

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
ESCUELA DE FORMACIÓN PROFESIONAL DE INGENIERÍA DE SISTEMAS
Y COMPUTACIÓN



T E S I S

**Modelo predictivo asociado al riesgo de anemia en niños menores
de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023**

**Para optar el título profesional de:
Ingeniero de Sistemas y Computación**

Autor:

Bach. Alexander Jesus CHACON ARTICA

Asesor:

Mg. Pit Frank ALANIA RICALDI

Cerro de Pasco – Perú – 2024

UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
ESCUELA DE FORMACIÓN PROFESIONAL DE INGENIERÍA DE SISTEMAS
Y COMPUTACIÓN



T E S I S

**Modelo predictivo asociado al riesgo de anemia en niños menores
de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023**

Sustentada y aprobada ante los miembros del jurado:

Mg. Oscar Cleverio CAMPOS SALVATIERRA
PRESIDENTE

Mg. Lisbeth Gisela NEGRETE CARHUARICRA
MIEMBRO

Mg. Jose Luis SOSA SANCHEZ
MIEMBRO



Universidad Nacional Daniel Alcides
Carrión Facultad de Ingeniería
Unidad de Investigación

INFORME DE ORIGINALIDAD N° 155-2024-UNDAC/UIFI

La Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión en mérito al artículo 23° del Reglamento General de Grados Académicos y Títulos Profesionales aprobado en Consejo Universitario del 21 de abril del 2022, La Tesis ha sido evaluado por el software antiplagio Turnitin Similarity, que a continuación se detalla:

Tesis:

**MODELO PREDICTIVO ASOCIADO AL RIESGO DE ANEMIA EN NIÑOS
MENORES DE 5 AÑOS EN LA PROVINCIA DANIEL ALCIDES CARRIÓN –
PASCO, 2023**

Apellidos y nombres de los tesistas:

Bach. CHACON ARTICA, Alexander Jesus

Apellidos y nombres del Asesor:

Mg. ALANIA RICALDI, Pit Frank

Escuela de Formación Profesional

Ingeniería de Sistemas y Computación

Índice de Similitud

26%

APROBADO

Se informa el Reporte de evaluación del software similitud para los fines pertinentes:

Cerro de Pasco, 15 de julio del 2024



Firmado digitalmente por MEJIA
CACERES Reynaldo FAU
20154020046.pdf
Motivo: Soy el autor del documento
Fecha: 06.11.2024 10:00:49 -05:00

DEDICATORIA

Dedico este artículo a Dios y a mis padres, a Dios por estar siempre conmigo en cada paso que doy, cuidándome y dándome fuerzas para continuar y a mis padres que siempre se han encargado de que mi felicidad y educación sean apoyadas en todo momento. confiar completamente en ellos con cada desafío que me lancen y nunca dudar de mis habilidades porque soy quien soy hoy gracias a ellos.

AGRADECIMIENTO

Primero que nada, doy gracias a Dios por ser mi mejor amigo, mi fortaleza, darme todo y no dejarme caer, gracias a mis padres por ser los mejores y estar conmigo incondicionalmente, gracias a las personas que amo y respeto, estudian y estudian.

RESUMEN

El trabajo de investigación que realice se titula: "MODELO PREDICTIVO ASOCIADO AL RIESGO DE ANEMIA EN NIÑOS MENORES DE 5 AÑOS EN LA PROVINCIA DANIEL ALCIDES CARRIÓN - PASCO, 2023" El objetivo principal fue Determinar el Modelo Predictivo que asocia al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023. El estudio utilizado fue con un diseño Correlacional-transversal Debido al uso de instrumentos y evidencia teórica relacionada con el tema de factores de riesgo relacionados con la anemia en niños menores de 5 años. La población en 2023 de la DIRESA SIEN, Departamento Daniel Alcides Carrión, Pasco, era de 1,028 niños menores de cinco años. "Con base en los resultados del ajuste del modelo, se puede estimar la contribución de los factores de riesgo de anemia. Utilizando el modelo de clasificación de muestra representativa, la prueba de aptitud física de Hosmer y Lemeshow, el nivel de significancia es 0,05, el valor $p = 0,088$ y la determinación del coeficiente de Nagelkerke es 70,7%. La tasa de clasificación de verdaderos negativos (predecir anemia versus ninguna anemia) fue del 92,8 %, mientras que la mejor tasa para clasificar correctamente a los niños con anemia (verdaderos positivos) fue del 91,6 %, lo que demuestra el fuerte poder predictivo del modelo".

Palabras Clave. Modelo Predictivo, Riesgo de anemia.

ABSTRACT

The research work I carried out is titled: "PREDICTIVE MODEL ASSOCIATED WITH THE RISK OF ANEMIA IN CHILDREN UNDER 5 YEARS OF AGE IN THE PROVINCE DANIEL ALCIDES CARRIÓN - PASCO, 2023" The main objective was to Determine the Predictive Model associated with the risk of anemia in children children under 5 years of age in the Daniel Alcides Carrión Province -Pasco, 2023. The study used was with a correlational-cross-sectional design due to the use of instruments and theoretical evidence related to the topic of risk factors related to anemia in children under 5 years of age. years. The population in 2023 of the DIRESA SIEN, Daniel Alcides Carrión Department, Pasco, was 1,028 children under five years of age. "Based on the model fitting results, the contribution of anemia risk factors can be estimated. Using the representative sample classification model, Hosmer and Lemeshow physical fitness test, the significance level is 0, 05, the p value = 0.088 and the determination of the Nagelkerke coefficient is 70.7% The classification rate of true negatives (predicting anemia versus no anemia) was 92.8%, while the best rate to correctly classify a. children with anemia (true positives) was 91.6%, which demonstrates the strong predictive power of the model."

Keywords. Predictive Model, Risk of anemia.

INTRODUCCIÓN

La anemia es un problema de salud global. En nuestro país, la anemia infantil es uno de los mayores problemas de salud pública que afecta a nuestra población, especialmente a los grupos más vulnerables como las mujeres embarazadas, los niños menores de dos años y las mujeres en edad reproductiva. Actualmente, el 43,5% de los niños peruanos entre 6 y 35 meses de edad padecen anemia, de los cuales el 51,1% se encuentran en el área rural y el 40% en el área urbana. "Eso es 620.000 niños anémicos menores de tres años de una población nacional de 1,6 millones". (Arroyo-Laguna, 2017)

"Vale la pena mencionar que la anemia crea una carga significativa para los individuos desde la primera infancia, afectando no sólo la vida de todos los pacientes con anemia, sino también las consecuencias sociales y económicas de toda la sociedad" (P. Dutta y Sengupta, 2017).

"La ingesta inadecuada de hierro y otros nutrientes puede provocar una deficiencia de hierro, que en las etapas más graves y persistentes puede provocar anemia. Se necesita un enfoque integrado de la anemia dentro del marco causal para no sólo aumentar la ingesta de hierro y micronutrientes de los niños, sino también reducir la carga de las enfermedades parasitarias, la diarrea y la malaria. (Zavalletta, 2017):

Capítulo 1: "Incluye los siguientes apartados: identificación y definición del problema, definición del estudio, planteamiento del problema, desarrollo de objetivos, justificación del estudio y limitaciones del estudio".

Capítulo 2: "Base teórica y científica, definición de conceptos, identificación de hipótesis y variables y finalmente definiciones operativas.

Capítulo 3: "Tipo de investigación, métodos de investigación, diseño de la investigación, principales conjuntos de datos y muestras, métodos y herramientas de recolección de datos, procesamiento de datos." y métodos de análisis, procesamiento

de datos estadísticos, selección de herramientas de investigación, validación y confiabilidad, ética y pautas de investigación.

El Capítulo 4: “Resultados y discusión” incluye las siguientes partes: descripción del trabajo, análisis e interpretación de resultados, prueba de hipótesis y discusión de resultados.

En conclusión, proporcionamos conclusiones, recomendaciones, referencias y apéndices.”.

El autor

ÍNDICE

DEDICATORIAI

AGRADECIMIENTO

RESUMEN

ABSTRACT

INTRODUCCIÓN

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1.	Identificación y determinación del problema	1
1.2.	Delimitación de la investigación.....	3
1.2.1.	Delimitación espacial	3
1.2.2.	Delimitación temporal	3
1.2.3.	Delimitación conceptual	3
1.3.	Formulación del problema	3
1.3.1.	Problema general	3
1.3.2.	Problemas específicos.....	3
1.4.	Formulación de objetivos.....	3
1.4.1.	Objetivo General.....	3
1.4.2.	Objetivos específicos	3
1.5.	Justificación de la investigación.....	4
1.6.	Limitaciones de la investigación	4

CAPITULO II
MARCO TEÓRICO

2.1.	Antecedentes de estudio	5
2.1.1.	A nivel Internacional	5
2.1.2.	A nivel Nacional	7
2.1.3.	A nivel Local	8
2.2.	Bases teóricas – científicas	9
2.2.1.	Modelo Predictivo	9
2.2.2.	Anemia	20
2.2.3.	Factores distales.....	24
2.3.	Definición de términos básicos	25
2.4.	Formulación de Hipótesis	27
2.4.1.	Hipótesis General	27
2.4.2.	Hipótesis Específicas.....	27
2.5.	Identificación de Variables.....	27
2.5.1.	Variables independientes.....	27
2.5.2.	Variables dependientes	27
2.6.	Definición Operacional de variables e indicadores	27

CAPITULO III

METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN

3.1.	Tipo de investigación.....	28
3.2.	Nivel de investigación.....	28
3.3.	Métodos de investigación	28
3.4.	Diseño de investigación.....	28
3.5.	Población y muestra	29
3.5.1.	Población.....	29
3.5.2.	Muestra	29
3.6.	Técnicas e instrumentos de recolección de datos	29

3.7.	Selección, validación y confiabilidad de los instrumentos de investigación.....	29
3.8.	Técnicas de procesamiento y análisis de datos.....	30
3.9.	Tratamiento Estadístico.....	31
3.10.	Orientación ética filosófica y epistémica	31

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1.	Descripción del trabajo de campo	32
4.2.	Presentación, análisis e interpretación de resultados	33
4.2.1.	Prevalencia de anemia	33
4.2.2.	Selección de variables por factor	34
4.2.3.	Estimación de modelo.....	37
4.2.4.	Verosimilitud del modelo.....	41
4.2.5.	Prueba de bondad de ajuste	42
4.3.	Prueba de Hipótesis	42
4.3.1.	Hipótesis específicas 1:	42
4.3.2.	Hipótesis específicas 2:	43
4.4.	Discusión de resultados	45

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

ÍNDICE DE TABLAS

Tabla 1. Definición Operacional de Variables	27
Tabla 2. Prevalencia de anemia según sexo	33
Tabla 3. Variables que ingresan al modelo	37
Tabla 4. Variables que ingresan al nuevo modelo	38
Tabla 5. Prueba del coeficiente de determinación	38
Tabla 6. Tabla de clasificación.....	39
Tabla 7. Tabla del área bajo la curva COR.....	41
Tabla 8. Prueba del coeficiente de intercepción.	41
Tabla 9. Prueba de bondad de ajuste.....	42
Tabla 10. Prueba de Prueba de Omnibus.....	43

ÍNDICE DE FIGURAS

Figura 1. Rangos normales y criterios para definir anemia en niños menores de seis años, propuestos por la OMS.....	21
Figura 2. Modelo causal de la Anemia en el Perú	23
Figura 3. Diagrama del árbol para el modelo de anemia según factores demográficos	35
Figura 4. Diagrama del árbol para el modelo de anemia según factores biológico	36

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Identificación y determinación del problema

Para la identificación y determinación de problemas relacionados con el "Modelo predictivo asociado al riesgo de anemia en niños menores de 5 años en la provincia Daniel Alcides Carrión - Pasco" en el contexto de Perú, se debe considerar una serie de factores a nivel nacional, ya que la anemia infantil es un problema prioritario en el país.

La anemia en niños menores de 5 años en Perú también está estrechamente vinculada a la desnutrición crónica, que afecta a un porcentaje significativo de la población infantil. Los niños con desnutrición crónica tienen mayor riesgo de desarrollar anemia y otros problemas de salud

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 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.

“La anemia es un importante problema de salud pública que afecta a la población de los países desarrollados y en desarrollo, no sólo porque es la discapacidad más común y común, sino también porque es más común entre grupos vulnerables como los niños y las mujeres. Afecta al 43% de los niños menores de 5 años, al 38% de las mujeres embarazadas y al 29% de las no embarazadas” (Zavaleta & Astete, 2017)

Según la OMS (Organización Mundial de la Salud), “En el mundo, 1.620 millones de personas padecen anemia, lo que supone el 24,8% de la población total, siendo la mayor prevalencia entre los niños en edad preescolar alcanzando el 47,4%” (OMS, 2020).

La prevalencia de la anemia es un mal que tiene consecuencias y consecuencias económicas y sociales, por lo tanto se deben tomar medidas para abordar el problema ya que afectará negativamente la vida futura de los niños.

La coordinación, conducción y seguimiento de la implementación del PMLCA es responsabilidad del MIDIS que hace las veces de Secretaría Técnica de la CIAS. La prevalencia de anemia en menores de 5 años en el Perú alcanza el 32%.

En Perú, la anemia afecta a una gran proporción de niños menores de 5 años, con tasas que superan el 40% en muchas regiones, y las zonas andinas como Pasco presentan incluso mayores porcentajes debido a la altitud y condiciones de vida.

En la región de Pasco, ENDES (2) registro que el 50,2% de niños entre 6 y 35 meses de edad presentaban anemia, valor por encima del promedio nacional 40,1%.

1.2. Delimitación de la investigación

1.2.1. Delimitación espacial

Con la presente investigación se diagnosticó, analizó, selecciono y utilizo el Modelo Predictivo asociado al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

1.2.2. Delimitación temporal

La investigación se llevó a cabo en el periodo 2023.

1.2.3. Delimitación conceptual

Concepto de Modelo Predictivo y riesgo de anemia.

1.3. Formulación del problema

1.3.1. Problema general

¿Cuál es el Modelo Predictivo que asocia al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023?

1.3.2. Problemas específicos

¿Cómo identificamos un conjunto de factores distales que influyan en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023?

¿Cómo estimamos un modelo predictivo para la probabilidad en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023?

1.4. Formulación de objetivos

1.4.1. Objetivo General

Determinar el Modelo Predictivo que asocia al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

1.4.2. Objetivos específicos

Identificar un conjunto de factores distales que influyan en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

Estimar un modelo predictivo para la probabilidad en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

1.5. Justificación de la investigación

En la región de Pasco y en todo el Perú los problemas de salud son uno de los factores que afectan el bienestar familiar. Una de las principales áreas de enfoque del Ministerio de Salud (MINSA) son los niños menores de 5 años, por lo que el papel continuo del MINSA en la mejora de la calidad de vida de los niños debe fortalecer las estrategias de prevención y control de la anemia.

De 1 a 5 años, el objetivo es proporcionar un modelo de regresión logística cuyas predicciones puedan conducir a decisiones coherentes. En este contexto, los modelos de regresión logística se aplican en las ciencias de la salud y permiten analizar los resultados de forma explicativa y predictiva, pudiendo comprender la fuerza de la asociación y comprender la previsibilidad utilizando factores de riesgo y consecuencias OR estudiados de forma independiente. el valor de cada uno de ellos o del modelo completo.

Antes de que se utilicen cada vez más, se debe considerar críticamente la selección de variables en los modelos y el análisis de sus resultados. Este trabajo de investigación utilizó modelos de regresión logística para identificar factores de riesgo asociados a la anemia en niños menores de 5 años de la población de estudio.

1.6. Limitaciones de la investigación

Un modelo predictivo basado en datos de 2023 puede volverse obsoleto rápidamente si no se actualiza periódicamente con nuevos datos o si las condiciones sociales y de salud cambian en la provincia.

CAPITULO II

MARCO TEÓRICO

2.1. Antecedentes de estudio

2.1.1. A nivel Internacional

Según Loaiza, Moreno y Ríos (2023) en su trabajo de investigación titulado “Proceso de analítica de datos aplicado a la desnutrición infantil en niños de 0 a 5 años en la Ciudad de Medellín” Su principal objetivo es implementar un proceso de análisis de datos para comprender la desnutrición infantil en niños de 0 a 5 años de la ciudad de Medellín. Está diseñado para abordar la siguiente hipótesis: “En el proceso de análisis de datos se puede obtener conocimientos sobre la desnutrición en niños de 0 a 5 años de la ciudad de Medellín”. Por lo tanto, para responder a esta pregunta, primero buscamos datos que cubran esta pregunta para obtener un conjunto de datos. Luego se definen los requisitos funcionales y arquitectónicos, especificando el proceso de limpieza y las herramientas de aplicación de análisis, al mismo tiempo que se definen las tecnologías a utilizar. Luego, los datos se limpian y transforman gradualmente, se codifican utilizando el lenguaje de programación de la tecnología elegida y se evalúa la idoneidad del modelo utilizado para derivar recomendaciones sobre la desnutrición infantil. Todo lo anterior hace referencia al planteamiento de la tesis, que es una adaptación híbrida entre procesos de ingeniería de software y

análisis de datos CRISP_DM y CATALYST. Finalmente, los resultados muestran una mejora en el proceso analítico, aunque son significativos, y aportan algunas recomendaciones para el informe sobre desnutrición en niños pequeños elaborado por el director de salud municipal.

Según Leal, Batista, Lira, Figueiroa, Osório (2020) en su trabajo de investigación titulado “Prevalencia de la anemia y factores asociados en niños de seis a 59 meses de Pernambuco, Noreste de Brasil” Objetivo: Evaluar la prevalencia de anemia e identificar factores asociados en niños de 6 a 59 meses. Métodos: Estudio transversal con datos del Estudio de Salud y Nutrición del Tercer País/Estado de Pernambuco (Noreste de Brasil), en una muestra representativa de 1403 niños urbanos y rurales. La anemia se diagnostica por la dosis de hemo. Se realizaron análisis multivariados basados en modelos jerárquicos utilizando regresión de Poisson con varianza robusta para estimar la prevalencia con base en las siguientes variables: biología infantil, morbilidad y estado nutricional, factores socioeconómicos, de vivienda, sanitarios y maternos. Resultados: La prevalencia ponderada de anemia fue de 32,8%, 31,5% en zona urbana y 36,6% en zona rural. En las zonas urbanas, las siguientes variables se asociaron significativamente con la anemia: educación materna, bienes de consumo, número de niños menores de cinco años en el hogar, tratamiento del agua potable, edad materna y anemia, y edad del niño. En las zonas rurales, sólo la edad materna y la edad infantil se asociaron significativamente con la anemia. Conclusión: la incidencia de anemia en niños urbanos y rurales es similar. Se deben tener en cuenta los factores asociados con la anemia al planificar medidas eficaces para controlar la anemia.

2.1.2. A nivel Nacional

Según Canaza (2021) en su trabajo de investigación titulado “Modelo Predictivo de Riesgo Asociado a la Anemia en niños menores de 5 años en la Microred Yauri Provincia de Espinar-Cusco,2019” El objetivo de este estudio fue determinar los patrones predictivos de riesgo asociados a la anemia en niños menores de 5 años en el año 2019 en MICRORED Yauri Cuzco, provincia de Espinal. El enfoque incluyó el uso de métodos científicos, investigación básica, diseños correlacionales, no experimentales y una muestra probabilística de 322 niños. Para lograr estos objetivos se analizó el sistema de información de la Autoridad Provincial de Salud del Cusco Espinal de Reza (SIEN) sobre el estado nutricional de niños menores de cinco años y mujeres embarazadas en el año 2019. El modelo de clasificación es el siguiente: , el coeficiente de determinación de Nagelkerke es del 85,8%. La tasa de clasificación negativa verdadera (predicción de anemia no anémica) fue del 93,8 %, mientras que la tasa de clasificación correcta más relevante (positivo verdadero) para niños anémicos fue del 96,2 %, lo que demuestra el fuerte poder predictivo del modelo.

Según Abimegireysch (2022) en su trabajo de investigación titulado “Factores de riesgo asociados a la anemia en niños menores de 5 años de edad, subanálisis ENDES 2020-2021” como objetivo determino los Los factores de riesgo asociados a la anemia en niños menores de 5 años se determinaron con base en los datos recopilados durante el período 2020-2021. de la Encuesta Anual de Demografía y Salud Familiar (ENDES). Utilizando la base de datos ENDES 2020-2021 se analizaron 32.285 niñas y niños de hasta 5 años. El procesamiento de datos y el análisis estadístico se realizaron mediante los programas de investigación SPSS v23 y R. Participaron del estudio un total de 32.285 niños, de los cuales el 68,84% (n=22.225) vivían en zonas urbanas y el 41,40% (n=13.367) en la región central del Perú. Así, el 66,18% (n=21.368) de los niños nacieron de madres con educación primaria. En el análisis multivariado

quedó claro que vivir en zonas rurales (OR: 1,55; IC 95%: 1,48-1,63), vivir en la zona sur del país (OR: 1,26; IC 95%: 1,19-1,33) y tener un índice de riqueza "muy pobre" (OR: 1,76; IC 95%: 1,68-1,85), edad materna ≤ 17 años (OR: 3, 18; IC 95%: 2, 34-4,32), sin edad ni educación de la madre (OR: 11,78; IC 95%: 9,65-14,39), usar queroseno para cocinar (OR: 8,64; IC 95%: 0,97 -7,28), edad ≤ 11 meses (OR: 4,85; IC 95%: 4,53) número de controles prenatales visitas ≤ 5 (OR: 1,58; IC 95%: 1,47-1,70) son factor de riesgo de anemia en niños menores de 5 años. Alfabetización o menos de 5 visitas prenatales.

2.1.3. A nivel Local

Según Huanca (2024) en su trabajo de investigación titulado "Implementación de un modelo de análisis predictivo en el control de nutrición de los niños menores de 5 años del Centro de Salud de Uliachin, Chaupimarca - Pasco, 2023" El objetivo principal fue aplicar un modelo de análisis predictivo al seguimiento nutricional de los niños menores de 5 años del Centro de Salud de Uliachin, Chaupimarca - Pasco, 2023. Por lo tanto, incluyendo la muestra a todos los niños hasta 5 años. se registró en el último control nutricional del año 2023, el cual fue a 87 niños del Centro de Salud Uliachin de Chaupimarca. Para ello se utilizaron datos del instrumento Encuesta y del mismo centro de salud. "El resultado fueron modelos predictivos y sistemas expertos que obtuvieron resultados positivos de acuerdo a los modelos predictivos aplicados en su investigación, por lo que el proyecto desarrollado mostró confiabilidad en los resultados y logró superar la proporción de expertos que es del 80%". Se logró medir el estado actual del control nutricional en el control nutricional de los niños del Sana Centro Uliachín Pasco, lo que permitió conocer los indicadores del organismo para realizar el control nutricional, y que además demostró que el índice más alto corresponde a la etapa crónica, también informaron que el uso

de modelos predictivos ayudaría enormemente a informar la implementación del control nutricional.

2.2. Bases teóricas – científicas

2.2.1. Modelo Predictivo

El proceso de creación de un modelo predictivo comienza con la recopilación de datos. Esto puede incluir una gran variedad de información tanto interna, externa y alternativa.

Una vez recopilados los datos, se formula un modelo estadístico. Este modelo se basa en el análisis de los datos recogidos y busca patrones y relaciones que puedan indicar comportamientos y tendencias hacia el futuro.

Después de formular el modelo es necesario realizar las predicciones correspondientes. Dichas predicciones se basan en los patrones y relaciones identificados por el modelo. Por ejemplo, un modelo predictivo puede predecir el riesgo de fraude basándose en patrones de datos.

Finalmente, el modelo se valida o se revisa. A medida que se recogen más datos, el modelo puede ajustarse para mejorar su precisión.

Estos modelos funcionan mediante la categorización de la información basada en datos históricos. Los modelos de clasificación se utilizan en diferentes industrias porque pueden ser fácilmente reentrenados con nuevos datos y pueden proporcionar un amplio análisis para responder a las preguntas de negocio:

Modelos de previsión

Modelos de clasificación

Modelos de clustering

Modelos de regresión

Modelos de datos atípicos (U outliers)

Modelos de series de tiempo

Árboles de decisión

Redes neuronales

Modelos de previsión

Este tipo de modelo se encarga de la predicción de métricas mediante la estimación de valores de nuevos datos, basándose en los aprendizajes de los datos históricos.

Suelen utilizarse para generar valores numéricos en los datos históricos cuando no existe alguno. Los modelos de previsión son populares porque son increíblemente versátiles.

Modelos de clasificación

Estos modelos funcionan mediante la categorización de la información basada en datos históricos. Los modelos de clasificación se utilizan en diferentes industrias porque pueden ser fácilmente reentrenados con nuevos datos y pueden proporcionar un amplio análisis para responder a las preguntas de negocio

Modelos de clustering

Los modelos de clustering, también conocidos como Modelos de Agrupamiento, utilizan minería de datos para agrupar objetos similares en función de sus características. Una empresa puede utilizarlos para segmentar a sus clientes en función de hábitos de compra.

Por ejemplo, retailers suelen utilizar clusterings para segmentar grupos de clientes con intereses similares y así mejorar la personalización de sus campañas de marketing.

Modelo de Regresión Logística

En los métodos de clasificación, una opción es intentar crear un modelo que explique los valores de las variables categóricas. Si por ej. Si quieres distinguir entre préstamos que se devuelven o préstamos que tienen problemas de recuperación, puedes agregar una nueva variable y a la base de datos que toma el valor 1 si el préstamo se devuelve sin problemas y 0 si el préstamo se

devuelve sin preguntas. Valor 1. El problema de discriminación pasa a ser predecir el valor de una variable ficticia y en el elemento de un nuevo vector variable x del que lo conocemos. Si el valor predicho está más cerca de cero que de uno, clasificaremos el elemento en la primera población. En caso contrario, lo hacemos en el segundo. Se utilizó un modelo de respuesta cualitativa para modelar este tipo de relación. La regresión logística es un tipo de análisis de regresión que se utiliza para predecir el resultado de una variable categórica o una variable de respuesta binaria, donde el resultado de la respuesta de cada ítem es "sí" o "no" según la variable. Esto es útil para modelar la probabilidad de un evento en función de otros factores. El análisis de regresión logística pertenece a la colección de modelos lineales generalizados (GLM), que utilizan la función logit como función de enlace. Modelar la probabilidad utilizando funciones logísticas que describen los posibles resultados de un solo experimento en función de variables explicativas.

La regresión logística es útil cuando se desea predecir si existe una función o un resultado en función de los valores de un conjunto de variables predictivas. Es similar al modelo de regresión lineal, pero es adecuado para modelos donde la variable dependiente es dicotómica. Los coeficientes de regresión logística se pueden utilizar para estimar el odds ratio para cada variable independiente del modelo. La regresión logística se puede aplicar a una gama más amplia de situaciones de investigación que el análisis discriminante. (Pérez, 2004)

La transformación logit

Sánchez-Cantalejo (2000), "La descripción es la siguiente: Considere el caso donde la variable Y es dicotómica, si el individuo tiene la característica de interés, entonces $Y = 1$, en caso contrario $Y = 0$. Por lo tanto, se supone que la variable respuesta sigue una distribución binomial con parámetros. 1 y p , donde p representa la probabilidad de que un individuo tenga la característica de

interés ($Y = 1$). Dado que la media de la variable binomial con parámetros n y p es $n \cdot p$, entonces la media en este caso será $1 \cdot p = p$. De esta forma, el modelo de regresión lineal se puede escribir de la siguiente forma”:

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

Los valores de los parámetros del modelo ya no están acotados y podemos estimar la probabilidad de que la función sea mayor que uno o negativa, lo cual no tiene ningún sentido, situación que hace que los intentos de utilizar métodos de regresión lineal sean ineficaces y binomiales

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.

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

Su valor posible puede ser cualquier número real, positivo o negativo, por lo que el problema anterior desaparece. Esta transformación de p se llama transformación logística de la probabilidad de p o transformación logit.

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

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.

Modelado de regresión logística binaria

Según Sánchez-Cantalejo (2000), “Actualmente, desde el punto de vista epidemiológico, sólo existe una variable predictiva, el factor de riesgo putativo. Según el modelo logístico propuesto anteriormente, esto se manifiesta en el caso de una única variable predictiva”:

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

Donde p Representa la probabilidad de que un individuo tenga el rasgo de interés, donde X es el único predictor. El primer término es equivalente al segundo”.

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

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 está 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 negativo, ya que el numerador siempre es menor que el denominador, el coeficiente variará de 0 a 1, por lo que el rango de valores está permitido por la probabilidad”.

Modelo de regresión logística binaria múltiple

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

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

Esto se llama función lógica. Los modelos de regresión logística modelan las probabilidades de un proceso binomial como una función logística de una combinación lineal de las variables dependientes. Un modelo de regresión logística binaria multivariante 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:

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

“El problema que surge en el contexto multivariado es determinar el mejor modelo explicativo, lo que significa evaluar las correlaciones entre las diversas variables independientes propuestas originalmente en el modelo de regresión logística y seleccionar la variable que mejor explica la variable dependiente. Para este fin se pueden utilizar varios métodos de selección de variables” (López-Roldán & Fachelli, 2015)

a. Proceso de análisis

En este proceso de análisis se consideran los siguientes aspectos:

1. Selección de variables del modelo.
 - Se debe definir un modelo analítico para justificar la dependencia del modelo. - Analizar las condiciones de uso.
 - Confirmar relaciones claras entre cada variable independiente y dependiente. - Seguir el proceso de selección de variables óptimas para ajustarse al mejor y más parsimonioso modelo.
2. Estimación del coeficiente de variable independiente
 - Utilizar estimación de máxima verosimilitud, un caso especial de mínimos cuadrados ordinarios (MCO).
3. Evaluación del modelo
 - Se trata del ajuste del modelo o de su poder explicativo: estadísticos de Nagelkerke R² o Cox y Snell R² y tests de Hosmer y Lemeshow.
 - Validez predictiva o poder discriminatorio del modelo: resulta de la intersección de las variables dependientes observadas y las variables predichas, la llamada tabla de clasificación o matriz de confusión. Este es el porcentaje de casos bien clasificados. Para ello se define un criterio de corte: la probabilidad de que la variable dependiente sea 1. Por defecto se considera un valor de 0,5, pero se puede ajustar en función del estudio específico. En particular, para este fin se puede utilizar el análisis de la curva denominada COR (Receiver Operating Characteristic o ROC en inglés).
 - Utiliza toda la información para interpretar los resultados, pero básicamente considera la tabla de probabilidades.

b. Condiciones de aplicación

La regresión logística no impone restricciones a su aplicación. En particular, no hay suposiciones sobre la distribución de las variables. Se aplican las siguientes condiciones:

1. El modelo deberá estar correctamente especificado y ser relevante en cuanto a su contenido.
2. No omitir variables independientes relevantes
3. No hay error en la medición de una variable independiente.
4. Las observaciones son independientes entre sí.
5. No hay colinealidad entre las variables independientes.
6. linealidad de variables cuantitativas
7. Relacionado con el número de muestras. Hosmer y Lemeshow recomiendan muestras de más de 400 cajas
8. El porcentaje de ejemplos donde la variable dependiente es 0 o 1 debe ser al menos del 10%.

c. Pruebas de significación

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

- Los coeficientes de regresión logística (si son significativamente diferentes de 0) se evalúan estadísticamente utilizando la estadística de Wald: sigue una distribución normal estándar. Los coeficientes significativos son aquellos que tienen una probabilidad inferior a 0,05.

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

- El cálculo del Pseudo R2 que determina la bondad de ajuste, cuánto mejora un modelo en relación al modelo, expresado en porcentaje:

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

Pseudo R2 de Cox y Snell:

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

Pseudo R2 de Nagelkerke:

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

- Es una estadística que además va de 0 a 1 con un valor ligeramente superior a la correlación anterior. Este y el anterior son indicadores de variabilidad explicada, que siempre dan valores muy bajos en comparación con indicadores homólogos en la regresión lineal clásica. Los resultados de 0,2 y 0,3 son comunes, pero un pseudo-R2 de 0,6 no lo es.
- Una estimación del estadístico de ajuste z2, donde Ri es el residual entre las probabilidades observadas y estimadas en el iésimo caso. Sigue una distribución chi-cuadrado. Si el modelo es significativo, la probabilidad de correlación es menor o igual a 0,05.

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

- Prueba de Hosmer y Lemeshow El modelo encaja bien. Se utilizó la siguiente comparación de distribuciones para confirmar el ajuste del modelo. Para hacer esto, se debe calcular la probabilidad del evento predicho para todos los individuos de la muestra y calcular la diferencia con el valor observado. Esta prueba implica dividir las rutas de probabilidad en deciles y comparar la distribución de frecuencia predicha con la distribución de frecuencia observada utilizando una prueba 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 diferencia entre los valores observados y predichos: ajuste del modelo. Por tanto, si el ajuste es bueno, el valor de probabilidad predicho será mayor o igual a 0,05. Esta es una prueba adecuada para muestras pequeñas y covariables continuas.

Según: (Berlanga, Rubio, & Vilà, 2013), “Un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (eventos) que pueden resultar de una decisión tomada en un momento determinado. Ante múltiples decisiones posibles, desde una perspectiva probabilística, nos ayudan a tomar la decisión “correcta”. Estos árboles le permiten inspeccionar los resultados y determinar visualmente cómo fluye el modelo. Visualizar sus resultados puede ayudarlo a

encontrar subgrupos y relaciones específicos que quizás no encuentre utilizando estadísticas más tradicionales”.

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

Los árboles de decisión construyen modelos de clasificación basados en diagramas de flujo. Agrupan casos o predicen el valor de la variable dependiente (criterio) en función del valor de la variable independiente (predictora).

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.

Modelos de serie de tiempo

Los modelos de serie de tiempo analizan datos a lo largo del tiempo para identificar patrones y tendencias. Estos modelos son particularmente útiles en el pronóstico de ventas y la gestión de inventarios.

Por ejemplo, una empresa minorista puede utilizar modelos de series de tiempo para predecir la demanda futura de sus productos y ajustar sus niveles de inventario en consecuencia, evitando así el exceso de stock o la falta de productos.

Árboles de decisión

Los árboles de decisión utilizan una estructura de árbol para representar posibles resultados y decisiones. Estos modelos permiten predecir la viabilidad de nuevos proyectos.

Por ejemplo, una empresa de energía puede utilizar árboles de decisión para evaluar diferentes opciones de inversión en proyectos de energías renovables (Considerando factores como costos y riesgos).

Redes neuronales

Las redes neuronales se basan en inteligencia artificial, ya que imita el funcionamiento del cerebro humano. Estos modelos pueden ser utilizados en diversas aplicaciones como el reconocimiento de patrones en grandes conjuntos de datos.

Por ejemplo, algunas empresas logísticas emplean redes neuronales para analizar datos de rutas y tiempos de entrega, y así optimizar sus operaciones y reducir costos.

2.2.2. Anemia

La anemia es una afección en la que la cantidad de glóbulos rojos (y, por lo tanto, la capacidad de la sangre para transportar oxígeno) es insuficiente para satisfacer las necesidades del cuerpo. La deficiencia de hierro generalmente se considera la causa más común de anemia, pero otras deficiencias nutricionales también pueden causarla (incluidos folatos, vitamina B12 y vitamina A), inflamación aguda y crónica, infecciones parasitarias y trastornos congénitos o adquiridos. Afecta la síntesis del hemo y la formación o supervivencia de los glóbulos rojos. (OMS, 2011).

La Organización Mundial de la Salud Definir la anemia como dos desviaciones estándar del nivel de hemo por debajo de las normas de edad y sexo. La anemia es causada por la incapacidad del cuerpo para producir hemoglobina y glóbulos rojos (las partes del sistema sanguíneo responsables de transportar oxígeno de una célula a otra).

La prevalencia de anemia es alta en las consultas de atención primaria de salud, pediatría y seguimiento del embarazo. La deficiencia de hierro (ID) es diferente de la anemia por deficiencia de hierro (IDA). La HAD es la causa más

común de anemia en todo el mundo y un importante problema de salud, principalmente en los países subdesarrollados. La HAD es muy común y afecta hasta el 5% de los niños y adolescentes, el 10% de las mujeres premenopáusicas y el 1% de los hombres, y hasta el 40% de los adultos mayores que se cuidan a sí mismos en el hogar. (Guzmán Llanos, Guzmán Zamudio, & Llanos de los Reyes, 2016).

Figura 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

Sin duda, es controvertido si esto es comprensible por la etiología y los factores de riesgo. Independientemente de la definición de actores de riesgo, la definición funcional de la causa de un resultado determinado se puede expresar de la siguiente manera: Cualquier factor, condición o característica, cuya eliminación impediría la posibilidad de este resultado, es un factor, condición o característica. es la causa del resultado. Por otro lado, hay factores que, si bien no son necesarios, contribuyen al efecto. Estos son los llamados factores de riesgo (FR) de la enfermedad o afección. Estos factores son modificables y, aunque no causantes en sí mismos, pueden contribuir a la acción del factor causante.

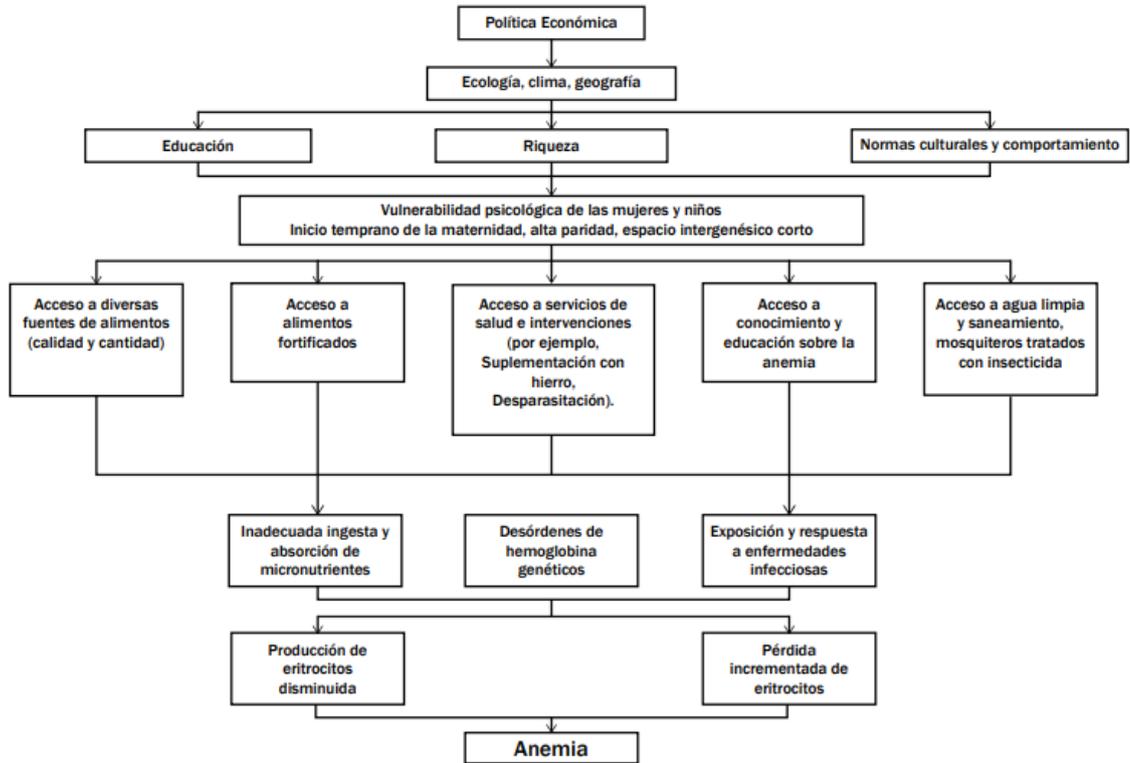
Según la OMS (2016), “Un factor de riesgo es cualquier rasgo, característica o exposición de un individuo que aumenta la probabilidad de sufrir una enfermedad o lesión. Los principales factores de riesgo son el peso insuficiente, las relaciones sexuales sin protección, la hipertensión arterial, el tabaquismo y el consumo de alcohol, el agua contaminada y las condiciones sanitarias e higiénicas deficientes”.

Factores determinantes de la anemia

Según MINSA (2017), La causa directa es la ingesta insuficiente de hierro y otros oligoelementos en los alimentos. La falta de hierro y vitaminas impedirá la formación normal de glóbulos rojos y hemoglobina. Otras causas directas de anemia son diarreas frecuentes, enfermedades parasitarias, infecciones como la malaria. Esta situación está asociada a prácticas limitadas de higiene, lavado de manos, acceso a agua potable e instalaciones sanitarias básicas.

En la literatura internacional se han propuesto modelos causales de anemia, como se muestra en la figura siguiente:

Figura 2. Modelo causal de la Anemia en el Perú



La anemia tiene varias causas (Figura 1), por lo que la deficiencia de hierro puede representar hasta el 60% de las causas de anemia en áreas sin malaria. La ingesta inadecuada de hierro y otros nutrientes es la principal causa de anemia. También se han identificado otras causas de anemia, por lo que un estudio realizado en nuestro país en 2015 encontró que las formas más comunes de anemia son la anemia parasitaria y la anemia causada por dos o más causas al mismo tiempo. (Zavaleta & Astete- Robilliard, 2017)

La anemia está asociada a diversos factores sociodemográficos y de salud infantil. Un análisis secundario en Perú (2016) encontró que en niños de 6 a 35 meses la anemia se asoció con dos factores sociodemográficos (residencia fuera de Lima, nivel socioeconómico más bajo, madres adolescentes con menor nivel de educación, menores de edad varones, edad menor a 24 meses). (Velásquez et al., 2016)

2.2.3. Factores distales

Una causa distal es la que da comienzo a una secuencia (cadena causal).

Los factores distales nos ayudaran a ver características que nos ayudaran en el modelo predictivo las causas varían depende del análisis.

- EESS

Establecimiento de salud

- Fecha

Se identifica la fecha en dd/mm/aaaa de la atención recibida por el niño menor de cinco años

- Hist_Clinica

Corresponde al número de historia clínica del niño que fue atendido en el establecimiento de salud

- 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

Hemoglobina corregida: 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.

- Peso

Corresponde al peso del niño en kilogramos

- Talla

Corresponde a la talla del niño en centímetros

- 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).

2.3. Definición de términos básicos

- **Brecha de pobreza:**

Esto es como el índice de la línea de pobreza, también conocido como déficit como porcentaje del ingreso promedio de los pobres, índice de brecha de ingresos o índice de brecha de pobreza. En términos econométricos, oscila entre 0 (nadie es pobre) y 1 (los pobres no tienen ingresos).

- **Desempleo:**

Esto se debe a la falta de trabajo remunerado. Los desempleados son personas que no tienen trabajo, es decir, personas que venden su mano de obra.

- **Factores distales:**

Los factores distales conducirían a los proximales a través de, al menos, tres procesos: configurando respuestas al entorno, formando creencias, esquemas y autoconceptos y posibilitando aprendizajes asociativos, operantes y observacionales.

- **Modelo Predictivo**

Los modelos predictivos, también conocidos como modelos de predicción, son un conjunto de herramientas y técnicas estadísticas que sirven para

pronosticar y predecir el comportamiento ante un evento. Y tienen el objetivo de predecir y pronosticar resultados probables a futuro.

- **Pobreza monetaria:**

Estas personas viven en hogares con un gasto per cápita insuficiente para comprar una canasta de alimentos y artículos básicos no alimentarios (ropa, educación, vivienda, transporte, salud, etc.)

- **Pobreza crónica:**

Es comprensible que la falta de seguridad básica afecte simultáneamente a varios aspectos de la vida de las personas, que persista durante mucho tiempo y reduzca gravemente las posibilidades de que las personas recuperen sus derechos y responsabilidades en un futuro próximo.

- **Pobreza extrema:**

Se la asocia con insatisfacción de necesidades mínimas de subsistencia.

- **Riesgo de anemia:**

Los siguientes factores pueden aumentar el riesgo de anemia: Una dieta pobre en vitaminas y minerales. No consumir suficiente hierro, vitamina B 12 y folato aumenta el riesgo de anemia. Problemas con el intestino delgado.

- **Regresión logística:**

La regresión logística es una técnica de análisis de datos que utiliza las matemáticas para encontrar las relaciones entre dos factores de datos. Luego, utiliza esta relación para predecir el valor de uno de esos factores basándose en el otro.

- **Transformación logit:**

La transformación Logit se utiliza principalmente para transformar datos de respuesta binarios, como supervivencia/no supervivencia o presencia/ausencia, para proporcionar un valor continuo en el rango

$(-\infty, \infty)$, donde p es la proporción de cada muestra que es 1 (o 0). La transformación inversa se muestra como p en términos de z .

2.4. Formulación de Hipótesis

2.4.1. Hipótesis General

El modelo Predictivo permite realizar la predicción asociado al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023.

2.4.2. Hipótesis Específicas

El conjunto de factores distales influye en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

El modelo predictivo permite realizar la probabilidad del riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

2.5. Identificación de Variables

2.5.1. Variables independientes

Modelo Predictivo.

2.5.2. Variables dependientes

Riesgo de anemia.

2.6. Definición Operacional de variables e indicadores

Tabla 1 *Definición Operacional de Variables*

VARIABLE	DIMENSIONES
INDEPENDIENTE	- Conjunto de factores distales.
Modelo Predictivo	- Análisis del modelo predictivo.
DEPENDIENTE	- Presencia de anemia
Riesgo de anemia	

CAPITULO III

METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN

3.1. Tipo de investigación

Este trabajo de investigación es de tipo Básica, según Carrasco (2013). Afirma que no tiene ningún interés aplicado ya que busca desarrollar y profundizar el conocimiento científico existente.

3.2. Nivel de investigación

El nivel descriptivo.(Hernández Sampieri & Mendoza Torres, 2018)

3.3. Métodos de investigación

El método analítico es definido como “Aplicar métodos científicos al discurso, entendiendo que el discurso es expresión de cualquier estructura subjetiva” (Manrique, Lopera, Pérez, Ramírez y Henao, 2016).

3.4. Diseño de investigación

Según (Hernández Sampieri & Mendoza Torres, 2018)Correlacional-transversal dado que se emplearon herramientas y pruebas teóricas relacionadas con los temas de factores de peligro asociados a la anemia en niños de edad inferior a 5 años.

3.5. Población y muestra

3.5.1. Población

Según Hernández Sampieri (2018) menciona que: “es un conjunto de individuos que se hallan en un determinado sector y que nos ayuda para adquirir la muestra y los resultados”.

La población está comprendida por los niños menores de cinco años registrados en el SIEN de la DIRESA provincia de Daniel Alcides Carrión correspondiente al año 2023, siendo un total de 1028 niños menores de 5 años.

3.5.2. Muestra

Una muestra intencionada (también conocida como muestra por conveniencia o muestra dirigida) es un tipo de muestreo no probabilístico donde los participantes se seleccionan intencionalmente según ciertos criterios preestablecidos.

Se usará toda la población de 1028 niños menores de 5 años.

3.6. Técnicas e instrumentos de recolección de datos

Técnica

Base de Datos

Instrumentos

Datos obtenidos SIEN de la DIRESA.

Aplicación

Modelo predictivo propuesto

3.7. Selección, validación y confiabilidad de los instrumentos de investigación

Se harán uso de la estadística inferencial específicamente con el Alfa de Cronbach para medir la confiabilidad, exactitud y veracidad de los datos presentados.

Estadísticas de fiabilidad

Alfa de Cronbach	N de elementos
,953	12

Procesado las 12 variables el alfa de Cronbach 0.953 con un porcentaje de excelente.

3.8. Técnicas de procesamiento y análisis de datos

Quando es necesario, los datos se codifican y se transfieren a una base de datos para su procesamiento utilizando Excel 2016 y SPSS versión 26. Luego se realizó un análisis descriptivo, seguido de un análisis de regresión logística binaria multivariante para su interpretación. Los hallazgos se elaboraron en tablas y figuras estadísticas, de conformidad con los propósitos de la investigación.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	sw	Diresa	Red	Microred	EESS	Dpto_EESS	Prov_EESS	Dist_EESS	Renipress	FechaAtenci	Pais	Sexo	FechaNacim	EdadMeses	UbigeoPN	Departamen	Provir
2	1	PASCO	PASCO	MESETA	I-2 - 000010	PASCO	PASCO	VICCO	1070	6/21/2023	PER	M	6/21/2022	12			
3	1	PASCO	PASCO	MESETA	I-2 - 000010	PASCO	PASCO	VICCO	1070	6/26/2023	PER	F	6/26/2022	12			
4	1	PASCO	OXAPAMPA	NO PERTENE II-1 - 000009	PASCO	OXAPAMPA	OXAPAMPA	OXAPAMPA	980	2/20/2023	PER	F	2/19/2020	36			
5	1	PASCO	OXAPAMPA	NO PERTENE II-1 - 000009	PASCO	OXAPAMPA	OXAPAMPA	OXAPAMPA	980	5/24/2023	PER	M	10/02/2020	31			
6	1	PASCO	PASCO	HUARIACA	I-2 - 000010	PASCO	PASCO	PALLANCHAY	1030	6/13/2023	PER	F	11/27/2019	43			
7	1	PASCO	DANIEL CARITUSI	I-2 - 000010	PASCO	DANIEL ALCII SANTA ANA	DANIEL ALCII SANTA ANA	DANIEL ALCII SANTA ANA	1006	6/17/2023	PER	F	5/30/2020	37	190206	PASCO	DANIEL CARITUSI
8	1	PASCO	DANIEL CARITUSI	I-2 - 000010	PASCO	DANIEL ALCII SANTA ANA	DANIEL ALCII SANTA ANA	DANIEL ALCII SANTA ANA	1006	1/17/2023	PER	F	11/10/2020	26	190206	PASCO	DANIEL CARITUSI
9	1	PASCO	DANIEL CARITUSI	I-2 - 000010	PASCO	DANIEL ALCII SANTA ANA	DANIEL ALCII SANTA ANA	DANIEL ALCII SANTA ANA	1006	4/05/2023	PER	M	10/05/2021	18	190206	PASCO	DANIEL CARITUSI
10	1	PASCO	PASCO	HUARIACA	I-1 - 000010	PASCO	PASCO	HUARIACA	1027	3/26/2023	PER	M	9/26/2020	30	190103	PASCO	PASCO
11	1	PASCO	PASCO	MESETA	I-3 - 000010	PASCO	PASCO	NINACACHA	1060	2/02/2023	PER	M	2/02/2021	24	190105	PASCO	PASCO
12	1	PASCO	DANIEL CARF CHAUPIHUAI	I-3 - 000012	PASCO	DANIEL ALCII TAPUC	DANIEL ALCII TAPUC	DANIEL ALCII TAPUC	1207	5/13/2023	PER	M	4/13/2021	25	190207	PASCO	DANIEL CARF CHAUPIHUAI
13	1	PASCO	DANIEL CARF YANAHUANG	I-1 - 000012	PASCO	DANIEL ALCII YANAHUANG	DANIEL ALCII YANAHUANG	DANIEL ALCII YANAHUANG	1221	4/19/2023	PER	F	4/19/2020	36	190201	PASCO	DANIEL CARF YANAHUANG
14	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	4/20/2023	PER	M	4/20/2020	36	190202	PASCO	DANIEL CARF CHAUPIHUAI
15	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	5/20/2023	PER	M	5/20/2020	36	190202	PASCO	DANIEL CARF CHAUPIHUAI
16	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	5/21/2023	PER	M	5/21/2020	36	190202	PASCO	DANIEL CARF CHAUPIHUAI
17	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	3/04/2023	PER	M	9/04/2020	30	190202	PASCO	DANIEL CARF CHAUPIHUAI
18	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	3/03/2023	PER	F	9/03/2020	30	190202	PASCO	DANIEL CARF CHAUPIHUAI
19	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	4/02/2023	PER	F	10/02/2020	30	190202	PASCO	DANIEL CARF CHAUPIHUAI
20	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	1/31/2023	PER	F	1/31/2021	24	190202	PASCO	DANIEL CARF CHAUPIHUAI
21	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	4/13/2023	PER	F	4/13/2021	24	190202	PASCO	DANIEL CARF CHAUPIHUAI
22	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	5/22/2023	PER	M	5/22/2021	24	190202	PASCO	DANIEL CARF CHAUPIHUAI
23	1	PASCO	DANIEL CARF CHAUPIHUAI	I-2 - 000009	PASCO	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	DANIEL ALCII CHACAYAN	992	3/15/2023	PER	M	9/15/2021	18	190202	PASCO	DANIEL CARF CHAUPIHUAI

3.9. Tratamiento Estadístico

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 26 para su tabulación. Posteriormente a los datos se les realizó un análisis descriptivo y luego se le realizo 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.

3.10. Orientación ética filosófica y epistémica

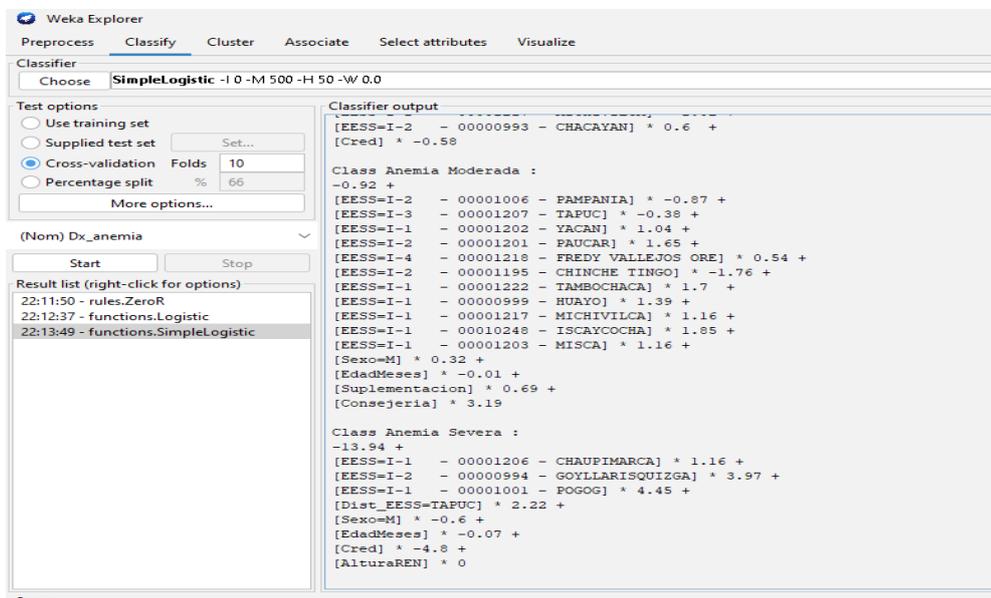
La información proporcionada por la base de datos ha sido utilizada exclusivamente para el propósito de esta investigación. El especialista se comprometió a asegurar la fiabilidad de los hallazgos y la confiabilidad de los datos proporcionados por la SIEN de la DIRESA.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. Descripción del trabajo de campo

Este trabajo se desarrolló en las entidades de salud de la Provincia Daniel Alcides Carrión – Pasco. La cual se analizó a los niños menores de 5 años que realicen su diagnóstico que tipo de anemia tienen.



The screenshot displays the Weka Explorer interface with the Simple Logistic classifier selected. The classifier output is shown for two classes: Anemia Moderada and Anemia Severa. The output lists various features and their coefficients, such as [EESS=I-2 - 00000993 - CHACAYAN] * 0.6 + and [Cred] * -0.58 for the Anemia Moderada class. The Anemia Severa class output includes features like [EESS=I-1 - 00001206 - CHAUPIMARCA] * 1.16 + and [EESS=I-2 - 00000994 - GOYLLARISQUIZGA] * 3.97 +.

```
Classifier: SimpleLogistic -I 0 -M 500 -H 50 -W 0.0

Test options
  Use training set
  Supplied test set
  Cross-validation Folds: 10
  Percentage split %: 66

(Nom) Dx_anemia
  Start Stop
Result list (right-click for options)
  22:11:50 - rules.ZeroR
  22:12:37 - functions.Logistic
  22:13:49 - functions.SimpleLogistic

Classifier output
[EESS=I-2 - 00000993 - CHACAYAN] * 0.6 +
[Cred] * -0.58

Class Anemia Moderada :
-0.92 +
[EESS=I-2 - 00001006 - PAMFANIA] * -0.87 +
[EESS=I-3 - 00001207 - TAPUC] * -0.38 +
[EESS=I-1 - 00001202 - YACAN] * 1.04 +
[EESS=I-2 - 00001201 - PAUCAR] * 1.65 +
[EESS=I-4 - 00001219 - FREDY VALLEJOS ORE] * 0.54 +
[EESS=I-2 - 00001195 - CHINCHE TINGO] * -1.76 +
[EESS=I-1 - 00001222 - TAMBOCHACA] * 1.7 +
[EESS=I-1 - 00000999 - HURAYO] * 1.39 +
[EESS=I-1 - 00001217 - MICHIVILCA] * 1.16 +
[EESS=I-1 - 00010248 - ISCAYCOCHA] * 1.85 +
[EESS=I-1 - 00001203 - MISCA] * 1.16 +
[Sexo=M] * 0.32 +
[EdadMeses] * -0.01 +
[Suplementacion] * 0.69 +
[Consejeria] * 3.19

Class Anemia Severa :
-13.94 +
[EESS=I-1 - 00001206 - CHAUPIMARCA] * 1.16 +
[EESS=I-2 - 00000994 - GOYLLARISQUIZGA] * 3.97 +
[EESS=I-1 - 00001001 - POGOG] * 4.45 +
[Dist_EESS=TAPUC] * 2.22 +
[Sexo=M] * -0.6 +
[EdadMeses] * -0.07 +
[Cred] * -4.8 +
[AlturaREN] * 0
```

```

=== Summary ===
Correctly Classified Instances      800          77.7454 %
Incorrectly Classified Instances    229          22.2546 %
Kappa statistic                    0.2534
Mean absolute error                 0.1341
Root mean squared error            0.2607
Relative absolute error             83.8273 %
Root relative squared error        92.4374 %
Total Number of Instances         1029
Ignored Class Unknown Instances     9

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.970   0.732   0.805     0.970   0.880     0.362   0.799    0.913    Normal
                0.239   0.036   0.515     0.239   0.327     0.286   0.740    0.301    Anemia Leve
                0.097   0.014   0.435     0.097   0.159     0.169   0.734    0.246    Anemia Moderada
                0.000   0.001   0.000     0.000   0.000     -0.002  0.937    0.047    Anemia Severa
                0.000   0.000   ?         0.000   ?         ?       0.453    0.002    1
Weighted Avg.   0.777   0.561   ?         0.777   ?         ?       0.785    0.758

=== Confusion Matrix ===
 a  b  c  d  e  <-- classified as
756 16  7  0  0 | a = Normal
102 34  6  0  0 | b = Anemia Leve
 78 14 10  1  0 | c = Anemia Moderada
  2  2  0  0  0 | d = Anemia Severa
  1  0  0  0  0 | e = 1

```

Analizando con el programa Weka la regresión logística es la más acertada ya que tiene un 80.5% de precisión.

4.2. Presentación, análisis e interpretación de resultados

4.2.1. Prevalencia de anemia

“Para determinar la prevalencia de anemia en niños menores de 5 años, se utilizó la muestra de los datos de la SIEN-DIRESA 2019, siendo un total de 1028 niños registrados de la Provincia Daniel Alcides Carrión – Pasco”.

Tabla 2. Prevalencia de anemia según sexo

		Dx_anemia				Total	
		Sin anemia		Con anemia		n	%
		n	%	n	%		
Sexo	F	398	38.7	110	10.7	508	49.4
	M	380	37.0	140	13.6	520	50.6
Total		778	75.7	250	24.3	1028	100.0

Los hallazgos obtenidos nos indican que la prevalencia de anemia en niñas es del 10.7% y en niños del 13.6%, lo cual es bastante significativo y, por

lo tanto, reflejando la problemática de la provincia Daniel Alcides Carrión. De igual modo, en general se encuentra una prevalencia de anemias del 24.3%.

4.2.2. Selección de variables por factor

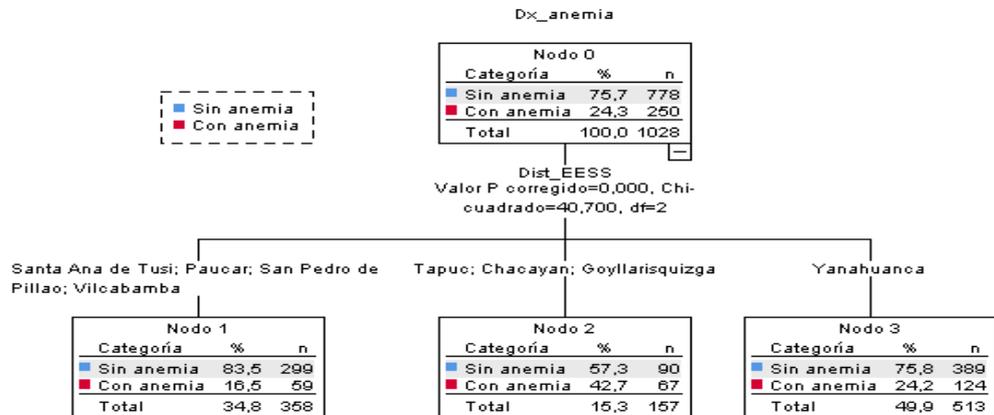
En este paso se utiliza la técnica del árbol de decisión porque este método es minería de datos, nos permite descubrir y explorar los datos para extraer la información almacenada en ellos (Berlanga, Rubio, & Vilà, 2013); razón por la cual, se trabajó con la totalidad de los datos

a. Factores demográficos

Los factores demográficos a considerar en la base de datos SIEN incluyen la región, el sitio y la elevación del sitio. En este caso, solo se consideran las variables de altitud de la región y la ubicación (Altitud_Loc), y el árbol de decisión como procedimiento excluirá automáticamente cualquier variable seleccionada que no tenga un efecto significativo en el modelo final.

En la figura, el nodo 0 describe las variables dependientes para el diagnóstico de anemia: el porcentaje sin anemia (75,7%) y con anemia (24,3%). La variable distrito es el idóneo para el diagnóstico de la anemia, la cual se clasifica en cinco grupos (nodos de 1 a 3).

Figura 3. Diagrama del árbol para el modelo de anemia según factores demográficos



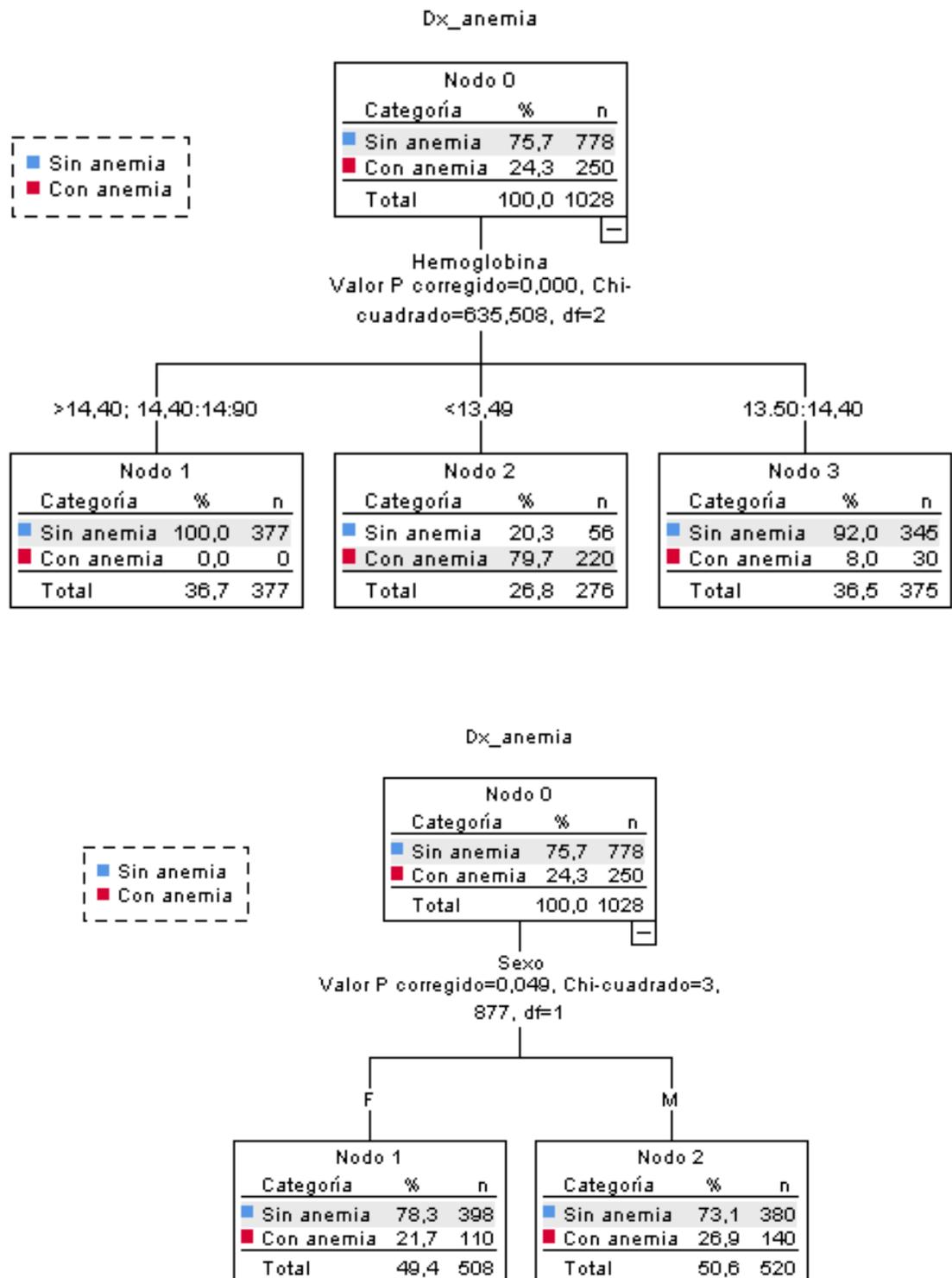
En resumen, los nodos que definen el perfil de los niños con anemia (variables demográficas que afectan la anemia): nodo 0 - nodo 2 - nodo 3. Es decir, influyen las siguientes variables: Distrito– Altitud_Loc.

b. Factores biológicos

Los factores biológicos incluyen la edad, el peso, la talla, la hemoglobina y el sexo. Elegiremos las variables cuya contribución al modelo es significativa utilizando la técnica del árbol de decisión.

En la figura, se puede ver que la variable hemoglobina, que se clasifica en cuatro grupos (nodos de 1 a 4), es el mejor predictor para el diagnóstico de anemia; las variables de peso se dividen en dos grupos (nodos 5 y 6), y las variables de tamaño también se dividen en dos grupos (nodos 9 y 10). Es importante recordar que la variable edad no se encuentra en la estructura del árbol, por lo que no afecta el diagnóstico de anemia.

Figura 4. Diagrama del árbol para el modelo de anemia según factores biológico



La mayor probabilidad de anemia (100%) en niños menores de 5 años se presenta en los varones con una hemoglobina inferior a 13,50 y un peso inferior a 9,29 kg. La hemoglobina está entre [14,4; 14.9], altura superior a 80,3 cm, mujer.

4.2.3. Estimación de modelo

Selección de variables

Después del análisis factorial y la posterior selección y recodificación de variables, se realizaron cálculos iniciales del modelo para todas las variables seleccionadas y se analizó su importancia en la determinación de la anemia en niños menores de 5 años. Para ello se utilizó la selección proporcional de estratos (distrito).

Tabla 3. Variables que ingresan al modelo

		Variables en la ecuación					
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	Sesion	-12,870	14059,038	,000	1	,999	,000
	AlturaREN	-4,612	14,483	,101	1	,750	,010
	Hemoglobina	3,841	,250	235,646	1	,000	46,575
	Sexo	,203	,244	,692	1	,405	1,225
	Juntos	-,099	,249	,156	1	,692	,906
	SIS	,031	,424	,005	1	,942	1,031
	Peso	-,012	,009	1,930	1	,165	,988
	Talla	-,295	,254	1,351	1	,245	,745
	Cred	-1,074	,423	6,435	1	,011	,342
	Suplementacion	,838	,297	7,965	1	,005	2,311
	Consejeria	4,364	9,922	,193	1	,660	78,567
	Constante	-5,031	14,490	,121	1	,728	,007

a. Variables especificadas en el paso 1: Sesion, AlturaREN, Hemoglobina, Sexo, Juntos, SIS, Peso, Talla, Cred, Suplementacion, Consejeria.

En la tabla 3, se observa las variables significativas para el modelo:

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

En la tabla 4, Se muestra claramente que las variables X1, X2, X3 y X10 en el nuevo modelo son significativas y, por lo tanto, tienen un impacto en el modelo. Cabe mencionar que, a pesar de las pruebas para excluir estas variables del modelo, su desempeño disminuyó, por lo que se mantuvieron.

Tabla 4. Variables que ingresan al nuevo modelo

		Variables en la ecuación							
		B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Paso 1 ^a	Sesion	-13,068	13970,360	,000	1	,999	,000	,000	.
	AlturaREN	-4,974	12,528	,158	1	,691	,007	,000	319089732,8
	Hemoglobina	3,828	,241	251,924	1	,000	45,964	28,651	73,740
	Suplementacion	,785	,282	7,741	1	,005	2,192	1,261	3,810
	Constante	-5,742	12,528	,210	1	,647	,003		

a. Variables especificadas en el paso 1: Sesion, AlturaREN, Hemoglobina, Suplementacion.

En la tabla 5, el coeficiente de determinación se observa, que para nuestro modelo es del 70.7%. Por lo tanto, las variables del modelo X1: Sector, X2: Altitud de la localidad, X3: Hemoglobina y X10: Número de visitas explican el 70.7 por ciento de la variación de la variable anemia en niños menores de 5 años (anémico o no anémico).

Tabla 5. 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	479,749 ^a	,474	,707

a. La estimación ha terminado en el número de iteración 20 porque se ha alcanzado el máximo de iteraciones. La solución final no se puede encontrar.

Tabla 6. Tabla de clasificación

Tabla de clasificación^a

Observado		Pronosticado		Porcentaje correcto
		Dx_anemia		
Paso 1	Dx_anemia	Sin anemia	Con anemia	
		Sin anemia	721	56
	Con anemia	30	220	88,0
	Porcentaje global			91,6

a. El valor de corte es ,500

En la tabla 5, Observamos la clasificación del modelo, que muestra una tasa de buena clasificación del 70.7%. Y en la tabla #6 de clasificación muestra que el modelo tiene una alta especificidad (92,8%) y una alta sensibilidad (88,0%). Se descubrió que el modelo tiene una efectividad del 91.6 % Variables predictoras X1: Sector, X2: Altura de la ciudad, X3: Fondo, X10: Número de visitas. Esto indica que el modelo tiene una buena capacidad de predicción con respecto a las variables que utiliza. Es importante tener en cuenta que el punto de corte es de 0,5

También de la tabla de clasificación podemos decir:

Variable Predictores: La condición de que un niño o niña menor de 5 años termine como anémico.

- Predicción correcta 88.0%.

La condición de que un niño o niña menor de 5 años termine sin anemia:

- Predicción correcta 92.8%.

La condición de que un niño o niña menor de 5 años termine con anemia;

- Predicción Incorrecta 2.8%.

La condición de que un niño o niña menor de 5 años termine sin anemia:

- Predicción Incorrecta 8.0%.

Para analizar el rendimiento (o eficiencia en el sentido más amplio) del modelo de clasificación, utilizamos curvas COR, donde comparamos el valor de la variable anemia con el valor de probabilidad predicho. La Figura 6 y la Tabla 10 muestran que nuestro modelo (representado por la curva azul) es 0,976 más alto que la curva aleatoria.

Para analizar el rendimiento (o eficacia, en el sentido más amplio) de nuestro modelo clasificador, utilizamos la curva COR, donde tomamos el valor de la variable anémica en comparación con el valor de probabilidad predicho. La figura y la tabla muestran que nuestro modelo, representado por la curva azul, está por encima de una curva aleatoria de área 0.976.

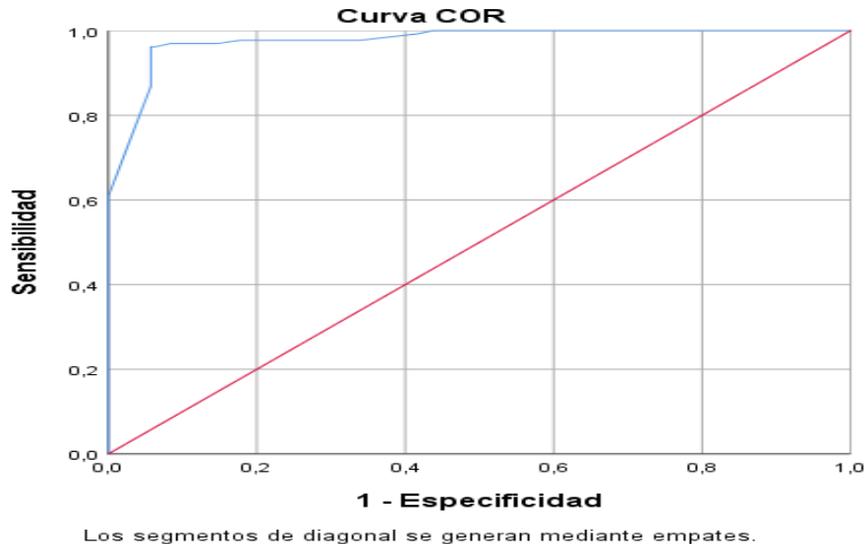


Tabla 7. Tabla del área bajo la curva COR.

Área bajo la curva				
Variables de resultado de prueba: Probabilidad pronosticada				
95% de intervalo de confianza asintótico				
Área	Desv. Error ^a	Significación asintótica ^b	Límite inferior	Límite superior
,976	,007	,000	,962	,990

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

Si el área que encierra la Curva COR es 1, hay un orden claro y las categorías van disminuyendo a medida que nos alejamos de los valores; Este es un buen modelo para clasificar a los niños ya que la línea COR es 0,976. Con punto de corte 0.5.

4.2.4. Verosimilitud del modelo

En la tabla 8, La significación estadística se demostró en la prueba de Wald. Un modelo consistente es importante. Por lo tanto, siempre debe incluirse en el modelo ya que provoca cambios en la variable anemia (estado del niño o niña con o sin anemia).

Tabla 8. Prueba del coeficiente de intercepción

Variables en la ecuación							
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 0	Constante	-1,134	,073	243,221	1	,000	,322

Al evaluar la hipótesis nula ($B_i = 0$) “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”:

$$p(x) = \frac{1}{1 + e^{[-5,742 - 13.068x_1 - 4,974x_2 + 3,828x_3 + 0.785x_{10}]}}$$

$$\text{Ln (odds)} = -5,742 - 13.068x_1 - 4,974x_2 + 3,828x_3 + 0.785x_{10}$$

4.2.5. Prueba de bondad de ajuste

La tabla 9, muestra la prueba de bondad de ajuste del modelo, el modelo es suficiente al nivel de significancia del 5 por ciento, es decir, el modelo de regresión se ajusta a los datos (p-valor = 0.088).

Tabla 9. Prueba de bondad de ajuste.

Prueba de Hosmer y Lemeshow

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

4.3. Prueba de Hipótesis

4.3.1. Hipótesis específicas 1:

H_0 : El conjunto de factores distales no influye en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

H_1 : El conjunto de factores distales influye en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

En la tabla 6, Observamos la clasificación del modelo, que muestra una tasa de buena clasificación del 70.7%. La tabla de clasificación muestra que el modelo tiene una alta especificidad (92,8%) y una alta sensibilidad (88,0%). Se descubrió que el modelo tiene una efectividad del 91.6 % Variables predictoras

X1: Sector, X2: Altura de la ciudad, X3: Fondo, X10: Número de visitas. Esto indica que el modelo tiene una buena capacidad de predicción con respecto a las variables que utiliza. Es importante tener en cuenta que el punto de corte es de 0,5

Interpretación:

El conjunto de factores distales influye en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.

4.3.2. Hipótesis específicas 2:

H0: El modelo predictivo no permite realizar la probabilidad del riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023.

H1: El modelo predictivo permite realizar la probabilidad del riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023.

Tabla 10. Prueba de Prueba de Omnibus

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	660,219	4	,000
	Bloque	660,219	4	,000
	Modelo	660,219	4	,000

Interpretación:

La primera línea, H0, indica que el coeficiente de la variable sumada es cero. En el nivel de significancia de 0,000, H0 es positiva, lo que significa que al menos uno de los coeficientes de la variable es diferente de cero. Señala que el modelo y sus variables asociadas son de gran importancia.

Predicción del riesgo de anemia

Usando los coeficientes estimados en la columna B del Cuadro 4 y el valor de la variable para cada caso utilizando un modelo de regresión, podemos estimar la probabilidad o el riesgo de anemia para cada niño o niña menor de 5 años. Por ejemplo, la siguiente tabla contiene dos preguntas (48 y 238 muestras), y sus valores para predecir el riesgo de anemia en niños menores de 5 años son los siguientes:

<i>Variables</i>	<i>Caso</i>	
	<i>Y₄₈</i>	<i>Y₂₃₈</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

Los factores distales son los factores genéticos y eventos adversos en la infancia.

En el caso Y48 muestra datos de un niño anémico ubicado en el Sector 4, correspondiente a la capital del estado, Daniel Alcides Carrión, con una altitud menor a 3998 metros. N. M., nivel de hemoglobina superior a 14. Nunca hemos recibido visitas de empleados de SIEN, que son 90 años y tienen un perfil infantil típico de esta región. Tenga en cuenta que la probabilidad predicha del modelo es menor que $P_c = 0,50$, por lo que este primer caso se clasificó como niño anémico. Para el caso Y238, que pertenece al Sector 4 correspondiente a los distritos de Tapúc, Chacayan y Goyllarisquizca, 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 con anemia.

4.4. Discusión de resultados

Los resultados del ajuste del modelo permitieron estimar la contribución de las causas de la anemia. El modelo que muestra el modelo de bondad de la prueba de Hosmer y Lemeshow tiene un valor $p = 0,088$ con un nivel de significancia de 0,05 y un coeficiente de determinación de Nagelkerke del 70,7%. La tasa de clasificaciones verdaderamente negativas (que predicen que la anemia no es anemia) fue del 92,8%; En este caso, el parámetro más importante es la clasificación correcta de los niños sin anemia (verdaderos positivos) con un 91,6%, lo que demuestra un fuerte poder predictivo del modelo.

En el trabajo de investigación de 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.

CONCLUSIONES

- Este trabajo de investigación ha identificado los factores de riesgo registrados en el SIEN de la DIRESA provincia Daniel Alcides Carrión – Pasco correspondiente al año 2023, se asocia con un alto riesgo de anemia en niños menores de 5 años y ha sido demostrado en la literatura como un factor de riesgo; La prevalencia global de anemia es del 24,3%.
- El ajuste del modelo permitió estimar la contribución de los factores de riesgo. La prueba de bondad de ajuste de Hosmer y Lemeshow, que es un modelo representativo, es el valor p del modelo en un nivel de significancia de 0,05. En este caso, la tasa más útil de clasificación correcta de niños con anemia (verdaderos positivos) fue del 91,6%, lo que indica una fuerte capacidad predictiva del modelo.

RECOMENDACIONES

- Con respecto a la MICRORED Daniel Alcides Carrión – Pasco Se recomienda enfocar campañas de concientización y/o jornadas demostrativas, así como visitas a los hogares de niños con hemoglobina baja que podría convertirse en otro caso de anemia, para los distritos con mayor número de casos. Son factores que combaten la anemia.
- Debido a la gran cantidad de datos que existen en las instituciones de salud de nuestra región y país, en ocasiones no resulta difícil organizarlos. Se recomienda utilizar la tecnología de Data Mining para encontrar formas de estudiar la dinámica de las enfermedades infecciosas en nuestros niños.

REFERENCIAS BIBLIOGRÁFICAS

- Abimegireysch Alcarraz, J. (2022). *Factores de riesgo asociados a la anemia en niños menores de 5 años de edad, subanálisis ENDES 2020-2021*. UNIVERSIDAD PRIVADA SAN JUAN BAUTISTA.
- Canaza, G. (2021). *Modelo Predictivo de Riesgo Asociado a la Anemia en niños menores de 5 años en la Microred Yauri Provincia de Espinar-Cusco, 2019*. UNIVERSIDAD NACIONAL DEL ALTIPLANO DE PUNO.
- Daniel, H. M. (2024). Implementación de un modelo de análisis predictivo en el control de nutrición de los niños menores de 5 años del Centro de Salud de Uliachin, Chaupimarca - Pasco, 2023. In *Interciencia* (Vol. 1, Issue 1). UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN.
- Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la Investigación*.
- Leal, L. P., Batista Filho, M., Lira, P. I. C. de, Figueiroa, J. N., & Osório, M. M. (2020). Prevalencia de la anemia y factores asociados en niños de seis a 59 meses de Pernambuco, Noreste de Brasil. *Revista de Saúde Pública*, 45(3), 457–466.
<https://doi.org/10.1590/S0034-89102011000300003>
- Loaiza, Á., Moreno, W., & Rios, J. (2023). Proceso de analítica de datos aplicado a la desnutrición infantil en niños de 0 a 5 años en la Ciudad de Medellín. In *Universidad San Buenaventura*. ITM Institución Universitaria.
- Zavaleta, N., & Astete, L. (2017). *Efecto de la anemia en el desarrollo infantil: consecuencias a largo plazo*. <https://doi.org/10.17843/rpmesp.2017.344.3251>

ANEXOS

Instrumentos de Recolección de datos

Datos.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Visible: 20 de 20 variables

	Red	Microred	EESS	Prov_EESS	Dist_EESS	Sexo	Ed
1	DANIEL CARRION	TUSI	I-2 - 00001006 - PAMPANIA	DANIEL ALCIDES CARRION	Santa Ana de Tusi	F	
2	DANIEL CARRION	TUSI	I-2 - 00001006 - PAMPANIA	DANIEL ALCIDES CARRION	Santa Ana de Tusi	F	
3	DANIEL CARRION	TUSI	I-2 - 00001006 - PAMPANIA	DANIEL ALCIDES CARRION	Santa Ana de Tusi	M	
4	DANIEL CARRION	CHAUPIHUARANGA	I-3 - 00001207 - TAPUC	DANIEL ALCIDES CARRION	Tapuc	M	
5	DANIEL CARRION	YANAHUANCA	I-1 - 00001221 - SANTIAGO PAMPA	DANIEL ALCIDES CARRION	Yanahuanca	F	
6	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	M	
7	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	M	
8	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	M	
9	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	M	
10	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	F	
11	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	F	
12	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	F	
13	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	F	
14	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	M	
15	DANIEL CARRION	CHAUPIHUARANGA	I-2 - 00000992 - CHANGO	DANIEL ALCIDES CARRION	Chacayan	M	
16	DANIEL CARRION	YANAHUANCA	I-1 - 00001204 - TANGOR	DANIEL ALCIDES CARRION	Paucar	M	
17	DANIEL CARRION	TUSI	I-2 - 00001007 - SANTA ROSA CHORA	DANIEL ALCIDES CARRION	Santa Ana de Tusi	F	
18	DANIEL CARRION	TUSI	I-2 - 00001007 - SANTA ROSA CHORA	DANIEL ALCIDES CARRION	Santa Ana de Tusi	M	
19	DANIEL CARRION	YANAHUANCA	I-1 - 00001228 - AYAYOG	DANIEL ALCIDES CARRION	Yanahuanca	F	
20	DANIEL CARRION	YANAHUANCA	I-1 - 00001224 - CHIPIPATA	DANIEL ALCIDES CARRION	Yanahuanca	F	
21	DANIEL CARRION	YANAHUANCA	I-1 - 00001192 - INDEPENDENCIA	DANIEL ALCIDES CARRION	Paucar	F	
22	DANIEL CARRION	YANAHUANCA	I-1 - 00001191 - OCHO DE DICIEMBRE	DANIEL ALCIDES CARRION	Paucar	F	
23	DANIEL CARRION	CHAUPIHUARANGA	I-1 - 00001206 - CHAUPIMARCA	DANIEL ALCIDES CARRION	Tapuc	M	

Vista de datos Vista de variables

Matriz de Consistencia

Tema: “Modelo Predictivo asociado al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023”

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLE INDEPENDIENTE	DIMENSIÓN	DISEÑO	POBLACIÓN Y MUESTRA
¿Cuál es el Modelo Predictivo que asocia al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023?	Determinar el Modelo Predictivo que asocia al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.	El modelo Predictivo permite realizar la predicción asociado al riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023.	Modelo Predictivo	- Conjunto de factores distales -Análisis del modelo predictivo.	Diseño: Correlacional-transversal Tipo de Investigación Básica	POBLACIÓN La población está comprendida por 1063 los niños menores de cinco años registrados en el SIEN de la DIRESA de la Provincia Daniel Alcides Carrión - Pasco, 2023.
PROBLEMA ESPECÍFICO	OBJETIVO ESPECÍFICO	HIPÓTESIS ESPECIFICA	VARIABLE DEPENDIENTE	DIMENSIÓN	MÉTODO DE INVESTIGACIÓN	TÉCNICAS - INSTRUMENTOS
¿Cómo identificamos un conjunto de factores distales que influyan en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023?	Identificar un conjunto de factores distales que influyan en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.	El conjunto de factores distales influye en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.	Riesgo de anemia.	- Presencia de anemia.	Método Analítico Enfoque Cuantitativo	Técnicas: -Registros
¿Cómo estimamos un modelo predictivo para la probabilidad en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión - Pasco, 2023?	Estimar un modelo predictivo para la probabilidad en riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.	El modelo predictivo permite realizar la probabilidad del riesgo de anemia en niños menores de 5 años en la Provincia Daniel Alcides Carrión -Pasco, 2023.				