,461	0	0	96
36	111	191	,467	24	0	80
37	179	188	,478	0	0	66
38	27	183	,479	14	0	118



39	10	66	,511	0	0	123
40	17	39	,516	11	0	105
41	54	63	,528	0	0	104
42	9	30	,534	17	0	65
43	41	80	,544	26	9	96
44	123	125	,561	0	0	48
45	37	44	,570	0	0	123
46	81	106	,576	0	31	101
47	72	108	,589	0	0	116
48	123	134	,607	44	1	109
49	141	163	,607	0	0	78
50	73	104	,611	22	0	74
51	58	161	,611	0	0	97
52	40	116	,615	0	0	85
53	16	20	,615	0	0	114
54	31	101	,618	19	0	88
55	24	76	,619	0	20	94
56	148	152	,634	15	0	90
57	136	145	,647	0	0	95
58	1	11	,649	0	0	172
59	149	193	,659	5	0	95
60	12	94	,661	33	0	148
61	175	177	,663	0	0	108
62	43	74	,682	0	0	112
63	79	128	,683	0	0	84
64	186	202	,685	0	0	99
65	9	59	,704	42	0	150



66	173	179	,706	0	37	129
67	126	199	,707	0	0	127
68	130	131	,712	0	0	129
69	119	129	,720	29	12	83
70	167	190	,743	0	0	127
71	34	95	,756	0	0	114
72	65	105	,759	8	0	130
73	127	162	,760	0	0	113
74	73	97	,774	50	0	143
75	51	100	,784	0	0	141
76	153	208	,791	0	0	102
77	137	144	,808	0	0	137
78	141	200	,821	49	0	152
79	71	117	,836	0	0	112
80	111	195	,861	36	0	106
81	118	174	,864	0	0	119
82	5	91	,885	0	3	153
83	119	201	,886	69	0	128
84	22	79	,894	18	63	105
85	35	40	,902	0	52	122
86	18	28	,914	0	0	177
87	8	23	,923	0	4	117
88	3	31	,936	23	54	133
89	33	110	,960	21	0	116
90	135	148	,980	0	56	135
91	25	26	,986	0	0	122
92	170	181	,987	16	0	162



93	138	142	,994	0	0	138
94	24	189	,997	55	0	118
95	136	149	1,048	57	59	107
96	21	41	1,054	35	43	139
97	14	58	1,064	0	51	136
98	150	178	1,081	0	0	138
99	172	186	1,089	0	64	144
100	159	180	1,101	0	0	140
101	81	182	1,120	46	0	107
102	153	165	1,125	76	0	149
103	83	99	1,133	25	0	126
104	54	140	1,144	41	0	124
105	17	22	1,153	40	84	154
106	111	122	1,178	80	0	137
107	81	136	1,192	101	95	173
108	57	175	1,218	34	61	143
109	123	196	1,219	48	7	128
110	32	103	1,264	0	0	178
111	42	67	1,267	0	0	164
112	43	71	1,327	62	79	134
113	115	127	1,327	0	73	130
114	16	34	1,381	53	71	131
115	19	93	1,418	0	0	159
116	33	72	1,422	89	47	158
117	8	36	1,453	87	0	150
118	24	27	1,465	94	38	151
119	118	207	1,478	81	0	125



120	164	206	1,493	0	0	169
121	46	75	1,521	0	0	160
122	25	35	1,542	91	85	163
123	10	37	1,573	39	45	154
124	54	107	1,577	104	0	139
125	87	118	1,599	0	119	162
126	7	83	1,608	30	103	161
127	126	167	1,617	67	70	149
128	119	123	1,683	83	109	153
129	130	173	1,696	68	66	145
130	65	115	1,696	72	113	176
131	16	62	1,703	114	0	167
132	68	176	1,715	0	0	167
133	3	38	1,739	88	0	151
134	43	171	1,772	112	0	157
135	109	135	1,783	0	90	181
136	14	56	1,792	97	27	170
137	111	137	1,861	106	77	145
138	138	150	1,907	93	98	157
139	21	54	1,966	96	124	168
140	159	187	2,063	100	0	155
141	49	51	2,107	0	75	172
142	4	102	2,111	0	0	180
143	57	73	2,123	108	74	165
144	158	172	2,130	0	99	183
145	111	130	2,166	137	129	173
146	132	168	2,175	0	0	182



147	154	184	2,210	0	0	189
148	12	146	2,253	60	0	171
149	126	153	2,266	127	102	169
150	8	9	2,297	117	65	164
151	3	24	2,299	133	118	163
152	113	141	2,311	0	78	174
153	5	119	2,399	82	128	168
154	10	17	2,474	123	105	161
155	147	159	2,558	0	140	166
156	2	82	2,574	0	0	199
157	43	138	2,576	134	138	182
158	33	77	2,608	116	32	177
159	19	61	2,675	115	0	195
160	46	55	2,734	121	0	202
161	7	10	2,793	126	154	178
162	87	170	2,824	125	92	187
163	3	25	2,826	151	122	187
164	8	42	2,900	150	111	175
165	57	185	3,009	143	0	170
166	147	151	3,055	155	0	183
167	16	68	3,104	131	132	180



168	5	21	3,141	153	139	181
169	126	164	3,201	149	120	184
170	14	57	3,223	136	165	191
171	12	133	3,231	148	0	185
172	1	49	3,232	58	141	196
173	81	111	3,240	107	145	188
174	53	113	3,265	0	152	185
175	8	86	3,454	164	0	196
176	65	120	3,569	130	0	186
177	18	33	3,675	86	158	186
178	7	32	3,686	161	110	192
179	155	166	3,859	0	0	194
180	4	16	3,956	142	167	190
181	5	109	3,966	168	135	188
182	43	132	4,092	157	146	191
183	147	158	4,140	166	144	198
184	126	198	4,302	169	0	193



185	12	53	4,544	171	174	197
186	18	65	4,762	177	176	192
187	3	87	4,833	163	162	200
188	5	81	4,882	181	173	201
189	154	197	5,236	147	0	198
190	4	6	5,278	180	28	200
191	14	43	5,458	170	182	193
192	7	18	5,569	178	186	195
193	14	126	5,688	191	184	197
194	155	169	5,993	179	0	204
195	7	19	6,877	192	159	199
196	1	8	7,195	172	175	201
197	12	14	7,234	185	193	207
198	147	154	7,738	183	189	204
199	2	7	7,759	156	195	202
200	3	4	7,904	187	190	203
201	1	5	8,612	196	188	203



202	2	46	9,246	199	160	205
203	1	3	9,987	201	200	205
204	147	155	10,352	198	194	206
205	1	2	11,017	203	202	206
206	1	147	14,077	205	204	207
207	1	12	17,045	206	197	0