



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



**MODELO PREDICTIVO DE RIESGO ASOCIADO A LA ANEMIA**  
**EN NIÑOS MENORES DE 5 AÑOS EN LA MICRORED YAURI**  
**PROVINCIA DE ESPINAR – CUSCO, 2019**

**TESIS**

**PRESENTADA POR:**

**Bach. GARY CANAZA ESPEZUA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**PUNO – PERÚ**

**2021**



## DEDICATORIA

*El presente trabajo se lo dedico a mi Padre (†)  
René Augusto por haber sido más que un padre  
un amigo, y a mi Madre Elvira que sin ella no lo  
hubiera logrado, ya que su persistencia es la que  
me llevo a seguir adelante.*

*A mis hermanos Nivardo y John que sin su cariño  
y apoyo no podría continuar en el duro trajinar  
de la vida.*

**GARY CANAZA ESPEZUA**



## AGRADECIMIENTO

- ✚ A la Escuela Profesional de Estadística E Informática de la Universidad Nacional del Altiplano por haberme dado la mejor etapa de mi vida como Estudiante y como vida en general.
- ✚ Al Ing. Alcides RAMOS CALCINA por brindarme sus amplios conocimientos, apoyo y orientación, y sobre todo por su paciencia para instruirme en el área de Estadística.
- ✚ A mi Director de Tesis y D.Sc. Percy HUATA PANCA, los miembros del Jurado M.C. Santos Octavio MORILLOS VALDERRAMA, Dr. Samuel Donato PEREZ QUISPE, M.Sc. Remo CHOQUEJAHUA ACERO, por sus valiosas aportaciones en pro de un mejor trabajo de investigación.
- ✚ Por último, agradecer a toda la plana docente de mi carrera, ya que sin ellas no tendría los conocimientos que ellos me inculcaron.



# ÍNDICE GENERAL

**DEDICATORIA**

**AGRADECIMIENTO**

**ÍNDICE GENERAL**

**ÍNDICE DE TABLAS**

**ÍNDICE DE FIGURAS**

**ÍNDICE DE ACRÓNIMOS**

**RESUMEN ..... 10**

**ABSTRACT..... 11**

## **CAPITULO I**

### **INTRODUCCIÓN**

**1.1. PLANTEAMIENTO DEL PROBLEMA ..... 13**

**1.2. FORMULACIÓN DEL PROBLEMA ..... 15**

**1.3. OBJETIVOS..... 15**

1.3.1. Objetivo General ..... 15

1.3.2. Objetivos Específicos..... 15

**1.4. HIPÓTESIS DE LA INVESTIGACIÓN ..... 15**

1.4.1. Hipótesis General: ..... 15

**1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN..... 16**

## **CAPITULO II**

### **REVISIÓN DE LITERATURA**

**2.1. ANTECEDENTES DE INVESTIGACIÓN ..... 17**

2.1.1. Antecedentes Internacionales..... 17

2.1.2. Antecedentes Nacionales ..... 19

**2.2. MARCO TEÓRICO ..... 21**

2.2.1. Anemia ..... 21

2.2.2 Factor de riesgo ..... 23

2.2.2.1. Factores determinantes de la anemia..... 24



2.2.3. Regresión Logística.....	25
2.2.3.1. La transformación logit .....	27
2.2.3.2 Modelo de regresión logística binaria .....	28
2.2.3.3 Modelo de regresión logística binaria múltiple.....	29
2.2.4 Selección de Variables con árboles de decisión.....	34
<b>2.3. DESCRIPCIÓN DE VARIABLES.....</b>	<b>35</b>

### **CAPITULO III**

#### **MATERIALES Y MÉTODOS**

<b>3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO .....</b>	<b>36</b>
<b>3.2. ÁMBITO DE ESTUDIO .....</b>	<b>36</b>
<b>3.3. TIPO Y DISEÑO DE INVESTIGACIÓN.....</b>	<b>37</b>
<b>3.4. POBLACIÓN Y MUESTRA .....</b>	<b>37</b>
3.4.1. Población.....	37
3.4.2. Muestra.....	37
<b>3.5. INSTRUMENTO DE RECOLECCIÓN Y PROCESAMIENTO DE DATOS</b>	<b>38</b>
3.5.1. Procedimiento de recolección de datos .....	39
3.5.2. Plan de procesamiento y análisis de datos .....	39
<b>3.6. OPERACIONALIZACIÓN DE VARIABLES.....</b>	<b>41</b>

### **CAPITULO IV**

#### **RESULTADOS Y DISCUSIÓN**

<b>4.1. IDENTIFICACIÓN DE FACTORES.....</b>	<b>43</b>
4.1.1. Prevalencia de anemia.....	43
4.1.2. Selección de variables por factor .....	44
<b>4.2. Estimación de modelo .....</b>	<b>52</b>
4.2.1. Selección de variables .....	52
4.2.2. Verosimilitud del modelo.....	58
4.2.3. Prueba de bondad de ajuste.....	59
<b>4.3. PREDICCIÓN DEL RIESGO DE ANEMIA.....</b>	<b>60</b>



<b>V. CONCLUSIONES.....</b>	<b>62</b>
<b>VI. RECOMENDACIONES .....</b>	<b>64</b>
<b>VII. REFERENCIAS BIBLIOGRAFICAS.....</b>	<b>65</b>
<b>ANEXO.....</b>	<b>70</b>

**Área : ESTADÍSTICA**

**Tema : REGRESIÓN LOGÍSTICA**

**FECHA DE SUSTENTACIÓN: 04 de marzo 2021**



## ÍNDICE DE TABLAS

<b>Tabla 1</b>	Rangos Normales Y Criterios Para Definir Anemia En Niños Menores De Seis Años, Propuestos Por La OMS. ....	22
<b>Tabla 2</b>	Valores normales de concentración de hemoglobina y grados de anemia en niñas y niños 6 a 59 meses (hasta 1000 msnm). ....	23
<b>Tabla 3.</b>	Operacionalización de Variables. ....	41
<b>Tabla 4.</b>	Prevalencia de anemia según sexo, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar – Cusco 2019.....	43
<b>Tabla 5.</b>	Recodificación de variables para el modelo logístico.....	51
<b>Tabla 6.</b>	Variables que ingresan al modelo .....	53
<b>Tabla 7.</b>	Variables que ingresan al nuevo modelo .....	54
<b>Tabla 8.</b>	Prueba del coeficiente de determinación .....	55
<b>Tabla 9.</b>	Tabla de clasificación .....	56
<b>Tabla 10.</b>	Tabla del área bajo la curva COR. ....	57
<b>Tabla 11.</b>	Prueba del coeficiente de intercepción. ....	58
<b>Tabla 12.</b>	Prueba de bondad de ajuste.....	59
<b>Tabla 13.</b>	Prueba de Omnibus. ....	59
<b>Tabla 14.</b>	Predicción de riesgo de anemia en niños menores de 5 años para dos casos. ....	60



## ÍNDICE DE FIGURAS

<b>Figura 1.</b>	Modelo causal de la Anemia en el Perú. ....	24
<b>Figura 2.</b>	Diagrama del árbol para el modelo de anemia según factores demográficos, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.....	45
<b>Figura 3.</b>	Diagrama del árbol para el modelo de anemia según factores biológico, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.....	47
<b>Figura 4.</b>	Diagrama del árbol para el modelo de anemia según factores programa social, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.....	48
<b>Figura 5.</b>	Diagrama del Árbol para el modelo de anemia según factores aplicación SIEN, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.....	50
<b>Figura 6.</b>	Gráfica de la curva COR.....	57





## ÍNDICE DE ACRÓNIMOS

<b>SIEN</b>	Sistema Integral De Estado Nutricional
<b>SIS</b>	Sistema Integral de Salud
<b>PIN</b>	Programa Integral de Nutrición
<b>MINSA</b>	Ministerio de Salud
<b>ENDES</b>	Encuesta Demográfica y de Salud Familiar
<b>INEI</b>	Instituto Nacional de Estadística e Informática
<b>OMS</b>	Organización Mundial de la Salud
<b>C.S.</b>	Centro de Salud
<b>CLAS</b>	Comité Local de Administración de Salud
<b>JUNTOS</b>	Programa de Apoyo Directo a los Más Pobres
<b>EESS</b>	Establecimiento de Salud



## RESUMEN

El problema de carencia de hierro constituye el principal problema de salud pública que afecta al país, por tanto, la anemia cobra vital importancia no solamente por su alta prevalencia sino por el impacto epidemiológico y social que tiene este problema en la salud de los niños. La presente investigación se propuso como objetivo el determinar un modelo predictivo de riesgo asociado a la anemia en niños menores de 5 años en la MICRORED Yauri, de la provincia de Espinar - Cusco 2019. Se planteo como hipótesis que el modelo predictivo permite realizar la predicción de riesgo asociado a la anemia en niños menores de 5 años. La metodología incluye el uso del método científico, siendo una investigación de tipo básica, nivel relacional, diseño no experimental, con una muestra probabilística de 322 niños. Para alcanzar los objetivos se analizó la base de datos proporcionada por el Sistema de Información del Estado Nutricional del niño menor de cinco años y Gestantes que acceden al Establecimiento de Salud (SIEN) de la DIRESA provincia de Espinar – Cusco correspondiente al año 2019. El modelo clasificador, es el siguiente:

$Ln(Y) = -8.304 + 0.582x_1 - 1.718x_2 + 4.534x_3 + 3.349x_{10}$ , cuya prueba de bondad de ajuste de Hosmer y Lemeshow, a un nivel de significación de 0.05, es de p-valor=0.088, con un coeficiente de determinación de Nagelkerke de 85.8%. La tasa de clasificación de verdaderos negativos (no anémico pronosticados como no anémicos) fue del 93.8%; mientras que la tasa más relevante en este caso, clasificar correctamente a los niños anémicos (verdaderos positivos) fue de 96.2%, demostrando gran capacidad predictiva del modelo.

**Palabras Clave:** Modelo de regresión logística, modelo predictivo, anemia, factor de riesgo, árboles de decisión.



## ABSTRACT

The problem of iron deficiency constitutes the main public health problem that affects the country, therefore, anemia is of vital importance not only because of its high prevalence but also because of the epidemiological and social impact that this problem has on the health of children. The objective of this research was to determine a predictive model of risk associated with anemia in children under 5 years of age in the Yauri MICRORED, Espinar - Cusco 2019. It was hypothesized that the predictive model allows the prediction of risk associated with anemia in children under 5 years of age. The methodology includes the use of the scientific method, being an investigation of a basic type, relational level, non-experimental design, with a probabilistic sample of 322 children. To achieve the objectives, the database provided by the Information System of the Nutritional Status of children under five years of age and pregnant women who access the Health Establishment (SIEN) of the DIRESA province of Espinar - Cusco corresponding to the year 2019 was analyzed. classifier model, is as follows:

$Ln(Y) = -8.304 + 0.582x_1 - 1.718x_2 + 4.534x_3 + 3.349x_{10}$ , whose Hosmer and Lemeshow goodness-of-fit test, at a significance level of 0.05, is p-value = 0.088, with a coefficient Nagelkerke's determination of 85.8%. The true negative classification rate (non-anemic predicted non-anemic) was 93.8%; while the most relevant rate in this case, correctly classifying anemic children (true positives) was 96.2%, showing the model's great predictive capacity.

**Keywords:** Logistic regression model, predictive model, anemia, risk factor, decision trees.



## CAPITULO I

### INTRODUCCIÓN

La anemia es un problema de salud mundial y, en nuestro país, la anemia infantil es uno de los principales problemas de salud pública que afecta a nuestra población, sobre todo a los sectores más vulnerables como las gestantes, los niños menores de dos años y las mujeres en edad fértil. Actualmente, en el Perú el 43,5% de los niños de 6 a 35 meses, tiene anemia, con una distribución de 51,1% en la zona rural y 40% en la zona urbana. Esto representa 620 mil niños menores de tres años anémicos en una población de 1,6 millones a nivel nacional. (Arroyo-Laguna, 2017)

Es importante mencionar que, la anemia representa una carga significativa del individuo desde temprana edad teniendo un efecto no solo en la vida de cada persona que la padece, sino también sobre la sociedad en su conjunto en términos sociales y económicos (P. Dutta & Sengupta, 2017). La inadecuada ingesta de hierro y otros nutrientes puede llevar a la deficiencia de hierro y que esta deficiencia en su fase más severa y prolongada ocasiona la anemia. Dentro del marco de causalidad, el enfoque de la anemia debe darse de una manera integral, para lograr no solo mejorar la ingesta de hierro y micronutrientes en el niño, sino reducir la carga de parasitosis, diarrea, y malaria. (Zavaleta, 2017)

Ante esta problemática, en los últimos años la utilización de los métodos estadísticos como su sofisticación han sufrido un importante incremento en el campo de la investigación sanitaria. Esta relación entre la estadística y la investigación en salud ha sido una motivación para su desarrollo y se tuvieron que proponer nuevos modelos; un ejemplo de esta relación es el modelo de regresión logística, siendo este actualmente, el



método más utilizado en el ámbito de la investigación sanitaria (Sánchez-Cantalejo, 2000).

En Salud Pública, es frecuente la situación en que se dispone de una variable con sólo dos posibles valores, es decir, una variable dicotómica; por tanto, la regresión logística es el método de análisis adecuado cuando se necesita modelizar una variable respuesta binaria, del tipo presencia o ausencia de enfermedad, y permite el uso conjunto de covariables de tipo categórico y continuo, proporcionando interpretación biológica a sus parámetros (Hosmer, Lemeshow, & Sturdivant, 2013).

En este sentido en el trabajo de investigación se planteó como objetivo general: Determinar un modelo predictivo de riesgo asociado a la anemia en niños menores de 5 años en la MICRORED Yauri, Espinar – Cusco 2019 y como objetivos específicos: Identificar un conjunto de factores distales que influyan en riesgo a anemia en niños menores de 5 años en la MICRORED Yauri, Espinar -Cusco; de igual manera se planteó estimar un modelo de regresión logístico para la predicción de la probabilidad de riesgo anemia en niños menores de 5 años. El tipo de estudio es de relación entre una variable dicotómica (dependiente) y variables continuas y categóricas (independientes), cuya metodología es el análisis de regresión logística para poder alcanzar los objetivos de la investigación.

## **1.1. PLANTEAMIENTO DEL PROBLEMA**

La anemia es el mayor problema de salud pública que afecta a la población en el mundo desarrollado y en desarrollo, no solo porque es el daño más común y el más



ampliamente distribuido, sino porque es el más prevalente entre los grupos vulnerables, como niños pequeños y mujeres en edad fértil. Afecta al 43% de los menores de 5 años, 38% de las gestantes y al 29% de las no gestantes. (Zavaleta & Astete-Robilliard, 2017)

Según la OMS (Organización Mundial de la Salud), la anemia afecta en todo el mundo a 1620 millones de personas, lo que representa al 24,8% de la población, mientras que, la máxima prevalencia se da en los niños en edad preescolar con 47,4%. (OMS, 2020)

La anemia en la Región Cusco en el 2018 afectó al 54.2% de niños menores de 36 meses (38284 niños) y en relación al año anterior se disminuyó en 1.1%. Según el Sistema de Información del Estado Nutricional 2018, todas las provincias presentan porcentajes por encima del 40% a excepción de Canchis (39.1%), La Convención (35.8%), Calca (29.8)% y Canas (20.9%); asimismo una de las provincias que presenta alto porcentaje y alberga mayor número de casos de anemia es la Provincia de Cusco (47.2%), en ese sentido la anemia afecta a zonas de pobreza y no pobreza; siendo la anemia por deficiencia de hierro en la dieta y parasitosis o enfermedades infecciosas las principales causas del problema. (DIRESA-Cusco, 2019)

La prevalencia de la anemia es un mal que acarrea repercusiones y consecuencias económicas, sociales y por ende se tiene que actuar en combatir este problema ya que la vida futura del niño se va verse afectada negativamente.



## **1.2. FORMULACIÓN DEL PROBLEMA**

¿Cuál es el modelo predictivo de riesgo asociado a la prevalencia de anemia en niños menores de 5 años en la MICRORED Yauri, provincia de Espinar – Cusco, en el año 2019?

## **1.3. OBJETIVOS**

### **1.3.1. Objetivo General**

Determinar un modelo predictivo de riesgo asociado a la anemia en niños menores de 5 años en el MICRORED de Yauri, provincia de Espinar - Cusco 2019.

### **1.3.2. Objetivos Específicos.**

- Identificar un conjunto de factores distales que influyan en riesgo de anemia en niños menores de 5 años en el MICRORED Yauri, provincia de Espinar - Cusco 2019.
- Estimar un modelo de regresión logístico para la predicción de la probabilidad en riesgo de anemia en niños menores de 5 años en el MICRORED Yauri, provincia de Espinar - Cusco 2019.

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

### **1.4.1. Hipótesis General:**

El modelo logístico permite realizar la predicción de riesgo asociado a la anemia en niños menores de 5 años en el MICRORED Yauri, provincia de Espinar – Cusco 2019



## 1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN

En la región Cusco y en todo el Perú, el tema de salud es una preocupación constante dentro de los diferentes factores que influyen en el bienestar de las familias. Una de las principales preocupaciones del Ministerio de Salud (MINSA) son los niños menores de 5 años, es por ello que el MINSA en su rol constante de mejorar la calidad de vida de los niños, requiere reforzar las estrategias de control y prevención de anemia en el periodo de 1 a 5 años, atendiendo esa necesidad, el trabajo pretende aportar un modelo de regresión logística cuyos resultados predictivos puedan llevar a tomar decisiones coherentes.

En este contexto, los modelos de regresión logística son aplicados a las ciencias de la salud los cuales, nos permiten el análisis de los resultados en términos explicativos y predictivos, pudiendo conocer la fuerza de asociación mediante los OR de los factores de riesgo con el efecto estudiado de una manera independiente y conocer el valor predictivo de cada uno de ellos o bien del modelo en su conjunto. Su uso, cada vez más frecuente debe ir precedido de una reflexión crítica tanto de la elección de las variables incluidas en el modelo como del análisis de sus resultados (Fiuza & Rodríguez, 2000; Solano et al, 2007).

Este trabajo de investigación usó los modelos de regresión logística para conocer cuáles son los factores de riesgo asociados a la anemia en niños menores de 5 años dentro de la población de estudio (Espinar, Cusco).





## CAPITULO II

### REVISIÓN DE LITERATURA

#### 2.1. ANTECEDENTES DE INVESTIGACIÓN

##### 2.1.1. Antecedentes Internacionales

Nusinovici *et al* (2020), evaluaron el rendimiento de los algoritmos de aprendizaje automático (ML) y compararlos con la regresión logística para la predicción del riesgo de enfermedades cardiovasculares (ECV), enfermedad renal crónica (ERC), diabetes (DM) e hipertensión (HTN) y en Estudio de cohorte prospectivo utilizando predictores clínicos simples. Encontraron que, la regresión logística alcanzó el área más alta, bajo la curva característica operativa del receptor para las predicciones de ERC (0.905 [0.88, 0.93]) y DM (0.768 [0.73, 0.81]). Para CVD y HTN, los mejores modelos fueron la red neuronal (0.753 [0.70, 0.81]) y la máquina de vectores de soporte (0.780 [0.747, 0.812]), respectivamente. Sin embargo, las diferencias con la regresión logística fueron pequeñas (menos del 1%) y no significativas. Concluyeron que, la regresión logística produce un rendimiento tan bueno como los modelos de ML para predecir el riesgo de enfermedades crónicas mayores con baja incidencia y predictores clínicos simples.

Dutta *et al* (2020), se centraron en estimar la prevalencia y los factores de riesgo asociados a nivel individual y comunitario entre niños de 6 a 59 meses en la India. Para explorar la asociación de la anemia infantil y algunos factores de riesgo potenciales con respecto a la jerarquía (nivel individual y comunitario) de datos, se realizó un modelo de regresión logística multinivel utilizando el software STATA-SE 14. Los resultados indicaron que algunos aspectos de la comunidad influyen en los riesgos de anemia de los niños. Concluyeron, incluir factores a nivel de la comunidad más allá de los factores a nivel



individual por parte de los formuladores de políticas para erradicar la carga de anemia en el país y los estados de alta prevalencia.

Ajakaye & Ibukunoluwa (2020), evaluaron la prevalencia y los factores de riesgo de infección por malaria, anemia y desnutrición entre los niños que viven en el campamento de desplazados internos (PDI) en el estado de Edo, Nigeria. El análisis de regresión logística se utilizó para determinar las asociaciones entre las variables predictoras y los resultados primarios. Tener 6-10 años (OR = 2.307,  $p = 0.015$ ) y tener infección de malaria (OR = 1.693,  $p = 0.048$ ) se identificaron como factores de riesgo significativos de anemia, mientras que estar en el grupo de edad de hasta 5 años fue el único significativo factor de riesgo (OR para el grupo de mayor edad = 0.251,  $p \leq 0.001$ ) asociado con la desnutrición. Concluyeron que, el control de la anemia y la desnutrición debe integrarse con el control de la malaria existente y debe incluir a los niños mayores de cinco años.

Ghazvini *et al* (2019), plantearon como objetivo determinar las variables que afectan la tuberculosis utilizando el modelo de predicción de regresión logística. La sensibilidad, la especificidad, la curva ROC (característica de funcionamiento del receptor) y los valores predictivos positivos y negativos se utilizaron para evaluar el poder predictivo del modelo de regresión logística. Los datos fueron analizados con el software SPSS versión 22 a través del modelo de regresión logística y prueba de Chi-cuadrado. Y el valor  $P < 0,05$  se consideró estadísticamente significativo. La sensibilidad y especificidad de este modelo para predecir la tuberculosis fueron 78% y 68%, respectivamente. Además, el área bajo la curva (Roc) fue de 0.821. Variables; vitamina D3 ( $p = 0.01$ ), hemoglobina ( $p = 0.01$ ) e índice de masa corporal (IMC) ( $p = 0.01$ ) significativamente asociados con tuberculosis.



Los resultados mostraron que las variables de vitamina D3, hemoglobina e índice de masa corporal (IMC) tienen una mejor predicción de TB en el modelo de regresión logística.

### **2.1.2. Antecedentes Nacionales**

Obregón (2018), analizó la base de datos de la Encuesta Demográfica y de Salud Familiar – ENDES 2015. Considerando la estructura anidada de la toma de datos en el muestreo y el modelo conceptual de los factores de riesgo de la anemia, se ajustó un modelo de regresión logístico multinivel con el objetivo de determinar la contribución de los factores de riesgo individual y contextual al mayor riesgo de anemia en niños menores de cinco años en el Perú. Los factores de riesgo de anemia: “lengua materna del niño” y “educación de la madre” según el modelo logístico, no resultaron significativos para el modelo multinivel. Las variables contextuales que influyen en el efecto de la región sobre el riesgo de anemia fueron el “Acceso sostenible a fuentes de abastecimiento de agua” (MOR= 7.47), la “Esperanza de vida al nacer” (MOR= 5.24), el “Acceso a servicios de saneamiento mejorados” (MOR= 5.01) y la variable “Niños con vacunas básicas completas para su edad” (MOR= 2.83).

Ríos (2017), se planteó conocer los factores que influyen en la presencia de la enfermedad de Gastritis y para ello se utilizó la regresión logística como técnica estadística. Los factores de riesgo fueron: tomar antiinflamatorios para el dolor y/o inflamación, beber agua potable sin hervir, consumir alimentos en lugares que no garantizan la calidad, consumir alimento hasta saciar el apetito, y la falta de tiempo para realizar tareas académicas universitarias. Los signos y síntomas característicos que influyeron en la presencia de la enfermedad de gastritis fueron: La acumulación de gases en el estómago (Aerofagia) y la presencia de vómitos con sangre (Hematemesis). Las características



sociodemográficas que influyen en la presencia de la enfermedad de gastritis fueron: la edad, los hábitos alimenticios y toxicológicos, así como tomar desayuno todos los días, consumo de alimentos mirando televisión, consumo de alimentos fritos o dorados que son cocidos en aceites, consumo frecuente de alcohol y consumo frecuente de tabaco.

Roque (2018), se planteó como 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. Como resultado se obtuvieron los siguientes modelos para los registros en conjunto  $g1(x) = -142.321 + 1.536 * FC$  para la calidad Baby y  $g2(x) = -47.578 + 0.554 * FC$  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. La probabilidad de que la calidad de fibra sea Baby va incrementando a medida que el Factor de confort es más alto.

Condori Apaza (2018), se planteó como objetivo: determinar los factores que determinan en nivel de riesgo de diabetes en pacientes de endocrinología del hospital regional Manuel Núñez Butrón – Puno 2017. El tipo de investigación es correlacional porque analiza causa – efecto, se extrajo una muestra de 65 pacientes utilizando muestreo no probabilístico juicio del investigador, se usó el Odds Ratio y la regresión logística binaria. La Regresión logística dio a conocer: los factores que determinan que un paciente tenga diabetes son pacientes con hipertensión arterial, parientes con primer grado con diabetes y con los triglicéridos altos.

Ramos (2020), tuvo como finalidad de determinar los factores que influyen en la calidad de servicio en salud de los usuarios atendidos en establecimientos regionales del Perú. La



investigación fue descriptiva, de tipo correlacional y explicativo. Se utilizó la regresión logística como técnica estadística. El modelo de los factores fue:  $\text{logit}(p) = -7,73 + 0,147X_{13} + 0,15X_{14} + 0,229X_1 + 0,28X_3 + 0,19X_5 + 0,330X_8 + 0,275X_9 + 0,574X_{10}$ . Se concluyó que los factores que influyeron en la calidad de servicio en salud de los usuarios atendidos en establecimientos regionales del Perú, 2016 fueron: Accesibilidad a los ambientes ( $P=1.158353$ ), Área de espera ( $P=1.161834$ ), amabilidad y cortesía por parte del personal administrativo ( $P=1.25734203$ ), interés y disposición por atender del personal administrativo ( $P=1.32312981$ ), de claridad de la información por parte del personal administrativo ( $P=1.20924959$ ), interés y disposición por atender por parte del médico tratante ( $P=1.39068128$ ), confianza y seguridad ( $P= 1.316530675$ ) y claridad de la información por parte del médico tratante ( $P= 1.775354285$ ).

## **2.2. MARCO TEÓRICO**

### **2.2.1. Anemia**

La anemia es un trastorno en el cual el número de eritrocitos (y, por consiguiente, la capacidad de transporte de oxígeno de la sangre) es insuficiente para satisfacer las necesidades del organismo. Se cree que, en conjunto, la carencia de hierro es la causa más común de anemia, pero pueden causarla otras carencias nutricionales (entre ellas, las de folato, vitamina B12 y vitamina A), la inflamación aguda y crónica, las parasitosis y las enfermedades hereditarias o adquiridas que afectan a la síntesis de hemoglobina y a la producción o la supervivencia de los eritrocitos. (OMS, 2011)

La OMS define la anemia como el descenso del nivel de hemoglobina dos desviaciones estándar por debajo de lo normal para la edad y el sexo. La anemia

provoca en el organismo la incapacidad de producir hemoglobina y de glóbulos rojos (componentes del sistema sanguíneo encargados de transportar el oxígeno de una célula a otra).

La anemia presenta una elevada prevalencia en las consultas de atención primaria, pediatría y en el control de la gestación. No es lo mismo la deficiencia de hierro (DH) que la anemia por deficiencia de hierro (ADH). La ADH es la causa más frecuente de anemia en todo el mundo y representa un importante problema de salud principalmente en los países subdesarrollados. La ADH es muy prevalente de forma que hasta un 5% de niños y adolescentes, un 10% de las mujeres pre menopáusicas y un 1% de los hombres tienen anemia por deficiencia de hierro; y puede llegar hasta un 40% de los ancianos que se cuidan en sus domicilios. (Guzmán Llanos, Guzmán Zamudio, & Llanos de los Reyes, 2016)

**Tabla 1. Rangos normales y criterios para definir anemia en niños menores de seis años, propuestos por la OMS.**

<i>Grupo de edad</i>	<i>Rango normal de hemoglobina (g/dl)</i>	<i>Anémico si la Hb es menor de: (g/dl)</i>
Al nacimiento	13.5 – 18.5	13.5
Niños 2 – 6 meses	9.5 – 13.5	9.5
Niños 6 meses – 6 años	11.0 – 14.0	11.0

Fuente: (OMS, 2011)

A continuación, se muestra la clasificación de la anemia según severidad:

**Tabla 2. Valores normales de concentración de hemoglobina y grados de anemia en niñas y niños 6 a 59 meses (hasta 1000 msnm).**

<i>Población</i>	<i>Normal (g/dl)</i>	<i>Anemia por niveles de hemoglobina (g/dl)</i>		
		<i>Leve</i>	<i>Moderada</i>	<i>Severa</i>
Niños de 6 a 59 meses de edad	11.0 a más	10.0 – 10.9	7.0 – 9.9	< 7.0

Fuente: (OMS, 2011)

### 2.2.2 Factor de riesgo

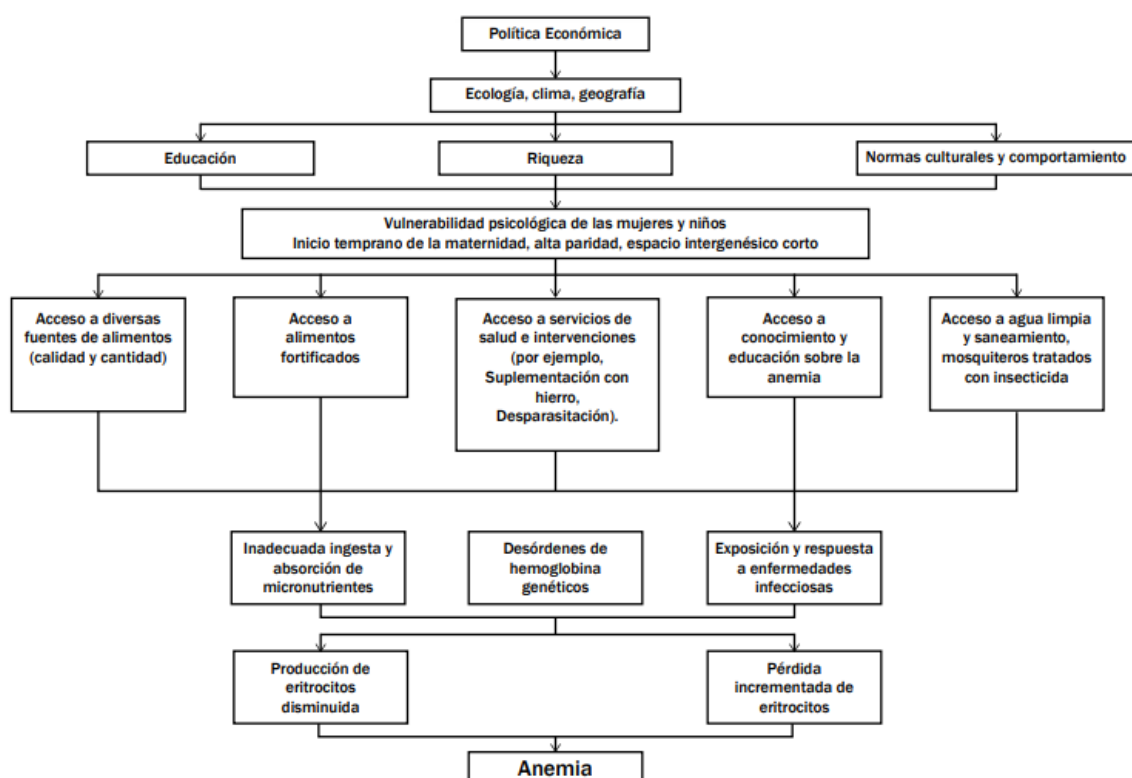
Que se entienda por un agente causal y por un factor de riesgo es sin duda polémica. Cualquiera que sea la definición de actor de riesgo, la definición funcional de la causa de un efecto dado puede expresarse del modo siguiente: cualquier factor, condición o característica, cuya supresión elimina la posibilidad de que se produzca el efecto, es una causa del mismo. Por otra parte, están los elementos que, sin ser necesarios, son propiciatorios del efecto. Se trata de los llamados factores de riesgo (FR) para desarrollar una enfermedad o una condición. Se trata de factores susceptibles de ser modificados que, sin ser causas propiamente, pueden favorecer que el agente causal actúe (Silva Ayçaguer, 2005).

Según la OMS (2016), Un factor de riesgo es cualquier rasgo, característica o exposición de un individuo que aumente su probabilidad de sufrir una enfermedad o lesión. Entre los factores de riesgo más importantes cabe citar la insuficiencia ponderal, las prácticas sexuales de riesgo, la hipertensión, el consumo de tabaco y alcohol, el agua insalubre, las deficiencias del saneamiento y la falta de higiene.

### 2.2.2.1. Factores determinantes de la anemia

Según MINSA (2017), entre las causas inmediatas se reconoce el consumo inadecuado de hierro y de otros micronutrientes a partir de los alimentos. Esta carencia de hierro y vitaminas no permitiría una apropiada formación de los glóbulos rojos y de la hemoglobina. Otras causas inmediatas de la anemia son la alta morbilidad por infecciones como la diarrea, parasitosis, malaria, etc. Esta situación está asociada a inadecuadas prácticas de higiene, de lavado de manos, limitado acceso a agua segura y saneamiento básico.

En la literatura internacional se plantea el modelo causal de la anemia, la cual se muestra en la siguiente figura:



**Figura 1.** Modelo causal de la Anemia en el Perú.

*Fuente:* (MINSA, 2017)





La anemia tiene diversas causas (Figura 1), así, en zonas de no malaria la deficiencia de hierro podría significar hasta el 60% de la causa de anemia. La inadecuada ingesta de hierro y otros nutrientes es una causa importante de anemia. Se han identificado otras causas de anemia, así en un estudio realizado en nuestro país en el año 2015, se determinó que las formas más frecuentes de anemia son las relacionadas a parasitosis y aquellas que tienen al mismo tiempo dos o más causas. (Zavaleta & Astete-Robilliard, 2017)

La anemia se asocia a diversos factores sociodemográficos, y factores relacionados al cuidado de la salud del niño. Un análisis secundario en el Perú (año 2016), encontró que en menores de 6 a 35 meses, la anemia está asociada tanto a factores sociodemográficos (lugar de residencia fuera de Lima, bajo nivel socioeconómico, madre adolescente y con poco nivel educativo, sexo masculino del menor, edad menor a 24 meses, fiebre reciente), como al cuidado madre-niño (falta de control prenatal y de tratamiento contra la anemia en la gestación, parto en el hogar, anemia materna durante la encuesta y ausencia de tratamiento antiparasitario en el menor). (Velásquez et al., 2016)

### **2.2.3. Regresión Logística**

Dentro de los métodos de clasificación, una posibilidad es intentar construir un modelo que explique los valores de la variable de clasificación. Por ejemplo, si se desea discriminar entre créditos que se devuelven o que presentan problemas para su cobro, puede añadirse a la base de datos una nueva variable, y, que tome el valor cero cuando el crédito se devuelve sin problemas y el valor uno en otro caso (riesgo



crediticio). El problema de discriminación se convierte en prever el valor de la variable ficticia,  $y$ , en un nuevo elemento del que conocemos el vector de variables  $x$ . Si el valor previsto está más próximo a cero que a uno, clasificaremos al elemento en la primera población. En otro caso, lo haremos en la segunda. Para modelar este tipo de relaciones se utilizan los modelos de respuesta cualitativa. (Peña, 2002)

La regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica o variable de respuesta binaria, para el cual el resultado de la respuesta para cada sujeto es un "éxito" o "fracaso" en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística pertenece al conjunto de Modelos Lineales Generalizados (GLM) que usa como función de enlace la función logit. Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística. (Agresti, 2007)

La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de variables predictoras. Es similar a un modelo de regresión lineal, pero está adaptado para modelos en los que la variable dependiente es dicotómica. Los coeficientes de regresión logística pueden utilizarse para estimar la razón de las ventajas (odds ratio) de cada variable independiente del modelo. La regresión logística se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante. (Pérez, 2004)

### 2.2.3.1. La transformación logit

Sánchez-Cantalejo (2000), describe lo siguiente: consideremos la situación en la que la variable  $Y$  es dicotómica tal que  $Y = 1$  si el individuo presenta la característica de interés e  $Y = 0$  en caso contrario. Por tanto, se puede admitir que la variable respuesta sigue una distribución binomial de parámetros  $1$  y  $p$ , donde  $p$  representa la probabilidad de que un individuo presente la característica de interés ( $Y = 1$ ). Como para una variable binomial de parámetros  $n$  y  $p$  su media es  $n.p$ , en este caso la media sería  $1.p = p$ . De esta forma el modelo de regresión lineal podría escribir de la forma:

$$E(Y) = p = \beta_0 + \beta_1 X$$

Ya no existe restricciones sobre los valores de los parámetros del modelo, podemos tener estimaciones de la probabilidad de presentar la característica o bien mayorea a la unidad o bien negativas, lo cual carece totalmente de sentido, esta situación invalida cualquier intento de utilizar la metodología de regresión lineal para el caso de que  $Y$  sea binomial.

Una medida muy utilizada en epidemiología es la que denominaremos ventaja u oportunidad, que deriva de la palabra anglosajona *odds*. Sea  $p$  la probabilidad definida anteriormente; el cociente

$$\frac{p}{1-p}$$

es decir, la probabilidad de presentar la característica ( $p$ ) dividida por la probabilidad no presentarla ( $1 - p = q$ ), se denomina ventaja de la característica presentada.

Si consideramos la transformación, mediante el logaritmo neperiano, de este parámetro

$$\log \frac{p}{1-p}$$

sus posibles valores pueden ser cualquiera número real, tanto positivo como negativo, con esto desaparece el problema anterior; esta transformación de  $p$  se denomina la transformación logística o *transformación logit* de la probabilidad de  $p$

$$\log \frac{p}{1-p} = \text{logit}(p)$$

De esta manera el modelo que nos puede permitir, en principio, resolver el problema que tenemos planteado puede representarse en la forma

$$\text{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m$$

Este modelo tiene como atractivo que, sus parámetros son interpretables como una medida de riesgo asociado a las predictoras.

### 2.2.3.2 Modelo de regresión logística binaria

Según Sánchez-Cantalejo (2000), se considera por ahora que se dispone de una sola variable predictora, dicho en términos epidemiológicos, un solo posible actor de riesgo. Según el modelo logístico antes propuesto, para el caso de una sola predictora, toma la forma

$$\text{logit}(p) = \log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X$$

donde  $p$  representa la probabilidad de que el individuo presente la característica de interés y  $X$  es la única predictora. La expresión anterior es equivalente a esta otra

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 X}$$

y despejando  $p$  obtenemos otra forma de escribir el modelo logístico

$$E(Y) = p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

se esta expresando la probabilidad como un cociente entre dos cantidades no negativas, pues sean cuales sean los valores de los parámetros, la cantidad  $e^{\beta_0 + \beta_1 X}$  no puede ser negativa; como, además, el numerador es siempre menor a igual que el denominador, el cociente va variar entre 0 y 1, el rango de valores permitido para una probabilidad.

### 2.2.3.3 Modelo de regresión logística binaria múltiple

En tanto Pérez (2004), plantea la generalización; a la función:

$$f(z) = \frac{1}{1 + e^{-z}}$$

le denomina función logística. El modelo de regresión logística modeliza la probabilidad de un proceso binomial como la función logística de una combinación lineal de la variable dependiente.

El modelo de regresión logística binaria múltiple tendrá la expresión:

$$p = \frac{e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

o su equivalente,

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$



La cuestión que se suscita en un contexto multivariable es la determinación del mejor modelo explicativo, lo que significa evaluar la pertinencia de las diferentes variables independientes propuestas inicialmente en el modelo de regresión logística y escoger aquellas que mejor explican la variable dependiente. A tal efecto se dispone de varios métodos de selección de variables. (López-Roldán & Fachelli, 2015)

#### a) **Proceso de análisis**

El proceso de análisis que se sigue tiene en cuenta los aspectos siguientes:

##### 1) Selección de las variables del modelo.

- Se requiere la definición de un modelo de análisis que justifique un modelo de dependencia.
- Analizar las condiciones de aplicación.
- Constatar manifiesta relación de cada variable independiente con la variable dependiente.
- Seguir un proceso de selección de las mejores variables con el objetivo de ajustar el mejor modelo lo más parsimonioso.

##### 2) Estimación de los coeficientes de las variables independientes

- Se aplica un método de estimación de máxima verosimilitud, un caso particular de Mínimos Cuadrados Ordinarios (MCO).

##### 3) Evaluación del modelo.

- Se trata de establecer la bondad de ajuste del modelo o capacidad explicativa: estadísticos de  $R^2$  de Nagelkerke o  $R^2$  de Cox y Snell, así como la prueba de Hosmer y Lemeshow.
- La eficacia predictiva o capacidad discriminadora del modelo: surge del



- cruce de la variable dependiente observada por la pronosticada, la llamada tabla de clasificación o matriz de confusión. Se trata de mirar el porcentaje de casos bien clasificados. Para ello se establece un criterio de corte: de probabilidad de que la variable dependiente sea 1. Por defecto se considera el valor 0,5, pero puede ajustarse con un estudio específico. Para ello se dispone en particular del denominado análisis de curvas COR (Característica de Operación del Receptor, en inglés ROC).
- Interpretación de los resultados con el conjunto de la información, pero fundamentalmente dando cuenta de la tabla de coeficientes.

## **b) Condiciones de aplicación**

La regresión logística no establece condiciones restrictivas de aplicación. En particular, no se imponen supuestos a la distribución de las variables.

Se establecen las siguientes condiciones:

- 1) El modelo debe estar correctamente especificado y ser relevante sustantivamente.
- 2) No se omiten variables independientes relevantes.
- 3) Las variables independientes se miden sin error.
- 4) Las observaciones son independientes entre sí.
- 5) Ausencia de colinealidad entre las variables independientes.
- 6) Linealidad de las variables cuantitativas.
- 7) En relación al tamaño de la muestra. Hosmer y Lemeshow recomiendan muestras mayores de 400 casos.



- 8) El tanto por ciento de casos que corresponden al 0 o al 1 de la variable dependiente debe ser del 10% al menos.

**c) Pruebas de significación**

En el análisis de regresión logística se consideran las siguientes pruebas estadísticas de significación:

- i) Evaluación estadística de los coeficientes de regresión logística, si son significativamente distintos de 0, mediante el **estadístico de Wald**:

$$wald = \frac{b^2}{s_b^2}$$

que sigue una distribución normal estándar. Los coeficientes significativos son los que tienen una probabilidad inferior a 0,05.

- ii) El cálculo del **Pseudo R<sup>2</sup>** que determina la bondad de ajuste, cuánto mejora un modelo en relación al modelo, expresado en porcentaje:

$$Pseudo R^2 = 1 - \frac{L_M^2}{L_0^2}$$

Se trata de medidas que evalúan el incremento de la verosimilitud del modelo: el cambio del estadístico  $-2\log(L)$  o bien de  $L$ , la razón de verosimilitud que varía entre 0 y 1. Si  $\Lambda = -2\log(L)$  entonces  $L = \exp(-\Lambda/2)$  y los pseudo R<sup>2</sup> se calculan de la forma siguiente en el caso del de Cox y Snell y del de Nagelkerke:

Pseudo R<sup>2</sup> de Cox y Snell:



$$R_{CS}^2 = 1 - \left( \frac{L_{constante}}{L_{modelo}} \right)^{\frac{2}{n}} = 1 - \exp\left( \frac{\Lambda_{modelo} - \Lambda_{constante}}{n} \right)$$

Estadístico que varía entre 0 y 1,  $0 \leq R^2 \leq 1$ , y donde  $L(a)$  es el modelo de la constante mientras que  $L(a, b_1, b_2, \dots, b_j)$  es el modelo completo considerado.

Pseudo  $R^2$  de Nagelkerke:

$$R_N^2 = \frac{1 - \left( \frac{L_{constante}}{L_{modelo}} \right)^{\frac{2}{n}}}{1 - \left( L_{constante} \right)^{\frac{2}{n}}} = \frac{1 - \exp\left( \frac{\Lambda_{modelo} - \Lambda_{constante}}{n} \right)}{1 - \exp\left( \frac{-\Lambda_{constante}}{n} \right)}$$

Con  $1 - \left( L_{constante} \right)^{\frac{2}{n}}$  que corresponde al  $R_{max}^2$ . Es un estadístico que varía también entre 0 y 1, con valores algo superiores en relación anterior. Tanto éste como el anterior son indicadores de la variabilidad explicada que siempre proporcionan valores muy bajos en relación a la medida homóloga en regresión lineal clásica. Es habitual encontrar resultados de 0,2 y 0,3, mientras que un pseudo  $R^2$  de 0,6 es poco habitual.

**iii) Evaluación del estadístico de bondad de ajuste  $z^2$ :**

$$z^2 = \sum_{i=1}^n \frac{(p_i - \hat{p}_i)^2}{\hat{p}_i(1 - \hat{p}_i)} = \sum_{i=1}^n \frac{R_i^2}{\hat{p}_i(1 - \hat{p}_i)}$$

donde  $R_i$  es el residuo entre la probabilidad observada y la probabilidad estimada del  $i$ -ésimo caso. Sigue una distribución de chi-cuadrado. Cuando el modelo es significativo la probabilidad asociada es menor o igual a 0,05.

**iv) Prueba de Hosmer y Lemeshow de bondad de ajuste del modelo. Para**

corroborar si el modelo se ajusta se utiliza este contraste de distribución. Para ello se calcula la probabilidad pronosticada del suceso para todos los individuos de la muestra y se calcula la diferencia con los valores observados. La prueba consiste en dividir el recorrido de la probabilidad en deciles y se comparan las distribuciones de frecuencias esperada y observada mediante un contraste de chi-cuadrado con 7 grados de libertad.

$$\chi_{HL}^2 = \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g \left(1 - \frac{E_g}{n_g}\right)}$$

La hipótesis nula establece que no hay diferencias entre los valores observados y pronosticados: el modelo ajusta. Por tanto, si el ajuste es bueno, se esperar un valor alto de probabilidad, superior o igual a 0,05. Se trata de una prueba adecuada para muestras pequeñas y covariables continuas.

#### 2.2.4 Selección de Variables con árboles de decisión

Según: (Berlanga, Rubio, & Vilà, 2013), un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (sucesos), que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión más “acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Estos árboles permiten examinar los resultados y determinar visualmente cómo fluye el modelo. Los resultados visuales ayudan a buscar subgrupos específicos y relaciones que tal vez no encontraríamos con estadísticos más tradicionales.

La función árboles de decisión crea árboles de clasificación y de decisión para identificar grupos, descubrir las relaciones entre grupos y predecir eventos futuros.



Existen diferentes tipos de árbol: CHAID, CHAID exhaustivo, CRT y QUEST, según el que mejor se ajuste a nuestros datos.

Los árboles de decisión crean un modelo de clasificación basado en diagramas de flujo. Clasifican casos en grupos o pronostican valores de una variable dependiente (criterio) basada en valores de variables independientes (predictoras).

Las ventajas de un árbol de decisión son:

- Facilita la interpretación de la decisión adoptada.
- Facilita la comprensión del conocimiento utilizado en la toma de decisiones.
- Explica el comportamiento respecto a una determinada decisión.
- Reduce el número de variables independientes.

### **2.3. DESCRIPCIÓN DE VARIABLES**

Ya que el estudio busca describir cómo la inclusión de variables influye en el comportamiento de la variable anemia, se ha tomado en cuenta la información proporcionada por Sistema de Información del estado nutricional del niño menor de cinco años y gestantes que acceden al establecimiento de salud (SIEN) de la MICRORED Yauri, provincia de Espinar – Cusco correspondiente al año 2019.

Las variables consideradas en el estudio se describen en el Anexo B.



## CAPITULO III

### MATERIALES Y MÉTODOS

#### 3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO

La provincia de Espinar es una de las trece que conforman el departamento del Cuzco en el Sur del Perú. Su nombre es un homenaje al teniente coronel Ladislao Espinar Carrera, héroe cuzqueño de la batalla de Dolores durante la guerra del Pacífico. Limita por el Norte con la provincia de Canas, por el Este con el departamento de Puno, por el Sur con el departamento de Arequipa y por el Oeste con la provincia de Chumbivilcas. Se encuentra en la región de la sierra a los 14° 47' 32" de latitud sur y 71° 24' 38" de longitud Oeste del meridiano de Greenwich. La provincia tiene una extensión de 5 311,09 km<sup>2</sup>. y una altura de 3928 metros sobre el nivel del mar, Geográficamente ubicada en una zona frígida, sus temperaturas oscilan entre -4 °C. y 19 °C. dependiendo de la estación.

#### 3.2. ÁMBITO DE ESTUDIO

El C.S. de Yauri fue fundada el 8 de setiembre del 1968 con la denominación de CLAS YAURI en la provincia de Espinar en el departamento de Cusco, y el área de Nutrición como lo hacen todas las Redes de Salud utiliza el Sistema Informático del SIEN, área que será nuestro ámbito de estudio, junto con el área de Estadística, quienes son responsables de dicho sistema SIEN.



### **3.3. TIPO Y DISEÑO DE INVESTIGACIÓN**

Este trabajo de investigación es de tipo Básica, según Carrasco (2013). Refiere que no tiene interés aplicativo ya que busca desarrollar y ahondar el conocimiento científico existente.

La investigación desarrollada corresponde al tipo de estudio correlacional y el diseño es no experimental transversal-analítico (Hernández, Fernandez, & Baptista, 2010), porque se utilizó herramientas y evidencias teóricas relacionadas con los temas de factores de riesgo asociados a la anemia en niños menores de 5 años.

### **3.4. POBLACIÓN Y MUESTRA**

#### **3.4.1. Población**

La población está comprendida por los niños menores de cinco años registrados en el SIEN de la DIRESA provincia de Espinar – Cusco correspondiente al año 2019, siendo un total de 2307 niños.

#### **3.4.2. Muestra**

Para determinar el tamaño de la muestra (Martinez, 2012), se calcula mediante el uso de la fórmula estadística para el caso en el que se conoce la proporción de anemia en niños menores niños menores a 5 años que acuden a los establecimientos de salud, 2019, el cual es de 41.8% en la región de Cusco. Según: (SIEN, 2019).



$$n = \frac{NZ_{\alpha/2}^2 pq}{(N-1)e^2 + Z_{\alpha/2}^2 pq}$$

Donde:

$Z_{\alpha/2} = 1.96$  : valor tipificado de la distribución normal para un nivel de confianza elegido (95%).

$p = 0.418$  : proporción de anemia en niños menores niños menores a 5 años que acuden a los establecimientos de salud, región Cusco de 2019 según (SIEN, 2019).

$q = 1 - 0.418 = 0.582$ : proporción de casos no favorables.

$e = 0.05$  ó 5% : Error máximo permisible.

$N = 2307$  : Tamaño de la población

Remplazando tenemos, un tamaño de muestra de  $n = 321.82 \approx 322$  niños.

Para la selección de los elementos que constituyeron la muestra de estudio, se utilizó el método del muestreo aleatorio simple, que consistió en listar del 001 al 2307 todos los nombres de los niños con sus respectivos códigos, y luego con el software SPSS versión 25 se seleccionaron cada historia clínica hasta completar la muestra.

### 3.5. INSTRUMENTO DE RECOLECCIÓN Y PROCESAMIENTO DE DATOS

El método que se utilizó fue análisis de registro de datos (ver Anexo A) proporcionado por el Sistema de Información del Estado Nutricional del niño menor de cinco años y Gestantes que acceden al Establecimiento de Salud (SIEN) de la



DIRESA provincia de Espinar – Cusco correspondiente al año 2019, dicha información es a nivel local, distrital y provincial, información útil para la toma de decisiones y la planificación de intervenciones principalmente a nivel local.

### **3.5.1. Procedimiento de recolección de datos**

Se realizó las coordinaciones, con el responsable del Área de Nutrición, como con el responsable de Estadística del Centro de Salud de Yauri, Obtenido la autorización para la revisión de la base de datos del Sistema SIEN.

### **3.5.2. Plan de procesamiento y análisis de datos**

Una vez solicitada los datos se sometieron a un proceso de codificación y transferidos a una base de datos, mediante el programa Excel 2016 y SPSS versión 25 para su tabulación. Posteriormente a los datos se les realizó un análisis descriptivo y luego se le realizó el análisis de regresión logística binaria múltiple para finalmente ser interpretados. Los resultados se organizaron en tablas y figuras estadísticas, de acuerdo a los objetivos de la investigación.

#### **Selección de variables a través de árboles de decisión**

Debido a la gran cantidad de variables y datos que se tuvieron, se optó por utilizar la técnica de los árboles de decisión con el propósito de clasificar las variables en subgrupos (factores de riesgo) de población que contienen elementos homogéneos dentro de ellos y heterogéneos entre los subgrupos con respecto a la variable a discriminar, en nuestro caso es la variable presencia de anemia.



Los árboles de decisión son una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas (Berlanga Silvente, Rubio Hurtado, & Vilà Baños, 2013). Así mismo, nos ayudó a tomar la decisión más “acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones.

### **Análisis estadístico**

La base de datos en estudio se construyó a partir de la información generada por el SIEN de la DIRESA provincia de Espinar – Cusco correspondiente al año 2019. Cada fila en la base de datos contiene información de un niño. Cada columna en la base de datos representa una variable en estudio.

Primeramente, se realizó un análisis descriptivo de datos y una revisión de los antecedentes del trabajo de investigación y de esta forma identificar los factores que influyen en el riesgo de anemia.

Posteriormente cada uno de los efectos fue contrastado con la variable dependiente “presencia de anemia en el niño” mediante un Modelo de Regresión Logística (MRL) donde se halló la fuerza de asociación (OR) e intervalo de confianza entre las variables y su significancia (p\_valor) mediante la prueba Wald como la prueba de Hosmer-Lemeshow. De lo que se trata es de plantear un modelo de regresión logística que estime la probabilidad del riesgo a presentar anemia en los niños menores a 5 años. Por tanto, el objetivo del trabajo se centró en explicar el comportamiento de una variable categórica con dos modalidades: niño con anemia o bien un niño normal.



### 3.6. OPERACIONALIZACIÓN DE VARIABLES

Los factores con su respectiva operacionalización de variables elegidos de la base de datos del SIEN para presente trabajo de investigación se muestran en la tabla 3. Algunas variables no fueron elegidas dado que se consideran como condición de estudio.

**Tabla 3. Operacionalización de Variables.**

<i>Variable</i>	<i>Dimensión</i>	<i>Indicadores</i>	<i>Medición</i>	
FACTORES DE RIESGO DE ANEMIA	1. Factores Demográficos	Sector	1. Alto Pichigua	
			2. Condorama	
			3. Coporaque	
			4. Espinar	
			5. Ocoruro	
			6. Pallpata	
			7. Pichigua	
			8. Suykutambo	
			Altitud_Loc	m.s.n.m.
			Hemoglobina	g/dl
2. Factores biológicos		Sexo	1. Femenino	
			2. Masculino	
		Edad	En meses	
		Peso	Kilogramos	
		Talla	Centímetros	



		N_CCred	N° de Controles
		N_Sachets	N° de Sachets
	3. Factores de aplicación SIEN	N_Consejerias	N° de Consejerías
		N_Sesiones	N° de Sesiones
		N_Visitas	N° de Visitas
		JUNTOS	0. No 1. Si
	4. Factores programas sociales	SIS	0. No 1. Si
		PIN	0. No 1. Si
VARIABLE DEPENDIENTE	Presencia de anemia en niños(as) menores de 5 años	Diagnóstico de anemia	0. No anémico 1. Anémico

**Fuente:** Elaboración propia.

## CAPITULO IV

### RESULTADOS Y DISCUSIÓN

#### 4.1. IDENTIFICACIÓN DE FACTORES

##### 4.1.1. Prevalencia de anemia

Para determinar la prevalencia de anemia en niños menores de 5 años, se utilizó el total de los datos de la SIEN-DIRESA 2019, siendo un total de 2307 niños registrados en la provincia de Espinar, Cusco.

**Tabla 4. Prevalencia de anemia según sexo, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar – Cusco 2019.**

Sexo	Anemia					
	No anémico		Anémico		Total	
	n	%	n	%	n	%
Femenino	712	30.9%	437	18.9%	1149	49.8%
Masculino	662	28.7%	496	21.5%	1158	50.2%
Total	1374	59.6%	933	40.4%	2307	100%

**Fuente:** Elaboración propia.

Los resultados nos muestran que, la prevalencia de anemia en niñas es de 18.9% y en niños de 21.5% siendo estos porcentajes bastante considerables y por tanto, reflejando la problemática de la provincia de Espinar. Así mismo, en general se tiene una prevalencia de Anemias del 40.4%.

#### 4.1.2. Selección de variables por factor

Para esta etapa se utilizó la técnica de árboles de decisión, dado que esta técnica pertenece a la minería de datos, nos permite sondear y explorar los datos para sacar la información oculta en ellos (Berlanga, Rubio, & Vilà, 2013); razón por la cual, se trabajó con la totalidad de los datos.

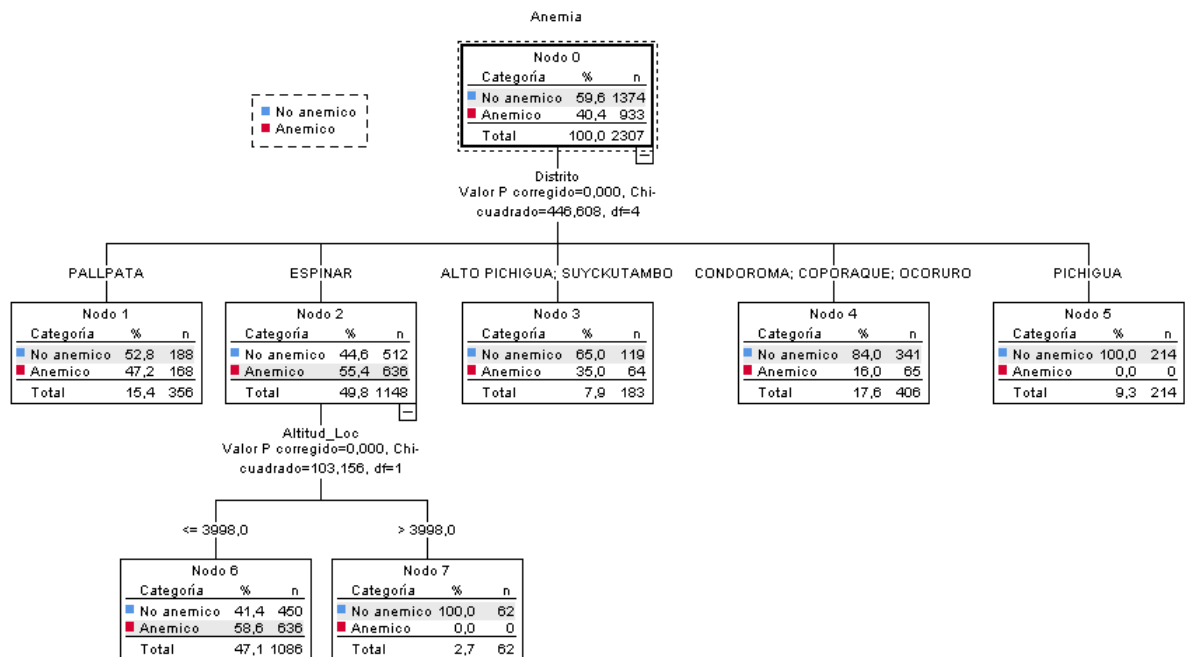
##### a) Factores demográficos

Los factores demográficos considerados en la base de datos SIEN son: distrito, localidad y altitud de la localidad. En este caso se considera solo las variables distrito y altitud de la localidad (Altitud\_Loc). El árbol de decisión como procedimiento, excluirá de forma automática cualquier variable de las seleccionadas cuya contribución al modelo final no sea significativa.

En la figura 2, el nodo 0 describe la variable dependiente, en nuestro caso diagnóstico de anemia: porcentaje de No anémico (59.6%) y anémico (40.4%). La variable distrito es el mejor predictor para el diagnóstico de la Anemia, la cual se clasifica en 5 grupos (nodos de 1 a 5) y la variable Altitud clasificada en dos grupos (nodos 6 y 7).

La probabilidad más alta de que un niño menor a 5 años tenga anemia (58.6%) se da entre los niños del distrito de Espinar y la localidad está a una altitud menor de 3998 msnm (ver figura 2). La probabilidad más baja (0.0%) se da entre los niños del distrito

de Pichigua y también entre los niños del distrito de Espinar y están en una altitud mayor a 3998 msnm.



**Figura 2.** Diagrama del árbol para el modelo de anemia según factores demográficos, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.

**Fuente:** Elaboración propia.

Por tanto, en resumen, los nodos que definen el perfil de los niños contraigan anemia (variables del factor Demográfico que influyen en la Anemia) son: Nodo 0 -Nodo 2 - Nodo 6. Es decir, influyen las siguientes variables: *Distrito* – *Altitud\_Loc*.

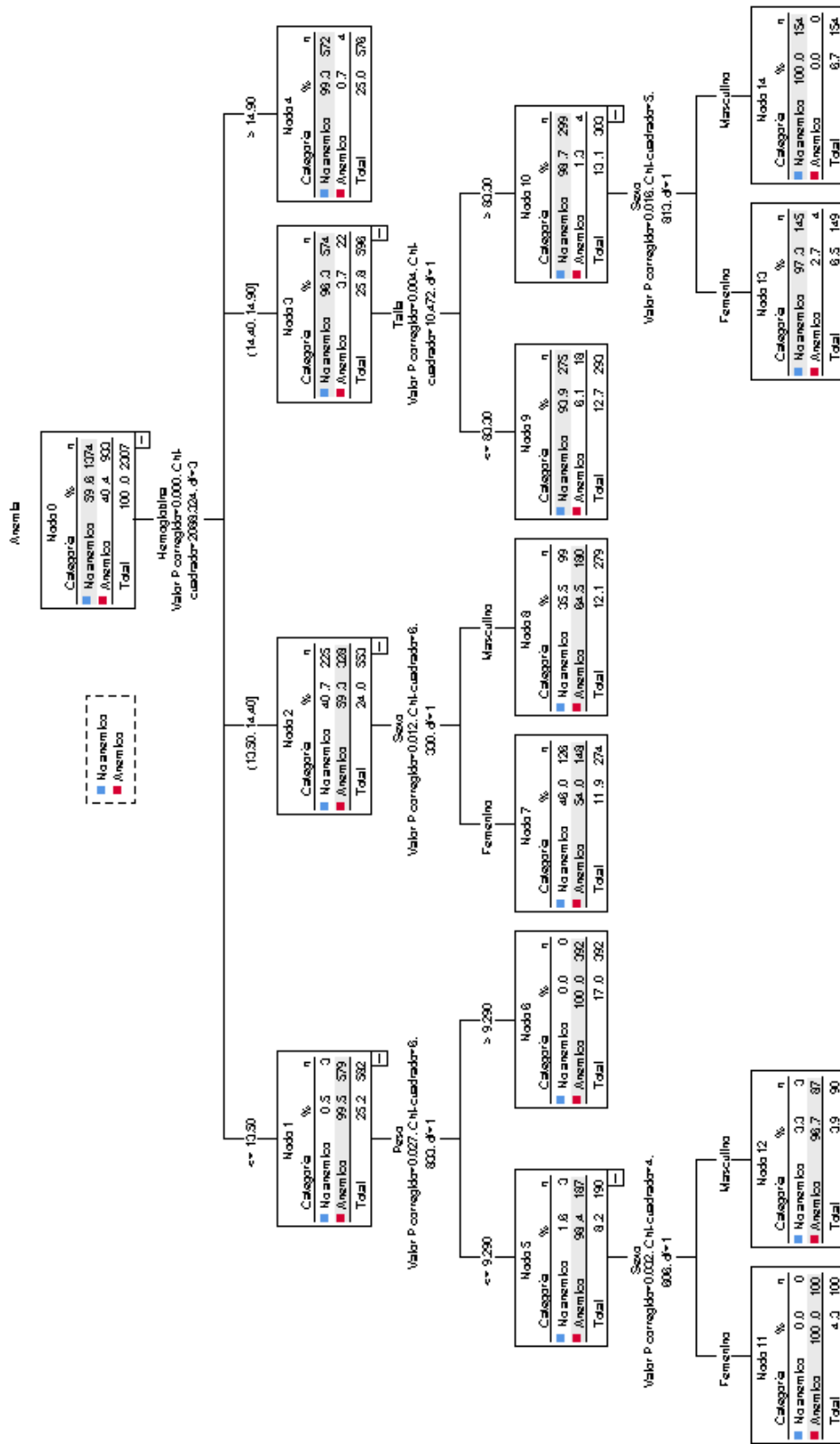
## b) Factores biológicos

Los factores biológicos son: Edad, Peso, Talla, Hemoglobina y Sexo. A través de la técnica de árbol de decisión elegiremos las variables cuya contribución al modelo sea significativa.



En la figura 3, se observa que, la variable hemoglobina es el mejor predictor para el diagnóstico de la Anemia, la cual se clasifica en 4 grupos (nodos de 1 a 4); la variable peso se clasifica en dos grupos (nodos 5 y 6) y la variable talla también clasificada en dos grupos (nodos 9 y 10). Es importante mencionar que, la variable edad no se muestra en la estructura del árbol, por tanto, esta no influye en el diagnóstico de anemia.

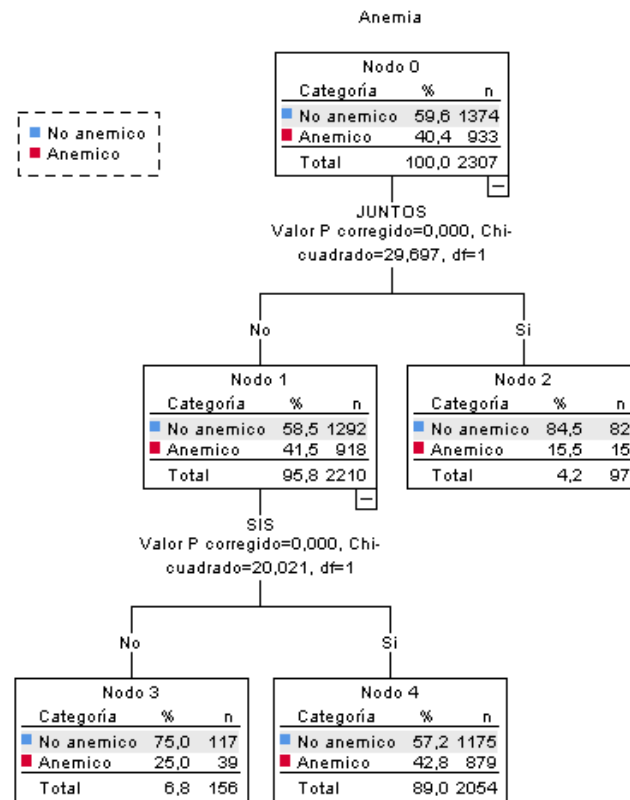
La probabilidad más alta de que un niño menor a 5 años tenga anemia (100%) se da entre los niños con hemoglobina menor igual a 13.50, con peso menor a 9.29 kg y sexo femenino. La probabilidad más baja (0.0%) se da entre los niños con hemoglobina entre  $(14.4 ; 14.9]$ , talla mayor a 80.3 cm y sexo masculino.



**Figura 3.** Diagrama del árbol para el modelo de anemia según factores biológico, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.

*Fuente:* Elaboración propia.

### c) Factores sociales



**Figura 4.** Diagrama del árbol para el modelo de anemia según factores programa social, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019.

**Fuente:** Elaboración propia.

Los factores sociales son: Pertenencia al programa social JUNTOS y PIN y pertenencia al SIS. A través de la técnica de árbol de decisión se eligieron las variables con contribución al modelo.

En la figura 4, se observa que, la variable de pertenencia al programa social JUNTOS es el mejor predictor para el diagnóstico de la anemia y la variable de pertenencia al SIS.



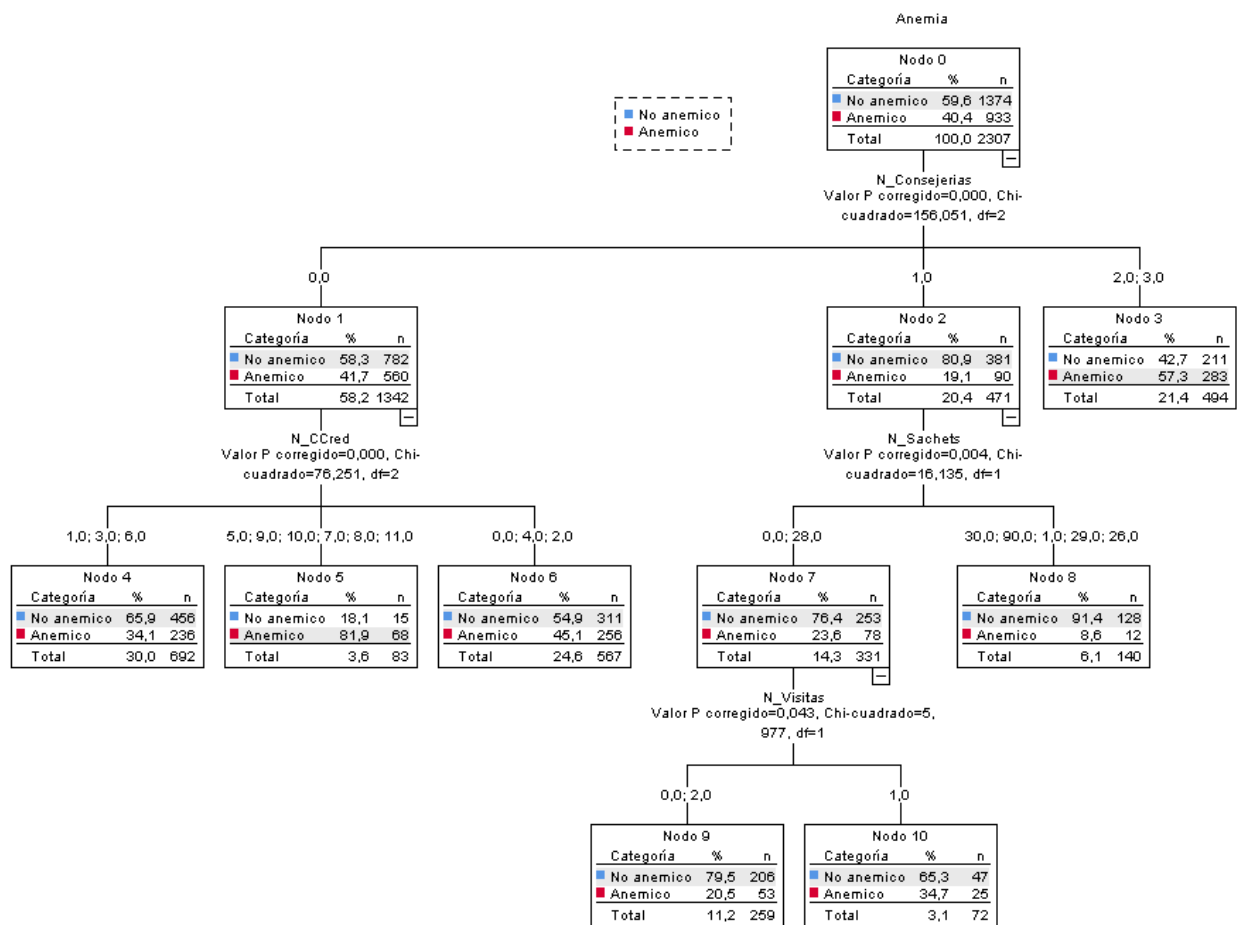


La probabilidad más alta de que un niño menor a 5 años tenga anemia (42.8%) se da entre los niños que no pertenecen al programa JUNTOS y si están en el SIS. La probabilidad más baja de ser anémico (15.5%) se da entre los niños que si pertenecen al programa JUNTOS.

#### **d) Factores de aplicación SIEN**

Los factores de aplicación del SIEN son: Número de controles, sachets, consejerías, sesiones y visitas. Las variables con mejor contribución al modelo se muestran en la figura 5, en esta se observa que, la variable número de consejerías es el mejor predictor para el diagnóstico de la anemia.

La probabilidad más alta de que un niño menor a 5 años tenga anemia (81.9%) se da entre los niños que no recibieron consejerías y tiene controles más de cinco. La probabilidad más baja de ser anémico (8.6%) se da entre los niños que tienen una consejería y sachets consumidos más de diez.



**Figura 5.** Diagrama del Árbol para el modelo de anemia según factores aplicación SIEN, en los niños menores de 5 años en la MICRORED Yauri de la provincia de Espinar, Cusco 2019

**Fuente:** Elaboración propia.

De acuerdo a este análisis las variables se recodifican de acuerdo a la clasificación que nos propuso el árbol de decisiones en cada factor del siguiente modo:

**Tabla 5. Recodificación de variables para el modelo logístico.**

<i>Variable</i>	<i>Dimensión</i>	<i>Indicadores</i>	<i>Categorías</i>
FACTORES DE RIESGO DE ANEMIA	1. Factores demográficos	X1: Sector	0. Pichigua 1. Condorama, Coporaque, Ocoruro 2. Alto Pichigua, Suyckutambo 3. Pallpata 4. Espinar
		X2: Altitud	0. > 3998 1. ≤ 3998
	2. Factores biológicos	X3: Hemoglobina	0. > 14.90 1. [ 14.40 ; 14.90 > 2. [ 13.50 ; 14.40 > 3. < 13.50
		X4: Sexo	0. Femenino 1. Masculino
		X5: Peso	0. > 9.29 1. ≤ 9.29
		X6: Talla	0. > 80.3 1. ≤ 80.3
	3. Factores de aplicación SIEN	X7: CCred	0. de 0 a 5 1. más de 5
		X8: Sachets	0. más de 20 1. menos de 20
		X9: Consejerías	0. Uno a más 1. Sin consejería
		X10: Visitas	0. Sin visitas 1. más de uno
	4. Factores sociales	X11: JUNTOS	1. No 0. Si
		X12: SIS	1. No 0. Si

**Fuente:** Elaboración propia.



## **4.2. ESTIMACIÓN DE MODELO**

### **4.2.1. Selección de variables**

Luego del análisis de los factores y su posterior selección y recodificación de variables, se procedió a la estimación inicial del modelo con todas las variables elegidas y analizar su significancia en la determinación de anemia en niños menores de 5 años, para esto se trabajó con la muestra seleccionada proporcional al estrato (distrito).

**Tabla 6. Variables que ingresan al modelo**

		Variables en la ecuación					
		Error					
		B	estándar	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup>	Sector	,561	,239	5,513	1	,019	1,752
	Altitud	-1,922	,914	4,419	1	,036	,146
	Hemoglobina	5,089	,757	45,197	1	,000	162,221
	Sexo	,037	,527	,005	1	,945	1,037
	Peso	-,731	,780	,877	1	,349	,482
	Talla	,641	,741	,749	1	,387	1,899
	CCred	-,496	1,132	,192	1	,661	,609
	Sachets	3,021	2,280	1,756	1	,185	20,521
	Consejeria	-,242	,758	,102	1	,750	,785
	Visitas	4,354	1,476	8,709	1	,003	77,821
	JUNTOS	,440	1,400	,099	1	,753	1,553
	SIS	1,230	1,136	1,173	1	,279	3,423
	Constante	-12,416	3,291	14,228	1	,000	,000

a. Variables especificadas en el paso 1: Sector, Altitud, Hemoglobina, Sexo, Peso, Talla, CCred, Sachets, Consejeria, Visitas, JUNTOS, SIS.

**Fuente:** Elaboración propia.

En la tabla 7, se observa las variables significativas para el modelo:

- X1: Sector
- X2: Altitud
- X3: Hemoglobina
- X10: Número de Visitas

En la tabla 7, se muestra la significancia de las variables X1, X2, X3 y X10 en el nuevo modelo y se observa claramente que son significativas, por tanto, influyen en el modelo. Cabe mencionar que se hicieron pruebas eliminando estas variables del modelo, sin embargo, su desempeño fue más pobre, por lo que se mantuvieron.

**Tabla 7. Variables que ingresan al nuevo modelo**

		Variables en la ecuación					95% C.I. para		
		B	Error estándar	Wald	gl	Sig.	Exp(B)	EXP(B)	
								Inferior	Superior
Paso	Sector	,582	,207	7,935	1	,005	1,790	1,194	2,685
1 <sup>a</sup>	Altitud	-1,718	,850	4,086	1	,043	,179	,034	,949
	Hemoglobina	4,534	,650	48,644	1	,000	93,122	26,045	332,956
	Visitas	3,349	1,378	5,909	1	,015	28,477	1,913	423,844
	Constante	-8,304	1,269	42,789	1	,000	,000		

a. Variables especificadas en el paso 1: Sector, Altitud, Hemoglobina, Visitas.

**Fuente:** Elaboración propia

En la tabla 8, se observa el coeficiente de determinación, para nuestro modelo es de 85.8 por ciento. Es decir, un 85.8 por ciento de la variación de la variable anemia en niños menores de 5 años (anémico o no anémico) es explicada por las variables incluidas en el modelo que son X1: Sector, X2: Altitud de la localidad, X3: Hemoglobina y X10: Número de visitas.

**Tabla 8. Prueba del coeficiente de determinación**

**Resumen del modelo**

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	109,419 <sup>a</sup>	,635	,858

- a. La estimación ha terminado en el número de iteración 8 porque las estimaciones de parámetro han cambiado en menos de ,001.

**Fuente:** Elaboración propia

En la tabla 9, observamos la clasificación para el modelo donde se obtiene una tasa de buena clasificación del 94.7 por ciento. En esta tabla de clasificación se comprueba que el modelo tiene una especificidad alta (93.8%) y una sensibilidad también alta de (96.2%). Con las variables predictoras: X1: Sector, X2: Altitud de la localidad, X3: Hemoglobina y X10: Número de visitas, se comprueba que el modelo tiene una efectividad de 94.7 por ciento, es decir posee una buena capacidad predictora con las variables incluidas en el modelo. Hay que tomar en cuenta que el punto de corte es de 0.5.

**Tabla 9. Tabla de clasificación**

**Tabla de clasificación<sup>a</sup>**

Observado		Pronosticado		
		Anemia		Porcentaje
		No anémico	Anémico	correcto
Paso 1	Anemia	180	12	93,8
	Anémico	5	125	96,2
Porcentaje global				94,7

a. El valor de corte es ,500

También de la tabla de clasificación podemos decir:

VP: La condición de que un niño o niña menor de 5 años termine como anémico – Predicción correcta 96.2%.

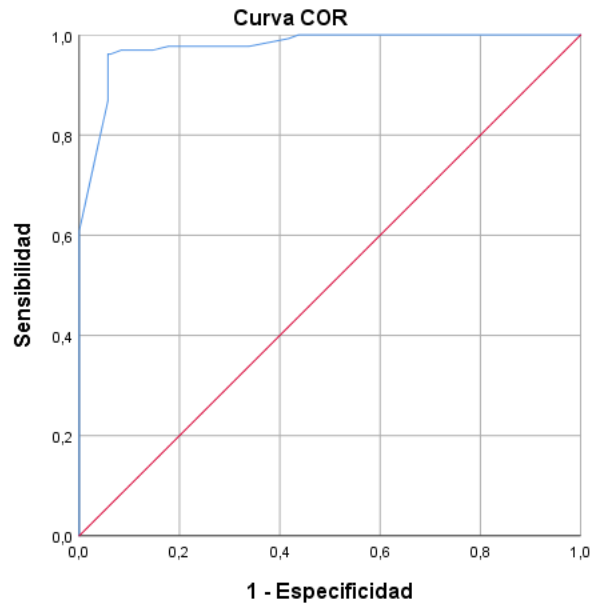
VN: La condición de que un niño o niña menor de 5 años termine como no anémico – Predicción correcta 93.8%.

FP: La condición de que un niño o niña menor de 5 años termine como anémico – Predicción Incorrecta 3.8%.

FN: La condición de que un niño o niña menor de 5 años termine como no anémico – Predicción Incorrecta 6.2%.

Para poder analizar el rendimiento (o la efectividad, en su más amplio sentido) de nuestro modelo clasificador, recurrimos a la curva COR, para lo cual tomamos los valores de la variable anemia frente a los valores de probabilidad pronosticados. En la figura 6 y tabla 10 se observa que nuestro modelo representado por la curva azul está por encima de la curva aleatoria cubriendo un área de 0,976.





Los segmentos de diagonal se generan mediante empates.

**Figura 6.** Gráfica de la curva COR

*Fuente:* Elaboración propia

**Tabla 10.** Tabla del área bajo la curva COR.

Área bajo la curva				
Variables de resultado de prueba: Probabilidad pronosticada				
95% de intervalo de confianza asintótico				
Área	Desv. Error <sup>a</sup>	Significación asintótica <sup>b</sup>	Límite inferior	Límite superior
			,976	,007

Las variables de resultado de prueba: Probabilidad pronosticada tienen, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Las estadísticas podrían estar sesgadas.

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

*Fuente:* Elaboración propia

Si el área que encierra la Curva COR es 1, se tendría una clasificación perfecta y conforme nos alejamos de este valor la clasificación disminuye; entonces como la curva COR es 0.976 hace de este un buen modelo para clasificar a niños. Con punto de corte 0.5.

#### 4.2.2. Verosimilitud del modelo

En la tabla 11, se observa la significación estadística con la prueba de Wald, la constante del modelo es significativa, por lo tanto, la constante se debe incluir en el modelo porque produce un cambio sobre la variable Anemia (estado del niño o niña como no anémico o anémico).

**Tabla 11. Prueba del coeficiente de intercepción.**

		Variables en la ecuación					
		Error					
	B	estándar	Wald	gl	Sig.	Exp(B)	
Paso 0	Constante	-,390	,114	11,788	1	,001	,677

**Fuente:** Elaboración propia

Al evaluar la hipótesis nula ( $\beta_i = 0$ ) la significancia estadística asociada, y el valor de la OR( $\exp(B)$ ), las variables que contribuyen al modelo son X1: Sector, X2: Altitud de la localidad, X3: Hemoglobina y X10: Número de visitas a un nivel de significación del 5 por ciento. Por tanto, el modelo estimado es el siguiente:

$$p(x) = \frac{1}{1 + e^{[-8.304 + 0.582x_1 - 1.718x_2 + 4.534x_3 + 3.349x_{10}]}}$$

$$\ln(\text{odds}) = -8.304 + 0.582x_1 - 1.718x_2 + 4.534x_3 + 3.349x_{10}$$

#### 4.2.3. Prueba de bondad de ajuste

La tabla 12, nos muestra la prueba de bondad de ajuste del modelo, a un nivel de significación de 5 por ciento, el modelo es el adecuado, es decir el modelo de regresión logística se ajusta a los datos (p-valor = 0.088).

**Tabla 12. Prueba de bondad de ajuste.**

##### Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	12,418	7	,088

*Fuente:* Elaboración propia.

También, en la tabla 13, se tiene la prueba de Ómnibus:

**Tabla 13. Prueba de Omnibus.**

##### Pruebas ómnibus de coeficientes de modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	324,955	4	,000
	Bloque	324,955	4	,000
	Modelo	324,955	4	,000

*Fuente:* Elaboración propia.

En la primera fila se contrasta la  $H_0$  de que los coeficientes de las variables añadidas son cero. Con la significancia de 0.000 se rechaza la  $H_0$ , es decir al menos uno de los coeficientes de las variables es diferente de cero.

Indica que el modelo con las variables pertenecientes al modelo tiene un ajuste significativo.

### 4.3. PREDICCIÓN DEL RIESGO DE ANEMIA

Si usamos los coeficientes estimados de la tabla 6, columna B, y los valores de las variables para cada caso en la base de datos de la muestra, podemos estimar la probabilidad o riesgo de anemia de cada niño o niña menor a 5 años mediante el modelo de regresión logística. Por ejemplo, en la tabla 14 se plantean dos casos (48 y 255 de la muestra), cuyos valores de predicción de riesgo de anemia en niños menores de 5 años son los siguientes:

**Tabla 14. Predicción de riesgo de anemia en niños menores de 5 años para dos casos.**

<i>Variables</i>	<i>Caso</i>	
	<i>Y<sub>79</sub></i>	<i>Y<sub>255</sub></i>
Y: Anemia	0	1
X1: Sector	4	1
X2: Altitud	1	1
X3: Hemoglobina	0	3
X10: Visitas	0	0
Predicción	0.00046	0.98467

**Fuente:** Elaboración propia.



En el caso Y<sub>79</sub> se representa datos de un niño no anémico, que pertenece al Sector 4 correspondiente a la capital de provincia Espinar, con una altitud menor o igual a 3998 m.s.n.m., niveles de hemoglobina mayores a 14.90 y no tuvo visitas del personal del SIEN, podemos decir que es un perfil típico de un niño de la zona. Observe que la probabilidad estimada por el modelo resultó menor que el  $P_c = 0.50$ , por lo que este primer caso es clasificado de manera correcta como niño no anémico. Para el caso Y<sub>255</sub>, que pertenece al Sector 4 correspondiente a los distritos de Condorama, Copraque y Ocoruro, con una altitud menor o igual a 3998 m.s.n.m., niveles de hemoglobina menores a 13.50 y no tuvo visitas del personal del SIEN es clasificado como anémico.

## V. CONCLUSIONES

**PRIMERO:** Este trabajo de investigación ha identificado los factores de riesgo registrados en el SIEN de la DIRESA provincia de Espinar – Cusco correspondiente al año 2019, que se asocian a un mayor riesgo de anemia en niños menores de cinco años y que la literatura refiere como factores de riesgo; encontrándose en general la prevalencia de anemia en un valor de 40.4%.

Con los resultados del ajuste del modelo se pudo estimar la contribución de los factores de riesgo de anemia. El modelo clasificador a través de una muestra representativa, cuya prueba de bondad de ajuste de Hosmer y Lemeshow, a un nivel de significación de 0.05, es de  $p\text{-valor}=0.088$ , con un coeficiente de determinación de Nagelkerke de 85.8%. La tasa de clasificación de verdaderos negativos (no anémico pronosticados como no anémicos) fue del 93.8%; mientras que la tasa más relevante en este caso, clasificar correctamente a los niños anémicos (verdaderos positivos) fue de 96.2%, demostrando gran capacidad predictiva del modelo.

**SEGUNDO:** Basado en la estructura de la base de datos de la SIEN y bajo el planteamiento del modelo conceptual de los factores de riesgo de la anemia, se logró identificar los factores de riesgo que contribuyen a un mayor riesgo de anemia en niños menores de 5 años en la provincia de Espinar utilizando la técnica de árboles de decisión que permitieron reducir y clasificar las variables siguientes: Dentro de los factores demográficos de riesgo corresponden el Distrito y Altitud de la localidad. Los factores de riesgo



biológicos tenemos niveles de hemoglobina, peso, talla y sexo. Los factores de aplicación del SIEN se tienen, número de consejerías, número de controles, número de sachet consumidos, y número de visitas y el último factor social se tiene la pertenencia a los programas JUNTOS y el SIS.

**TERCERO:** Los factores de riesgo que influyen en la presencia de anemia en niños menores de 5 años en la provincia de Espinar fueron: X1: Sector, X2: Altitud de la localidad, X3: Niveles de hemoglobina y X10: Número de visitas; se logró construir el modelo que determinaría la contribución de los factores de riesgo de anemia en niños.

$$\ln(odds) = -8.304 + 0.582x_1 - 1.718x_2 + 4.534x_3 + 3.349x_{10}$$

Con los resultados del ajuste del modelo se pudo estimar la contribución de los factores de riesgo de anemia. El modelo clasificador a través de una muestra representativa, cuya prueba de bondad de ajuste de Hosmer y Lemeshow, a un nivel de significación de 0.05, es de p-valor=0.088, con un coeficiente de determinación de Nagelkerke de 85.8%. La tasa de clasificación de verdaderos negativos (no anémico pronosticados como no anémicos) fue del 93.8%; mientras que la tasa más relevante en este caso, clasificar correctamente a los niños anémicos (verdaderos positivos) fue de 96.2%, demostrando gran capacidad predictiva del modelo.



## VI. RECOMENDACIONES

Los hallazgos que se puede destacar al realizar este estudio, nos permiten recomendar:

- Con respecto a la MICRORED Yauri se recomienda que focalicen las campañas de sensibilización y/o campañas de sesiones demostrativas así como la visitas a los domicilios de los niños que presentan bajo hemoglobina que pueda convertirse en un caso más de anemia, para los distritos con mayor numero de casos. Ya que con demostramos con el presente estudio que son factores que logran combatir la anemia.
- Debido a la generación masiva de datos en las instituciones de salud en nuestras regiones y el país, disponemos de tanta información, que a veces no es imposible organizarla con efectividad. Por ello, se recomienda utilizar la técnica de Minería de Datos para descubrir patrones o algoritmos que permitan estudiar la dinámica de las enfermedades que afectan a nuestra niñez en el Perú.
- Tomando en cuenta que la anemia en niños menores a 5 años aqueja a todos el país, la Universidad Nacional del Altiplano y los puestos de salud promuevan la realización de campañas educativas en medios de comunicación; como también la preparación y ejecución de talleres didácticos dirigidos a la comunidad con la finalidad de identificar los factores de riesgo de la anemia en niños y madres gestantes.





## VII. REFERENCIAS BIBLIOGRAFICAS

- Ajakaye, O. G., & Ibukunoluwa, M. R. (2020). Prevalence and risk of malaria, anemia and malnutrition among children in IDPs camp in Edo State, Nigeria. *Parasite Epidemiology and Control*, 8, e00127. <https://doi.org/10.1016/j.parepi.2019.e00127>
- Arroyo-Laguna, J. (2017). Hacia un Perú sin anemia. *Revista Peruana de Medicina Experimental y Salud Publica*, 34(4), 586–587. <https://doi.org/10.17843/rpmesp.2017.344.3279>
- Berlanga Silvente, V., Rubio Hurtado, J. M., & Vilà Baños, R. (2013). Cómo aplicar árboles de decisión en SPSS. *REIRE Revista de Innovación e Investigación En Educación*, 6, 65–79. <https://doi.org/10.1344/reire2013.6.1615>
- Carrasco Diaz, S. O. (2013). *Metodología de la investigación científica* (Quinta edi). Perú: Editorial San Marcos.
- Condori Apaza, V. (2018). *Factores que determinan la Deabetes en pacientes de endocrinología del Hospital Regional Manuel Nuñez Butrón - Puno 2017*. Universidad Nacional del Altiplano. Retrieved from <http://repositorio.unap.edu.pe/handle/UNAP/9276>
- DIRESA-Cusco. (2019). Dirección Regional de Salud y Educación articulan esfuerzos y dan inicio a la semana de la lucha contra la anemia. Retrieved April 26, 2020, from <http://www.diresacusco.gob.pe/new/archivos/3086>
- Dutta, M., Bhise, M., Prashad, L., Chaurasia, H., & Debnath, P. (2020). Prevalence and risk factors of anemia among children 6–59 months in India: A multilevel analysis. *Clinical Epidemiology and Global Health*. <https://doi.org/10.1016/j.cegh.2020.02.015>



Dutta, P., & Sengupta, B. (2017). Socioeconomic Determinants of Maternal Anemia: a Disaggregated Level Analysis From Assam, India, *3*(2015), 71–87. <https://doi.org/10.17501/icoph.2017.3109>

Fiuza, D., & Rodríguez, J. C. (2000). La regresión logística: una herramienta versátil. *Sociedad Española de Nefrología*, Vol. 20., 477–565. Retrieved from [www.revistanefrologia.com/es-relacionados-la-regresion-logistica-una-herramienta-versatil-articulo-X0211699500035664](http://www.revistanefrologia.com/es-relacionados-la-regresion-logistica-una-herramienta-versatil-articulo-X0211699500035664)

Ghazvini, K., Yousefi, M., Firoozeh, F., & Mansouri, S. (2019). Predictors of tuberculosis: Application of a logistic regression model. *Gene Reports*, *17*, 100527. <https://doi.org/10.1016/j.genrep.2019.100527>

Guzmán Llanos, M. J., Guzmán Zamudio, J. L., & Llanos de los Reyes-García, M. (2016). Significado de la anemia en las diferentes etapas de la vida. *Enfermería Global*, *15*(3), 407. <https://doi.org/10.6018/eglobal.15.3.248221>

Hernández, R., Fernández, C., & Baptista, M. del P. (2010). *Metodología de la Investigación*. (McGraw-Hill, Ed.) (Quinta Edi).

Hosmer, J., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. (John Wiley, Ed.).

López-Roldán, P., & Fachelli, S. (2015). *Metodología de la Investigación Social Cuantitativa* (Primera ed). Barcelona - España: Universitat Autònoma de Barcelona. Retrieved from <https://ddd.uab.cat/record/129382>

Martínez, C. (2012). *Estadística y Muestreo* (Decima ter). Bogota - Colombia: ECOE Ediciones.



MINSA. (2017). *Plan Nacional para la reducción y control de la anemia materna Infantil y la Desnutrición Crónica Infantil en el Perú: 2017-2021*. Retrieved from <http://www.minsa.gob.pe/>

Nusinovici, S., Tham, Y. C., Chak Yan, M. Y., Wei Ting, D. S., Li, J., Sabanayagam, C., ... Cheng, C.-Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2020.03.002>

Obregón Cahuaya, C. E. (2018). *Contribución de los factores de riesgo individual y contextual al mayor riesgo de anemia en niños menores de cinco años en el Perú*. Universidad Nacional Mayor de San Marcos. Retrieved from <http://cybertesis.unmsm.edu.pe/handle/cybertesis/9694>

OMS. (2011). Concentraciones de hemoglobina para diagnosticar la anemia y evaluar su gravedad. Ginebra, Organización Mundial de la Salud, 2011 (WHO/NMH/NHD/MNM/11.1). *Vmnis*, 1–7. Retrieved from [https://www.who.int/vmnis/indicators/haemoglobin\\_es.pdf](https://www.who.int/vmnis/indicators/haemoglobin_es.pdf)

OMS. (2016). Temas de salud - Factores de riesgo. Retrieved October 2, 2020, from [https://www.who.int/topics/risk\\_factors/es/](https://www.who.int/topics/risk_factors/es/)

OMS. (2020). Sistema de Información Nutricional sobre Vitaminas y Minerales (VMNIS). Retrieved December 2, 2020, from [https://www.who.int/vmnis/database/anaemia/anaemia\\_data\\_status\\_t2/es/](https://www.who.int/vmnis/database/anaemia/anaemia_data_status_t2/es/)

Peña, D. (2002). *Análisis de datos multivariantes*. Madrid - España: University Carlos III de Madrid.



Pérez, C. (2004). *Técnica de análisis multivariante de Datos: aplicaciones con SPSS*. España: PEARSON EDUCACIÓN, S.A.

Ramos Arohuanca, M. M. (2020). *Modelo para determinar los factores que influyen en la calidad de servicio en la salud de los usuarios atendidos en establecimientos del Perú - 2016*. Universidad Nacional del Altiplano. Retrieved from <http://repositorio.unap.edu.pe/handle/UNAP/13819>

Ríos, K. D. (2017). *Regresión logística de los factores que influyen en la gastritis, hospital reional Manuel Nuñez Butrón - Puno*. Universidades Nacional del Altiplano.

Roque Cruz, M. J. (2018). *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*. Universidad Nacional del Altiplano. Retrieved from <http://repositorio.unap.edu.pe/handle/UNAP/7755>

Sánchez-Cantalejo, E. (2000). *Regresión Logística en Salud Pública*. Granada - España: Escuela Andaluza de Salud Pública.

SIEN. (2018). *Estado Nutricional en Niños y Gestantes de los Establecimientos de Salud del Ministerio de Salud*. (MINSa, Ed.). Lima, Perú: Dirección Ejecutiva de Vigilancia Alimentaria y Nutricional. Retrieved from <http://www.ins.gob.pe>

Silva Ayçaguer, L. C. (2005). Una ceremonia estadística para identificar factores de riesgo. *Salud Colectiva*, 1(3), 309. <https://doi.org/10.18294/sc.2005.49>

Solano, D. O., Ramírez, A., Bartolo, F., Giraldo, O., & Salinas, A. (2007). Análisis de diagnóstico en el modelo de regresión logística: una aplicación. *Pesquimat*, 10(1), 55–70.

Velásquez, J., Rodríguez, Y., Gonzáles, M., Astete, L., Loyola, J., Vigo, W., & Rosas, Á. (2016). Factores asociados con la anemia en niños menores de tres años en Perú: análisis



de la Encuesta Demográfica y de Salud Familiar, 2007-2013. revista Biomédica [revista en internet] 2016 [acceso 30 de septiembre]; 36; [220-9]. <http://dx.doi.org/10.7705/biomedica.v3>. *Biomédica*, 36, 220–229. Retrieved from file:///C:/Users/user/Downloads/2896-Texto del manuscrito completo (cuadros y figuras insertos)-15095-1-10-20160523.pdf

Zavaleta, N. (2017). Anemia infantil: Retos y oportunidades al 2021. *Rev Peru Med Exp Salud Publica*, 34(4 SUPPL.), 588–589. <https://doi.org/10.17843/rpmesp.2017.344.3281>

Zavaleta, N., & Astete-Robilliard, L. (2017). Efecto de la anemia en el desarrollo infantil: Consecuencias a largo plazo. *Revista Peruana de Medicina Experimental y Salud Publica*, 34(4), 716–722. <https://doi.org/10.17843/rpmesp.2017.344.3251>



**ANEXO A:** Registro SIEN para el periodo 2019, provincia de Espinar – Cusco.

EESS	X1	X2	X4	X5	X6	X3	X11	X12	X8	X9	X10	Anemia	PRE_1	
1	3	1	0	0	0	2	1	1	1	1	0	1	0.68854	
1	3	1	0	0	1	1	1	1	1	1	0	1	0.02319	
1	3	1	0	0	0	3	1	1	1	1	0	1	0.99517	
1	3	1	1	1	0	3	1	1	1	1	0	1	0.99517	
1	3	1	1	0	1	1	1	1	1	1	0	0	0.02319	
1	3	1	1	0	1	0	1	1	1	1	0	0	0.00025	
1	3	1	0	0	1	0	1	1	1	1	0	0	0.00025	
1	3	1	1	0	0	0	1	0	1	1	0	0	0.00025	
1	3	1	1	1	1	0	2	1	0	1	1	0	1	0.68854
1	3	1	1	0	1	0	1	0	1	1	0	0	0.00025	
1	3	1	1	0	1	2	1	0	1	1	0	1	0.68854	
1	3	1	1	0	1	0	1	0	1	1	0	0	0.00025	
1	3	1	0	0	1	1	1	0	1	1	0	0	0.02319	
1	3	1	0	0	1	0	1	0	1	1	0	0	0.00025	
1	3	1	0	0	1	0	1	0	1	1	0	0	0.68854	
1	3	1	0	0	0	2	1	0	1	1	0	1	0.68854	
1	3	1	0	0	0	2	1	0	1	1	0	1	0.68854	
1	3	1	1	0	1	2	1	0	1	1	0	1	0.68854	
1	3	1	1	0	1	1	1	0	0	1	0	0	0.02319	
1	3	1	0	0	1	1	1	0	1	1	0	0	0.02319	
1	3	1	1	0	1	1	1	0	1	1	0	0	0.02319	
1	3	1	0	0	1	0	1	0	1	1	0	0	0.00025	
2	4	1	0	0	1	3	1	0	1	0	0	1	0.99729	
2	4	1	1	0	0	2	1	0	1	0	0	1	0.79830	
2	4	1	0	1	0	0	1	0	1	0	0	0	0.00046	
2	4	1	1	0	0	3	1	0	1	1	0	1	0.99729	
2	4	1	1	0	0	3	1	0	1	0	0	1	0.99729	
5	1	1	0	0	0	0	1	0	1	1	0	0	0.00008	
5	1	1	1	1	0	1	1	1	1	1	0	0	0.00735	
5	1	1	1	0	1	3	1	0	1	1	0	1	0.98467	
5	1	1	0	0	1	0	1	0	1	1	0	0	0.00008	
5	1	1	1	0	1	2	1	0	1	1	0	0	0.40818	
5	1	1	0	0	1	1	1	0	1	1	0	0	0.00735	
5	1	1	1	1	0	0	0	0	1	1	0	0	0.00008	
5	1	1	0	0	1	1	1	0	1	1	0	0	0.00735	
5	1	1	1	1	0	3	1	0	1	1	0	1	0.98467	
5	1	1	1	1	0	1	1	0	1	1	0	0	0.00735	
5	1	1	1	0	1	1	1	0	1	1	0	0	0.00735	
5	1	1	1	1	0	2	1	0	1	1	0	0	0.40818	
6	1	1	0	0	0	2	0	0	1	1	0	0	0.40818	
6	1	1	1	0	1	1	0	0	1	1	0	0	0.00735	
6	1	1	0	0	1	1	0	0	1	1	0	0	0.00735	
6	1	1	0	1	0	2	1	0	1	1	1	1	0.95155	
6	1	1	0	0	1	0	0	0	1	1	0	0	0.00008	
6	1	1	0	1	0	2	0	0	1	1	0	0	0.40818	
6	1	1	0	0	1	3	1	1	1	1	0	1	0.98467	
6	1	1	1	1	0	2	0	0	1	1	1	1	0.95155	
11	1	1	1	1	0	1	1	0	0	1	0	0	0.00735	
11	1	1	1	1	0	3	1	0	1	1	0	1	0.98467	
10	2	0	1	0	1	0	0	0	1	1	0	0	0.00079	
10	2	0	0	0	0	0	0	0	1	1	0	0	0.00079	
10	2	0	1	0	0	1	1	0	1	1	0	1	0.06881	
10	2	0	0	0	0	2	1	0	1	1	0	1	0.87311	
10	2	0	0	0	1	0	1	0	1	1	0	0	0.00079	



**ANEXO B:** Descripción de variables del SIEN - DIRESA provincia de Canchis.

<i>Variable</i>	<i>Descripción</i>
EESS	<b>Establecimiento de salud:</b> Corresponde a la identificación de los 12 establecimientos de salud que cuenta el Ministerio de Salud para las atenciones en niños y gestantes en la MICRORED de Yauri.
Fecha	Se identifica la fecha en dd/mm/aaaa de la atención recibida por el niño menor de cinco años.
Hist_Clinica	<b>Historia Clínica:</b> Corresponde al número de historia clínica del niño que fue atendido en el establecimiento de salud.
Provincia	Corresponde a la identificación de la provincia, para nuestro caso ESPINAR, donde reside el niño, de acuerdo a lo señalado por el INEI.
Distrito	Corresponde a la identificación del distrito, de los 8 distritos, donde reside el niño, de acuerdo a lo señalado por el INEI.
Localidad	Corresponde a la identificación de la localidad donde reside el niño.
Hemoglobina	Corresponde al resultado de la hemoglobina (Hb) obtenido del análisis sanguíneo respectivo en el niño, expresado en g/dl.



---

Altitud_Loc	Corresponde a la altitud de la localidad o distrito donde reside el niño, en metros sobre el nivel del mar (msnm) identificados por INEI.
Hbc	<b>Hemoglobina corregida:</b> Corresponde a la cifra de hemoglobina aplicando los factores de corrección de acuerdo a la altitud, y se expresa en g/dl.
Sexo	Se identifica el sexo del niño, puede ser masculino o femenino.
Fec_Nacimiento	<b>Fecha de nacimiento:</b> Corresponde a la fecha del nacimiento del niño en dd/mm/aaaa.
Edad	Corresponde a la edad del niño en meses y fracción de mes.
Peso	Corresponde al peso del niño en kilogramos.
Talla	Corresponde a la talla del niño en centímetros.
Dx_P/T	<b>Peso para la talla (P/T):</b> Este índice compara el peso del niño con el peso esperado para su talla y permite establecer si ha ocurrido una pérdida reciente de peso corporal. Identifica desnutrición aguda y en el patrón OMS son los niños menores de cinco años que se encuentran entre los z-Score P/T entre menor a -2 hasta -5. En caso de sobrepeso los z-Score P/T son mayor a +2 hasta +3 y en caso de obesidad son z-Score mayor a +3 (hasta +5).
Dx_T/E	<b>Talla para la edad (T/E):</b> Este índice compara la longitud o talla del niño con la longitud o talla esperada para su edad y permite

---





---

establecer si está ocurriendo un retraso en el crecimiento. Identifica el retardo de crecimiento (desnutrición crónica) y en el patrón OMS son los niños menores de cinco años que se encuentran entre los z-Score T/E menor a -2 a -3 y es severo si el z-Score es menor de -3.

---

**Dx\_P/E** **Peso para la edad (P/E):** Este índice compara el peso del niño con el peso esperado para su edad y permite establecer si está ocurriendo desnutrición. Identifica la desnutrición global y en el patrón OMS son los niños menores de cinco años que se encuentran entre los z-Score P/E menor a -2. En caso de sobrepeso los z-Score son mayores a +2.

---

**Dx\_Anemia** **Diagnóstico de Anemia:** Es la clasificación en función de la cifra de hemoglobina, de acuerdo a lo establecido por la OMS, que considera anemia si la hemoglobina es menor de 11.0 g/dl, para los niños a partir de los 6 meses de edad; se clasifica en Normal ( $Hb \geq 11,0$  g/dl), anemia leve ( $Hb$  entre 10,0 – 10,9 g/dl), anemia moderada ( $Hb$  entre 7,0 – 9,9 g/dl) y anemia severa ( $Hb < 7,0$  g/dl).

---

**N\_CCred** **Número de Controles CRED:** Es el número de controles de Crecimiento y Desarrollo (CRED) realizados al niño menor de cinco años, que tiene a la fecha de la atención, incluyendo el control que se realiza en ese momento.

---

**N\_Sachets** **Número de Sachets consumidos:** Es el número de sachets de multimicronutrientes consumidos por el niño menor de cinco

---



---

	años desde la consulta anterior a la fecha de la atención de ese momento.
N_Consejerías	<b>Número de consejerías:</b> Es el número de consejerías nutricionales recibidas por las madres de los niños menores de cinco años desde la consulta anterior a la fecha de atención de ese momento.
N_Sesiones	<b>Número de Sesiones:</b> Es el número de sesiones demostrativas recibidas por las madres de los niños menores de cinco años desde la consulta anterior a la fecha de atención de ese momento.
N_Visitas	<b>Número de visitas domiciliarias:</b> Es el número de visitas domiciliarias recibidas por las familias con niños menores de cinco años desde la consulta anterior a la fecha de atención de ese momento.
JUNTOS	Se identifica si la familia del niño, está afiliado al programa JUNTOS.
SIS	Se identifica si el niño, está afiliado al Seguro Integral de Salud - SIS.
PIN	Se identifica si el niño, está afiliado al Subprograma Infantil del Programa Integral de Nutrición - PIN.

---

Fuente: (SIEN, 2019)