

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



**“MODELOS DE REGRESIÓN LOGÍSTICA MULTINOMIAL DE LA CALIDAD DE  
FIBRA DE ALPACA HUACAYA EN FUNCIÓN DE SUS CARACTERÍSTICAS:  
SEXO Y EDAD - CORANI, CARABAYA, PUNO - 2017”**

**TESIS**

**PRESENTADA POR:**

**Bach. MARIA JUAQUINA ROQUE CRUZ**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**PUNO – PERÚ**

**2018**

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

**“MODELOS DE REGRESIÓN LOGÍSTICA MULTINOMIAL DE LA  
CALIDAD DE FIBRA DE ALPACA HUACAYA EN FUNCIÓN DE  
SUS CARACTERÍSTICAS: SEXO Y EDAD - CORANI, CARABAYA,  
PUNO - 2017”**

**TESIS PRESENTADA POR:**

Bach. MARÍA JUAQUINA ROQUE CRUZ

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

INGENIERO ESTADÍSTICO E INFORMÁTICO



**APROBADO POR EL JURADO REVISOR CONFORMADO POR:**

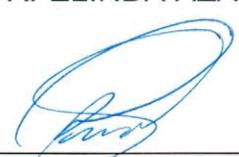
**PRESIDENTE**

:   
Dr. JUAN REYNALDO PAREDES QUISPE

**PRIMER MIEMBRO**

:   
Mg. EMMA ORFELINDA AZAÑERO DE AGUIRRE

**SEGUNDO MIEMBRO**

:   
M. Sc. SAMUEL DONATO PEREZ QUISPE

**DIRECTOR / ASESOR**

:   
Dr. EDGAR ELOY CARPIO VARGAS

**Área** : Estadística

**Tema** : Regresión logística multinomial

**Fecha de sustentación** : 13/07/2018

## DEDICATORIAS

A mis padres Paulino y Estefa por su confianza, comprensión, apoyo y amor incondicional que hicieron posible la culminación de mi carrera profesional.

A Roxana, Maritza, Patricia, Marisol, Alex y Ricardo, mis hermanos por motivarme a cumplir con cada objetivo trazado y estar siempre pendientes de mí.

A Miguel por su amor incondicional e infinita comprensión que me hacen mejorar como persona, agradezco haberte conocido.

A mi amiga yesi que es como mi hermana por brindarme su confianza y siempre impulsándome a continuar con mi formación personal y profesional.

## AGRADECIMIENTOS

A mis padres y hermanos que con paciencia y comprensión hicieron realidad este anhelado sueño.

A la facultad de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano – Puno, donde forman profesionales de alto nivel competitivo y capaces de afrontar retos satisfactoriamente.

Mi más sincero agradecimiento al Dr. Edgar Eloy Carpio Vargas que con sus ideas y ganas de apoyar a los estudiantes no solo dentro de las aulas si no también fuera de ellas de forma incondicional fomentando la unión y apoyo de toda la familia que conforman la Facultad de Ingeniería Estadística e Informática.

Un agradecimiento muy especial a la Mg. Emma Orfelinda Azañero de Aguirre por preocuparse de mi formación integra como profesional.

A mis jurados Dr. Juan Reynaldo Paredes Quispe y M. Sc. Samuel Donato Pérez Quispe que aportaron a través de sus observaciones y con su dirección hicieron posible la conclusión de la presente investigación.

## ÍNDICE GENERAL

|  |           |
|--|-----------|
| <b>RESUMEN.....</b>  | <b>12</b> |
| <b>ABSTRACT .....</b>  | <b>13</b> |
| <b>CAPITULO I INTRODUCCIÓN .....</b>   | <b>14</b> |
| 1.1. PLANTEAMIENTO DEL PROBLEMA .....  | 15        |
| 1.2. JUSTIFICACIÓN DE LA INVESTIGACIÓN .....   | 18        |
| 1.3. OBJETIVOS DE LA INVESTIGACIÓN .....   | 19        |
| 1.3.1. OBJETIVO GENERAL.....   | 19        |
| 1.3.2. OBJETIVOS ESPECÍFICOS.....  | 19        |
| 1.4. HIPÓTESIS DE LA INVESTIGACIÓN .....   | 20        |
| 1.4.1. HIPÓTESIS GENERAL .....   | 20        |
| 1.4.2. HIPÓTESIS ESPECÍFICAS .....   | 20        |
| <b>CAPITULO II REVISIÓN DE LITERATURA .....</b>  | <b>21</b> |
| 2.1. ANTECEDENTES DE LA INVESTIGACIÓN .....  | 21        |
| 2.2. MARCO TEÓRICO .....   | 27        |
| 2.2.1. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL .....                                 | 27        |
| 2.2.2. FORMULACIÓN E INTERPRETACIÓN DEL MODELO .....                                   | 29        |
| 2.2.3. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD .....                                       | 34        |
| 2.2.4. INFERENCIA EN REGRESIÓN LOGÍSTICA MULTINOMIAL .....                             | 46        |
| 2.3. MARCO CONCEPTUAL .....  | 54        |
| 2.3.1. ALPACA .....  | 58        |
| 2.3.2. FIBRA DE ALPACA .....   | 60        |
| 2.3.3. CARACTERÍSTICAS DE LA FIBRA DE ALPACA.....                                      | 61        |
| <b>CAPITULO III MATERIALES Y MÉTODOS .....</b>   | <b>73</b> |
| 3.1. MÉTODO DE INVESTIGACIÓN .....   | 73        |
| 3.2. DETERMINACIÓN DE LA POBLACIÓN Y MUESTRA.....                                      | 74        |
| 3.3. ÁMBITO DE ESTUDIO .....   | 74        |
| 3.4. PERTINENCIA DE LA METODOLOGÍA DE INVESTIGACIÓN .....                              | 75        |
| 3.4.1. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES DEFINIDAS .....                           | 75        |
| 3.4.2. METODOLOGÍA PARA DETERMINAR LA DIFERENCIA SIGNIFICATIVA<br>ENTRE FACTORES ..... | 76        |

|  |            |
|--|------------|
| 3.4.3. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL, FACTORES ASOCIADOS A LA CALIDAD DE FIBRA DE ALPACA HUACAYA. ....                         | 79         |
| 3.5. OPERACIONALIZACIÓN DE VARIABLES .....   | 84         |
| <b>CAPITULO IV RESULTADOS Y DISCUSIÓN.....</b>   | <b>86</b>  |
| 4.1. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES.....  | 86         |
| 4.2. ANÁLISIS DE VARIANZA PARA LOS FACTORES SEXO Y EDAD. ....  | 89         |
| 4.2.1. EFECTOS DE LA EDAD Y SEXO SOBRE EL DIÁMETRO DE FIBRA. ....  | 89         |
| 4.3. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA LA CALIDAD DE FIBRA DE ALPACA HUACAYA EN FUNCIÓN A SUS CARACTERÍSTICAS SEGÚN SEXO..... | 90         |
| 4.3.1. ESTIMACIÓN E INTERPRETACIÓN DEL MODELO PARA HEMBRAS.....  | 90         |
| 4.3.2. ESTIMACIÓN E INTERPRETACIÓN DEL MODELO PARA MACHOS.....   | 93         |
| 4.4. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA LA CALIDAD DE FIBRA DE ALPACA HUACAYA EN FUNCIÓN A SUS CARACTERÍSTICAS SEGÚN EDAD..... | 96         |
| 4.4.1. ESTIMACIÓN DEL MODELO PARA DIENTES DE LECHE.....  | 97         |
| 4.4.2. ESTIMACIÓN DEL MODELO PARA 2 DIENTES .....  | 99         |
| 4.4.3. ESTIMACIÓN DEL MODELO PARA 4 DIENTES .....  | 102        |
| 4.4.4. ESTIMACIÓN DEL MODELO PARA BOCA LLENA .....   | 106        |
| 4.5. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA DETERMINAR LA CALIDAD DE FIBRA DE ALPACA HUCAYA .....                                  | 109        |
| 4.5.1. ESTIMACIÓN DEL MODELO E INTERPRETACIÓN DEL MODELO .....   | 110        |
| 4.5.2. BONDAD DEL AJUSTE.....  | 110        |
| 4.5.3. CALIDAD DE AJUSTE DEL MODELO .....  | 111        |
| 4.5.4. VALIDACIÓN Y PREDICCIÓN DE LOS MODELOS.....   | 112        |
| <b>CAPITULO V CONCLUSIONES.....</b>  | <b>114</b> |
| <b>CAPITULO VI RECOMENDACIONES .....</b>   | <b>116</b> |
| <b>CAPITULO VII REFERENCIAS.....</b>   | <b>118</b> |
| <b>ANEXOS .....</b>  | <b>124</b> |

**ÍNDICE DE FIGURAS**

|   |    |
|---|----|
| Figura 1: Población mundial de alpacas.....   | 59 |
| Figura 2: Población Nacional de Alpacas, 2015 .....   | 60 |
| Figura 3: Interacción entre el tejido, los terminales de la fibra o lana y la piel, que muestran la importancia del factor de confort ..... | 64 |
| Figura 4: Representación bidimensional de la forma de una fibra de lana .....   | 66 |
| Figura 5: Población de Alpacas en el distrito de Corani. ....   | 74 |
| Figura 6: Ubicación del distrito de Corani.....   | 75 |
| Figura 7: Histograma de factor de confort .....   | 87 |
| Figura 8: Histograma de corrección de Índice de Curvatura .....   | 88 |
| Figura 9: Histograma de corrección de longitud de mecha .....   | 88 |

**ÍNDICE DE TABLAS**

|   |     |
|---|-----|
| Tabla 1: Operacionalización de variables.....   | 84  |
| Tabla 2: Cantidad y porcentaje de alpacas por comunidad.....  | 86  |
| Tabla 3: Estadísticos descriptivos de las características de la fibra de alpaca de raza Huacaya .....           | 87  |
| Tabla 4: Análisis de varianza según edad y sexo.....  | 90  |
| Tabla 5: Estimación de los parámetros, hembras. ....  | 91  |
| Tabla 6: Estadísticos sobre la bondad de ajuste del modelo para hembras .....                                   | 92  |
| Tabla 7 Matriz de clasificación observados-predichos para el modelo estimado, para hembras.....                 | 92  |
| Tabla 8: Coeficientes de medición de ajuste Pseudo R-cuadrado, para hembras. ....                               | 93  |
| Tabla 9: Estimación de los parámetros, machos.....  | 94  |
| Tabla 10: Estadísticos sobre la bondad de ajuste del modelo según sexo, alpacas machos.....                     | 95  |
| Tabla 11: Matriz de clasificación observados-predichos para el modelo estimado según sexo, alpacas machos ..... | 95  |
| Tabla 12: Coeficientes de medición de ajuste Pseudo R-cuadrado, para machos. ....                               | 96  |
| Tabla 13: Estimación de los parámetros, Dientes de leche.....   | 97  |
| Tabla 14: Estadísticos sobre la bondad de ajuste del modelo según edad, DL ...                                  | 98  |
| Tabla 15: Matriz de clasificación observados-predichos para el modelo estimado, según edad DL .....             | 98  |
| Tabla 16: Coeficientes de medición, ajuste Pseudo R-cuadrado, según edad DL. ....                               | 99  |
| Tabla 17: Estimación de los parámetros para el modelo según edad, 2D. ....                                      | 100 |

|  |     |
|--|-----|
| Tabla 18: Estadísticos sobre la bondad de ajuste del modelo según edad, 2D .                           | 101 |
| Tabla 19: Matriz de clasificación observados-predichos para el modelo estimado,<br>Según edad 2D.....  | 101 |
| Tabla 20: Coeficientes de medición, ajuste Pseudo R-cuadrado, para la edad 2D.<br>.....                | 102 |
| Tabla 21: Estimación de los parámetros, para la edad 4D.....   | 103 |
| Tabla 22: Estadísticos sobre la bondad de ajuste del modelo según edad 4D. .                           | 104 |
| Tabla 23: Matriz de clasificación observados-predichos para el modelo estimado,<br>Según edad 4D.....  | 105 |
| Tabla 24: Coeficientes de medición, ajuste Pseudo R-cuadrado, según edad 4D.<br>.....                  | 105 |
| Tabla 25: Estimación de los parámetros según edad BLL. ....  | 106 |
| Tabla 26: Estadísticos sobre la bondad de ajuste del modelo según edad BLL                             | 107 |
| Tabla 27: Matriz de clasificación observados-predichos para el modelo estimado,<br>Según edad BLL..... | 108 |
| Tabla 28: Coeficientes de medición de ajuste Pseudo R-cuadrado para la edad BLL.<br>.....              | 109 |
| Tabla 29: Estimación de los parámetros para el modelo general.....                                     | 109 |
| Tabla 30: Estadísticos sobre la bondad de ajuste del modelo general.....                               | 110 |
| Tabla 31: Matriz de clasificación para el modelo general estimado. ....                                | 111 |
| Tabla 32: Coeficiente de medición de ajuste pseudo R-cuadrado para el modelo<br>general .....          | 111 |
| Tabla 33: Estimación de las probabilidades para cada modelo general.....                               | 112 |

**ÍNDICE DE ACRÓNIMOS**

|          |  |
|----------|--|
| FAO      | Organización de las Naciones Unidas para la Agricultura y la Alimentación                      |
| INIA     | Instituto Nacional de Innovación Agraria   |
| CENAGRO  | Censo Nacional Agropecuario  |
| MINAGRI  | Ministerio de Agricultura y Riego  |
| ADEVA    | Análisis de Varianza   |
| MDF      | Media de Diámetro de Fibra   |
| LM       | Longitud de Mecha  |
| FC       | Factor de Confort  |
| IC       | Índice de Curvatura  |
| PECSA    | Proyecto Especial Camélidos Sudamericanos  |
| FH       | Finura al Hilado   |
| INDECOPI | Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual |
| OFDA     | Optical Fibre Diameter Analyser  |

## RESUMEN

La crianza de alpacas constituye una actividad económica importante para el poblador alto andino debido a las condiciones climáticas y adaptación de este animal a condiciones extremas de frío, por otra parte, las familias están organizadas en comunidades bajo un sistema de crianza de rebaño mixto familiar (alpacas, llamas y ovinos), es por esta razón que se planteó el siguiente objetivo: Determinar los modelos de regresión logística multinomial para la calidad de fibra de alpaca Huacaya en función a sus características del distrito de Corani, en el estudio se consideró tres categorías que son el tipo de fibra Baby, Fleece y mediana, las características fueron: Factor de Confort, Índice de Curvatura y Longitud de Mecha, tomando muestras de fibra de alpaca de la raza Huacaya que fueron analizadas mediante el OFDA 2000 por el Proyecto Especial Camélidos Sudamericanos, la muestra está constituida por 1526 registros de las comunidades de Quelcaya y Chimboya pertenecientes al distrito de Corani. Como resultado se obtuvieron los siguientes modelos para los registros en conjunto  $g_1(x) = -142.321 + 1.536 * FC$  para la calidad Baby y  $g_2(x) = -47.578 + 0.554 * FC$  para la calidad Fleece, según los resultados la variable que en mayor medida permite clasificar a una determinada alpaca con la calidad de fibra Baby, Fleece y media es el Factor de Confort, exceptuando el caso para las alpacas de 4 dientes que en el modelo consideramos el Factor de Confort e Índice de Curvatura, indicando que a medida que el factor de confort incrementa es 4.647 veces más probable que la calidad sea Baby a que la calidad sea Fleece, con una tasa de clasificación del 88.4% indicando que el modelo posee una buena predicción. La probabilidad de que la calidad de fibra sea Baby va incrementando a medida que el Factor de confort es más alto.

**Palabras Clave:** Regresión logística, Calidad, Fibra, Características.

## ABSTRACT

The alpacas breeding is an important economic activity for the Andean high population due to climatic conditions and adaptation of this animal to extreme cold conditions, another hand, families are organized in communities under a breeding mixed family herd system ( alpacas, llamas and sheeps), it is for this reason that the following objective was established: To determine the multinomial logistic regression models for the quality of Huacaya alpaca fiber according to its characteristics of the Corani district, in the study three categories were considered type of fiber Baby, Fleece and medium, the characteristics were: Comfort Factor, Curvature Index and Wick Length, taking samples of alpaca fiber of the Huacaya breed that were analyzed by OFDA 2000 by the Special Project of South American Camelids, the sample is constituted by 1526 records of the communities of Quelcaya and Chimboya belonging it is to the district of Corani. As a result, the following models were obtained for the records as a whole  $g_1(x) = -142.321 + 1.536 * FC$  for the quality Baby and  $g_2(x) = -47.578 + 0.554 * FC$  for the Fleece quality, depending on the results the variable that to a greater extent it allows to classify a certain alpaca with the quality of Baby fiber, Fleece and average is the Comfort Factor, except for the case of the alpacas of 4 teeth that in the model we consider the Comfort Factor and Curvature Index, indicating that as the comfort factor increases, it is 4.647 times more likely that the quality is Baby than the quality is Fleece, with a classification rate of 88.4%, indicating that the model has a good prediction. The probability that the fiber quality is Baby is increasing as the comfort factor is higher.

**Key Words:** Logistic regression, Quality, Fiber, Characteristics.

## CAPITULO I

### INTRODUCCIÓN

A nivel nacional se cuenta con 3,978,290 alpacas, de estos el 47.27% pertenece a la Región Puno, el 78.98% es de raza Huacaya (AGROPUNO, 2013), distribuida a nivel de la sierra con pequeños productores cuyo 80% cuenta con un promedio de 50 cabezas de alpacas por familia (INFOALPACAS, 2012).

La actividad pecuaria de camélidos es principalmente desarrollada por comunidades campesinas donde se encuentra la población más pobre del país su actividad principal es la ganadería, siendo la agricultura la actividad complementaria, sin embargo, el sistema de crianza continua es tradicional, por lo tanto los índices productivos y reproductivos no son competitivos ni generan ingresos atractivos para los criadores, por lo tanto resulta siendo una actividad de subsistencia (Huanca, 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 altoandinos 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.

En la actualidad existen proyectos que se encargan específicamente al apoyo en la crianza de alpacas en la Región Puno sin embargo aún no se cuenta con estudios que puedan hacer un análisis de las características de la fibra en alpacas Huacaya con respecto a la longitud de mecha, factor de confort e índice de curvatura, estas variables son de mucha importancia desde el punto de vista de la industria textil, teniendo en cuenta que la calidad de la fibra producida ha decrecido notoriamente, trayendo como consecuencia un progresivo deterioro en su precio y por consiguiente, en los ingresos de los criadores altoandinos (Ormachea & Olarte, 2010).

En este sentido en el presente estudio se planteó los siguientes objetivos: Determinar el modelo de regresión logística multinomial para la calidad de fibra de alpaca Huacaya en función a sus características del distrito de Corani, 2017, de igual manera se planteó determinar la calidad de fibra de alpaca Huacaya en función a la longitud de mecha, índices de curvatura y factor de confort y, según el sexo y edad. El tipo de estudio es de relación entre una variable categórica (dependiente) y variables continuas (independientes), cuya metodología es el análisis de regresión logística multinomial para poder alcanzar los objetivos de la investigación.

### **1.1. PLANTEAMIENTO DEL PROBLEMA**

La región de Puno, donde se realiza este trabajo, que cuenta con la mayor población de cabezas de alpaca siendo fuente principal de ingresos de los productores alpaqueros es la comercialización de la fibra de alpaca sin

previa clasificación y poco conocimiento de sus características. Parte de las limitaciones de los productores alpaqueros es la clasificación de su fibra de alpaca en Baby, Fleece, mediana fleece, huarizo, gruesa y corta (INDECOPI, 2010) para ello es necesario encontrar un modelo de regresión logística multinomial para determinar la calidad de la fibra.

La crianza de alpaca constituye una importante actividad económica del poblador alto andino, las familias están organizadas en comunidades, bajo un sistema de crianza de rebaño mixto familiar (alpacas, llamas y ovinos). Se estima más de 60 comunidades alpaqueras, que agrupan a 3300 familias aproximadamente a nivel regional; en el distrito de Corani la población alpaquera está distribuida principalmente en cinco comunidades estas son: Chimboya, Quelcaya, Chacaconiza, Isivilla y Corani Aconsaya, para este estudio se considera los distritos de Chimboya y Quelcaya que en total tienen 1526 alpacas de raza Huacaya de color blanco.

El sexo influye categóricamente sobre el diámetro de fibra, lo cual se debe a que las hembras tienen requerimientos nutricionales más altos por las diferentes condiciones fisiológicas difíciles que pasan (lactación y preñez) las cuales tienen impacto en el perfil de diámetro de fibra. Los machos poseen mayor diámetro de fibra en comparación con las hembras, esto probablemente se debe a factores medio ecológicos, grado de mejoramiento genético, por la técnica usada para la determinación del diámetro de fibra. El diámetro de fibra se incrementa significativamente conforme a la edad del animal obteniendo los siguientes valores:  $19.6\mu$ ,  $21.07\mu$  y  $22.28\mu$  en alpacas

de la categoría dos, cuatro y boca llena respectivamente (Ormachea et al., 2013).

Para que el proceso de comercialización de la fibra sea el más adecuado y que los productores puedan tener conocimiento sobre cuales características son las que indican que la fibra de alpaca es buena según su sexo y edad considerando sus características se realizó este proyecto de forma que los criterios de venta de los productores sean los adecuados y así poder elevar su precio base, para la clasificación de la fibra es necesario que los productores tomen en cuenta sus características por la edad y sexo.

Por ello, la presente investigación pretende analizar las características de la fibra de alpaca de raza Huacaya color blanco como son; factor de confort (FC), índice de curvatura (IC) y longitud de mecha (LM) según sexo y grupo etario de acuerdo a la dentadura, comenzando por recoger muestras de fibra y posteriormente analizarlas en los laboratorios de Gobierno Regional del Proyecto Especial Camélidos Sudamericanos (PECSA), para poder ver si existe diferencias de las medias del diámetro de fibra según sexo y edad y llegar a modelos de regresión logística multinomial para cada caso.

Así ante lo expuesto se responderá a lo siguiente:

### **Pregunta general**

¿Cuáles son los modelos que explican la calidad de fibra de alpaca de raza Huacaya en función a sus características del distrito de Corani, 2017?

### **Preguntas específicas**

¿Qué modelos explican la calidad de fibra de alpaca de raza Huacaya en función a sus características según sexo del distrito de Corani, 2017?

¿Qué modelos explican la calidad de fibra de alpaca de raza Huacaya en función a sus características según edad del distrito de Corani, 2017?

## **1.2. JUSTIFICACIÓN DE LA INVESTIGACIÓN**

El presente trabajo de investigación es necesario debido a que en la actualidad no se cuenta con estudios que analicen la calidad de fibra de alpaca ya constituye la principal actividad económica, el sistema de comercialización de la fibra de alpaca ha sido siempre un problema debido a la participación de los famosos “compadres” (intermediarios que compran fibra a bajos precios) y, por otro lado, por el bajo nivel de organización de los productores locales que no permite negociar la venta de fibra en mejores condiciones, la venta de fibra de alpaca sin clasificar puede llegar a representar el 30% de su valor comercial real, ya que la fibra no se categoriza ni se clasifica antes de ser comercializada (Fernández, 2015).

Los productores alpaqueros al poder clasificar la fibra y tener precedentes para poder predecir qué calidad pudiera tener a medida que va transcurriendo los años, ya que no sería necesario el análisis de los vellones cuyo costo es de 50 soles por vellón y si los productores tendrían conocimiento de que calidad tendrá para los posteriores años, reducirían costos, así mismo se podría proporcionar a los acopiadores un producto clasificado como

también puede hacer más simple el proceso de clasificación para las maestras clasificadoras.

Por las razones expuestas se ha creído conveniente realiza el presente estudio con la finalidad de ayudar a establecer las bases para una mejor negociación con los acopiadores y así permitir mejorar los ingresos con fibras clasificadas, con lo cual mediante un análisis de regresión logística multinomial se determinará la calidad de la fibra de alpaca según sexo y edad en las comunidades de Quelcaya y Chimboya del distrito de Corani.

### **1.3. OBJETIVOS DE LA INVESTIGACIÓN**

#### **1.3.1. OBJETIVO GENERAL**

Determinar los modelos de regresión logística multinomial para la calidad de fibra de alpaca Huacaya en función a sus características del distrito de Corani,2017.

#### **1.3.2. OBJETIVOS ESPECÍFICOS**

- Explicar los modelos de regresión logística multinomial para la calidad de fibra de alpaca Huacaya en función a sus características según sexo del distrito de Corani, 2017.
- Explicar los modelos de regresión logística multinomial para la calidad de fibra de alpaca Huacaya en función a sus características según edad del distrito de Corani, 2017.

## **1.4. HIPÓTESIS DE LA INVESTIGACIÓN**

### **1.4.1. HIPÓTESIS GENERAL**

Los modelos de regresión logística multinomial explican adecuadamente la calidad de fibra de alpaca Huacaya.

### **1.4.2. HIPÓTESIS ESPECÍFICAS**

- Los modelos de regresión logística multinomial explican adecuadamente la calidad de fibra de alpaca Huacaya por sexo.
- Los modelos de regresión logística multinomial explican adecuadamente la calidad de fibra de alpaca Huacaya por edad.

## CAPITULO II

### REVISIÓN DE LITERATURA

#### 2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Como antecedentes existen investigaciones que se relacionan con el análisis de las características de la fibra de alpaca y otras investigaciones que se realizan aplicando la regresión logística multinomial.

Aucancela (2015), tiene como objetivo general Caracterizar la fibra de alpaca (*Vicugna pacos*) de la Provincia de Chimborazo, para lo cual se llevó a cabo una ADEVA multifactorial donde a cada muestra se determinó la longitud absoluta , longitud relativa , numero de rizos, diámetro y tasa de medulación, con estos factores se evaluó el efecto del sexo, localización y edad, donde nos muestra que el sexo y edad presentaron efectos sobre el diámetro, longitud absoluta, longitud relativa y número de rizos mas no en la medulación y que la comunidad de origen de los animales afectó significativamente la calidad de la fibra únicamente en hembras, en este estudio se puede apreciar que las fibras más gruesas corresponden a hembras adultas y las fibras más finas corresponden a machos jóvenes con 23.17  $\mu\text{m}$  y 13.60 $\mu\text{m}$  respectivamente.

Vásquez, Gómez, & Quispe (2015), busco estimar las características tecnológicas de la fibra de alpaca Huacaya color blanco en una comunidad altoandina de Apurimac, según sexo y grupo etario para lo cual las variables se evaluaron a través de un análisis de varianza con un diseño completamente al azar, en donde los resultados obtenidos fueron que el diámetro de fibra es menor en machos que en hembras en estos valores pueden incrementar conforme incrementa la edad en los grupos etarios con una probabilidad menor del 0.05 la cual indica diferencia significativa entre las variables, la finura en promedio en machos es de 19.6  $\mu\text{m}$  y en hembras es de 20.1  $\mu\text{m}$  con estos valores se indicó que se tiene buena calidad de fibra en esta zona del país, en cuanto a la variación del diámetro de fibra no existe diferencias significativas tanto en sexo y edad, en el índice de confort si se pudo apreciar diferencia entre sexos y grupos etarios, el índice de curvatura indica diferencia según grupo etario mas no en los sexos y por último la finura al hilado fue diferente entre sexos y grupos etarios.

Sangurima & Sandoya (2015), el estudio “Uso de regresión multinomial logística para determinar el perfil del estudiante de la ESPOL, según preferencias alimenticias” se propone discutir, que de algunas de las variables utilizadas mediante una encuesta realizada a los estudiantes de la ESPOL, cuales verdaderamente influyen en el rendimiento académico del estudiante. En primer lugar se realizó el análisis univariado, para luego continuar con el análisis multivariado: en donde se hace un análisis de tablas de contingencia, llegando al análisis de regresión logística multinomial haciendo una tabla de bondad de ajuste con distribuciones chi-cuadrado y ajustar el modelo con las variables que contribuyen, para llegar a las conclusiones: De acuerdo a las

enfermedades que mayormente padecen los estudiantes se encontró, la gastritis con un 41.5% quienes la padecen y las afecciones a la garganta, así lo muestran los resultados con un 78.5% de los estudiantes registrados en el primer término del 2015, el análisis de Regresión Multinomial Logística es una herramienta estadística de mejor capacidad cuando se requiere de una variable de respuesta de más de dos categorías, en estudios como el presente es bastante útil pues, en este caso se tenía variables como rendimiento académico, promedio, etc. Se concluyó mediante este análisis que las variables significativas y que aportaron en gran medida para determinar el perfil del estudiante según preferencias alimenticias, fueron: carrera, edad, si trabajan o no, como consideraban la alimentación para el desempeño de su rol de estudiante, si consumían carne roja y cuantas veces comía el estudiante en el día; descartando el resto de variables por no aportar gran influencia en el modelo según una significancia 0.05.

Rozas (2014), en su trabajo de investigación tiene como objetivo determinar el diámetro de fibra, finura del hilado, índice de curvatura y factor de confort en función al lugar de procedencia, sexo y raza de las alpacas para ello la metodología empleada fue de un análisis experimental con diseño completamente al azar y el coeficiente de correlación de Pearson. Los resultados demostraron que tanto el sexo como el lugar de procedencia no influyen en el diámetro de fibra, las alpacas Huacaya presentan menor finura en comparación con alpacas Suri, en cuanto al factor de confort las alpacas de raza Huacaya brindan mayor factor de confort con 98.76% en comparación a las alpacas de raza Suri con 95.58% al igual que el índice de curvatura con 41.47 grad/mm en Huacaya y 18.14 grad/mm en Suris, el diámetro de fibra y

el factor de confort tienen correlación negativa en Huacaya ( $r = -0.88895$ ) y en Suris ( $r = -0.888895$ ).

Mamani (2012), cuyo objetivo fue determinar las características físicas de la fibra de alpaca Huacaya, según su edad y sexo. Para ello se analizó un diseño completamente al azar con un arreglo factorial de  $2 \times 4$ , los resultados encontrados fueron que el diámetro de fibra de las hembras fue de  $22.55 \pm 2.73 \mu\text{m}$  y para machos  $23.45 \pm 2.73 \mu\text{m}$  lo cual muestra que no existe diferencia estadística, en cuanto a las edades si existe diferencias altamente significativas, el menor diámetro de fibra encontrado fue en alpacas dientes de leche con  $22.87 \pm 2.61 \mu\text{m}$  y el mayor pertenece a las alpacas boca llena con  $24.74 \pm 2.86$ , para la variable densidad de fibra se encontró que diferencia estadística entre hembras y machos, para la edad las alpacas de dientes de leche son las que se muestran más densas, en la variable número de rizos en 2 cm de longitud de mecha no se encontró diferencia significativa para hembras y machos, en cuanto a la edad fue mayor en alpacas de dos dientes y en las otras edades fue igual, la variable longitud de mecha no se encuentra diferencia significativa según sexo, en cuanto a la edad las de dos dientes son las de mayor longitud seguido de las alpacas de cuatro dientes, al último se encuentran las alpacas de dientes de leche que poseen menor longitud de mecha.

Ormachea & Olarte (2010), el objetivo de este estudio fue determinar las características las características textiles de la fibra como el diámetro de fibra, índice de confort y el índice de curvatura de fibra de alpacas Huacaya de dos comunidades de Carabaya, cuyo análisis estadístico fue conducido en un

diseño completo al azar bajo un arreglo factorial de  $2 \times 2 \times 3$  y la comparación de promedios fue mediante la prueba de significación múltiple de Duncan con  $\alpha=0.05$ . Los resultados obtenidos fueron que el factor sexo no influye en el diámetro de fibra,  $20.69 \pm 2.69 \mu\text{m}$  en hembras y  $21.28 \pm 2.55 \mu\text{m}$  en machos, el factor de confort en alpacas según sus edades si cuenta con diferencia significativa cuyos valores resultantes fueron 97.50%, 95.85% y 93.43 % en alpacas de dos, tres y cuatro años de edad, y el factor de confort en hembras es superior con respecto a machos y por último para el índice de curvatura tanto el sexo como la edad no presentan diferencias significativas.

Sánchez & Gómez (2008), el estudio “Modelacion logística multinomial para clasificar los hogares de el salvador por nivel de pobreza”, el objetivo principal fue de clasificar los hogares salvadoreños, utilizando un análisis de regresión logística multinomial. Realizo la inferencia en regresión logística multinomial hallando los intervalos de confianza al 95% para el valor exponente base para finalmente interpretar el modelo con los Odd\_ratios de cada variable, llegando a las siguientes conclusiones: el modelo de regresión logística multinomial permite relacionar la variable tipo de pobreza del hogar con las variables referidas al hogar su jefatura como lo son: la Región, el Área, el Ingreso Familiar, el Número de Miembros del hogar, si posee electricidad el hogar, si posee lavadora, el Sexo del Jefe(a) de Hogar, el estado Familiar del Jefe(a) de Hogar y la Edad del Jefe(a) de Hogar en la Encuesta de Hogares de Propósitos Múltiples del año 2010, Las variables que reducen tanto la probabilidad de que el hogar sea clasificado como “Pobre Relativo” y la probabilidad de que el hogar sea clasificado como “Pobre Extremo” son: el Ingreso Familiar (un mayor ingreso familiar aumenta la probabilidad de que el

hogar sea “No Pobre”), también si el hogar posee Energía Eléctrica y si el hogar posee Lavadora. La Regresión Logística Multinomial es una de las herramientas estadísticas con mejor capacidad para el análisis, cuando la variable dependiente es categórica, sirve para determinar los factores de riesgo y factores de prevención frente a la situación de pobreza que enfrentan los hogares. En esta investigación se obtiene un modelo que cumple con los supuestos requeridos y que cumple con todos los test estadísticos requeridos.

Quispe (2007), en su investigación tiene por objetivo caracterizar, relacionar e identificar los efectos de la edad, localidad y sexo sobre el peso vivo, diámetro de fibra y peso vivo al nacimiento, mediante utilizando un modelo aditivo lineal de efectos principales. Para la caracterización de las variables se utilizó estadística descriptiva, para evaluar el efecto se realizó el análisis de varianza y prueba de medias; y para determinar la relación se hizo uso del coeficiente de correlación de Pearson y el coeficiente de correlación parcial. Los resultados obtenidos fueron que cuando se realiza una comparación a mayor edad mayor peso vivo, con respecto al diámetro de fibra de los animales de dientes de leche tienen mejor finura mientras que las hembras exhiben fibra más gruesa que esta entre  $22.63 \mu\text{m}$  y  $21.68 \mu\text{m}$  contra los machos que su fibra esta entre  $20.52 \mu\text{m}$  -  $21.76 \mu\text{m}$  esto muestra que tanto el sexo como la edad tienen efectos altamente significativos esto con respecto al diámetro de fibra comportándose similar con el peso vivo, sobre el peso vivo al nacer, el sexo no tuvo efecto. En cuanto a la correlación se encontró que existe una correlación baja entre el peso vivo y el diámetro de fibra que las variables intervinientes Sexo, Edad y Locación, y se obtiene una correlación parcial, resulta que no existe relación entre dichas variables, por

lo tanto, cuando se haga necesario modelar para fines de evaluación genética, será necesario considerar como factores fijos a las variables Sexo, Edad y Locación, además de otras que es necesario investigar.

## **2.2. MARCO TEÓRICO**

### **2.2.1. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL**

Los modelos de elección dicotómica o binaria se pueden generalizar para el caso de más de dos opciones, dando origen a los Modelos de Respuesta Múltiple, donde se generaliza el proceso de elección de tal forma que el agente económico se enfrenta a varias alternativas posibles (Gómez & Palacios, 2013). Los modelos de regresión logística son modelos estadísticos en los que se pretende conocer la relación entre una variable dependiente cualitativa, dicotómica (regresión logística binaria o binomial) o con más de dos categorías (regresión logística multinomial) y entre variables explicativas independientes, que pueden ser cualitativas o cuantitativas (Pando Fernández & Fernández San Martín, 2004).

Los modelos de regresión logística son modelos estadísticos en los que se pretende conocer la relación entre una variable dependiente cualitativa, dicotómica (regresión logística binaria o binomial) y entre variables explicativas independientes, que pueden ser cualitativas o cuantitativas. Las covariables cualitativas que sean dicotómicas, es aconsejable que se codifiquen tomando valores 0 para una de las categorías o para su ausencia, y 1 para la otra categoría o para su presencia (esta codificación es importante ya que cualquier otra codificación podría provocar modificaciones en la interpretación del modelo). Pero si la covariable cualitativa tuviera más de dos

categorías, se realiza una transformación, para poderla incluir en el modelo. Esta transformación consiste en crear varias variables cualitativas dicotómicas ficticias o de diseño, llamadas variables dummies, de forma que una de las variables se tomaría como referencia y cada una de las variables creadas entraría en el modelo de forma individual. En general, si la covariable cualitativa posee  $c$  categorías, habrá que realizar  $c-1$  covariables ficticias.

La regresión logística multinomial es utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica. Las variables independientes pueden ser tanto continuas (covariables) como categóricas o mayormente factores (Gonzales, 2015).

Tradicionalmente las variables dependientes politómicas han sido modeladas mediante análisis discriminante pero, gracias al creciente desarrollo de las técnicas de cálculo, cada vez es más habitual el uso de modelos de regresión logística multinomial (Gómez & Palacios, 2013), implementados en paquetes estadísticos, debido a la mejor interpretabilidad de los resultados que proporciona. La regresión logística multinomial es utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica. Las variables independientes pueden ser tanto continuas (covariables) como categóricas (factores).

Estos modelos se analizan eligiendo una categoría como referencia de la variable respuesta y se modelan varias ecuaciones simultáneamente, una para cada una de las categorías respecto a la de referencia.

### 2.2.2. FORMULACIÓN E INTERPRETACIÓN DEL MODELO

En el modelo de regresión logística se codifican los valores de la variable dependiente como 0 y 1, lo que da como resultado que la media de la variable represente la proporción de casos que ocurren en una de sus dos categorías (en el caso binomial) o en una de sus múltiples categorías (en el caso multinomial). El valor predicho de la probabilidad por el modelo según la categoría puede ser interpretado como la probabilidad de que un caso caiga en esa categoría. Un modelo lineal no se ajusta apropiadamente a variables binomiales, dado que los valores predichos de la variable dependiente con este modelo (ajustados mediante la ecuación de una recta), pueden tomar valores imposibles de probabilidad mayores que 1 o menores que 0, a pesar de que los valores observados estén entre 0 y 1. La misma situación se extiende a variables multinomiales. El mejor modelo que linealiza la relación entre variable dependiente e independiente es el modelo logit, construido a través de regresión logística (según libro de Debella-Gilo et al., 2007). En una variable dependiente binomial ( $Y=0$ ;  $Y=1$ ), si se conoce la probabilidad de pertenecer a una clase ( $Y=0$ ), se puede conocer la probabilidad de pertenecer a la otra clase ( $Y=1$ ), es decir:

$$P(Y = 1) = [1 - P(Y = 0)]$$

Se puede tratar de aplicar el modelo lineal de probabilidad, expresado como:

$$P(Y = 1) = \beta_0 + \beta_1 X$$

Donde:

$P(Y = 1)$ , es la probabilidad asociada a la variable predicha ( $Y$ ),

X, es la variable independiente o predictora.

$\beta$ , son los parámetros de la población a ser estimados.

Este modelo de probabilidad presenta el problema de no linealidad, con valores predichos que pueden ser menores que cero o mayores que uno.

#### 2.2.2.1. Formulación del modelo logístico multinomial

Para los modelos de regresión logística binaria, si tenemos una variable dependiente Y, que toma valores Y=1 (presencia de una característica u otra categoría de la variable) y Y=0 (ausencia de la característica o la otra categoría de la variable), la ecuación de partida del modelo viene dada por:

$$P[Y = 1|X] = \frac{\exp(b_0 + \sum_{s=1}^n b_s x_s)}{1 + \exp(b_0 + \sum_{s=1}^n b_s x_s)}$$

donde  $P[Y=1|X]$  es la probabilidad de que, Y tome el valor 1, en presencia de las covariables X, que se denota por  $p(X)$ .

X es un conjunto de n covariables  $\{x_1, 2, \dots, x_n\}$  que forman parte del modelo;  $b_0$  es la constante del modelo o término independiente y los  $b_i$  son los coeficientes de las covariables.

Esta ecuación inicial del modelo es de tipo exponencial, pero se realiza su transformación logarítmica (logit), dada por:

$$\ln \left[ \frac{p(x)}{1 - p(x)} \right] = b_0 + \sum_{s=1}^n b_s x_s$$

De esta forma, se permite su uso como una función lineal y es de más fácil interpretación. Para el caso de que la variable dependiente presente más de dos categorías, se utiliza el modelo de regresión logística multinomial que se modela, como se indicó anteriormente, mediante varios logit simultáneamente, uno para cada una de las restantes categorías respecto a la categoría de referencia que se haya considerado de la variable dependiente.

Consideremos una variable de respuesta politómica  $Y$  con más de dos categorías de respuesta que denotaremos por  $Y_1, Y_2, \dots, Y_k$ .

Se pretende explicar la probabilidad de cada categoría de respuesta en función de un conjunto de covariables  $X = \{x_1, x_2, \dots, x_n\}$  observadas. Es decir, ajustar un modelo de la forma  $p_j(x) = P[Y = Y_j | X = x] = f_j(x) \forall j = 1, \dots, k$ . Para cada vector  $x$  de valores observados de las variables explicativas  $X$ .

En el caso de una variable de respuesta binaria, su distribución condicionada a cada combinación de valores observados de las covariables sigue una Bernouilli.

Cuando la variable de respuesta es politómica, la distribución de Bernouilli se convierte en una distribución multinomial de parámetros las probabilidades de cada una de las categorías de respuesta. Es decir,  $(Y|X = x) \rightarrow M(1; p_1(x), \dots, p_k(x))$ , siendo  $\sum_{j=1}^k p_j(x) = 1$

Así que para obtener un modelo lineal, obtendremos  $\binom{k}{2}$  transformaciones logit para comparar cada par de categorías de la variable respuesta, que sería de este tipo:

$$\ln \left[ \frac{\frac{p_i(x)}{p_i(x) + p_j(x)}}{\frac{p_j(x)}{p_i(x) + p_j(x)}} \right] = \ln \left[ \frac{p_i(x)}{p_j(x)} \right], \quad \forall i, j = 1, \dots, k (i \neq j)$$

Que representan el logaritmo de la ventaja de respuesta  $Y_i$  frente a  $Y_j$  condicionado a las observaciones de las variables independientes que caen en uno de ambos niveles. Pero para construir el modelo logit de respuesta multinomial bastaría con considerar  $(k - 1)$  transformaciones logit básicas, definidas con respecto a una categoría de referencia. Tomando como categoría de referencia la última  $Y_k$ . Así las transformaciones logit generalizadas se definen como  $L_j(x) = \ln \left[ \frac{p_j(x)}{p_k(x)} \right] \forall j=1, \dots, k-1$ , siendo  $L_j(x)$  el logaritmo de la ventaja de respuesta  $Y_j$  dado que las observaciones de las variables independientes caen en la categoría  $Y_j$  o en la  $Y_k$ .

El modelo lineal para cada una de las transformaciones logit generalizadas, para  $n$  variables explicativas, es de la siguiente forma:

$$L_j(x) = \sum_{s=0}^n b_{sj} x_s = x' b_j \quad \forall j = 1, \dots, k - 1$$

para cada vector de valores observados de las variables explicativas  $x=(x_0, x_1, x_2, \dots, x_n)'$  con  $x_0=1$  y  $b_j=(b_{0j}, b_{1j}, \dots, b_{nj})'$  el vector de parámetros asociado a la categoría  $Y_j$ .

Para las probabilidades de respuesta, podemos escribir el modelo de la siguiente forma:

$$p_j(x) = \frac{\exp(\sum_{s=0}^n b_{sj} x_s)}{1 + \sum_{j=1}^{k-1} \exp(\sum_{s=0}^n b_{sj} x_s)} \quad \forall j = 1, \dots, k - 1$$

$$p_k(x) = \frac{1}{1 + \sum_{j=1}^{k-1} \exp(\sum_{s=0}^n b_{sj}x_s)}$$

Siendo  $b_{sk}=0 \forall s=0, 1, \dots, n$

#### 2.2.2.2. Interpretación del modelo

A continuación, mostramos la interpretación de los parámetros del modelo, pero distinguiendo los casos según de qué tipo son las variables explicativas, cuantitativas o cualitativas.

##### - Una variable predictora cuantitativa X.

Si en el modelo tenemos sólo una única covariable cuantitativa X, el modelo para cada valor observado x de la variable X viene dado por:

$$L_j(x) = a_j + b_j x, \forall j = 1, \dots, k-1$$

A continuación, mostramos la exponencial de los parámetros  $b_j$  asociados a cada categoría de la variable dependiente, que se interpreta en términos de cocientes de ventajas (odds ratio):

$$\theta_j(\Delta X = 1) = \frac{\frac{p_i(x+1)}{p_i(x)}}{\frac{p_j(x)}{p_k(x)}} = \frac{\exp(a_j + b_j(x+1))}{\exp(a_j + b_j x)} = \exp(b_j) \quad \forall j = 1, \dots, k-1$$

$(\Delta X=1)$  es el cociente de ventajas de respuesta  $Y_j$  frente a la última categoría,  $Y_k$  cuando aumenta en una unidad la variable X.

### - Más de una variable predictora cuantitativa.

Para el modelo logit generalizado múltiple, los cocientes de ventajas se definen incrementando una de las variables y controlando fijas las demás.

$$\theta_j(\Delta X_r = 1 / X_s = x_s, s \neq r) = \frac{\frac{P[Y=Y_j / X_r=x_1+1, X_s=x_s, s \neq r]}{P[Y=Y_k / X_r=x_1+1, X_s=x_s, s \neq r]}}{\frac{P[Y=Y_j / X_r=x_1+1, X_s=x_s, s \neq r]}{P[Y=Y_k / X_r=x_1+1, X_s=x_s, s \neq r]}} = \exp(b_{rj})$$

$$\forall j = 1, \dots, k-1$$

siendo  $\theta_j(\Delta X_r=1 / X_s= x_s, s \neq r)$  el cociente de ventajas de respuesta  $Y_j$  frente a la última categoría,  $Y_k$  cuando aumenta en una unidad la variable  $X_r$  y las demás se controlan fijas.

### 2.2.3. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD

Para la estimación de los coeficientes del modelo y de sus errores estándar se utiliza la estimación por máxima verosimilitud, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente y proporcionados por los datos de la muestra. Al contrario de lo que ocurre con la estimación de los coeficientes de regresión lineal múltiple que se utiliza el método de los mínimos cuadrados, los cálculos para las estimaciones de los coeficientes de la regresión logística multinomial no son directos, hay que llevar a cabo métodos iterativos, como el método de Newton–Raphson.

Al aplicar estos métodos además de obtener las estimaciones de los coeficientes de regresión, se obtienen sus errores estándar y las covarianzas entre las covariables del modelo. Se describe el método de estimación de

máxima verosimilitud para el cálculo de los coeficientes del modelo de regresión logística multinomial.

Supongamos que se compara muestra aleatoria de tamaño  $N$  con  $Q$  combinaciones diferentes de valores de las variables explicativas  $X_1, \dots, X_n$ . Se denota a cada combinación de valores de las variables explicativas por  $x_q = (x_{q0}, x_{q1}, \dots, x_{qn})'$  con  $x_{q0} = 1 \quad \forall \quad q=1, \dots, Q$ . En cada una de estas combinaciones se tiene una muestra aleatoria de  $d_q$  observaciones independientes de la variable de respuesta politémica  $Y$ , de entre las cuales denotamos por  $y_{j/q}$  al número de observaciones que caen en la categoría de respuesta  $Y_j \quad \forall \quad j = 1, \dots, k$ .

Así que, se verifica que  $\sum_{j=1}^k y_{j/q} = d_q$  y  $\sum_{q=1}^Q d_q = N$ .

Los vectores  $(y_{1/q}, \dots, y_{k/q})' \quad \forall \quad q = 1, \dots, Q$  siguen una distribución de probabilidad multinomiales independientes,  $M(d_q; p_{1/q}, \dots, p_{k/q})$ , siendo  $p_{j/q} = P[Y=Y_j/X=x_q]$  y verificando que  $\sum_{j=1}^k y_{j/q} = d_q$

Por tanto, la función de verosimilitud de los datos viene dada por:

$$V = \prod_{q=1}^Q \left( \frac{d_q!}{\prod_{j=1}^k (y_{j/q})!} \prod_{j=1}^k p_{j/q}^{y_{j/q}} \right)$$

Así que el núcleo de la log-verosimilitud es:  $K = \sum_{q=1}^Q \sum_{j=1}^k y_{j/q} \ln(p_{j/q})$

Normalmente, en vez de utilizar la función de verosimilitud se utiliza la siguiente función auxiliar:

$$\Lambda = -2 \ln(V)$$

Por lo que el problema de maximizar la verosimilitud equivale al de minimizar esta función auxiliar.

Teniendo en cuenta la ecuación del modelo logit generalizado multinomial, y sustituyendo en la expresión anterior, se obtiene la siguiente expresión del núcleo de la log-verosimilitud:

$$K = \sum_{q=1}^Q \sum_{j=1}^k y_{j/q} \left( \sum_{s=0}^n b_{sj} x_{qs} \right) - \sum_{q=1}^Q \left( \sum_{j=1}^k y_{j/q} \right) \ln \left( \sum_{j=1}^k \exp \left( \sum_{s=0}^n b_{sj} x_{qs} \right) \right)$$

$$= \sum_{q=1}^Q \sum_{j=1}^k y_{j/q} \left( \sum_{s=0}^n b_{sj} x_{qs} \right) - \sum_{q=1}^Q n_q \ln \left( \sum_{j=1}^k \exp \left( \sum_{s=0}^n b_{sj} x_{qs} \right) \right)$$

Derivando respecto de los parámetros tenemos:

$$\frac{\Delta K}{b_{sj}} = \sum_{q=1}^Q y_{jq} x_{qs} - \sum_{q=1}^Q n_q x_{qs} \frac{\exp(\sum_{s=0}^n b_{sj} x_{qs})}{\sum_{j=1}^k \exp(\sum_{s=0}^n b_{sj} x_{qs})}$$

Así, se obtiene las ecuaciones de verosimilitud con forma matricial:

$$X'_{((n+1) \times Q)} Y_{j(Q \times 1)} = X'_{((n+1) \times Q)} \hat{m}_{j(Q \times 1)} \quad \forall j=1, \dots, -1,$$

siendo  $y_j = (y_{j/1}, \dots, y_{j/q})'$  y  $\hat{m}_j = (\hat{m}_{j/1}, \dots, \hat{m}_{j/q})'$  con  $\hat{m}_{j/q}$  la frecuencia esperada de respuesta  $Y_j$  en la combinación  $xq$  de valores observados de las variables predictoras, estimada bajo el modelo y definida como  $\hat{m}_{j/q} = d_q \hat{p}_{j/q}$ .

Para obtener los estimadores de máxima verosimilitud hay que resolver k-1 sistemas de n+1 ecuaciones no lineales. Así que para resolverlo se utilizara el método iterativo de Newton-Raphson.

Con este método se obtendrá el estimador de los parámetros, que es una matriz de dimensión  $(n+1) \times (k-1)$  formado por las siguientes columnas:

$\hat{b} = (\hat{b}'_1, \hat{b}'_2, \dots, \hat{b}'_{k-1})'$  siendo  $\hat{b}'_j$  el estimador de máxima verosimilitud del vector de parámetros asociado a la categoría de la variable respuesta  $Y_j$ .

Se obtendrá la matriz de covarianzas, que es la inversa de la matriz de información de Fisher. La matriz de covarianzas de cada vector de parámetros  $\hat{b}'_j$ . Para ello hay que calcular las derivadas segundas de K con  $r \neq s$ :

$$\frac{\Delta^2 K}{\Delta b_{rj} \Delta b_{sj}} = \sum_{q=1}^Q n_q x_{qs} x_{qr} \frac{\exp(\sum_{s=0}^n b_{sj} x_{qs}) [\sum_{j=1}^k \exp(\sum_{s=0}^n b_{sj} x_{qs}) - \exp(\sum_{s=0}^n b_{sj} x_{qs})]}{[\sum_{j=1}^k \exp(\sum_{s=0}^n b_{sj} x_{qs})]^2}$$

Así que la matriz de covarianzas viene dada por:

$$\text{cov}(\hat{b}_j) = \left[ -E \left( \frac{\Delta^2 K}{\Delta b_{rj} \Delta b_{sj}} \right) \right]^{-1} = [X' \text{diag}[d_q p_{j/q} (1 - p_{j/q})] X]^{-1}$$

Las matrices de covarianzas cruzadas entre cada par de estimadores  $\hat{b}_j$  y  $\hat{b}_i$  ( $i \neq j$ ). Para ello se calcula las siguientes derivadas segundas de K con  $r \neq s$  y  $j \neq i$ .

$$\frac{\Delta^2 K}{\Delta b_{ri} \Delta b_{sj}} = \sum_{q=1}^Q n_q x_{qs} x_{qr} \frac{-\exp(\sum_{s=0}^n b_{sj} x_{qs}) \exp(\sum_{s=0}^n b_{si} x_{qs})}{[\sum_{j=1}^k \exp(\sum_{s=0}^n b_{sj} x_{qs})]^2}$$

Dando lugar a la siguiente expresión de la matriz de covarianzas:

$$\text{cov}(\hat{b}_j, \hat{b}_i) = \left[ -E \left( \frac{\Delta^2 K}{\Delta b_{ri} \Delta b_{sj}} \right) \right]^{-1} = [-X' \text{diag}[d_q p_{j/q} p_{i/q}] X]^{-1}$$

Finalmente, se tiene que la matriz de covarianzas del estimador es:

$$\text{cov}(\hat{b}) = \begin{pmatrix} \text{cov}(\hat{b}_1) & \text{cov}(\hat{b}_1, \hat{b}_2) & \dots & \text{cov}(\hat{b}_1, \hat{b}_{k-1}) \\ \text{cov}(\hat{b}_1, \hat{b}_2) & \text{cov}(\hat{b}_2) & \dots & \text{cov}(\hat{b}_2, \hat{b}_{k-1}) \\ \dots & \dots & \dots & \dots \\ \text{cov}(\hat{b}_1, \hat{b}_{k-1}) & \text{cov}(\hat{b}_2, \hat{b}_{k-1}) & \dots & \text{cov}(\hat{b}_{k-1}) \end{pmatrix}$$

### 2.2.3.1. Bondad de ajuste del modelo

Uno de los primeros indicadores de importancia para apreciar el ajuste del modelo logístico multinomial es el doble logaritmo del estadístico de verosimilitud (likelihood), que se verá posteriormente. Se trata de un estadístico que sigue una distribución similar a  $\chi^2$ .

Sea  $y_{j/q}$  el número de observaciones que caen en la categoría de respuesta  $Y_j \forall j = 1, \dots, k$ , y sean las  $d_q$  observaciones correspondientes a la  $q$ -ésima combinación de valores de las variables explicativas.

Denotamos por  $\hat{m}_{j/q}$  la frecuencia esperada de respuesta  $Y_j$  en la combinación  $x_q$  de valores observados de las variables predictoras, estimada bajo el modelo y definida como  $\hat{m}_{j/q} = d_q \hat{p}_{1/q}$ .

Para contrastar la bondad del ajuste global del modelo cuando el número de observaciones en cada combinación de valores de las variables explicativas es grande se utiliza el estadístico chi-cuadrado de Pearson y el estadístico de Wilks de razón de verosimilitudes.

El test global de bondad de ajuste del modelo de regresión logística multinomial múltiple contrasta la siguiente hipótesis:

$$H_0: p_{j/q} = \frac{\exp(\sum_{s=0}^n b_{sj} x_{qs})}{1 + \exp(\sum_{s=0}^n b_{sj} x_{qs})} \quad \forall q = 1, \dots, Q; \quad \forall j = 1, \dots, k$$

$$H_1: p_{j/q} \neq \frac{\exp(\sum_{s=0}^n b_{sj} x_{qs})}{1 + \exp(\sum_{s=0}^n b_{sj} x_{qs})} \quad \text{para algún } q \text{ y } j.$$

### a. Test chi-cuadrado de Pearson

El estadístico chi-cuadrado de Pearson de bondad de ajuste a un modelo de regresión logística multinomial, M de la forma anterior viene dado por:

$$X^2(M) = \sum_{q=1}^Q \sum_{j=1}^k \frac{(y_{j/q} d_q \hat{p}_{j/q})^2}{d_q \hat{p}_{j/q}}$$

siendo  $\hat{p}_{j/q}$  la estimación por máxima verosimilitud de  $p_{j/q}$ .

Este estadístico tiene distribución asintótica chi-cuadrado con grados de libertad obtenidos como la diferencia entre el número de parámetros  $p_{j/q}$  y el número de parámetros independientes en el modelo,  $Q \cdot (n+1) \cdot (k-1)$ . Es decir,  $X^2(M) \xrightarrow{d} \chi^2_{Q \cdot (n+1) \cdot (k-1)}$ , si  $dq \rightarrow \infty$ .

Así que se rechaza la hipótesis nula con un nivel de significación  $\alpha$  cuando  $X^2(M)_{obs} \geq X^2_{Q \cdot (n+1) \cdot (k-1); \alpha}$ . O equivalentemente podemos definir el p-valor del contraste como la probabilidad acumulada a la derecha del valor observado:  $p\text{-valor} = P[X^2(M) \geq X^2(M)_{obs}]$ , se rechaza la hipótesis nula cuando  $p\text{-valor} \leq \alpha$ .

**b. Test chi-cuadrado de razón de verosimilitudes. Estadístico de Wilks. Devianza**

El estadístico de Wilks de razón de verosimilitudes para el contraste de bondad de ajuste del modelo de regresión logística multinomial M se obtiene como menos dos veces el logaritmo del cociente entre el supremo de la verosimilitud bajo la hipótesis nula y el supremo de la verosimilitud en la población. A partir de esta expresión operando se obtiene la expresión de este estadístico que viene dada por:

$$G^2(M) = 2 \left[ \sum_{q=1}^Q \sum_{j=1}^k y_{j/q} \ln \left( \frac{y_{j/q}}{\hat{m}_{j/q}} \right) \right]$$

Este estadístico tiene distribución asintótica chi-cuadrado con grados de libertad la diferencia entre la dimensión del espacio paramétrico y la dimensión de este espacio bajo la hipótesis nula. Para un modelo de regresión logística multinomial los grados de libertad es la diferencia entre el número de parámetros  $p_{1/q}$  y el número de parámetros  $b_{sj}$  bajo el modelo, es decir,  $Q-(n+1) \times (k-1)$  grados de libertad  $G^2(M)^d \rightarrow X_{Q-(n+1) \times (k-1)}^2$ , si  $dq \rightarrow \infty$ .

Así que se rechaza la hipótesis nula con un nivel de significación  $\alpha$  cuando

$$G^2(M)_{obs} \geq X_{Q-(n+1) \times (k-1); \alpha}^2. \text{ O equivalentemente cuando } p$$

$$p - \text{valor} = p[G^2(M) \geq G^2(M)_{obs}] \leq \alpha.$$

Al estadístico de Wilk,  $G^2(M)$ , se le denomina devianza.

### 2.2.3.2. Calidad de ajuste

Además de los contrastes mencionados anteriormente, podemos calcular otras medidas que nos dan información sobre la calidad del modelo, al igual que en la regresión lineal que se utiliza la medida de  $R^2$ .

En los modelos de regresión logística binaria, la calidad del ajuste se mide mediante coeficientes de determinación conocidos como Pseudo- $R^2$ , para la regresión logística multinomial también se utilizan estos coeficientes. De entre todos los que existen, los más usados son el de Mc-Fadden, el de Cox-Snell y el de Nagelkerke.

#### a. Coeficiente pseudo- $R^2$ de Mc-Fadden.

Si se tiene  $\Lambda = -2\ln(V)$ , se identifica  $\Lambda_0$  que es el valor inicial de esta función, es decir el mínimo  $\Lambda$  bajo el modelo nulo dado sólo por un término constante y por  $\Lambda_f$  el mínimo de  $\Lambda$  bajo el modelo ajustado con todos los parámetros, obtenemos la siguiente expresión del pseudo- $R^2$  de Mc-Fadden:

$$R_{MF}^2 = 1 - \frac{\Lambda_f}{\Lambda_0}.$$

Siendo su rango teórico de valores  $0 \leq R_{MF}^2 \leq 1$ , pero muy raramente su valor se aproxima a 1. Suele considerarse una buena calidad del ajuste cuando  $0.1 \leq R_{MF}^2 \leq 0.4$  y excelente para valores superiores.

#### b. Coeficiente pseudo- $R^2$ de Cox-Snell.

En este caso se utiliza directamente la función de verosimilitud  $V$ , y no la función auxiliar  $\Lambda$ . Por lo que se denota por  $V_0 = \exp(-\Lambda_0/2)$  el máximo de

verosimilitud bajo el modelo nulo dado sólo por un término constante y por  $V_f = \exp(-\Lambda_f/2)$  el máximo de verosimilitud bajo el modelo ajustado con todos los parámetros, se define el coeficiente pseudo- $R^2$  de Cox-Snell como:

$$R_{CS}^2 = 1 - \left(\frac{V_0}{V_f}\right)^{\frac{2}{N}} = 1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right)$$

El rango teórico de valores para el coeficiente es  $0 \leq R_{CS}^2 \leq 1$ , lo que le hace poco interpretable al depender de  $V_0$ . Ya que puede ser próximo a cero cuando hay pocos datos. Por ello es preferible utilizar el siguiente coeficiente como medida de bondad de ajuste.

**c. Coeficiente pseudo- $R^2$  de Nagelkerke.**

Viene dado por la siguiente expresión:

$$R_N^2 = \frac{R_{CS}^2}{1 - V_0^{\frac{2}{N}}} = \frac{1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right)}{1 - \exp\left(\frac{-\Lambda_0}{N}\right)}$$

Y en este caso, su rango de valores es  $0 \leq R_N^2 \leq 1$ , por lo que puede interpretarse del mismo modo que el coeficiente de determinación de la regresión lineal clásica, aunque es más difícil que alcance valores cercanos a 1.

Para comparar modelos de regresión logística multinomial con diferente número de variables predictoras suelen introducirse coeficientes Pseudo- $R^2$  ajustados. El más conocido es el de Mc-Fadden, definido como:

$$Adj - R_{MF}^2 = 1 - \frac{0.5\Lambda_f + n + 1}{0.5\Lambda_0 + 1}, \text{ siendo } n \text{ el número de variables predictoras.}$$

### **2.2.3.3. Tasa de clasificaciones correctas**

Para cuantificar la bondad del ajuste global del modelo se dispone también de otra medida como es la tasa de clasificaciones correctas. Es decir, a partir del modelo ajustado, se clasifica cada observación en la categoría más probable, construyendo así una matriz de clasificación observados-predichos y se utiliza el porcentaje de clasificaciones correctas como una medida de la calidad de predicción, del mismo modo que se hace en el análisis discriminante. Se define como la proporción de individuos clasificados correctamente por el modelo y se calcula como el cociente entre el número de observaciones clasificadas correctamente y el tamaño muestral  $N$ . Un individuo es clasificado correctamente por el modelo cuando su valor observado de la variable respuesta  $Y$  ( $Y_1, Y_2, \dots, Y_k$ ) coincide con su valor estimado por el modelo.

### **2.2.3.4. Contrastes sobre los parámetros del modelo**

Además de construir el modelo y ajustarlo y habiendo obtenido las estimaciones, lo siguiente será comprobar la significación estadística de cada uno de los coeficientes de regresión en el modelo. Para ello se pueden emplear básicamente dos métodos para los modelos de regresión logística multinomial: el estadístico de Wald y el estadístico condicional de razón de verosimilitud.

Así que se plantea contrastar si un subconjunto de los parámetros del modelo de regresión logística multinomial, que se demuestra por  $b=(b_1, \dots, b_r)'$ , es nulo. Así que se plantea el contraste de hipótesis:

$$H_0: b=0$$

$$H_1: b \neq 0$$

Se ve dos tipos de contrastes mencionados anteriormente que se utiliza para contrastar esta hipótesis.

**a. Contrastes de Wald.**

Se basan en la normalidad asintótica de los estimadores de máxima verosimilitud.

El estimador de máxima verosimilitud de  $b$ ,  $\hat{b}$ , tiene distribución normal asintótica de media  $b$  y matriz de covarianzas estimada  $\widehat{cov}(\hat{b})$  obtenida a partir de la matriz de covarianza  $cov(\hat{b})$ . Así que el estadístico de Wald presenta la forma cuadrática:  $\hat{b}'[\widehat{cov}(\hat{b})]^{-1}\hat{b}$ , que tiene distribución chi-cuadrado asintótica con  $r$  grados de libertad (número de parámetros nulos bajo la hipótesis nula).

Así que se rechaza la hipótesis nula al nivel de significación  $\alpha$  cuando el valor observado de este estadístico sea mayor o igual que el cuantil de orden  $(1 - \alpha)$  de la distribución  $\chi_r^2$ .

Su valor para un coeficiente concreto viene dado por el cociente entre el valor del coeficiente y su correspondiente error estándar. Es decir, si se quiere contrastar:

$$H_0: b_{sj}=0$$

$$H_1: b_{sj} \neq 0$$

el estadístico será:  $w = \frac{\hat{b}_{sj}^2}{\hat{\sigma}^2(\hat{b}_{sj})}$ , que tiene distribución chi-cuadrado asintótica con un grado de libertad. Así que se rechaza la hipótesis nula con nivel de confianza  $1-\alpha$  si  $W_{Obs} \geq \chi_{1;\alpha}^2$ ;

Es decir, la obtención de significación indica que dicho coeficiente es diferente de 0 y merece la pena su conservación en el modelo.

En modelos con errores estándar grandes, el estadístico de Wald puede proporcionar falsas ausencias de significación. Tampoco es recomendable su uso si se están empleando variables de diseño. En estos casos se recomienda el uso del test de razón de verosimilitudes.

#### **b. Contrastes condicionales de razón de verosimilitud**

Se trata de ir contrastando cada modelo que surge de eliminar de forma aislada cada una de las covariables frente al modelo completo. La ausencia de significación implica que el modelo sin la covariable no empeora respecto al modelo completo (es decir, da igual su presencia o su ausencia), por lo que según la estrategia de obtención del modelo más reducido (principio de parsimonia), dicha covariable debe ser eliminada del modelo ya que no aporta nada al mismo.

Se tiene un modelo de regresión logística multinomial MG que se ajusta bien y se desea contrastar si un subconjunto de parámetros,  $b=(b_1, \dots, b_r)$ , son nulos. Sea MP el modelo con ese subconjunto de parámetros ceros. Así que  $M_P$  está anidado en el modelo general  $M_G$ . Así que se plantea el contraste:

$H_0: b=0$  ( $M_P$  se verifica)

$H_1: b \neq 0$  (asumiendo cierto  $M_G$ )

Si asumimos que  $M_G$  se verifica, el estadístico del test de razón de verosimilitudes para contrastar si  $M_P$  se verifica es:  $G^2(M_P|M_G) = -2(L_P - L_G) = G^2(M_P) - G^2(M_G)$ , siendo  $L_P$  y  $L_G$  los máximos de la log-verosimilitud bajo la suposición de que se verifican los modelos saturados,  $M_P$  y  $M_G$ , respectivamente. Es decir, el test de razón de verosimilitud para contrastar dos modelos anidados es la diferencia de los contrastes de razón de verosimilitudes de bondad de ajuste para cada modelo.

El estadístico  $G^2(M_P|M_G)$  tiene distribución chi-cuadrado con grados de libertad la diferencia entre los grados de libertad de las distribuciones chi-cuadrado asintóticas de  $G^2(M_P)$  y  $G^2(M_G)$ , es decir, el número de parámetros que se anulan para  $H_0$ .

Así que se rechaza la hipótesis nula al nivel de significación  $\alpha$  cuando

$$G_{obs}^2(M_P|M_G) \geq X_{r;\alpha}^2.$$

#### 2.2.4. INFERENCIA EN REGRESIÓN LOGÍSTICA MULTINOMIAL

Lo principal que se pretende cuando se realiza un modelo estadístico a través de los datos procedentes de una muestra, es extrapolar los resultados muestrales a la población general, es por ello se de haber estimados los parámetros del modelo de regresión logística multinomial con ello se hace la inferencia.

### 2.2.4.1. Intervalos de confianza

Basándonos en la normalidad asintótica de los estimadores de máxima verosimilitud se pueden construir intervalos de confianza asintóticos para cada uno de los parámetros del modelo, utilizando la distribución normal, y mediante las transformaciones correspondientes, intervalos de confianza para las odds ratio.

- **Intervalos de confianza para los parámetros**

Se construye un intervalo de confianza con nivel de confianza  $1-\alpha$  para cada parámetro del modelo de regresión logística multinomial,  $b_{sj}$  con  $j=1, \dots, k$ . La distribución asintótica de  $\hat{b}_{sj}$  es  $N(b_{sj}, \hat{\sigma}^2(\hat{b}_{sj}))$ , donde  $\hat{\sigma}(\hat{b}_{sj})$  es el valor correspondiente al error estándar del estimador del parámetro  $b_{sj}$ .

Así que se tiene:  $P[-Z_{\alpha/2} \leq \frac{\hat{b}_{sj} - b_{sj}}{\hat{\sigma}(\hat{b}_{sj})} \leq Z_{\alpha/2}] = 1 - \alpha$ . Por lo que se obtiene así el intervalo de confianza aproximado para  $b_{sj}$  al nivel  $1 - \alpha$ :

$$IC(b_{sj}) = (\hat{b}_{sj} \pm Z_{\alpha/2} \hat{\sigma}(\hat{b}_{sj}))$$

- **Intervalos de confianza para las odds ratio**

Los cocientes de ventajas vienen dados por:

$$\theta_j(\Delta X_r = 1/X_s = x_s, s \neq r) = \exp(b_{sj}) \quad \forall r = 1, \dots, n; \quad \forall j = 1, \dots, k - 1$$

Por lo tanto, el intervalo de confianza para los cocientes de ventajas se calcula tomando exponenciales en el intervalo de confianza obtenido

anteriormente para cada uno de los parámetros  $b_{sj}$ . Así que el intervalo de confianza para  $\exp(b_{sj})$  al nivel de confianza  $1-\alpha$ , viene dado por:

$$IC(\exp(b_{rj})) = \exp(\hat{b}_{sj} \pm Z_{\alpha/2} \hat{\sigma}(\hat{b}_{sj}))$$

#### 2.2.4.2. Métodos de selección del modelo

Una vez conocido el procedimiento de ajuste de modelos de regresión logística multinomial, el siguiente paso es el desarrollo de estrategias para seleccionar las variables que mejor explican a la variable de respuesta. Para ello se adoptará el principio de parsimonia que consiste en seleccionar el modelo que con menor número de parámetros se ajuste bien a los datos y lleve a una interpretación sencilla en términos de cocientes de ventajas.

Hay que tener especial atención a las covariables cualitativas que se transforman en varias variables dummies. Siempre que se incluya o excluya una de estas variables, todas las demás categorías deben ser incluidas o excluidas en bloque. Si no se tiene en cuenta esta consideración, implicaría que se habría recodificado la variable, y por tanto la interpretación de la misma no sería la correcta. Además, hay que tener en cuenta la significación que pudiera tener cada variable dummy. No siempre todas las categorías de una covariable son significativas, o todas no significativas. Por lo que, cuando ocurra esta situación es recomendable contrastar el modelo completo frente al modelo sin la covariable mediante la prueba de razón de verosimilitud, decidiendo incluir o excluir la covariable dependiendo del resultado de la prueba y del interés clínico de la covariable. Si se obtiene significación en este contraste, la variable permanecería en el modelo, si no se obtiene significación

y la covariable es de interés clínico, su inclusión en el modelo es a criterio del investigador.

**a. Hacia adelante**

1. Se inicia con un modelo vacío (sólo la constante).
2. Se ajusta un modelo y se calcula el p-valor del contraste de razón de verosimilitud que resulta de incluir cada variable por separado.
3. Se selecciona el modelo con el p-valor más significativo.
4. Se ajusta de nuevo un modelo con la(s) variable(s) seleccionada(s) y se calcula el p-valor de añadir cada variable no seleccionada anteriormente por separado.
5. Se selecciona el modelo con el más significativo.
6. Se repite 4 -- 5 hasta que no queden variables significativas para incluir.

**b. Hacia atrás**

1. Se inicia con un modelo con todas las variables candidatas.
2. Se eliminan, una a una, cada variable y se calcula la pérdida de ajuste al eliminar.
3. Se selecciona para eliminar la menos significativa.
4. Se repite 2 – 3 hasta que todas las variables incluidas sean significativas y no pueda eliminarse ninguna sin que se pierda ajuste.

### c. Stepwise

En este método se combinan los métodos adelante y atrás. Puede empezarse por el modelo vacío o por el completo, pero en cada paso se exploran las variables incluidas, por si deben salir y las no seleccionadas, por si deben entrar. Pero no todos los métodos llegan a la misma solución necesariamente. El método de stepwise, está basado en contrastes condicionales de razón de verosimilitudes.

Si se parte del modelo vacío, sólo con la constante, este método consiste en partir de ese modelo inicial, y en cada paso se ajustarán todos aquellos modelos que resultan de incluir cada una de las variables explicativas que no están en el modelo seleccionado en el paso anterior. Entonces se llevan a cabo contrastes condicionales de razón de verosimilitudes que tienen en la hipótesis nula el modelo seleccionado en el paso anterior y en la hipótesis alternativa el modelo resultante de la inclusión de cada variable. De este modo se seleccionarán las variables para las que el contraste sea significativo, y se incluiría en el modelo aquella variable asociada al mínimo p-valor de entre todos los menores o iguales que  $\alpha_1$ . La inclusión de variables mediante este método continúa hasta que ninguno de estos contrastes condicionales sea significativo.

Por otra parte, a la misma vez, se considera en cada paso la posibilidad de eliminar alguno de los parámetros del modelo seleccionado en el paso anterior (método hacia atrás). Pero no se puede eliminar en un paso la variable que acaba de entrar en el paso anterior, por lo que se fijará para la eliminación de variables un nivel de significación  $\alpha_2$  mayor que  $\alpha_1$ . Al igual que antes, para

la eliminación de variables se realizarán contrastes condicionales de razón de verosimilitudes que tienen en la hipótesis nula el modelo que resulta de la eliminación de cada variable y en la hipótesis alternativa el modelo seleccionado en el paso anterior. Así, las variables candidatas a eliminar serán aquellas cuyo p-valor sea mayor de  $\alpha_2$  y se eliminará la variable con el mayor p-valor de éstos. La eliminación de variables continúa hasta que todos estos contrastes condicionales resulten significativos.

#### 2.2.4.3. Validación del modelo

Una vez utilizado el Test de chi-cuadrado de Pearson  $X^2$ , o el test de chi-cuadrado de razón de verosimilitudes  $G^2$ , se debe estudiar la bondad del ajuste de cada observación, para comprobar si una observación es influyente o no. Para ello una de las técnicas para estudiar esta cuestión es el estudio de los residuos que comparan el número observado de éxitos, en cada combinación de valores de las variables predictoras, con su valor ajustado por el modelo. Veamos a continuación los tipos de residuos más habituales basándose en los estadísticos  $X^2$  y  $G^2$ , que se definen en cada combinación de valores  $x_q$  de las variables explicativas.

#### 2.2.4.4. Residuos de Pearson o residuos estandarizados

Vienen dado por la siguiente expresión:

$$r_{j/q} = \frac{y_{j/q} - d_q \hat{p}_{j/q}}{[d_q \hat{p}_{j/q}]^{\frac{1}{2}}}$$

Con esta expresión, podemos definir el estadístico chi-cuadrado de Pearson como:

$$X^2 = \sum_{q=1}^Q \sum_{j=1}^k r_{j/q}^2$$

Para contrastar la significación estadística de los residuos planteamos el contraste:

$$H_0: r_{j/q} = 0$$

$$H_1: r_{j/q} \neq 0$$

Bajo la hipótesis nula  $r_{j/q}$  tiene una distribución asintótica normal con media cero y varianza estimada  $\hat{\sigma}^2(\hat{b}_{sj}) < 1$ , es decir que los residuos tienen menor variabilidad que una variable aleatoria estándar, pero suelen ser tratados como normales estándar, considerándose significativos cuando sus valores absolutos son mayores que dos (falta de ajuste).

Para evitar este problema se definen los residuos de Pearson ajustados que presentan distribuciones asintóticas normales estándar y vienen dados por:

$$r_{j/q}^s = \frac{r_{j/q}}{\hat{\sigma}(r_{j/q})}$$

También se puede tomar el cuadrado de  $r_{j/q}^s$  que tiene distribución chi-cuadrado con un grado de libertad.

Así que se rechaza la hipótesis nula con un nivel de significación  $\alpha$  cuando

$$|r_{j/q}^s| \geq Z_{\alpha/2}$$

### Residuos de la devianza.

Se definen como:

$$d_{j/q} = \left( 2 \left[ y_{j/q} \ln \left( \frac{y_{j/q}}{\hat{m}_{j/q}} \right) \right] \right)^{\frac{1}{2}}$$

Con esta expresión se puede definir el estadístico de chi-cuadrado de razón de verosimilitudes como:  $G^2 \sum_{q=1}^Q \sum_{j=1}^k d_{j/q}^2$  igual que en el caso anterior, bajo la hipótesis nula  $H_0: d_{j/q} = 0$  el residuo  $d_{j/q}$  tiene distribución asintóticamente normal con media 0 y varianza estimada  $\hat{\sigma}^2(r_{j/q}) < 1$ . En este caso, se consideran significativos cuando el valor absoluto es mayor que 4, y se considera que la observación correspondiente es anormal.

Al igual que antes, se definen los residuos de la devianza ajustados o estandarizados,

$$d_{j/q}^s = \frac{d_{j/q}}{\hat{\sigma}(d_{j/q})}$$

que tiene distribución asintótica normal estándar.

Así que rechazamos la hipótesis nula con un nivel de significación  $\alpha$  cuando

$$|d_{j/q}^s| \geq Z_{\alpha/2}.$$

La diferencia entre los dos tipos de residuos es que los de la devianza convergen más rápidamente a la distribución normal que los de Pearson.

## 2.3. MARCO CONCEPTUAL

### REGRESIÓN

El análisis de la regresión es un proceso estadístico para estimar las relaciones entre variables. Incluye muchas técnicas para el modelado y análisis de diversas variables, cuando la atención se centra en la relación entre una variable dependiente y una o más variables independientes (o predictoras). Más específicamente, el análisis de regresión ayuda a entender cómo el valor de la variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las otras variables independientes fijas. Más comúnmente, el análisis de regresión estima la esperanza condicional de la variable dependiente dadas las variables independientes - es decir, el valor promedio de la variable dependiente cuando se fijan las variables independientes.

### REGRESIÓN LOGÍSTICA

En este tema se considera un modelo de regresión logística donde la variable dependiente tiene más de dos categorías. La respuesta puede o bien ser nominal o bien ordinal. A su vez, las variables explicativas pueden ser categóricas o cuantitativas.

### MODELO LOGÍSTICO MULTINOMIAL

Modelo estadístico cuyo objetivo principal es conocer la relación entre una variable dependiente cualitativa con más de dos respuestas (multinomial) y una o más variables explicativas independientes (o covariables) las cuales pueden ser cualitativas o cuantitativas.

## **DEVIANZA**

La devianza residual es un criterio de bondad de ajuste sumamente útil para evaluar la contribución de las variables durante su proceso de selección como para evaluar el ajuste general del modelo. En los modelos generalizados no se obtiene un valor  $R^2$  pero un resultado semejante residual en relación con la devianza nula del modelo

## **CRITERIO DE INFORMACIÓN DE AKAIKE (AIC)**

Índice estadístico que evalúa tanto el ajuste del modelo a los datos como la complejidad del modelo. Cuanto más pequeño es el AIC mejor será el ajuste del modelo estudiado.

## **DIÁMETRO**

Es el grosor o finura de la fibra, se mide en micras ( $\mu$ ). El diámetro de fibra es un parámetro físico que determina el uso de una fibra textil. El diámetro de la fibra tiene efecto sobre el nivel de afieltramiento de la fibra durante el lavado, por ejemplo, para una masa de unidad dada, las lanas más finas tienen una mayor superficie, más grasa y suciedad, y un mayor número de fibras, por lo que requieren una mayor cantidad de detergente para eliminar los contaminantes, especialmente las grasas. Además, para un volumen dado, las lanas más finas tendrán más contacto fibra/fibra, lo que ocasionaría un mayor afieltramiento. Las lanas más finas por lo tanto se afieltran más (Liu et al., 2010). En cuanto a la suciedad presente sobre la fibra, debemos tener presente que, cuanta más gruesa es la fibra, menor cantidad de ella contiene,

por lo que las lanas finas están más perjudicadas en este aspecto (Codina, 1973).

## **LONGITUD**

Es una característica de importancia, conjuntamente con el diámetro determinan las propiedades manufactureras del material textil. Este parámetro tipifica la materia prima para la hilandería y junto con el diámetro de las fibras define el precio final de la lana peinada. Representa el promedio de longitud de las mechas en el lote y se mide en milímetros. En contraste con el diámetro, el largo de fibra promedio cambia continuamente durante el proceso, en la cual el 40% de las fibras se rompe durante el cardado. La cantidad de rupturas depende en la mayor parte de la resistencia de la fibra y de la calidad del lavado.

## **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 (Antúnez et al., 1996).

## **CAMÉLIDOS SUDAMERICANOS**

Los camélidos (Camelidae) son una familia de mamíferos artiodáctilos del suborden ruminantes. 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 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.

### **CATEGORIZACIÓN**

Durante la categorización, se forman grupos de vellones buscando estandarizar las categorías acordes a la tabla, teniendo en cuenta básicamente la finura y uniformidad. Actividad debe ser ejecutada por personal especializado, que tenga principios, muy buen criterio y conocimiento del material textil y de las características del mismo; además deberá tener un amplio conocimiento de las normas técnicas imperantes y su aplicación respectiva.

### **CLASIFICACIÓN**

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.

### **EDAD**

La edad la definiremos como la cantidad de años transcurridos desde el día de nacimiento de las alpacas hasta la fecha actual (el día de hoy, en nuestro caso es el día de la medición de los vellones de las alpacas).

## **SEXO**

Características biológicas que definen a una alpaca si es hembra o macho.

## **CONFORT**

Es aquello que brinda comodidad, el mercado internacional desea fibras con un alto grado de confort que se logra con las fibras más finas.

## **CARACTERÍSTICAS 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: Factor de confort (FC) y índice de curvatura (IC), y Longitud de mecha (LM).

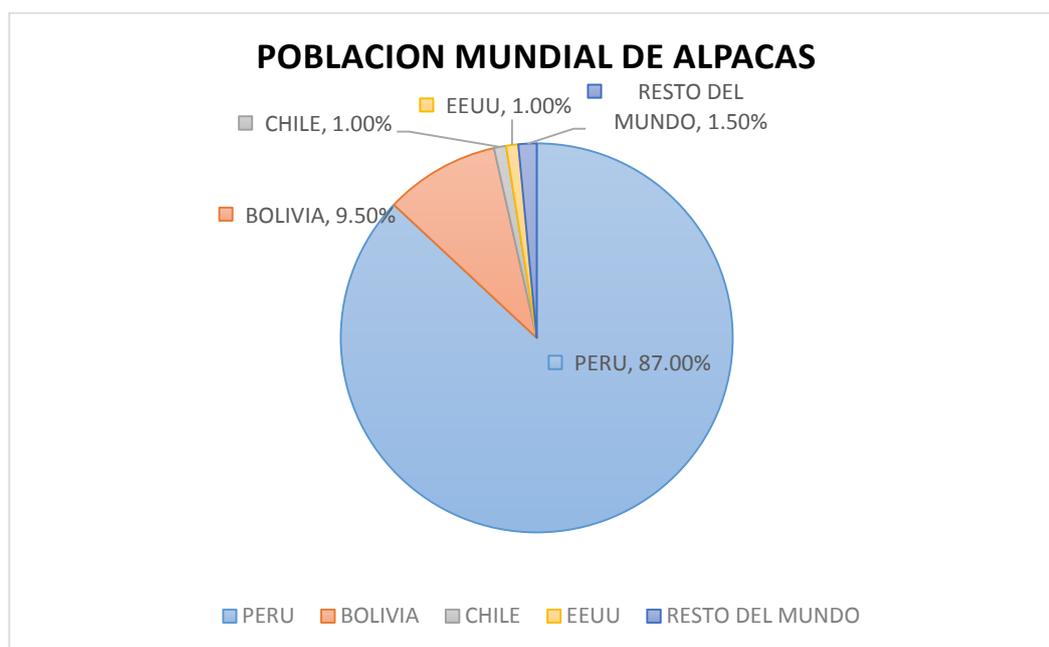
### **2.3.1. ALPACA**

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, 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.).

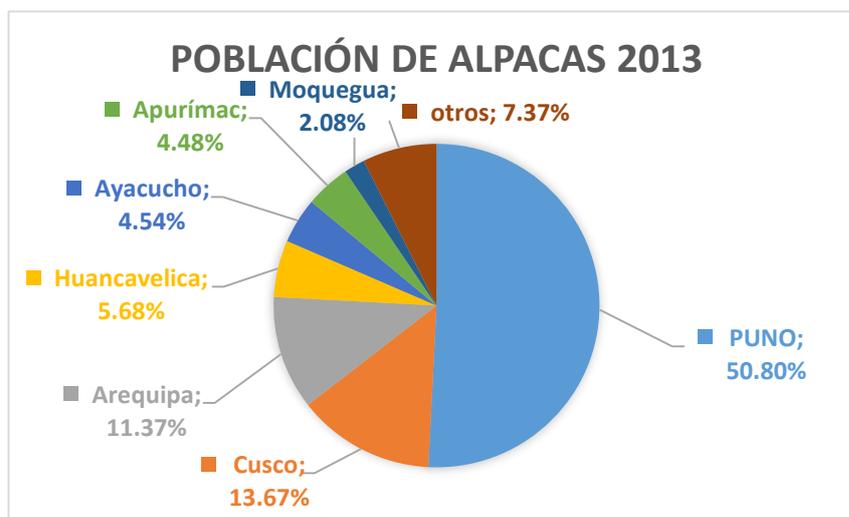


**Figura 1: Población mundial de alpacas**

Fuente: Dirección General de Negocios Agrarios (Digna) del Ministerio de Agricultura y Riego (Minagri) - 2015  
Elaboración propia

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 2: Población Nacional de Alpacas, 2015**

Fuente: Dirección General de Negocios Agrarios (Digna) del Ministerio de Agricultura y Riego (Minagri) - 2015  
Elaboración propia

Existen dos razas de alpacas:

**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 mechaz rizadas, que crecen paralelas al cuerpo (INIA, 2014).

### 2.3.2. FIBRA DE ALPACA

La fibra de alpaca es considerada una de las fibras más lujosas y finas del mundo, es tres veces más fuerte que de la oveja y siete veces más caliente, es resistente y liviana, pero aun así es muy caliente haciéndola muy escasa en el mercado, haciéndola mucho más exclusiva, con poca

probabilidad de producir alérgicas y las prendas que se confeccionan con ellas son bastante agradables a la observación; asimismo los vestidos exhiben unos excelentes pliegues, apariencia, caída y lustrosidad los cuales dan la sensación de ser nuevos a pesar que puedan tener tiempo de uso. En alusión a las fibras de alpaca y vicuña, además que resaltan por su suavidad (Wang et al., 2004), muestran alta resistencia a la tracción lo cual tiene gran importancia en los procesos textiles. Su resistencia a la tracción se mantiene mientras más Fleece sea, volviéndola ideal para procesos industriales, debido a que tienen un alto grado de limpieza, su rendimiento también resulta ser alto (entre el 87% y 95% versus un 43% a 76% de la lana de oveja), además que su proceso es más fácil y barato debido a la ausencia de grasa o lanolina en dichas fibras, lo que les permite ser trabajada en los sistemas de peinado o cardado, pudiendo obtenerse telas entre tweeds gruesa y gabardina Fleece. Estas fibras no se rompen, no se deshilachan, no manchan o crean estática y resultan fáciles de lavar.

### **2.3.3. CARACTERÍSTICAS 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:

#### **2.3.3.1. DIÁMETRO MEDIO DE FIBRA (MDF)**

La finura o diámetro medio de la fibra es la característica más importante y se mide en micras (Zarate, 1992) de igual forma esta representa una de las

características máspreciadas para su apreciación, siendo determinante para su clasificación.

El diámetro de la fibra es uno de los factores más importantes en la clasificación de la misma, con el cual se determina el precio de la fibra en el mercado. La comercialización usualmente se realiza por peso del mismo (Villarroel, 1963), aunque se otorgan incentivos por finura de la fibra de alpaca. En la actualidad, con el avance de la tecnología en temas de mejoramiento genético y con el apoyo que vienen dando los gobiernos nacionales, regionales y locales, en asistencia técnica, capacitación y desarrollo en toda la cadena productiva de la fibra, es posible afirmar que los productores de fibra de alpaca tienen mayor accesibilidad para determinar objetivamente el diámetro de la fibra.

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 finalmente 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,

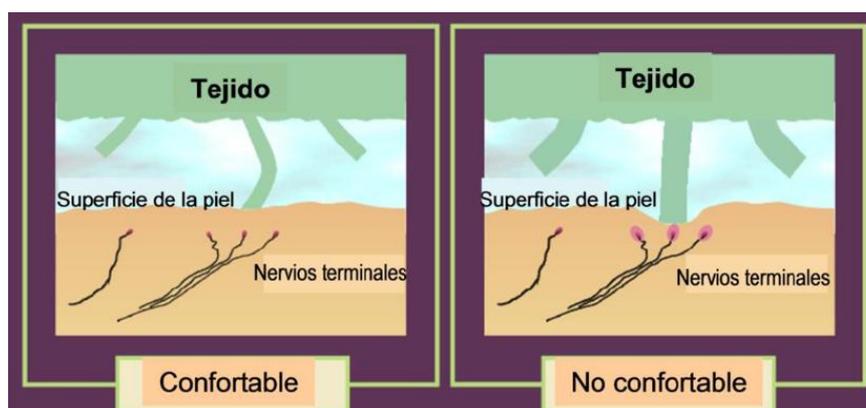
sexo, raza, temperatura, radiación solar fotoperiodo, estrés, época del año, época de empadre, época de esquila, sanidad y otros factores característicos del medio ambiente alto andino (Solis, 1991).

Con respecto al diámetro de la fibra resulta importante el trabajo de Lupton, McColl, & Stobart, 2006 quien analizó 585 muestras de vellón de alpacas norteamericanas de distintos sexos y edades, encontrando diámetros de fibra de 26.7  $\mu\text{m}$  para hembras y 27.1  $\mu\text{m}$  para machos; con respecto a la edad, encontró valores de 24.3  $\mu\text{m}$ , 26.5  $\mu\text{m}$  y 30.1  $\mu\text{m}$  para alpacas de 1, 2 y 3 ó más años de edad, respectivamente. Por otra parte, McGregor (2006) al estudiar alpacas criadas en Australia encontró que el 10% de alpacas Huacaya presentan un diámetro medio de 24  $\mu\text{m}$  y más del 50% estaban en 29.9  $\mu\text{m}$ . A su vez, Wang et al. (2004), también refieren medias de diámetro similares a los descritos por McGregor.

En alpacas, Hack (1999), McGregor (2006), Lupton (2006), Morante R., Goyache F., Burgos A., Cervantes I. (2009), Quispe et al. (2009a) y Quispe (2010) obtuvieron resultados de CVDF de 24.40, 27.00, 23.30, 23.60 y 21.4 % respectivamente, los cuales, si bien resultan un tanto elevados, muestran una alta variabilidad de los animales que resulta conveniente para programas de mejora genética. Asimismo, casi todos los resultados (a excepción de lo encontrado por McGregor (2002), no superan el 24%, que representa el límite para rendimientos textiles acorde a su diámetro, y que se encuentra asociado al rendimiento del hilado, propiedad conocida también como finura al hilado.

### Factor de confort (FC) y factor de picazón (FP).

El factor de confort se define como el porcentaje de las fibras menores a 30 micras y se conoce también como factor de comodidad (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 3: Interacción entre el tejido, los terminales de la fibra o lana y la piel, que muestran la importancia del factor de confort**

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

Sacchero, (2008), indica que el factor de picazón no es un carácter técnico de la fibra, sino que más bien tiene relación con el grado mayor o menor de confort que brindan las prendas fabricadas con fibra de alpaca sobre el usuario. Se tiene conocimiento que mientras menor diámetro tiene las fibras el confort es mayor. McGregor y Butler, (2004), al realizar estudios en alpacas criadas en Australia, obtiene un factor de picazón de 44.42 % y un índice de

confort de 55.58 %. Ponzoni R.W., R.J. Grimson, J.A. Hill, D.J. Hubbard, B.A. McGregor, A. Howse & Judson., (1999), en un estudio realizado en alpacas al sur de Australia muestran un índice de confort de 75.49 %, mientras que Lupton et al. (2006), trabajando en alpacas criados en EEUU sobre la evaluación de las características de la fibra de alpaca Huacaya, con una muestra representativa de 585 animales determinó un índice de confort de  $68.39 \pm 25.05$  %.

### **Índice de curvatura de la fibra**

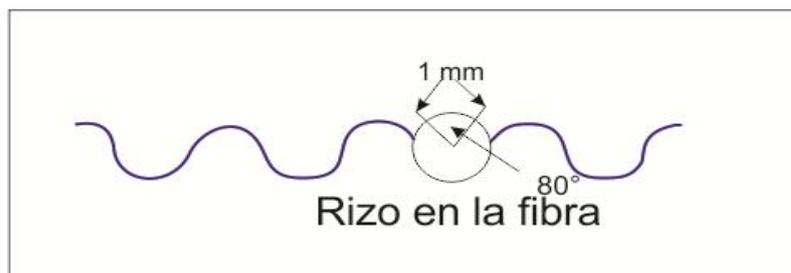
El índice de curvatura (IC) de la fibra es una característica textil adicional que puede ser utilizado para describir la propiedad espacial de una masa de fibras de lana. Esta propiedad, que es común a todas las fibras textiles, es de interés para los fabricantes de alfombras y prendas de vestir. Los fabricantes de fibras sintéticas introducen rizos a sus fibras y filamentos con el fin de mejorar la densidad de sus productos textiles (Fish, Mahar, & Crook, 1999).

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 final. 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 4: Representación bidimensional de la forma de una fibra de lana.**

Tomado de: 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)

### **Finura al hilado**

La finura al hilado (FH) expresada en  $\mu\text{m}$  (spinning fineness), provee una estimación del rendimiento de la muestra cuando es hilada y convertida en hilo. Su estimación proviene de la combinación de la media del diámetro de fibra (MDF) y el coeficiente de variación (CVMDF). La idea original viene de Martindale (1945), que fue analizada y planteada por Anderson (1976) como “effective fineness” y que, luego fue modificada por una ecuación práctica llamándose a dicho valor finura al hilado McGregor B.A. y Butler K.L., (2004) y es una característica heredable. La ecuación se normaliza bajo un coeficiente de variación del 24% en la cual la finura al hilado es lo mismo que la media del diámetro de fibra previa al procesamiento (Lupton et al., 2006).

La expresión:

$$F = (MDF) \times \sqrt{1 + 5\left(\frac{CVMDF}{100}\right)^2}$$

Donde:

MDF: Media del Diámetro de Fibra

CVMDF: Coeficiente de Variabilidad de la Media del Diámetro de Fibra.

Finura efectiva (effective fineness) y lo usó con la finalidad de demostrar la influencia de los cambios de la MDF y el CVMDF sobre la uniformidad de los hilados. Dos tops con diferentes MDF y CVMDF pueden producir hilados de la misma uniformidad, si sus finuras efectivas tienen el mismo valor al utilizar la fórmula anteriormente descrita. Por ejemplo, un top con MDF y CVMDF de 21.5  $\mu\text{m}$  y 20.0% respectivamente, produce un hilado más uniforme que otro top con MDF y CVMDF de 20.2 y 27% respectivamente (De Groot, 1995).

La finura efectiva sólo depende de la MDF y del CVMDF y es siempre numéricamente mayor que la MDF, aunque esto puede corregirse normalizando la finura efectiva mediante la aplicación de la ecuación:

$$F \approx (0.881 \times \text{MDF}) \times \sqrt{1 + 5 \left( \frac{\text{CVMDF}}{100} \right)^2} \dots, \text{ (McGregor B.A. y Butler K.L., 2004).}$$

En alpacas Huacaya de color blanco, Quispe (2010) encontró una finura al hilado de 20.9 $\mu\text{m}$  observando que animales jóvenes tienen menor FH que animales adultos y que los animales menores de 18 meses son los que exhiben una mejor FH; asimismo, encontró efectos altamente significativos de factores como año y comunidad, sobre dicha finura.

### **2.3.3.2. FACTORES QUE INFLUYEN SOBRE LAS CARACTERÍSTICAS DE LA FIBRA DE ALPACA**

Los factores que influyen en la cantidad y la calidad de la producción de fibra en camélidos sudamericanos se clasifican en factores medioambientales externos y factores genéticos o internos. Los factores externos que modifican la respuesta productiva en alpacas son la alimentación, la locación geográfica

o lugar de pastoreo y, en el caso del peso de vellón, es particularmente relevante considerar la frecuencia, año de esquila y la precipitación pluvial. La altitud no ejerce influencia ni sobre la cantidad ni la calidad de fibra. Entre los factores internos que afectan el diámetro de fibra y peso de vellón resaltan el sexo, la edad, la raza, sanidad, estado fisiológico, condición corporal y color de vellón (FLORES QUISPE, 2017) .

#### **2.3.3.2.1. SEXO**

Estudios realizados en fibra de alpacas muestran que los machos producen vellones más pesados que las hembras (Wuliji et al., 2000; McGregor, 2006; Lupton et al., 2006; y Quispe et al., 2007); sin embargo, existen desacuerdos sobre el efecto del sexo en el diámetro de la fibra, debido a que algunos investigadores como Morante et al. (2009), Quispe et al. (2009) y Montes et al. (2008), han encontrado que alpacas machos tienen fibras más finas que las hembras. Otros como Lupton et al. (2006) han reportado lo contrario, debido probablemente a que las hembras priorizan el uso de los aminoácidos ingeridos hacia la producción (preñez y lactación) en vez del abastecimiento del bulbo piloso para su excreción como fibra. Sin embargo, Bustinza (2001), Wuliji et al. (2000) y McGregor y Butler (2004) consideran que no existe efecto del sexo sobre el diámetro de fibra.

#### **2.3.3.2.2. EDAD**

Se conoce que, a medida que aumenta la edad, se incrementa el diámetro (Wuliji et al., 2000; McGregor y Butler, 2004; Quispe et al., 2009a). Las alpacas jóvenes producen vellones menos pesados que las adultas, por tener una menor superficie corporal, sin embargo, producen vellones con

fibras más finas, debido a que las esquilas tienen el efecto de incrementar el funcionamiento folicular (Siguayro, 2009).

#### **2.3.3.2.3. ALIMENTACIÓN**

En trabajos realizados en alpacas, Franco & San Martín (2007) cuando suplementaron dietas con bajo contenido nutricional, encontraron que la producción de fibra disminuye debido a la disminución de la tasa de crecimiento y del diámetro de la fibra, observándose también que estos animales producen fibras más finas.

#### **2.3.3.3. EQUIPOS PARA MEDIR LAS CARACTERÍSTICAS DE LA FIBRA DE ALPACA**

Como se mencionó anteriormente, el diámetro de fibra constituye la medida objetiva de mayor importancia que define el destino industrial de la fibra. De ahí que se ha invertido mucho esfuerzo en el desarrollo de métodos de medición. En un principio se empleaban los microscopios de proyección (lanómetros), sin embargo, a su mayor laboriosidad en la medición se buscaron otros métodos más precisos y rápidos. El Air Flow fue un avance importante en este sentido. Pero a pesar de su rapidez y precisión, este no informaba la frecuencia de los distintos diámetros presentes en la muestra. En los últimos años, se ha extendido el uso de nuevos instrumentos de medición, el Laserscan y el OFDA. Estos instrumentos además de ser rápidos y precisos, proporcionan una información adicional sobre la frecuencia de los diámetros y su variabilidad.

#### **2.3.3.3.1. MÉTODO DEL OFDA (OPTICAL FIBRE DIAMETER ANALYSER)**

El OFDA es un instrumento que se basa en la tecnología de digitalización de imágenes y el análisis óptico de las mismas, que permite medir las características de las fibras de lana y otras fibras animales a lo largo de las mechas sucias en tiempo real y si uno lo requiere en el propio galpón de esquila.

Esencialmente, el OFDA 2000 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).

#### **2.3.3.3.2. MÉTODO AIRFLOW**

El método Airflow se basa en la medición del descenso de la presión del flujo de aire a través de una masa estándar de fibras (Grishanov, S.A., Harwood, R.J., 2006). En este método, las muestras de lana son previamente mezcladas, lavadas, secadas y luego acondicionadas en una atmosfera estándar. Luego, éstas son colocadas en una cámara donde se fuerza el paso

de una corriente de aire a través de ellas (masa de fibras). El valor del flujo de aire en lana sucia se determina midiendo la tasa de flujo a una presión estándar. Este método mide el área superficial de fibra por unidad de masa. De ahí que sea un método indirecto y debe ser calibrado utilizando lanas o fibras de diámetro conocido (Baxter, 1994).

El método Airflow no proporciona mediciones de desviación estándar ni coeficiente de variación del diámetro de fibra, limitándose únicamente al promedio del diámetro de fibra (Sommerville, 2000). Es uno de los métodos antiguos que fue empleado comercialmente por muchos años.

## CAPITULO III

### MATERIALES Y MÉTODOS

#### 3.1. MÉTODO DE INVESTIGACIÓN

Para el desarrollo del presente trabajo de investigación se desarrolla el estudio explicativo utilizara los siguientes métodos:

Método analítico. Este método servirá para determinar las similitudes entre las unidades de análisis originados por los mismos factores y estarán presentes en todas las fases del desarrollo del modelo.

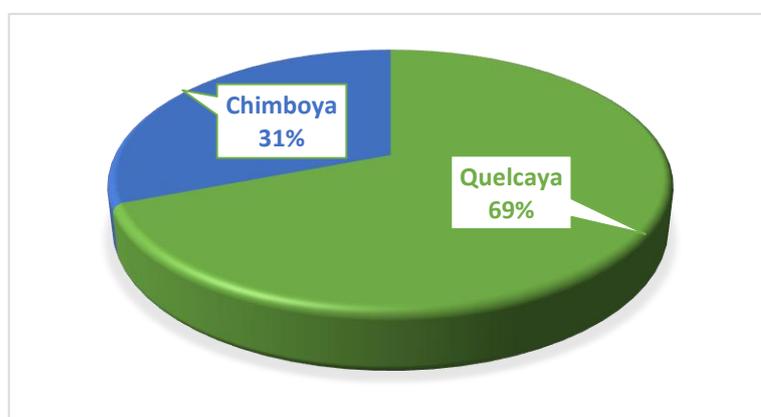
Método descriptivo. Este método se utilizará para describir cada una de estas partes, el mismo que nos permitirá realizar comparaciones de las características cuantitativas y cualitativas.

Método experimental El presente método nos permite la consideración y la utilización del método analítico (inductivo) y sintético (deductivo); en donde nos permitirá desarrollar las características de los principales factores el cual nos facilita la explicación, de esta manera se puede arribar a conclusiones convincentes sean específicos y /o globales. Su esencia es la de someter el objeto de estudio a la influencia de ciertas variables en condiciones controladas y conocidas por el investigador.

### 3.2. DETERMINACIÓN DE LA POBLACIÓN Y MUESTRA

Población: La población está constituida por todas las alpacas de las comunidades de Chimboya y Quelcaya.

Muestra: La muestra está constituida por 1526 registros de alpaca Huacaya de las comunidades Quelcaya y Chimboya. Los datos fueron del gobierno regional y el PECSA. Los porcentajes correspondientes se pueden ver en la figura 6, considerando por sexo (hembras y machos) y por edad (DL, 2D, 4D y BLL).



**Figura 5: Población de Alpacas en el distrito de Corani.**

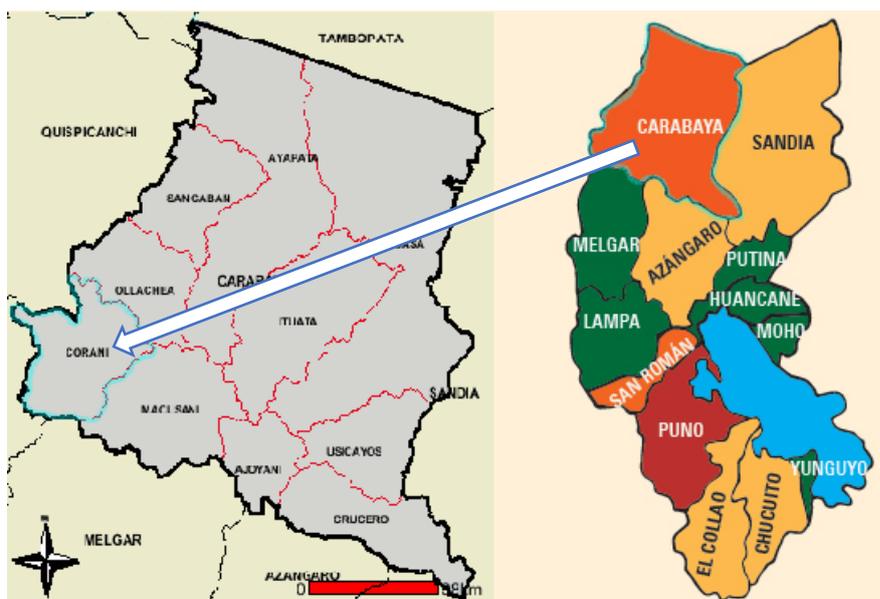
Fuente: Gobierno Regional Puno.  
Elaboración propia

La figura muestra que la población de alpacas en la comunidad de Quelcaya es mayor con 69% y Chimboya con 31% de alpacas de raza Huacaya color blanco.

### 3.3. ÁMBITO DE ESTUDIO

El distrito de Corani es uno de los 10 distritos que conforman la Provincia de Carabaya, ubicada en el Departamento de Puno, perteneciente a la Región Puno, en el sudeste de Perú.

El trabajo de investigación se realizó en 2 comunidades del distrito de Corani; cuya altitud varía de 4800 a 5350 m, siendo sus coordenadas 13° 55'54 latitud Sur y de 70° 51'04 longitud Oeste; el área de estudio pertenece a la región natural de Puna por ende su clima es seco y frío con dos épocas bien definidas.



**Figura 6: Ubicación del distrito de Corani**

Fuente: Elaboración propia

### 3.4. PERTINENCIA DE LA METODOLOGÍA DE INVESTIGACIÓN

#### 3.4.1. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES DEFINIDAS

Antes de realizar el análisis estadístico para conocer la relación entre las distintas variables, se realizará un análisis descriptivo unidimensional de las variables definidas anteriormente teniendo en cuenta los resultados del mensurando escogido, que en este caso será calidad de fibra a fin de conocer un poco su comportamiento.

### **3.4.2. METODOLOGÍA PARA DETERMINAR LA DIFERENCIA SIGNIFICATIVA ENTRE FACTORES**

#### **3.4.2.1. Identificación de alpacas**

Durante la identificación de alpacas se consideró las características zootécnicas y los requisitos que el proyecto (fortalecimiento del proceso productivo de la crianza de alpacas – del distrito de Corani) estableció para el ingreso de un animal al plan de mejora genética, considerando alpacas Huacaya, de color blanco, de ambos sexos, de cuatro edades y que sean libres de defectos genéticos (ojos zarco, manchados, prognatismo inferior o superior, etc.); para la identificación de las alpacas seleccionadas se utilizó aretes de plástico con ayuda de un aretador, y con el número de codificación correspondiente a cada alpaca.

#### **3.4.2.2. Obtención de vellones de fibra de alpaca**

La toma de muestra de fibra se realizó antes que los animales sean esquilados, en la cantidad de 3.0 gr. de la región del costillar medio considerada como la zona representativa (Aylan Parker y McGregor, 2001). Las muestras fueron colocadas en bolsas de polietileno, debidamente identificadas rotulándose con datos del propietario, comunidad, número de arete, sexo, edad de la alpaca, fecha de obtención de la muestra las que fueron analizados en el laboratorio de fibras del PECSA.

### 3.4.2.3. Procedimiento del análisis del diámetro de fibra, factor de confort e índice de curvatura.

Las muestras fueron analizadas con la finalidad de determinar el diámetro de fibra; factor de confort e índice de curvatura de la fibra, para lo cual se utilizó el equipo OFDA 2000, mediante el siguiente procedimiento:

- Se realizó el calibrado el equipo con el slide usando patrones de fibra poliéster estándar para fibra de alpaca.
- Para determinar el factor de corrección por grasa primero se realizó la identificación de 24 muestras de fibra en sucio, que representa un mínimo de la muestra total, luego se procedió con la lectura de las 24 muestras procediendo con el lavado y pasar con la segunda lectura, determinándose el factor de corrección por grasa que fue de  $0.7\mu$ .
- Posteriormente las muestras restantes de fibra sucia fueron colocadas en el slide uno por uno para su posterior análisis, en OFDA 2000 de aplicar la corrección de grasa de forma automática para la determinación del diámetro de fibra, factor de confort e índice de curvatura.

### 3.4.2.4. Diseño experimental.

El trabajo de investigación fue conducido con Diseño Completo al Azar bajo un arreglo factorial de  $2 \times 3$  (sexo x edad) siendo el modelo aditivo lineal el siguiente:

$$Y_{ijk} = \mu + S_j + E_k + (SE)_{jk} + E_{ijk}$$

$Y_{ijkl}$ : variable respuesta

$\mu$  : Media general

$S_j$ : Efecto del j-ésimo sexo.

$E_k$  : Efecto de k-ésimo edad.

$(SE)_{jk}$  : Efecto de la interacción del j-ésimo factor sexo con el k-ésimo factor edad.

$E_{ijkl}$  : Error experimental.

Para determinar el coeficiente de correlación del diámetro de fibra entre índice de curvatura, factor de confort y longitud de mecha, se utilizó el método de Pearson, con la siguiente formula:

$$r = \frac{\sum xy - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

Dónde:

X = Variable del diámetro de fibra.

Y = Variable de correlacionada: índice de curvatura, factor de confort o índice de curvatura.

$\sum xy$  = Sumatoria de las variables del diámetro de fibra e índice de curvatura, factor de confort y longitud de mecha.

#### **3.4.2.5. Prueba de significancia.**

La comparación de promedios de las variables diámetro de fibra, factor de confort e índice de curvatura de fibra para efectos de los factores edad y sexo de los animales, se ha realizado mediante la prueba de Significación Múltiple de Duncan con  $\alpha=0.05$ .

#### **3.4.3. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL, FACTORES ASOCIADOS A LA CALIDAD DE FIBRA DE ALPACA HUACAYA.**

Como resultado y para alcanzar el objetivo principal se realiza el ajuste de un modelo de regresión logística multinomial se tiene que la variable dependiente las calidades de fibra, en la que se toma como categoría de referencia “mediana” y se introduce en el modelo como un factor. Las variables independientes como son LN, IC y FC. Una vez seleccionado el modelo final, el más parsimonioso, se contrastará los parámetros del modelo mediante el contraste de Wald, calculando, además las odds ratio de los coeficientes, sus intervalos de confianza y los p-valores, interpretándose el modelo final. Se medirá la bondad del ajuste global del modelo mediante el test de Chi-cuadrado de razón de verosimilitudes y la tasa de clasificaciones correctas. La calidad del ajuste se realizará mediante los parámetros pseudo R-cuadrado de Cox-Snell, Nagelkerke y MCFadden.

## ESPECIFICACIÓN DE UN MODELO LOGÍSTICO MULTINOMIAL

En este tipo de modelos las alternativas de la variable respuesta indican la pertenencia de las observaciones a un determinado grupo. La formulación de un Logístico Multinomial queda recogida a través de la siguiente ecuación:

$$Prob(Y_{i=j}) = P_{ij} = \frac{e^{\beta_j X_i}}{\sum_{j=0} e^{\beta_j X_i}}$$

Donde para el caso sencillo de un modelo en el que la variable dependiente presenta tres posibles alternativas de elección, las probabilidades asociadas a cada una de las alternativas posibles tomarían las siguientes expresiones:

$$P_0 = \frac{1}{1 + e^{\alpha_1 + \beta_1 X_i} + e^{\alpha_2 + \beta_2 X_i}}$$

$$P_1 = \frac{e^{\alpha_1 + \beta_1 X_i}}{1 + e^{\alpha_1 + \beta_1 X_i} + e^{\alpha_2 + \beta_2 X_i}}$$

$$P_2 = \frac{e^{\alpha_2 + \beta_2 X_i}}{1 + e^{\alpha_1 + \beta_1 X_i} + e^{\alpha_2 + \beta_2 X_i}}$$

Con

$$P_0 + P_1 + P_2 = 1$$

## FORMULACIÓN DEL MODELO LOGÍSTICO MULTINOMIAL

La variable dependiente Y será la calidad de fibra de alpaca (Baby, Fleece y mediana), se tomará como base la calidad mediana, por lo tanto, si Y=1 se entenderá como la probabilidad de obtener un resultado satisfactorio, si Y = 0 será comprendido como la ausencia de esto.

La variable dependiente (calidad de fibra) tiene 3 categorías BABY / FLEECE / MEDIANA, Por lo tanto, se modelan 2 logit simultáneamente  $\text{logit}(\text{BABY}/ \text{MEDIANA } X) = a_1 + b_1X_i$   $\text{logit}(\text{FLEECE}/ \text{MEDIANA } X) = a_2 + b_2X_i$ . Las variables  $X$  son comunes en los dos logit sin embargo, se estiman coeficientes  $b_i$  diferentes para cada logit, e incluso las constantes. Para la selección del mejor modelo y evaluar la bondad de ajuste se analizará la devianza del modelo nulo con los modelos propuestos, de la siguiente forma:

$$\%D^2 = \frac{\text{devianza nula} - \text{devianza residual}}{\text{devianza nula}} \times 100$$

La fórmula anterior expresa el porcentaje de la variabilidad explicada por el modelo. Entre más pequeño sea el valor de la devianza residual mejor será el ajuste del modelo, por tanto entre más alto el porcentaje  $D^2$  habrá mayor proporción de datos que es explicada por el modelo. Una vez se obtienen las devianzas de los modelos establecidos se compara cada una de ellas con la devianza del modelo nulo. Adicionalmente a la devianza se utilizará el Criterio de Información de Akaike (AIC del inglés Akaike Information Criterion), el cual evalúa el ajuste del modelo a los datos como la complejidad del modelo. Cuanto más pequeño es el AIC mejor será su ajuste.

Se evaluará la sobre-dispersión de los modelos utilizando la estadística ji-cuadrado y sus grados de libertad. Con estos tres criterios se escogerá el modelo que mejor explique los datos analizados.

#### **3.4.3.1. Selección del modelo**

Para la selección del modelo, que resultará ser el que contenga el menor número de variables y sea un modelo explicativo, se realizará mediante el

método de stepwise, por lo que se detalla cada uno de los pasos que se seguirán, comenzando con el modelo inicial que contiene sólo la constante, en cada paso se analizará la inclusión o no de alguna de las variables mediante contrastes de razón de verosimilitudes, considerándose además en cada paso la posibilidad de eliminar alguna variable que se incluyó en los pasos anteriores.

#### **3.4.3.2. Odds ratios e Intervalos de Confianza**

Se calcula las exponenciales de los coeficientes, las odds ratios, y los intervalos de confianza de éstos al 95%, ya que el modelo se interpretará en base a éstas, por ser más común en la aplicación dar los resultados en función de los cocientes de ventajas.

#### **Odds ratios**

El Odd es la probabilidad de que suceda un evento dividido por la probabilidad de que no suceda, es una medida de asociación entre dos variables, lo utilizamos cuando relacionamos a dos variables, como sabemos los modelos de regresión logística están basados en probabilidades, y concretamente en logaritmos, esto para saber si las variables independientes están relacionadas con las variables dependientes.

#### **3.4.3.3. Contraste sobre los parámetros.**

#### **Contraste de Wald**

Se contrasta los parámetros del modelo mediante el contraste de Wald, es decir, el cociente entre el coeficiente al cuadrado y su error estándar, que

tiene distribución chi-cuadrado asintótica con un grado de libertad. En este caso, al obtener del modelo el valor del estadístico como el valor del coeficiente entre el error estándar, se tiene que sigue una distribución normal de media 0 y desviación típica 1, por lo que calcularemos las probabilidades de estos estadísticos mediante una distribución normal.

Así se obtendrá los niveles de significación de este estadístico para contrastar si conservamos en el modelo cada uno de los coeficientes de las variables explicativas considerando un nivel de significación de 0.05.

### **Contraste condicional de razón de verosimilitud**

Por otra parte, al realizar un contraste de razón de verosimilitud, para contrastar el efecto conjunto de las variables predictoras, comparando las devianzas del modelo Final con el modelo con sólo la constante.

#### **3.4.3.4. Interpretación del modelo**

Una vez estudiado el ajuste del modelo se presenta el modelo Final con los parámetros (B), sus errores estándares (EE), las odds ratio (OR), el estadístico de Wald, el intervalo de confianza al 95% de las OR (IC 95%) y el p-valor. Posteriormente interpretamos sus parámetros en términos de cociente de ventajas.

#### **3.4.3.5. Ajuste global del modelo**

Para realizar el ajuste global del modelo, realizamos el test de chi-cuadrado de la razón de verosimilitud comparando el modelo ajustado con el

modelo saturado, que resulta de multiplicar todas las variables, la ausencia de significación de este test significa que el ajuste del modelo es bueno.

### Calidad del ajuste del modelo

Para medir la calidad del ajuste del modelo se utiliza los coeficientes Pseudo-R<sup>2</sup> de Mc-Fadden, de Cox-Snell y de Nagelkerke. Los cálculos de éstos dependen del valor de las devianzas del modelo final y del modelo inicial con sólo la constante.

### Validación del modelo

Por último, realizamos la validación del modelo mediante los residuos de la devianza, considerando que los residuos que indican una falta de ajuste global son aquellos cuyo valor absoluto son mayores que 4.

## 3.5. OPERACIONALIZACIÓN DE VARIABLES

Tabla 1: Operacionalización de variables

| VARIABLE                 | INDICADOR                   | ÍNDICE  |
|--------------------------|-----------------------------|---------|
| Variable Dependiente     | Media del diámetro de fibra | Baby    |
|                          |                             | Fleece  |
|                          |                             | Mediana |
| Variables Independientes | Longitud de Mecha (LM)      | mm      |
|                          | Índice de curvatura (IC)    | °/mm    |
|                          | Factor de Confort (FC)      | %       |

El análisis de regresión logística multinomial requiere de una variable categórica dependiente, que para este caso es la media del diámetro de fibra, que esta categorizada por su calidad, al separar los grupos por sexo y por edad, para cada modelo se considera las variables independientes, en la tabla 01 se muestran los indicadores e índices.

| Variable              | Indicador | índice                |
|-----------------------|-----------|-----------------------|
| Variables por modelos | Sexo      | Hembras               |
|                       |           | Machos                |
|                       |           | Dientes de Leche (DL) |
|                       | Edad      | 2 Dientes (2D)        |
|                       |           | 4 Dientes (4D)        |
|                       |           | Boca llena (BLL)      |

Para poder hacer el análisis de regresión logística multinomial se agrupan alpacas por sexo y por edad debido a que en estudios anteriores hallaron diferencias significativas en cuanto a sus calidades de.

## CAPITULO IV

### RESULTADOS Y DISCUSIÓN

#### 4.1. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES

Calidad de Fibra de alpaca: la calidad de fibra según Frank 2005, clasifica a la calidad en cuatro grupos diferenciados, Baby, Fleece, mediana y gruesa. Esta clasificación se realiza mediante el diámetro de fibra.

**Tabla 2: Cantidad y porcentaje de alpacas por comunidad**

|          | Frecuencia | Porcentaje | Porcentaje<br>válido | Porcentaje<br>acumulado |
|----------|------------|------------|----------------------|-------------------------|
| CHIMBOYA | 1054       | 69.1       | 69.1                 | 69.1                    |
| QUELCAYA | 472        | 30.9       | 30.9                 | 100.0                   |
| Total    | 1526       | 100.0      | 100.0                |                         |

Fuente: Gobierno Regional Puno, 2017  
Elaboración Propia

Como se observa en la tabla 2 el 30.9 % de alpacas representa a la comunidad de Quelcaya y el 69.1% a la comunidad de Chimboya. La calidad de fibra de alpaca está distribuida en tres categorías, la categoría con mayor porcentaje es la Fleece con 43.7%, para analizar los modelos se tendrá de dividir los datos, primero por sexo y luego por edad (ver anexo).

**Tabla 3: Estadísticos descriptivos de las características de la fibra de alpaca de raza Huacaya**

|            | FC*     | LM**    | IC*** |      |
|------------|---------|---------|-------|------|
| N          | Válidos | 1526    | 1526  | 1526 |
| Media      | 92,76   | 102,51  | 44,26 |      |
| Mediana    | 96,00   | 100,00  | 44,00 |      |
| Moda       | 99,00   | 100,00  | 47,00 |      |
| Desv. típ. | 8,21    | 36,12   | 8,32  |      |
| Varianza   | 67,50   | 1304,68 | 69,36 |      |
| Mínimo     | 30,0    | 10,00   | 14,00 |      |
| Máximo     | 100,0   | 400,00  | 89,00 |      |

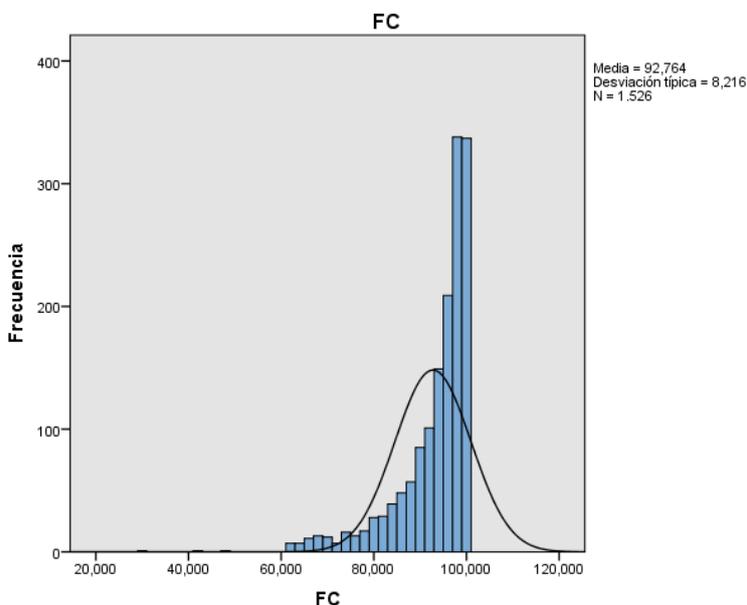
\*Factor de Confort

\*\*Longitud de Mecha

\*\*\*Índice de curvatura

Elaborado por los ejecutores en base a los datos

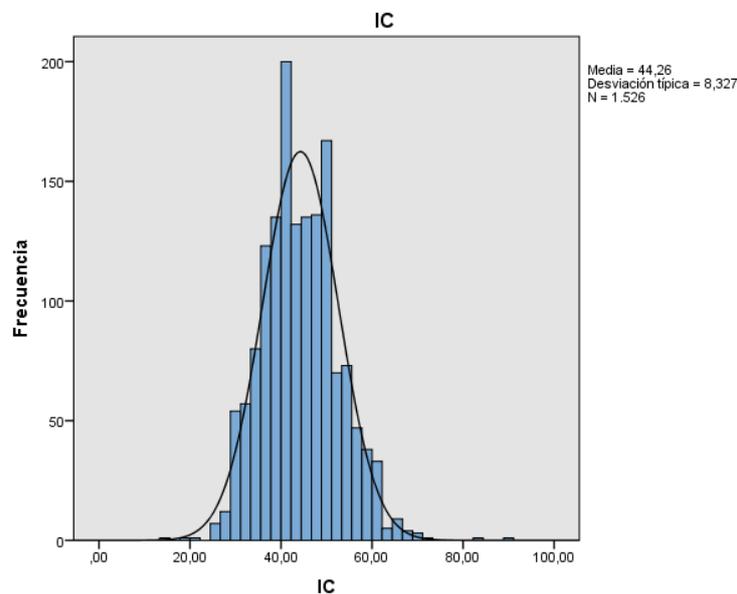
La tabla muestra los estadísticos de las características de fibra de alpaca que son el factor de confort, longitud de mecha e índice de curvatura, en promedio el FC es 92.76%, la LM es 100 mm y el IC de las alpacas es 44.52 %/mm.



**Figura 7: Histograma de factor de confort**

Elaborado por los ejecutores en base a los datos

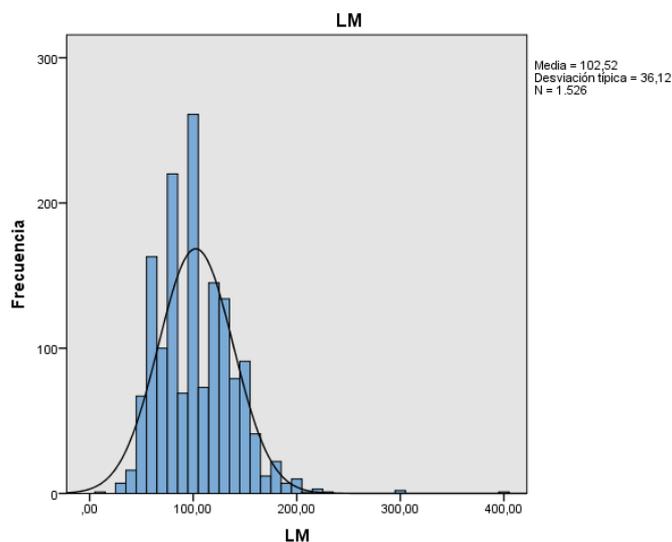
Se puede observar como el histograma para la variable factor de confort presenta una morfología parecida a la distribución de poisson, la distribución coincide con la distribución normal cuando realizamos la aproximación.



**Figura 8: Histograma de corrección de Índice de Curvatura**

Elaborado por los ejecutores en base a los datos

Se puede observar como el histograma para la variable índice de curvatura presenta una morfología muy parecida a una curva de Gauss, con poca asimetría, por tanto, podemos concluir que sigue una distribución normal.



**Figura 9: Histograma de corrección de longitud de mecha**

Elaborado por los ejecutores en base a los datos

Se puede observar como el histograma para la variable longitud de mecha presenta una morfología muy parecida a una curva de Gauss, con asimetría positiva, con una media 102.52 mm y una variación esperada con respecto a la media aritmética de 36,12.

#### **4.2. ANÁLISIS DE VARIANZA PARA LOS FACTORES SEXO Y EDAD.**

##### **4.2.1. EFECTOS DE LA EDAD Y SEXO SOBRE EL DIÁMETRO DE FIBRA.**

Se mide la diferencia significativa de la media del diámetro de fibra de alpaca mediante un análisis de varianza para hallar los modelos de regresión logística multinomial por factor.

Tomando las hipótesis:

$H_0$ : todos los  $A_i$ , los  $B_j$  y los  $AB_{ij}$  son cero.

$H_1$ : alguno de ellos no es cero.

La hipótesis nula supone que, si todos los efectos de todos los niveles de los factores son nulos, todas las medias de todos los grupos considerados son iguales entre sí e iguales a la media general y no hay interacción.

En la tabla 4, el valor de significancia es 0.022 y 0.000 que son menor al 5%, es por ello que se rechaza la hipótesis nula y se acepta la hipótesis alterna, con esta prueba asegura que es posible hallar modelos por sexo y por edad.

En la tabla 4 podemos observar que el sexo y edad de las alpacas tienen efectos altamente significativos respecto a la media del diámetro de fibra (MDF), con ello podemos indicar que podemos realizar los modelos tanto por sexo y por edad.

En la tabla podemos observar que los factores tienen significancia menor al 5%, debido a ello indicamos que influye el sexo y la edad en la MDF

**Tabla 4: Análisis de varianza según edad y sexo**

| Number of obs | =          | 1,526   | R-squared     | =      | 0.2334 |
|---------------|------------|---------|---------------|--------|--------|
| Root MSE      | =          | 2.58209 | Adj R-squared | =      | 0.2314 |
| Source        | Partial SS | df      | MS            | F      | Prob>F |
| Model         | 3087.9506  | 4       | 771.99        | 115.79 | 0.0000 |
| SEXO          | 34.920365  | 1       | 34.920        | 5.24   | 0.0222 |
| EDAD          | 2554.3242  | 3       | 851.44        | 127.71 | 0.0000 |
| Residual      | 10140.774  | 1,521   | 6.6671        |        |        |
| Total         | 13228.724  | 1,525   | 8.6745        |        |        |

Fuente: Gobierno Regional Puno  
Elaborado por los ejecutores en base a los datos

El ajuste del modelo es bajo con un  $R^2 = 0.23$  sin embargo las probabilidades son significativas, es por ello que se realiza un análisis separado tanto para la edad y por sexo.

### **4.3. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA LA CALIDAD DE FIBRA DE ALPACA HUACAYA EN FUNCIÓN A SUS CARACTERÍSTICAS SEGÚN SEXO.**

#### **4.3.1. ESTIMACIÓN E INTERPRETACIÓN DEL MODELO PARA HEMBRAS.**

En la tabla 5 se realiza un ajuste del modelo de regresión logística multinomial que explica la calidad de fibra de alpaca según sus características, que son índice de curvatura, factor de confort y longitud de mecha.

**Tabla 5: Estimación de los parámetros, hembras.**

| Calidad de fibra |    | B        | Error típ. | Wald    | gl | Sig.  | Exp(B) | Intervalo de confianza al 95% para Exp(B) |            |
|------------------|----|----------|------------|---------|----|-------|--------|---|------------|
|                  |    |          |            |         |    |       |        | Límite inf                                | límite sup |
| Baby             | I* | -126.133 | 6.854      | 338.650 | 1  | 0.000 |        |   |            |
|                  | FC | 1.357    | 0.073      | 347.706 | 1  | 0.000 | 3.885  | 3.368                                     | 4.480      |
|                  | LM | -0.008   | 0.004      | 3.182   | 1  | 0.074 | 0.992  | 0.984                                     | 1.001      |
|                  | IC | -0.002   | 0.011      | 0.018   | 1  | 0.893 | 0.998  | 0.977                                     | 1.021      |
| Fleece           | I* | -44.104  | 3.531      | 155.979 | 1  | 0.000 |        |   |            |
|                  | FC | 0.513    | 0.040      | 164.639 | 1  | 0.000 | 1.670  | 1.544                                     | 1.806      |
|                  | LM | -0.005   | 0.003      | 2.230   | 1  | 0.135 | 0.995  | 0.988                                     | 1.002      |
|                  | IC | -0.007   | 0.005      | 1.584   | 1  | 0.208 | 0.993  | 0.983                                     | 1.004      |

\*I: Intersección

Elaborado por los ejecutores en base a los datos

Para las alpacas hembras, existe solo una variable significativa para el modelo “Baby” y “Fleece”, que es el factor de confort, con esta variable se halla el modelo para cada calidad tomando como referencia la calidad “mediana”.

- Modelo para la calidad Baby tomando como referencia la calidad mediana

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -126.133 + 1.357 * FC$$

Según Exp(B), para la categoría “Baby”  $B_1 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 3,885 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana.

- Modelo para la calidad Fleece tomando como referencia la calidad mediana

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -44.104 + 0.513 * FC$$

Para el modelo de la categoría Fleece que, a medida que incremente el factor de confort es 1.670 veces más probable que la calidad de fibra sea Fleece a que a calidad de fibra sea mediana.

BONDAD DEL AJUSTE

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 6: Estadísticos sobre la bondad de ajuste del modelo para hembras**

|                | Chi-cuadrado | gl   | Sig.  |
|----------------|--------------|------|-------|
| <b>Hembras</b> |              |      |       |
| Pearson        | 2,594E+037   | 2056 | 0,930 |
| Desviación     | 996,024      | 2056 | 0,990 |

Elaborado por los ejecutores en base a los datos

En la tabla de bondad de ajuste del modelo las significancias son mayores que 0.05 por lo que las variables que se redujeron están adecuadamente ajustadas al modelo.

TASA DE CLASIFICACIONES

Se puede cuantificar la bondad del ajuste de modelo utilizando la tasa de clasificaciones correctas, que nos permite clasificar cada observación en la categoría más probable, construyendo así una matriz de clasificación.

Con los registros obtenidos se hallaron 51 casos que están clasificados como Fleece y 2 como mediana siendo estos de categoría Baby, para la categoría Fleece se clasificaron 40 registros como Baby y 10 como mediana.

**Tabla 7 Matriz de clasificación observados-predichos para el modelo estimado, para hembras.**

| Observado         | Pronosticado |        |         | % correcto |
|-------------------|--------------|--------|---------|------------|
|                   | Baby         | Fleece | mediana |            |
| Baby              | 305          | 51     | 2       | 85,2%      |
| Fleece            | 40           | 478    | 10      | 90,5%      |
| mediana           | 1            | 32     | 231     | 87,5%      |
| Porcentaje global | 30,1%        | 48,8%  | 21,1%   | 88,2%      |

Elaborado por los ejecutores en base a los datos

Se concluye la eficacia predictiva del modelo, que al ser de 88.2%, se indica que el modelo posee una muy buena predicción de los casos analizados logrando ser correctamente clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

#### CALIDAD DE AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes **Pseudo-R2 de Mc-Fadden, de Cox-Snell y de Nagelkerke**. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo final y del modelo inicial con solo la constante.

**Tabla 8: Coeficientes de medición de ajuste Pseudo R-cuadrado, para hembras.**

| <i>Pseudo R-cuadrado</i> | Hembras |
|--------------------------|---------|
| Cox y Snell              | 0,696   |
| Nagelkerke               | 0,791   |
| MCFadden                 | 0,562   |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta buena calidad de ajuste.

#### 4.3.2. ESTIMACIÓN E INTERPRETACIÓN DEL MODELO PARA MACHOS

En la estimación de los modelos la variable significativa es el factor de confort con niveles de sig=0.000.

**Tabla 9: Estimación de los parámetros, machos**

| Calidad de fibra |    | B        | Error típ. | Wald   | gl | Sig.  | Exp(B) | Intervalo de confianza al 95% para Exp(B) |            |
|------------------|----|----------|------------|--------|----|-------|--------|---|------------|
|                  |    |          |            |        |    |       |        | Límite inf                                | Límite sup |
| Baby             | I* | -251.199 | 31.002     | 65.655 | 1  | 0     |        |   |            |
|                  | FC | 2.711    | 0.338      | 64.473 | 1  | 0     | 15.041 | 7.761                                     | 29.15      |
|                  | LM | -0.023   | 0.018      | 1.684  | 1  | 0.194 | 0.977  | 0.944                                     | 1.012      |
|                  | IC | 0.035    | 0.079      | 0.199  | 1  | 0.655 | 1.036  | 0.887                                     | 1.209      |
| Fleece           | I* | -99.661  | 25.606     | 15.148 | 1  | 0     |        |   |            |
|                  | FC | 1.128    | 0.284      | 15.803 | 1  | 0     | 3.088  | 1.771                                     | 5.385      |
|                  | LM | -0.019   | 0.017      | 1.239  | 1  | 0.266 | 0.982  | 0.95                                      | 1.014      |
|                  | IC | 0.07     | 0.075      | 0.864  | 1  | 0.352 | 1.073  | 0.925                                     | 1.243      |

**I: Intersección**

Elaborado por los ejecutores en base a los datos

- Modelo para la calidad Baby tomando como referencia la calidad mediana

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -251.199 + 2.711 * FC$$

Para la categoría “Baby”  $B_1 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 15,041 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana.

- Modelo para la calidad Fleece tomando como referencia la calidad mediana

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -99.661 + 1.128 * FC$$

Del mismo modo el Exp(B) muestra para el modelo con la calidad Fleece que, si el factor de confort incrementa en una unidad es 3,088 veces más probable que la calidad de fibra sea Fleece a que a calidad de fibra sea media.

**BONDAD DEL AJUSTE**

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 10: Estadísticos sobre la bondad de ajuste del modelo según  
sexo, alpacas machos**

|               | Chi-<br>cuadrado | gl  | Sig.  |
|---------------|------------------|-----|-------|
| <b>Machos</b> |                  |     |       |
| Pearson       | 227,054          | 702 | 0,800 |
| Desviación    | 196,130          | 702 | 0,890 |

Elaborado por los ejecutores en base a los datos

#### TASA DE CLASIFICACIONES CORRECTAS

Según la tabla 11 se han clasificado 16 registros como calidad Fleece siendo de calidad Baby, para la calidad Fleece se han clasificado 23 registros como Fleece y 2 registros como mediana, 4 registros fueron clasificados en la categoría Fleece siendo de la categoría mediana.

**Tabla 11: Matriz de clasificación observados-predichos para el modelo  
estimado según sexo, alpacas machos**

| Observado            | Pronosticado |        |             | % correcto |
|----------------------|--------------|--------|-------------|------------|
|                      | Baby         | Fleece | media<br>na |            |
| Baby                 | 178          | 16     | 0           | 91,8%      |
| Fleece               | 23           | 117    | 2           | 82,4%      |
| mediana              | 0            | 4      | 36          | 90,0%      |
| Porcentaje<br>global | 53,5%        | 36,4%  | 10,1%       | 88,0%      |

Elaborado por los ejecutores en base a los datos

Al ser de 88,0% el porcentaje global, el modelo posee una muy buena predicción, que los casos analizados logran ser correctamente clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

## CALIDAD DE AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes **Pseudo-R<sup>2</sup> de Mc-Fadden, de Cox-Snell** y de **Nagelkerke**. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo final y del modelo inicial con solo la constante.

**Tabla 12: Coeficientes de medición de ajuste Pseudo R-cuadrado, para machos.**

| <i>Pseudo R-cuadrado</i> | Machos |
|--------------------------|--------|
| Cox y Snell              | 0,742  |
| Nagelkerke               | 0,859  |
| MCFadden                 | 0,681  |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta muy buena calidad de ajuste.

#### 4.4. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA LA CALIDAD DE FIBRA DE ALPACA HUACAYA EN FUNCIÓN A SUS CARACTERÍSTICAS SEGÚN EDAD.

Se realiza un ajuste de un modelo de regresión logística multinomial que explique la calidad de fibra de alpaca según sus características, se empieza estimando el modelo con las variables explicativas que son el factor de confort, longitud de mecha e índice de curvatura.

#### 4.4.1. ESTIMACIÓN DEL MODELO PARA DIENTES DE LECHE

En la tabla13 se realiza un ajuste del modelo de regresión logística multinomial que explica la calidad de fibra de alpaca según sus características, que son índice de curvatura, factor de confort y longitud de mecha, para la selección del modelo se toma la significancia del 5%.

**Tabla 13: Estimación de los parámetros, Dientes de leche.**

| Calidad de fibra | B  | Error típ. | Wald   | gl      | Sig. | Exp(B) | Intervalo de confianza al 95% para Exp(B) |            |
|------------------|----|------------|--------|---------|------|--------|---|------------|
|                  |    |            |        |         |      |        | Límite inf                                | Límite sup |
| Baby             | I* | -136,524   | 11,075 | 151,948 | 1    | 0,000  |   |            |
|                  | FC | 1,446      | 0,117  | 153,563 | 1    | 0,000  | 4,247                                     | 3,379      |
|                  | LM | -0,008     | 0,009  | 0,832   | 1    | 0,362  | 00,992                                    | ,010       |
|                  | IC | 0,070      | 0,064  | 1,216   | 1    | 0,270  | 1,073                                     | 0,947      |
| Fleece           | I* | -38,992    | 6,883  | 32,092  | 1    | 0,000  |   |            |
|                  | FC | 0,437      | 0,073  | 36,124  | 1    | 0,000  | 1,547                                     | 1,342      |
|                  | LM | -0,003     | 0,009  | 0,122   | 1    | 0,727  | 0,997                                     | 0,981      |
|                  | IC | 0,058      | 0,061  | 0,918   | 1    | 0,338  | 1,060                                     | 0,941      |

\*I: Intersección

Elaborado por los ejecutores en base a los datos

- Modelo para la calidad Baby tomando como referencia la calidad mediana

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -136.524 + 1.446 * FC$$

Para la categoría “Baby”  $B_1 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 4,247 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana

- Modelo para la calidad Fleece tomando como referencia la calidad mediana

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -38.992 + 0.437 * FC$$

Para el modelo de calidad Fleece el  $\text{Exp}(B) > 0$ , podemos indicar que por cada unidad que incremente el factor de confort es 1.547 veces más probable que la calidad sea Fleece con respecto a que la calidad sea mediana.

### BONDAD DE AJUSTE

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 14: Estadísticos sobre la bondad de ajuste del modelo según edad, DL**

| DL         | Chi-cuadrado | gl        | Sig.  |
|------------|--------------|-----------|-------|
|            | Pearson      | 1,496E+28 | 992   |
| Desviación | 414,527      | 992       | 0,990 |

Elaborado por los ejecutores en base a los datos

En la tabla de bondad de ajuste del modelo las significancias son mayores que 0.05 por lo que las variables que se redujeron están adecuadamente ajustadas al modelo.

### TASA DE CLASIFICACIONES

**Tabla 15: Matriz de clasificación observados-predichos para el modelo estimado, según edad DL**

| Observado         | Pronosticado |        |        | % correcto |
|-------------------|--------------|--------|--------|------------|
|                   | Baby         | Fleece | median |            |
| Baby              | 338          | 13     | 1      | 96,0%      |
| Fleece            | 50           | 150    | 0      | 75,0%      |
| mediana           | 0            | 5      | 25     | 83,3%      |
| Porcentaje global | 66,7%        | 28,9%  | 4,5%   | 88,1%      |

Elaborado por los ejecutores en base a los datos

Se observa que la tasa de clasificaciones correctas es de 88,1%, a través de la misma se puede concluir acerca de la eficacia predictiva del modelo, que al ser de 88,1%, el modelo posee una muy buena predicción, los casos analizados logran ser correctamente clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

#### CALIDAD DE AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes **Pseudo-R<sup>2</sup> de Mc-Fadden, de Cox-Snell** y de **Nagelkerke**. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo Final y del modelo inicial con solo la constante.

**Tabla 16: Coeficientes de medición, ajuste Pseudo R-cuadrado, según edad DL.**

| <i>Pseudo R-cuadrado</i> | Dientes de Leche |
|--------------------------|------------------|
| Cox y Snell              | 0,581            |
| Nagelkerke               | 0,720            |
| MCFadden                 | 0,528            |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta muy buena calidad de ajuste.

#### 4.4.2. ESTIMACIÓN DEL MODELO PARA 2 DIENTES

En la tabla 16 vemos que las variables factor de confort, longitud de mecha e índice de curvatura son seleccionadas mediante la significancia.

**Tabla 17: Estimación de los parámetros para el modelo según edad, 2D.**

| Calidad de fibra |    | B        | Error típ. | Wald   | gl | Sig.  | Exp(B) | Intervalo de confianza al 95% para Exp(B) |            |
|------------------|----|----------|------------|--------|----|-------|--------|---|------------|
|                  |    |          |            |        |    |       |        | Límite inf                                | Límite sup |
| Baby             | I* | -186,511 | 24,068     | 60,054 | 1  | 0,000 |        |   |            |
|                  | FC | 2,003    | 0,257      | 60,706 | 1  | 0,000 | 7,413  | 4,479                                     | 12,270     |
|                  | LM | -0,008   | 0,012      | 0,410  | 1  | 0,522 | 0,992  | 0,969                                     | 1,016      |
|                  | IC | ,002     | 0,017      | 0,010  | 1  | 0,919 | 1,002  | 0,968                                     | 1,037      |
| Fleece           | I* | -68,272  | 14,982     | 20,766 | 1  | 0,000 |        |   |            |
|                  | FC | 0,800    | 0,170      | 22,165 | 1  | 0,000 | 2,225  | 1,595                                     | 3,105      |
|                  | LM | -0,012   | 0,010      | 1,501  | 1  | 0,221 | 0,988  | 0,969                                     | 1,007      |
|                  | IC | -0,014   | 0,011      | 1,564  | 1  | 0,211 | 0,986  | 0,964                                     | 1,008      |

\*I: Intersección

Elaborado por los ejecutores en base a los datos

- Modelo para la calidad Baby tomando como referencia la calidad mediana.

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -186.511 + 2.003 * FC$$

Para la categoría “Baby”  $B_1, B_3 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 7,413 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana.

- Modelo para la calidad Fleece tomando como referencia la calidad mediana

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -68.272 + 0.8 * FC$$

Del mismo modo se puede decir para el modelo con calidad Fleece, si el factor de confort incrementa en una unidad es 2,225 veces más probable que la calidad de fibra sea Fleece a que a calidad de fibra sea media

## BONDAD DE AJUSTE

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 18: Estadísticos sobre la bondad de ajuste del modelo según edad, 2D**

| 2D         | Chi-cuadrado | gl      | Sig.  |
|------------|--------------|---------|-------|
|            | Pearson      | 217,218 | 470   |
| Desviación | 186,825      | 470     | 1,000 |

Elaborado por los ejecutores en base a los datos

En la tabla de bondad de ajuste del modelo las significancias son mayores que 0.05 por lo que las variables que se redujeron están adecuadamente ajustadas al modelo.

## TASA DE CLASIFICACIONES

En la tabla N°18 vemos que 20 registros son clasificados como calidad Fleece siendo de calidad Baby, para la calidad Fleece 18 registros son clasificados como Baby y 4 registros como calidad mediana, para los registros observados como mediana se clasificaron a 6 en la calidad Fleece.

**Tabla 19: Matriz de clasificación observados - predichos para el modelo estimado, Según edad 2D.**

| Observado         | Pronosticado |        |         | % correcto |
|-------------------|--------------|--------|---------|------------|
|                   | Baby         | Fleece | mediana |            |
| Baby              | 63           | 20     | 0       | 75,9%      |
| Fleece            | 18           | 105    | 4       | 82,7%      |
| mediana           | 0            | 6      | 36      | 85,7%      |
| Porcentaje global | 32,1%        | 52,0%  | 15,9%   | 81,0%      |

Elaborado por los ejecutores en base a los datos

El porcentaje global es de 81,0%, el modelo posee una buena predicción, los casos analizados logran ser correctamente clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

#### CALIDAD DE AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes Pseudo-R<sup>2</sup> de Mc-Fadden, de Cox-Snell y de Nagelkerke. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo Final y del modelo inicial con solo la constante.

**Tabla 20: Coeficientes de medición, ajuste Pseudo R-cuadrado, para la edad 2D.**

| <i>Pseudo R-cuadrado</i> | <b>2 Dientes</b> |
|--------------------------|------------------|
| Cox y Snell              | 0,712            |
| Nagelkerke               | 0,821            |
| McFadden                 | 0,617            |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta muy buena calidad de ajuste.

#### 4.4.3. ESTIMACIÓN DEL MODELO PARA 4 DIENTES

Los modelos de regresión estimados por calidad se han seleccionado dos variables según el nivel de significancia menores a 0.05 las cuales son factor de confort e índice de curvatura dejando de lado la variable longitud de mecha.

**Tabla 21: Estimación de los parámetros, para la edad 4D.**

| Calidad de fibra | B  | Error típ. | Wald  | gl     | Sig. | Exp(B) | Intervalo de confianza al<br>95% para Exp(B) |             |
|------------------|----|------------|-------|--------|------|--------|--|-------------|
|                  |    |            |       |        |      |        | Límite inf                                   | Límite sup  |
| Baby             | I* | -96,227    | 9,763 | 97,142 | 1    | 0,000  |  |             |
|                  | FC | 0,953      | 0,105 | 81,771 | 1    | 0,000  | 2,594  | 2,110 3,189 |
|                  | LM | -0,013     | 0,008 | 2,564  | 1    | 0,109  | 0,987  | 0,971 1,003 |
|                  | IC | 0,196      | 0,050 | 15,716 | 1    | 0,000  | 1,217  | 1,104 1,341 |
| Fleece           | I* | -45,739    | 5,843 | 61,281 | 1    | 0,000  |  |             |
|                  | FC | 0,481      | 0,066 | 53,118 | 1    | 0,000  | 1,618  | 1,422 1,842 |
|                  | LM | -0,012     | 0,006 | 3,445  | 1    | 0,063  | 0,988  | 0,976 1,001 |
|                  | IC | 0,111      | 0,041 | 7,372  | 1    | 0,007  | 1,117  | 1,031 1,210 |

\*I: Intersección

Elaborado por los ejecutores en base a los datos

Para las alpacas de 4 dientes, existe dos variables significativas para el modelo “Baby” y “Fleece”, que es el factor de confort e índice de curvatura, con estas variables se halla el modelo para cada calidad tomando como referencia la calidad “mediana”.

- Modelo para la calidad Baby tomando como referencia la calidad mediana

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -96.227 + 0.953 * FC + 0,196 * IC$$

Para la categoría “Baby”  $B_1, B_2 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 2,594 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana, para el índice de curvatura indicamos que por cada unidad que incremente el índice de curvatura es 1,217 veces más probable que sea de la calidad Baby comparado con que la calidad sea mediana.

- Modelo para la calidad Fleece tomando como referencia la calidad mediana

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -45.739 + 0.481 * FC + 0,111 * IC$$

Del mismo indicamos para el modelo con calidad Fleece con lo cual se indica que, si el factor de confort incrementa en una unidad es 1,422 veces más probable que la calidad de fibra sea Fleece a que a calidad de fibra sea media y si el índice de curvatura incrementa en una unidad, es 1.031 veces más probable que la calidad de fibra sea Fleece con respecto a que la calidad de fibra sea media.

#### BONDAD DE AJUSTE DEL MODELO

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 22: Estadísticos sobre la bondad de ajuste del modelo según edad 4D.**

|            | Chi-<br>cuadrado | gl  | Sig.  |
|------------|------------------|-----|-------|
| <b>4D</b>  |                  |     |       |
| Pearson    | 1,599E+24        | 710 | 0,870 |
| Desviación | 348,886          | 710 | 0,940 |

Elaborado por los ejecutores en base a los datos

En la tabla de bondad de ajuste del modelo las significancias son mayores que 0.05 por lo que las variables que se redujeron están adecuadamente ajustadas al modelo.

#### TASA DE CLASIFICACIONES CORRECTAS

Según la tasa de clasificación se observa que 23 registros son clasificados como Fleece y 1 registro como mediana siendo de la calidad

Baby, 7 registros son clasificados como Baby y 9 como mediana siendo de calidad Fleece, para la calidad mediana se han clasificado como Fleece 11 registros que son de calidad mediana.

**Tabla 23: Matriz de clasificación observados - predichos para el modelo estimado, Según edad 4D.**

| Observado         | Pronosticado |        |         | % correcto |
|-------------------|--------------|--------|---------|------------|
|                   | Baby         | Fleece | mediana |            |
| Baby              | 50           | 23     | 1       | 67,6%      |
| Fleece            | 7            | 177    | 9       | 91,7%      |
| Mediana           | 0            | 11     | 95      | 89,6%      |
| Porcentaje global | 15,3%        | 56,6%  | 28,2%   | 86,3%      |

Elaborado por los ejecutores en base a los datos

El porcentaje de clasificación global al ser de 86,3%, se puede decir que el modelo posee una buena predicción, que los casos analizados logran ser clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

#### CALIDAD DE AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes **Pseudo-R2 de Mc-Fadden, de Cox-Snell** y de **Nagelkerke**. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo Final y del modelo inicial con solo la constante.

**Tabla 24: Coeficientes de medición, ajuste Pseudo R-cuadrado, según edad 4D.**

| <i>Pseudo R-cuadrado</i> | 4 Dientes |
|--------------------------|-----------|
| Cox y Snell              | 0,658     |
| Nagelkerke               | 0,757     |
| McFadden                 | 0,527     |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta muy buena calidad de ajuste.

#### 4.4.4. ESTIMACIÓN DEL MODELO PARA BOCA LLENA

Para los modelos de alpacas de boca llena según el nivel de significancia (sig.<0.05) se considera el factor de confort con el nivel de significancia 0.000 para las demás variables su significancia es mayor al 0.05 por ello no se considera en el modelo.

**Tabla 25: Estimación de los parámetros según edad BLL.**

| Calidad de fibra | B  | Error típ. | Wald   | gl     | Sig. | Exp(B) | Intervalo de confianza al 95% para Exp(B) |              |
|------------------|----|------------|--------|--------|------|--------|---|--------------|
|                  |    |            |        |        |      |        | Límite inf                                | Límite sup   |
| Baby             | I* | -266,342   | 46,491 | 32,821 | 1    | 0,000  |   |              |
|                  | FC | 2,714      | 0,482  | 31,715 | 1    | 0,000  | 15,094                                    | 5,869 38,820 |
|                  | LM | -0,010     | 0,013  | 0,607  | 1    | 0,436  | 0,990                                     | 0,965 1,015  |
|                  | IC | 0,161      | 0,068  | 5,589  | 1    | 0,118  | 1,175                                     | 1,028 1,343  |
| Fleece           | I* | -52,275    | 7,026  | 55,357 | 1    | 0,000  |   |              |
|                  | FC | 0,563      | 0,080  | 49,649 | 1    | 0,000  | 1,756                                     | 1,501 2,053  |
|                  | LM | 0,002      | 0,006  | 0,062  | 1    | 0,804  | 1,002                                     | 0,989 1,014  |
|                  | IC | 0,057      | 0,042  | 1,823  | 1    | 0,177  | 1,059                                     | 0,975 1,150  |

\*I: Intersección

Elaborado por los ejecutores en base a los datos

Para las alpacas de boca llena, existe una variable significativa para el modelo “Baby” y “Fleece”, la variable que influye positivamente es el factor de confort, con estas variables se halla el modelo para cada calidad tomando como referencia la calidad “mediana”.

- Modelo para la calidad Baby tomando como referencia la calidad mediana

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -266.342 + 2,714 * FC$$

Para la categoría “Baby”  $B_1 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 15,094 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana.

- Modelo para la calidad Fleece tomando como referencia la calidad mediana

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -52.275 + 0.563 * FC$$

Para el modelo de calidad Fleece decimos, si el factor de confort incrementa en una unidad es 1.756 veces más probable que la calidad de fibra sea Fleece a que a calidad de fibra sea media

#### BONDAD DEL AJUSTE

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 26: Estadísticos sobre la bondad de ajuste del modelo según edad BLL**

|            | Chi-<br>cuadrado | gl  | Sig.  |
|------------|------------------|-----|-------|
| <b>BLL</b> |                  |     |       |
| Pearson    | 9491,674         | 608 | 0,980 |
| Desviación | 179,633          | 608 | 1,000 |

Elaborado por los ejecutores en base a los datos

En la tabla de bondad de ajuste del modelo las significancias son mayores que 0.05 por lo que las variables que se redujeron están adecuadamente ajustadas al modelo.

## TASA DE CLASIFICACIONES CORRECTAS

En la tabla de tasas de clasificación 9 registros son clasificados como Fleece siendo Baby, para la calidad Fleece 1 registro es clasificado como Baby y 9 registros como mediana, en cuanto a la calidad mediana 12 registros son clasificados como Fleece siendo de calidad mediana.

**Tabla 27: Matriz de clasificación observados-predichos para el modelo estimado, Según edad BLL**

| Observado            | Pronosticado |        |             | % correcto |
|----------------------|--------------|--------|-------------|------------|
|                      | Baby         | Fleece | media<br>na |            |
| Baby                 | 34           | 9      | 0           | 79,1%      |
| Fleece               | 1            | 140    | 9           | 93,3%      |
| mediana              | 0            | 12     | 114         | 90,5%      |
| Porcentaje<br>global | 11,0%        | 50,5%  | 38,6%       | 90,3%      |

Elaborado por los ejecutores en base a los datos

Se observa que la tasa de clasificaciones correctas es de 91 %, a través de la misma se puede concluir acerca de la eficacia predictiva del modelo, que al ser de 91%, se puede decir que el modelo posee una muy buena predicción, que un 91% de los casos analizados logran ser correctamente clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

## CALIDAD DE AJUSTE DEL MODELO

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes **Pseudo-R2 de Mc-Fadden, de Cox-Snell** y de **Nagelkerke**. El cálculo de éstos depende del valor de las funciones de log-verosimilitud del modelo Final y del modelo inicial con solo la constante.

**Tabla 28: Coeficientes de medición de ajuste Pseudo R-cuadrado para la edad BLL**

| <i>Pseudo R-cuadrado</i> | Boca Llena |
|--------------------------|------------|
| Cox y Snell              | 0,754      |
| Nagelkerke               | 0,874      |
| McFadden                 | 0,707      |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta muy buena calidad de ajuste.

#### 4.5. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL PARA DETERMINAR LA CALIDAD DE FIBRA DE ALPACA HUCAYA

Como resultado Final y para alcanzar el objetivo principal de este estudio, se realiza un ajuste de un modelo de regresión logística multinomial a nivel general de las alpacas de raza Huacaya.

**Tabla 29: Estimación de los parámetros para el modelo general**

| Calidad de fibra | B  | Error típ. | Wald  | gl      | Sig. | Exp(B) | Intervalo de confianza al 95% para Exp(B) |             |
|------------------|----|------------|-------|---------|------|--------|---|-------------|
|                  |    |            |       |         |      |        | Límite inf                                | Límite sup  |
| Baby             | I* | -142,321   | 6,659 | 456,779 | 1    | 0,000  |   |             |
|                  | FC | 1,536      | 0,071 | 466,828 | 1    | 0,000  | 4,647                                     | 4,042 5,341 |
|                  | LM | -0,009     | 0,004 | 4,659   | 1    | 0,061  | 0,991                                     | 0,983 0,999 |
|                  | IC | -0,014     | 0,011 | 1,617   | 1    | 0,203  | 0,986                                     | 0,964 1,008 |
| Fleece           | I* | -47,578    | 3,518 | 182,909 | 1    | 0,000  |   |             |
|                  | FC | 0,554      | 0,040 | 193,105 | 1    | 0,000  | 1,740                                     | 1,609 1,881 |
|                  | LM | -0,006     | 0,003 | 3,381   | 1    | 0,066  | 0,994                                     | 0,988 1,000 |
|                  | IC | -0,007     | 0,005 | 1,768   | 1    | 0,184  | 0,993                                     | 0,982 1,003 |

I\*: Intersección

Elaborado por los ejecutores en base a los datos

Para las alpacas de raza Huacaya de las comunidades de Quelcaya y Chimboya, existe solo una variable significativa para el modelo con la categoría “Baby” y “Fleece” que es el factor de confort e influye positivamente en el modelo, con estas variables se halla el modelo para cada calidad tomando como referencia la calidad “mediana”.

#### 4.5.1. ESTIMACIÓN DEL MODELO E INTERPRETACIÓN DEL MODELO

Para la categoría “Baby”  $B_1 > 0$ , por lo tanto, por cada unidad que aumente el factor de confort, es aproximadamente 4,647 veces más probable que la calidad fibra sea Baby a que la calidad sea mediana. Del mismo modo para el modelo de calidad Fleece se puede decir que, si el factor de confort incrementa en una unidad es 1,740 veces más probable que la calidad de fibra sea Fleece a que a calidad de fibra sea media.

$$g_1(x) = \ln\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i3}}\right) = -142.321 + 1.536 * FC$$

$$g_2(x) = \ln\left(\frac{\hat{\pi}_{i2}}{\hat{\pi}_{i3}}\right) = -47.578 + 0.554 * FC$$

#### 4.5.2. BONDAD DEL AJUSTE

La bondad y ajuste de este modelo logístico multinomial se puede contrastar con la información de la siguiente tabla.

**Tabla 30: Estadísticos sobre la bondad de ajuste del modelo general**

|            | Chi-cuadrado | gl   | Sig.  |
|------------|--------------|------|-------|
| Pearson    | 3,256E+042   | 2610 | 0,000 |
| Desviación | 1177,680     | 2610 | 0,980 |

Elaborado por los ejecutores en base a los datos

En la tabla de bondad de ajuste del modelo las significancias son mayores que 0.05 por lo que las variables que se redujeron están adecuadamente ajustadas al modelo.

## TASA DE CLASIFICACIONES CORRECTAS

**Tabla 31: Matriz de clasificación para el modelo general estimado**

| Observado            | Pronosticado |        |         |                        |
|----------------------|--------------|--------|---------|------------------------|
|                      | Baby         | Fleece | mediana | Porcentaje<br>correcto |
| Baby                 | 499          | 51     | 2       | 90,4%                  |
| Fleece               | 73           | 585    | 12      | 87,3%                  |
| mediana              | 1            | 38     | 265     | 87,2%                  |
| Porcentaje<br>global | 37,5%        | 44,2%  | 18,3%   | 88,4%                  |

Elaborado por los ejecutores en base a los datos

Se observa que la tasa de clasificaciones correctas es de 88.4%, a través de la misma se puede concluir, que al ser de 88.4%, el modelo posee una buena predicción, que los casos analizados logran ser clasificados, al coincidir la calidad de fibra observado con el pronosticado por el modelo.

**4.5.3. CALIDAD DE AJUSTE DEL MODELO**

Para medir la calidad del ajuste del modelo se utiliza como se muestra, los coeficientes Pseudo-R<sup>2</sup> de Mc-Fadden, de Cox-Snell y de Nagelkrke.

**Tabla 32: Coeficiente de medición de ajuste pseudo R-cuadrado para el modelo general**

| Pseudo R-cuadrado |       |
|-------------------|-------|
| Cox y Snell       | 0,712 |
| Nagelkerke        | 0,812 |
| MCFadden          | 0,793 |

Elaborado por los ejecutores en base a los datos

Se puede observar que los coeficientes pseudo-R<sup>2</sup>, alcanzan valores altos cercanos a 1 por lo que se puede concluir que el modelo presenta muy buena calidad de ajuste.

#### 4.5.4. VALIDACIÓN Y PREDICCIÓN DE LOS MODELOS

Un modelo predictivo trata de ver si puede predecir bien la variable dependiente en un modelo de calidad de fibra de alpaca de raza Huacaya. Por ello se realiza predicciones con probabilidades según a los modelos hallados, si consideramos el FC de 95% y para el caso de 4D se toma el índice de curvatura de 38 °/mm para cada modelo tenemos las siguientes probabilidades:

**Tabla 33: Estimación de las probabilidades para cada modelo general**

| Calidad de fibra<br>(Probabilidad) | Hembras | Machos | DL   | 2D   | 4D   | BLL  | Modelo General |
|------------------------------------|---------|--------|------|------|------|------|----------------|
| Baby                               | 0.13    | 0.24   | 0.14 | 0.02 | 0.00 | 0.00 | 0.19           |
| Fleece                             | 0.86    | 0.76   | 0.80 | 0.98 | 0.98 | 0.77 | 0.81           |
| Mediana                            | 0.01    | 0.00   | 0.06 | 0.00 | 0.02 | 0.23 | 0.01           |

Elaborado por los ejecutores en base a los datos

En la tabla N°33 vemos la estimación para cada modelo, mostrando que, si tenemos un factor de confort del 95% y para el caso del modelo de 4D el IC de 38 °/mm, existe según el modelo general mayor probabilidad de que las alpacas estén consideradas con calidad de fibra Fleece, según la estimación si se analiza por modelo hay menores probabilidades en alpacas machos y de BLL.

#### ANÁLISIS

El modelo de regresión logística es de gran importancia en la parte aplicativa debido a que facilita la modelación de una variable de naturaleza cualitativa en términos de un conjunto que tiene más de dos categorías de carácter nominal. Como se utiliza el modelo de regresión logística multinomial,

se deben estimar tantos modelos como categorías menos una tenga las variables respuesta, digamos  $g$ , y en consecuencia se deben estimar  $g-1$  vectores regresores.

Para los modelos según sexo se consideró la variable factor de confort con ello se realiza la estimación de probabilidades de que la calidad de fibra sea Baby son 99% y 82% para hembras y machos respectivamente considerando un factor de confort del 99%, en alpacas hembras existe mayor probabilidad que sean de calidad Baby, en los estudios de Quispe (2015) y Mamani (2012) muestran que no hay diferencia entre las calidades de fibra en hembras y machos.

El factor de confort es la variable significativa según edad DL, 2D, 4D y BLL excepto para la edad de 4D que a su vez incluye la variable índice de curvatura en su modelo; si consideramos el FC 99%, las probabilidades de que su calidad de fibra sea Baby son 90%, 80%, 71% y 24% por edad respectivamente según lo estimado por los modelos, en relación con el estudio de Ormachea (2010) en donde indica que diámetro de fibra incrementa conforme la edad del animal avanza, se hallaron resultados similares ya que se tiene mayor probabilidad de obtener fibra con calidad Baby en alpacas jóvenes de DLL y va disminuyendo su probabilidad conforme avanza la edad.

## CAPITULO V

### CONCLUSIONES

1. Los modelos obtenidos para la fibra de alpaca tomando como referencia la calidad de fibra mediana son: hembras fibra Baby  $g_1(x) = -126.133 + 1.357 * FC$ , hembras fibra Fleece  $g_2(x) = -44,104 + 0.513 * FC$ . machos fibra Baby  $g_1(x) = -251.199 + 2.711 * FC$ , Machos fibra Fleece  $g_2(x) = -99.661 + 1.128 * FC$ , con probabilidades al 0.05 por lo que los modelos tienen buen ajuste con tasas de clasificación 88.2% y 88% respectivamente.
2. Los modelos obtenidos por edad son:  $g_1(x) = -136.524 + 1.446 * FC$  y  $g_2(x) = -38.992 + 0.437 * FC$  para la edad dientes de leche.  $g_1(x) = -186.511 + 2.003 * FC$  y  $g_2(x) = -68.272 + 0.8 * FC$  para la edad de 2 dientes.  $g_1(x) = -96.227 + 0.953 * FC + 0,196 * IC$  y  $g_2(x) = -45.739 + 0.481 * FC + 0,111 * IC$  para la edad de 4 dientes. Para las alpacas de boca llena los modelos son  $g_1(x) = -266.342 + 2,714 * FC$  y  $g_2(x) = -52.275 + 0.563 * FC$  con probabilidades mayores 0.05, tasas de clasificación de 88.1%, 81%, 81%, 86.3% y 91% por edad respectivamente, disminuyendo las probabilidades conforma incrementa la edad.
3. El modelo de regresión logística multinomial permite relacionar la variable calidad de fibra con las variables independientes referidas al factor de confort, longitud de mecha e índice de confort, hallando modelos por sexo y edad de las alpacas y modelos para todos los datos en conjunto los cuales son:

$g_1(x) = -142.321 + 1.536 * FC$  para la calidad Baby y  $g_2(x) = -47.578 + 0.554 * FC$  para la calidad Fleece, según los resultados la variable que en mayor medida permite clasificar a una determinada alpaca con la calidad de fibra Baby, Fleece y media es el Factor de Confort, exceptuando el caso para las alpacas de 4 dientes que en el modelo consideramos el Factor de Confort e Índice de Curvatura, indicando que a medida que el factor de confort incrementa es 4.647 veces más probable que la calidad sea Baby a que la calidad sea Fleece que es 1.740 veces más probable. La probabilidad de que la calidad de fibra sea Baby va incrementando a medida que el Factor de confort es más alto.

4. Mediante el análisis de regresión logística multinomial resultan modelos específicos por sexo y por edad con ellos es posible estimar la calidad de fibra a medida que incrementa la edad del animal si consideramos las variables significativas.

## CAPITULO VI

### RECOMENDACIONES

1. Se recomienda utilizar los resultados obtenidos como referencia, para clasificar las calidades de fibra en las comunidades de Quealcaya y Chimboya y así mejorar los niveles de ingresos.
2. La categorización y clasificación de la calidad de fibra por sexo beneficia a las unidades familiares alpaqueras sin embargo es necesario realizar estudios considerando factores productivos y biológicos para tener una mejor clasificación. Contemplar en próximos estudios otras características como rendimiento al lavado y contenido de materia vegetal.
3. Para estudios posteriores se recomienda estimar el factor de confort por edad y con ello hacer inferencia conforme incrementa la edad de la alpaca ya que con esto se podría reducir costos de los análisis de fibra. Debido a que el mercado internacional requiere fibras con alto confort que se logra con fibras más Finas.
4. Se recomienda realizar futuras investigaciones en las que se puedan evaluar con la misma metodología, los parámetros de otras zonas y poder contrastar los resultados y medir si existen diferencias.

5. Se debe incentivar a los estudiantes a la utilización del análisis multivariado aplicado debido a que se puede considerar gran cantidad de variables y así poder analizar mejor los estudios para contrastar la teoría.

## CAPITULO VII

### REFERENCIAS

- AGROPUNO. (2013). Dirección De Estadística Agraria e Informática. Recuperado de <http://www.agropuno.gob.pe/?q=node/1041>
- Aucancela Quishpi, B. (2015). *Caracterización de la fibra de Vicugna Pacos (Alpaca) de la parroquia San Juan, Provincia de Chimborazo.*
- Bustanza, A. (2001). *La alpaca, conocimiento del gran potencial andino.* Puno. Univ. Nacional del Altiplano.
- Calle Escobar, R. (1982). *Producción y mejoramiento de la producción de la alpaca.* Fondo del Libro Banco Agrario del Perú. Lima.
- Canaza, A. W. (2009). *Evaluación Cualitativa y Cuantitativa de la Fibra de Alpaca Mediante Espectroscopia de Reflectancia en el Infrarrojo Cercano (NIRS).*
- CENAGRO. (2012). IV Censo Nacional Agropecuario.
- De Groot, G. J. (1995). The Effect of coefficient of Variation of Fibre diameter in Wool tops on Yarn and Fabric Properties.
- Dueñas, M. (2015). *Modelos de respuesta discreta en R y aplicación con datos reales.*

FAO. (2005). Situación actual de los camélidos sudamericanos en Perú, 63.

Fernández, A. (2015). No más intermediación en Corani: los productores acopian y negocian directamente fibra de alpaca. Recuperado de: <http://infoalpacas.com.pe/wp-content/uploads/2015/03/no-mas-intermediacion-en-corani-los-productores-acopian-y-negocian-directamente-fibra-de-alpaca.pdf>

Fish, V. E., Mahar, T. J., & Crook, B. J. (1999). *Fibre curvature morphometry and measurement. International Wool Textile Organization*. [https://doi.org/Nice Meeting](https://doi.org/NiceMeeting)

FLORES QUISPE, W. (2017). *Perfil de fibra, índice de confort e índice de curvatura en alpacas Huacaya del distrito de Corani-Carabaya*. UNIVERSIDAD NACIONAL DEL ALTIPLANO. Recuperado de <http://repository.javeriana.edu.co/>

Franco, F., & San Martín, F. (2007). Efecto del Nivel Alimenticio sobre el rendimiento y calidad de fibra en alpacas. Sistema de revisiones en Investigación Veterinaria en UNMSM. Recuperado de <https://books.google.com.pe/books?id=coEbKTddsoC&pg=PA64&lpg=PA64&dq=Russel+A.J.+y+Redden+H.L.+1997.+The+effect+of+nutrition+on+fibre+growth+in+the+alpaca>.

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.

- Gómez, S., & Palacios, D. (2013). *Modelacion logística multinomial para clasificar los hogares de el salvador*. Universidad de El Salvador.
- Gonzales, A. (2015). *Selección de variables: Una revisión de métodos existentes*. Universidad de Coruña.
- 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.
- Hack, W. & R. I. R. and D. C. (1999). Australian alpaca fibre : improving productivity and marketing : a report for the Rural Industries Research and Development Corporation.
- Huanca, T. (2003). Resúmenes de proyectos de investigación en camélidos sudamericanos periodo 1997-2003, 1–5.
- INDECOPI, C. de N. y de F. de B. C. N. A. (2010). Tecnología Pecuaria, Buenas prácticas de esquila y manejo del vellón de la fibra de alpaca.
- INFOALPACAS. (2012). Comunidades Alpaqueras. Recuperado de <http://infoalpacas.com.pe/comunidades-alpaqueras/>
- INIA. (2014). Camélidos. *Camelidos Sudamericanos*. Recuperado de <http://www.inia.gob.pe/files/crianzas/camelidos.pdf>
- Lupton, C. J., McColl, A., & Stobart, R. H. (2006). Fiber characteristics of the Huacaya Alpaca. *Small Ruminant Research*, 64(3), 211–224. <https://doi.org/10.1016/j.smallrumres.2005.04.023>

- Mamani Siña, M. A. (2015). *Características físicas de la fibra en alpacas Huacaya del distrito de Susapaya, Provincia de Tarata*. UNIVERSIDAD NACIONAL JORGE BASADRE GROHMANN – TACNA.
- McColl, A. (2004). Methods for measuring microns. *Alpacas Magazine, Herd Sire*.
- McGregor, B. A. (2006). Production attributes and relative value of alpaca fleeces in southern Australia and implications for industry development. *Small Rumin. Res.*, 61: 93-111.
- McGregor B.A. y Butler K.L. (2004). Sources of variation in fibre diameter attributes of Australian alpacas and implications for fleece evaluation and animal selection., 433–442.
- Morante R., Goyache F., Burgos A., Cervantes I., P.-C. M. A. y G. J. P. (2009). Genetic improvement for alpaca fibre production in the Peruvian Altiplano: the Pacamarca experience.
- Mueller, P. (2007). Novedades en la determinación de diámetros de fibra y su relevancia en programas de selección INTA Bariloche.
- Ormachea, E., & Olarte, U. (2010). Características textiles de la fibra en alpacas Huacaya del distrito de Corani, Carabaya - Puno, 5–10.
- Pando Fernández, V., & Fernández San Martín, R. (2004). Regresión logística multinomial, *000*, 323–327.
- Ponzoni R.W., R.J. Grimson, J.A. Hill, D.J. Hubbard, B.A. McGregor, A. Howse, I., & Judson., C. y G. J. (1999). The inheritance of and association among some production traits in young Australian alpacas.

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

Rozas Diaz, A. J. (2014). Principales características de la fibra de alpacas Huacaya y suri del sector Chocoquilla - Carabaya.

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

Sánchez, E., & Gómez, D. (2008). *Modelos de eleccion discreta, Revisión y aplicación mediante cuadratura gaussiana*. Universidad Eafit.

Sangurima, Y., & Sandoya, F. (2015). Uso de regresion multinomial logistica para determinar el perfil del estudiante de la ESPOL, segun preferencias alimenticias.

Siguayro, R. (2009). *Comparación de las características físicas de las fibras de llama cha'ku (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.

Sommerville, P. (2000). Measuring Micron: Different Methods can mean different results.

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

Vásquez, R., Gómez, O., & Quispe, E. (2015). Características Tecnológicas de la

Fibra Blanca de Alpaca Huacaya en la Zona Altoandina de Apurímac.

Villarroel, J. (1963). Un estudio de la fibra de alpaca. *Anales Científicos*.

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

Zarate, A. (1992). *Estudio Tecnológico de la lana y fibra, Pieles y Cueros del Programa de Ovinos y Camelidos Americanos*. Universidad Agraria La Molina.

# ANEXOS

### Anexo A: Análisis de varianza con interacción para determinar si existe diferencia significativa

|                |            |               |           |        |        |
|----------------|------------|---------------|-----------|--------|--------|
| Number of obs= | 1,526      | R-squared     | =         | 0.2356 |        |
| Root MSE =     | 2.58098    | Adj R-squared | =         | 0.2321 |        |
| Source         | Partial SS | dF            | MS        | F      | Prob>F |
| Model          | 3116.6089  | 7             | 445.23    | 66.84  | 0.0000 |
| SEXO           | 39.754859  | 1             | 39.75     | 5.97   | 0.0147 |
| EDAD           | 1894.0903  | 3             | 631.36    | 94.78  | 0.0000 |
| SEXO#EDAD      | 28.658352  | 3             | 9.55      | 1.43   | 0.2311 |
| Residual       | 10112.115  | 1,518         | 6.66      |        |        |
| Total          | 13228.724  | 1,525         | 8.6745732 |        |        |

### Anexo B: Tabla de frecuencia: calidad de fibra de alpaca de raza Huacaya color blanco.

| Calidad de fibra | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|------------------|------------|------------|-------------------|----------------------|
| Baby             | 552        | 36,2       | 36,2              | 36,2                 |
| Fleece           | 670        | 43,9       | 43,9              | 80,1                 |
| mediana          | 304        | 19,9       | 19,9              | 100,0                |
| Total            | 1526       | 100,0      | 100,0             |                      |

### Anexo C: Tabla de frecuencia según sexo de alpacas de raza Huacaya color blanco.

| Sexo  | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|-------|------------|------------|-------------------|----------------------|
| H     | 1150       | 75,4       | 75,4              | 75,4                 |
| M     | 376        | 24,6       | 26,4              | 100,0                |
| Total | 1526       | 100,0      | 100,0             |                      |

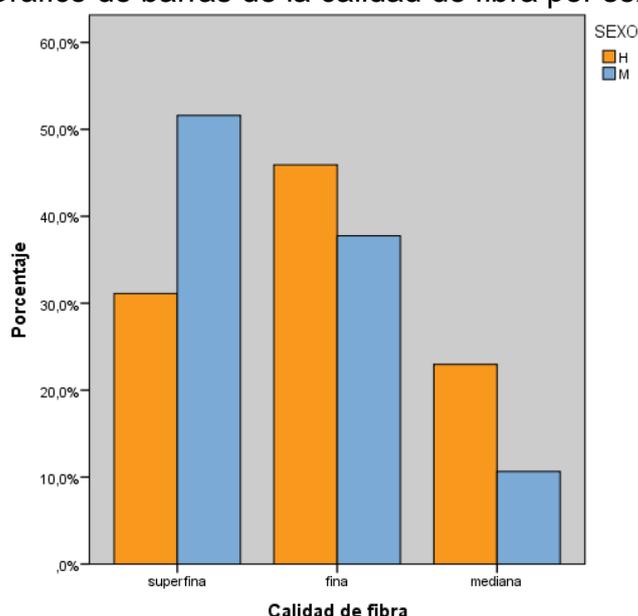
### Anexo D: Tabla de frecuencia según edad de alpacas de raza Huacaya color blanco.

| Edad  | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|-------|------------|------------|-------------------|----------------------|
| DL    | 582        | 38,1       | 38,1              | 38,1                 |
| 2D    | 252        | 16,5       | 16,5              | 54,7                 |
| 4D    | 373        | 24,4       | 24,4              | 79,1                 |
| BLL   | 319        | 20,9       | 20,9              | 100,0                |
| Total | 1526       | 100,0      | 100,0             |                      |

Anexo E: Tabla de contingencia de la calidad de fibra con el diámetro de fibra y sexo

|                            |         | SEXO |     | Total |
|----------------------------|---------|------|-----|-------|
|                            |         | H    | M   |       |
| Calidad de fibra de alpaca | Baby    | 358  | 194 | 552   |
|                            | Fleece  | 528  | 142 | 670   |
|                            | mediana | 264  | 40  | 304   |
| Total                      |         | 1150 | 376 | 1526  |

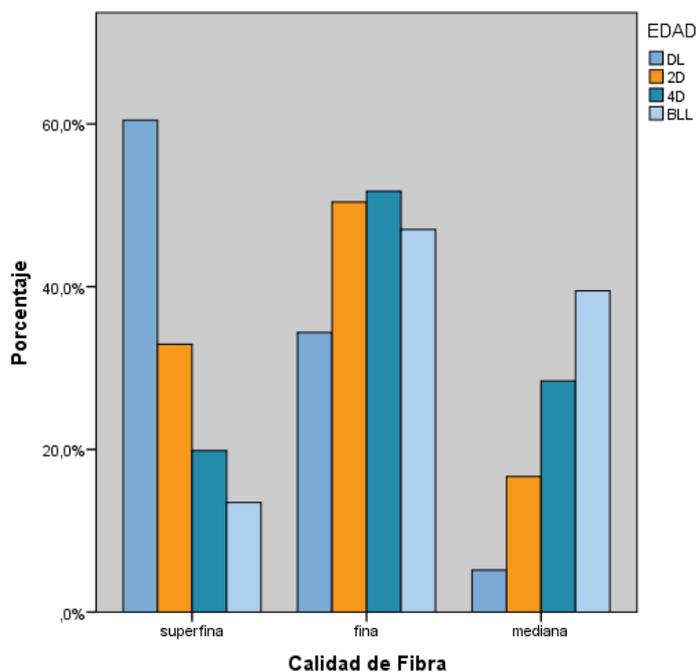
Anexo F: Gráfico de barras de la calidad de fibra por sexo en porcentaje



Anexo G: Tabla de contingencia de la calidad de fibra con el diámetro de fibra y edad

|                  |         | EDAD |     |     |     | Total |
|------------------|---------|------|-----|-----|-----|-------|
|                  |         | DL   | 2D  | 4D  | BLL |       |
| Calidad de Fibra | Baby    | 352  | 83  | 74  | 43  | 552   |
|                  | Fleece  | 200  | 127 | 193 | 150 | 670   |
|                  | mediana | 30   | 42  | 106 | 126 | 304   |
| Total            |         | 582  | 252 | 373 | 319 | 1526  |

Anexo H: Gráfico de barras de la calidad de fibra por edad en porcentaje



Anexo I: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya machos.

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |      |      |      |      |      |      |      |      |
|------------------------------------|-------------------|------|------|------|------|------|------|------|------|------|------|
|                                    | 90                | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  |
| Superfina                          | 0                 | 0    | 0.01 | 0.03 | 0.06 | 0.13 | 0.27 | 0.46 | 0.66 | 0.82 | 0.91 |
| Fina                               | 0.89              | 0.92 | 0.95 | 0.95 | 0.92 | 0.86 | 0.73 | 0.54 | 0.34 | 0.18 | 0.09 |
| Mediana                            | 0.11              | 0.07 | 0.04 | 0.03 | 0.02 | 0.01 | 0    | 0    | 0    | 0    | 0    |

Anexo J: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya hembras

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |      |      |      |      |      |      |      |      |
|------------------------------------|-------------------|------|------|------|------|------|------|------|------|------|------|
|                                    | 90                | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  |
| Superfina                          | 0.00              | 0.00 | 0.00 | 0.01 | 0.06 | 0.24 | 0.61 | 0.88 | 0.97 | 0.99 | 1.00 |
| Fina                               | 0.87              | 0.95 | 0.98 | 0.98 | 0.94 | 0.76 | 0.39 | 0.12 | 0.03 | 0.01 | 0.00 |
| Mediana                            | 0.13              | 0.05 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Anexo K: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya según edad: dientes de leche

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |      |      |      |      |      |      |      |      |
|------------------------------------|-------------------|------|------|------|------|------|------|------|------|------|------|
|                                    | 90                | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  |
| Superfina                          | 0                 | 0    | 0.01 | 0.02 | 0.05 | 0.14 | 0.31 | 0.55 | 0.77 | 0.9  | 0.96 |
| Fina                               | 0.58              | 0.68 | 0.77 | 0.82 | 0.84 | 0.8  | 0.66 | 0.43 | 0.22 | 0.09 | 0.04 |
| Mediana                            | 0.42              | 0.31 | 0.23 | 0.16 | 0.1  | 0.06 | 0.03 | 0.01 | 0    | 0    | 0    |

Anexo L: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya según edad: 2 dientes

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |    |      |      |      |      |      |     |      |
|------------------------------------|-------------------|------|------|----|------|------|------|------|------|-----|------|
|                                    | 90                | 91   | 92   | 93 | 94   | 95   | 96   | 97   | 98   | 99  | 100  |
| Superfina                          | 0                 | 0    | 0    | 0  | 0.01 | 0.02 | 0.06 | 0.18 | 0.41 | 0.8 | 0.89 |
| Fina                               | 0.98              | 0.99 | 0.99 | 1  | 0.99 | 0.98 | 0.94 | 0.82 | 0.59 | 0.2 | 0.11 |
| Mediana                            | 0.02              | 0.01 | 0    | 0  | 0    | 0    | 0    | 0    | 0    | 0   | 0    |

Anexo M: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya según edad: 4 dientes

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |      |      |      |      |      |      |      |     |
|------------------------------------|-------------------|------|------|------|------|------|------|------|------|------|-----|
|                                    | 90                | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100 |
| Superfina                          | 0.03              | 0.05 | 0.08 | 0.13 | 0.19 | 0.27 | 0.38 | 0.49 | 0.61 | 0.71 | 0.8 |
| Fina                               | 0.94              | 0.93 | 0.91 | 0.87 | 0.81 | 0.72 | 0.62 | 0.51 | 0.39 | 0.29 | 0.2 |
| Mediana                            | 0.02              | 0.01 | 0.01 | 0.01 | 0    | 0    | 0    | 0    | 0    | 0    | 0   |

Anexo N: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya según edad: boca llena

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |      |      |      |      |      |      |      |      |
|------------------------------------|-------------------|------|------|------|------|------|------|------|------|------|------|
|                                    | 90                | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  |
| Superfina                          | 0.00              | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.24 | 0.73 |
| Fina                               | 0.17              | 0.26 | 0.38 | 0.52 | 0.66 | 0.77 | 0.85 | 0.91 | 0.91 | 0.74 | 0.26 |
| Mediana                            | 0.83              | 0.74 | 0.62 | 0.48 | 0.34 | 0.23 | 0.15 | 0.09 | 0.05 | 0.02 | 0.00 |

Anexo O: Probabilidades estimadas de la calidad de fibra en alpacas de raza Huacaya para el modelo general

| Calidad de fibra<br>(Probabilidad) | Factor de confort |      |      |      |      |      |      |      |      |      |      |
|------------------------------------|-------------------|------|------|------|------|------|------|------|------|------|------|
|                                    | 90                | 91   | 92   | 93   | 94   | 95   | 96   | 97   | 98   | 99   | 100  |
| Superfina                          | 0                 | 0    | 0.01 | 0.03 | 0.08 | 0.19 | 0.38 | 0.62 | 0.82 | 0.92 | 0.97 |
| Fina                               | 0.91              | 0.94 | 0.96 | 0.95 | 0.91 | 0.81 | 0.61 | 0.37 | 0.18 | 0.08 | 0.03 |
| Mediana                            | 0.09              | 0.06 | 0.03 | 0.02 | 0.01 | 0.01 | 0    | 0    | 0    | 0    | 0    |