

# UCUENCA

## Universidad de Cuenca

Facultad de Ciencias Químicas

Carrera de Ingeniería Industrial

**Modelo de predicción para las variables físico químicas que generan devolución del producto leche semidescremada UHT polietileno 1L, a partir del uso de técnicas de regresión logística para el área de calidad en el sector lácteo**


Trabajo de titulación previo a la obtención del título de Ingeniera Industrial

**Autora:**

Paula Karina Lanchimba Parra

**Directora:**

Paulina Alejandra Echeverría Paredes

ORCID:  0000-0001-9487-6940

**Cuenca, Ecuador**

2023-07-17

## Resumen

El rechazo de un producto por parte de los consumidores genera consecuencias financieras y un aumento de productos no conformes dentro de una empresa dedicada a la producción de lácteos; es por ello que, para cumplir con las condiciones de seguridad alimentaria del producto se deben establecer los controles de calidad adecuados. La presente investigación se basó en identificar al producto con mayor cantidad de devoluciones realizados por los consumidores al Área de Gestión de Calidad en la empresa Lácteos San Antonio, siendo la Leche Semidescremada UHT polietileno 1L. Se estudiaron las variables físico-químicas con el objetivo de plantear un modelo de predicción que permitiera disminuir la cantidad de producto no conforme. Con esta finalidad, se aplicó una metodología de investigación deductiva y correlacional, partiendo del análisis de los productos devueltos con las condiciones más representativas por rechazo, siendo las fundas de polietileno cortadas, hinchadas y abolladas, con el 43%, 20% y 14%, respectivamente. Se realizó un análisis ANOVA para las variables de pH, acidez y crioscopía, el cual determinó que la crioscopía no tuvo influencia para el desarrollo del modelo de predicción. Además, el análisis de la regresión logística llevó a determinar que el sólo el pH es la variable predictora. Finalmente, a partir del modelo de predicción se ajustó el rango inferior de pH desde 6,55 a 6,60 y se obtuvo una disminución del porcentaje de probabilidad de rechazo del producto de 0,5% a 0,25%.

*Palabras clave:* calidad de productos lácteos, gestión de calidad, seguridad alimentaria



The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

**Institutional Repository:** <https://dspace.ucuenca.edu.ec/>

### Abstract

The rejection of a product by consumers generates financial consequences and an increase of nonconforming products within a company dedicated to the production of dairy products; that is why, in order to comply with the food safety conditions of the product, adequate quality controls must be established. This research was based on identifying the product with the highest number of returns made by consumers to the Quality Management Area of the company Lácteos San Antonio, being the Semi-skimmed Milk UHT polyethylene 1L. The physicochemical variables were studied in order to propose a prediction model that would allow reducing the amount of nonconforming product. To this end, a deductive and correlational research methodology was applied, starting from the analysis of the returned products with the most representative conditions for rejection, being the polyethylene sleeves cut, swollen and dented, with 43%, 20% and 14%, respectively. An ANOVA analysis was performed for the pH, acidity and cryoscopy variables, which determined that cryoscopy had no influence on the development of the prediction model. In addition, logistic regression analysis led to the determination that pH alone is the predictor variable. Finally, from the prediction model, the lower pH range was adjusted from 6.55 to 6.60 and a decrease in the percentage probability of product rejection from 0.5% to 0.25% was obtained.

**Keywords:** dairy product quality, quality management, food safety



The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

**Institutional Repository:** <https://dspace.ucuenca.edu.ec/>

## Índice de contenido

1.	Identificación del Problema y Justificación del Ensayo Académico .....	10
2.	Estado del Arte .....	13
2.1.	Generalidades de la Leche Semidescremada .....	13
2.2.	Control de procesos y calidad .....	15
2.3.	Estadística en el control de procesos.....	17
2.3.1.	Regresión logística.....	19
2.3.2.	Regresión logística simple .....	21
2.3.3.	Modelo de predicción .....	24
2.3.4.	Evaluación del modelo de predicción.....	24
2.3.5.	Herramientas de control de calidad .....	26
3.	Hipótesis y/o preguntas de investigación .....	28
4.	Objetivos.....	28
4.1.	Objetivo General .....	28
4.2.	Objetivos específicos .....	28
5.	Descripción de la Metodología .....	29
5.1.	Análisis del proceso de devoluciones dentro del Área de Gestión de Calidad de la empresa Lácteos San Antonio. ....	29
5.2.	Definición de la muestra de estudio .....	30
5.3.	Identificación de las principales causas de devolución del producto leche semidescremada UHT polietileno 1L, y cuáles son sujetas a predicción .....	30
5.4.	Recolección de datos .....	31
5.5.	Definición de los factores que determinan el modelo de predicción mediante técnicas de inferencia estadística .....	33
5.6.	Obtención de un modelo de predicción mediante el uso de regresión logística .....	34
5.7.	Aplicación de la herramienta de gestión de calidad para el reajuste de las líneas de control del producto lácteo.....	36
6.	Resultados y Discusión .....	37
6.1.	Definición de la muestra de estudio .....	37
6.2.	Principales causas de devolución del producto leche semidescremada UHT polietileno 1L.....	39
6.3.	Resultados del análisis bromatológico .....	40
6.4.	Definición de los factores que determinan el modelo de predicción por medio de técnicas de inferencia estadística .....	44

Análisis ANOVA en RStudio.....	44
6.5. Creación del Modelo de Predicción para las variables físico químicas que generan devolución del producto en RStudio .....	46
6.5.1. Modelo de Regresión Logística Lineal .....	46
6.5.2. Evaluación del modelo de regresión logística.....	52
6.6. Aplicación de los resultados del modelo de predicción para el recálculo de las líneas de control de una herramienta de gestión de calidad.....	54
7. Conclusiones.....	60
Referencias.....	62

**Índice de figuras**

Ilustración 1 Porcentaje de la producción de leche a nivel nacional. Fuente: (MAG, 2020) Elaborado por: Paula Lanchimba P. ....	10
Ilustración 2 Principales Industrias del Sector Lácteo. Fuente: (Superintendencia de Control del Poder de Mercado, 2017). Elaborado por: Paula Lanchimba P. ....	11
Ilustración 3 Matriz de Confusión Binaria. Fuente: (Barrios, 2019), Elaborado por: Paula Lanchimba P. ....	25
Ilustración 4 Ecuaciones de la matriz de confusión. Fuente: (Barrios, 2019). Elaborado por: Paula Lanchimba P. ....	26
Ilustración 5 Diagrama de flujo para el proceso de recepción de devoluciones. Elaborado por: Paula Lanchimba P. ....	29
Ilustración 7 Diagrama de flujo por fases del modelo de predicción. Elaborado por: Paula Lanchimba P. ....	35
Ilustración 8 Indicadores de la Matriz de Quejas y Reclamos. Fuente: (Lácteos San Antonio, 2022). Elaborado por: Paula Lanchimba P. ....	37
Ilustración 9 Definición de la muestra de estudio por medio del Diagrama Pareto. Fuente: (Lácteos San Antonio, 2022). Elaborado por: Paula Lanchimba P. ....	38
Ilustración 10 Causas de rechazo. Fuente: (Lácteos San Antonio, 2022). Elaborado por: Paula Lanchimba P. ....	39
Ilustración 11 Diagrama de caja pH. Elaborado por: Paula Lanchimba P. ....	47
Ilustración 12 Diagrama de caja Acidez. Fuente: Elaboración propia. ....	47
Ilustración 13 Análisis multivariado, pH. Fuente: Elaboración propia. ....	48
Ilustración 14 Análisis multivariado, Acidez. Fuente: Elaboración propia. ....	49
Ilustración 15 Regresión Logística para el pH. Fuente: Elaboración propia. ....	51
Ilustración 16 Matriz de confusión. Elaborado por: Paula Lanchimba P. ....	53
Ilustración 17 Límites de los valores físico químicos. Fuente: (Lácteos San Antonio, 2022) .....	55
Ilustración 18 Carta de Control pH, rango actual. Elaborado por: Paula Lanchimba P. ....	55
Ilustración 19 Carta de Control pH, rango propuesto. Elaborado por: Paula Lanchimba P. .....	59
Ilustración 20 Muestra de leche semidescremada .....	73
Ilustración 21 Muestras de leche semidescremada.....	73

**Índice de tablas**

Tabla 1 Estudio bromatológico de las variables físico-químicas de los productos devueltos. Fuente: Elaboración propia.....	41
Tabla 2 Estudio bromatológico de las variables físico-químicas de los productos no devueltos. Fuente: Elaboración propia. ....	42
Tabla 3 Estudio bromatológico de las variables físico-químicas de los productos comprados. Fuente: Elaboración propia.....	43
Tabla 4 Resultados ANOVA, pH vs Categoría. Elaborado por: Paula Lanchimba P. ....	44
Tabla 5 Resultados prueba post hoc de la variable pH. Elaborado por: Paula Lanchimba P. ....	44
Tabla 6 Resultados ANOVA, Acidez vs Categoría. Elaborado por: Paula Lanchimba P. ..	45
Tabla 7 Resultados de prueba post hoc de la variable acidez. Elaborado por: Paula Lanchimba P.....	45
Tabla 8 Resultados ANOVA, Crioscopía vs Categoría. Elaborado por: Paula Lanchimba P. ....	46
Tabla 9 Análisis Multivariado, pH vs Acidez. Elaborado por: Paula Lanchimba P. ....	49
Tabla 10 Resultados de desviación. Elaborado por: Paula Lanchimba P. ....	50
Tabla 11 Resultados de los valores de la distribución del pH. Elaborado por: Paula Lanchimba P.....	50
Tabla 12 Resultados del Pseudo Estadístico R <sup>2</sup> de McFadden. Elaborado por: Paula Lanchimba P.....	52
Tabla 13 Resultados de la matriz de confusión. Elaborado por: Paula Lanchimba P.....	52
Tabla 14 Resultados de las métricas. Elaborado por: Paula Lanchimba P.....	53
Tabla 15 Probabilidades de devolución. Elaborado por: Paula Lanchimba P. ....	57
Tabla 16 Indicador de la meta de devoluciones de la Leche Semidescremada UHT polietileno 1L. Fuente: (Lácteos San Antonio, 2023). Elaborado por: Paula Lanchimba P. ....	58

## Dedicatoria

A mis abuelitas, Victoria y Mercedes, por su constante amor incondicional.

Son las personas más fuertes y luchadoras que puedo conocer. Me enseñaron lo que es el amor, a ser fuerte y valerosa. Ustedes son mi equilibrio y nunca dejaré de estar agradecida por todo.



## Agradecimientos

A Dios por haberme acompañado y guiado a lo largo de mi vida, por darme la fuerza y valentía en momentos de debilidad y sobre todo por su amor.

A mi tutora, Ing. Paulina Echeverría por compartir sus valiosos conocimientos, su paciencia y constancia durante el desarrollo del trabajo de titulación.

Al Ing. Franklin Tusa por abrirme las puertas a nuevos conocimientos y experiencias profesionales, y por su apoyo y tiempo durante el desarrollo del trabajo de titulación.

A los docentes, sus conocimientos rigurosos y palabras sabias me permitieron desarrollarme como profesional. Gracias por su espíritu de docencia, por compartir sus conocimientos de manera profesional, por su dedicación y perseverancia.

A mis padres Cecilia y Patricio por ser siempre una roca en mi vida y ser un apoyo indispensable durante el camino de mi carrera como profesional. Gracias por ser quienes son y por creer en mí. Gracias por su cariño y amor. Los amo con todo mi corazón.

A mis hermanos Vanesa, Andrés y Carolina por ser una guía en mi camino y brindarme su cariño.

A mi novio Benjamín, a mi compañero de vida, quien siempre me animó a seguir mis sueños y cumplir mis metas. Gracias por tu apoyo constante durante todo el proceso de formación académica y personal. Gracias por ser mi guía y mi mejor amigo. Eres mi fuerza y vida. Te amo.

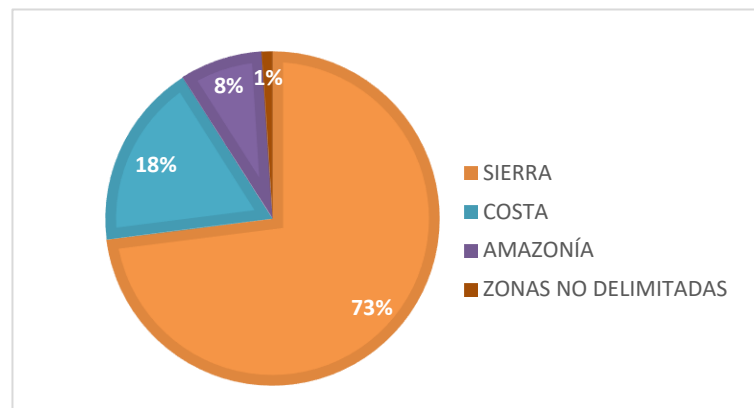
A mis compañeros y amigos Paúl Hernández, Lucas Villalta, Melina Sarmiento, Paula Cañar y Doménica Crespo por su apoyo y cariño cuando más lo necesite.

## 1. Identificación del Problema y Justificación del Ensayo Académico

Para Ecuador la fertilidad de su tierra y la variedad de climas permiten desempeñar las labores ganaderas con las condiciones adecuadas para el cuidado de ganado productor de leche. El sector lechero a través de los años ha ido tecnificándose y especializándose con el propósito de alcanzar niveles de calidad cada vez mejores, a fin de llegar a competir en mercados internacionales; beneficiando así a muchas familias que dependen de este sector (Superintendencia de Control del Poder de Mercado, 2017).

La importancia del sector lácteo en la economía ecuatoriana es notable además de fundamental para desarrollar nuevas fuentes de empleo, tanto directo como indirecto, así como también de valor agregado. Según datos del Instituto Nacional de Estadísticas y Censos (INEC, 2020) la producción de leche es de aproximadamente 6.15 millones de litros diarios, representando el 4% del PIB agroalimentario del país.

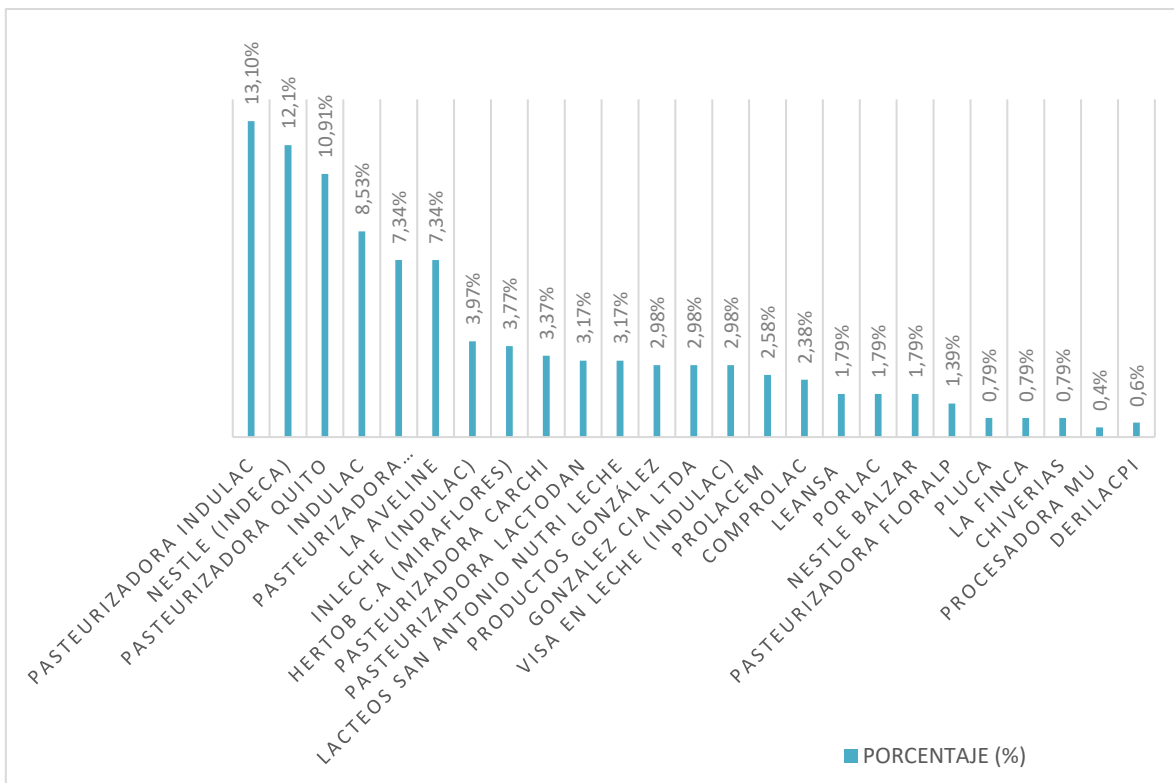
En este contexto, la región Sierra es pionera en la producción de leche, concentrando el 73% de la producción nacional, en segundo lugar, la región Costa con un 18%, la región de la Amazonía con un aporte nacional del 8% y un valor menor al 1% correspondiente a zonas no delimitadas, como se observa en la *Ilustración 1* (MAG, 2020).



La producción de leche en el país ha demostrado ser eficiente para el abastecimiento local y es además fundamental para el sistema económico del país. Este, además, representa una fuente de ingresos para los actores de la red de comercialización, producción y distribución de lácteos, centrándose principalmente en los productores del sector rural que

se encuentra en el país (Bonilla, 2019). Algunas cadenas presentan importantes niveles de organización, por lo cual estas deben llevar procesos controlados para mantener sus estándares de calidad y desarrollar un bien inocuo (Boquero & Paredes, 2010).

De acuerdo a la información obtenida de la Superintendencia de Compañías y Seguros (2017), se registran 65 operadores económicos en la actividad de pasteurizadoras de leche. Los principales actores del sector que se han dedicado a la elaboración de productos lácteos se encuentran en la *Ilustración 2*, se observa el porcentaje correspondiente a las pasteurizadoras nacionales.



Como se puede observar la organización Lácteos San Antonio C.A. es representativa con respecto a las industrias con mayor capacidad de producción en el país, lleva más de 45 años colaborando con alimentos que asisten al bienestar de los ciudadanos. Además, el origen de sus productos considera primordialmente la salud y la naturalidad durante el proceso de fabricación de cada actividad productiva destinada a todo el país.

La empresa Lácteos San Antonio enfoca sus estándares de calidad en un estricto cumplimiento de las normativas vigentes, en donde cada proceso es controlado, partiendo desde la recolección de la leche con el productor ganadero hasta el proceso de empaclado del producto lácteo. Es por ello que para el Departamento de Gestión de Calidad es importante controlar a las variables físico-químicas de la mejor manera para que cumplan con los requisitos de salud e inocuidad de sus productos. En Azuay, la empresa gestiona su producción gracias a las buenas prácticas de manufactura (BPM) por medio de la acreditación de la Norma FSSC 22000, la misma que abarca al sistema agrícola, de producción y de distribución, considerando la inocuidad y los puntos críticos de control (Lácteos San Antonio, 2022).

Para Lácteos San Antonio, una inconformidad en la gestión de calidad del producto genera rechazo por parte de los consumidores, causando pérdidas económicas para la cadena lechera. Todos estos datos son manejados por el Departamento de Gestión de Calidad, quienes llevan registro de las devoluciones por mes, el tiempo de espera para resolver el caso, el recuento del producto por descripción de rechazo y el total de reclamos generados por producto. Con base en esa información se generan indicadores de calidad que reportan la disminución de consumo del producto, representado el 0,5% de las ventas, así como también de pérdidas económicas y de materia prima, generando a su vez un desperfecto en la cadena de suministro y un impacto en el sistema financiero. En el presente estudio se analizó la matriz de quejas y reclamos de Lácteos San Antonio y los indicadores correspondientes, con el fin de identificar las principales razones para el rechazo o devolución de la leche (Lácteos San Antonio, 2022).

La propuesta de un modelo de predicción para las variables físico-químicas que generan devoluciones con bases técnicas de inferencia estadística, permitirá beneficiar a la empresa Lácteos San Antonio. Este estudio permitirá brindar una posible alternativa para la disminución del rechazo hacia el producto a partir de la generación de una herramienta de control, que sea favorable para la productividad de la organización y ofreciendo al mercado un producto inocuo y apto para su consumo.

La empresa mantiene una cantidad numerosa de productos lácteos, en donde su mix de productos más destacado es la leche, la misma viene en diferentes presentaciones, ya sea en Polietileno o Tetra Pak, estos a su vez presentan diferentes procesos de elaboración, lo

que les permite extenderse a Leche Entera, Leche Semidescremada, Leche Descremada y Leche Deslactosada.

En base al estudio se obtuvo el producto con mayor cantidad de devoluciones y se analizaron estadísticamente las variables físico químicas que generan devolución de la leche semidescremada UHT polietileno 1L para la empresa Lácteos San Antonio. El diseño del modelo de predicción se orienta a la prevención del rechazo y mejora de calidad, abarcando aspectos productivos en el sector lácteo. El estudio se define ante las necesidades generadas en la empresa, por lo que es elemental para la mejora del proceso productivo y la estandarización de los límites de control de la industria.

## **2. Estado del Arte**

### **2.1. Generalidades de la Leche Semidescremada**

El análisis realizado por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, 2011) destacan a la calidad y a las grasas que se obtienen de la leche de vaca como indispensables fuentes de nutrientes para la alimentación, puesto que brindan calcio, magnesio, selenio, vitamina B12, entre otros.

La riqueza de nutrientes es relevante y el consumirla aporta beneficiosamente a la dieta de la población. La leche de vaca consiste principalmente de agua en un 87,5%, incluyendo diferentes componentes: 3,5% de Proteínas, 5% Lactosa, Glúcidos 4,8%, Lípidos 3,5%, Prótidos 3,5% y Sales Minerales 0,7% (FAO, 2011).

Los elementos naturales que constituyen la leche se identifican por medio de variables que influyen en mayor proporción la calidad de la leche, entre ellas: el pH, acidez titulable, grasa, sólidos totales, proteína, densidad relativa y el punto de congelación (crioscópico). Cada una de ellas resulta ser necesaria para analizar las condiciones en las que se encuentra el producto (Navas, 2016). El pH representa la acidez o basicidad de la leche, cuando su valor es cercano a 7 se considera que el valor es neutro. Según la norma (NTE INEN 10:2009) para la leche semidescremada a 20°C en proceso se establecen los valores de mínimo y máximo de 6,4 a 6,80, respectivamente; siendo esta una de las variables más importantes de controlar para la inocuidad del producto.

Esta característica de la leche es fundamental en la industrialización de lácteos, esta influye de manera directa en el tratamiento y en la adaptación de los componentes de la leche (García, 2014). Además, el mantener estabilizado el pH permite un sabor, olor y textura

apto para el consumo humano, a diferencia de las medidas que se encuentran sobre el valor neutro del pH que generan una textura arenosa, junto con un sabor amargo de la leche (Torres, 2022).

La calidad del producto terminado depende sustancialmente de las características físico químicas, en donde, un valor de pH que no cumpla de acuerdo a lo establecido a la norma generará un producto no conforme y consecuentemente la devolución o rechazo del mismo.

Además del pH, la acidez estabilizada dentro de su rango también es un actor principal para que el producto final no se pueda alterar. La acidez se mide con los grados de acidez Dornic (°D) y para la leche semidescremada deben mantenerse sus valores entre 0,14 y 0,17 (NTE INEN 10:2009). El monitoreo y control de la acidez permite determinar la calidad de los productos lácteos, pues esta característica físico-química indica el equilibrio de componentes ácidos que contiene la leche que está siendo analizada (Torres, 2022).

El exceso o falta de la acidez en la leche puede causar una desestabilización proteica, causando sabores agrios que no son deseados, así también la coagulación de proteínas, proporcionando texturas ásperas y grumos, adicionalmente el color y el olor característico del producto también son afectados (Ariza, 2017).

En relación a las principales variables físico químicas, la crioscopía es un parámetro que se considera para la liberación del producto y el control de la calidad, se basa en el punto de congelación de la leche en relación con el porcentaje de agua adicionada. Este rango se encuentra entre los valores de 0,560 a - 0,530 (NTE INEN 10:2009). El punto crioscópico altera ciertos factores de la leche que influyen directamente a la generación de la fermentación (Contexto ganadero, 2022). Al presentarse dicha fermentación generará un desequilibrio en la calidad del producto y a su vez si este es liberado presentará rechazo de los lotes afectados.

Consecuentemente, dichas variables críticas afectan a la calidad, siendo valoradas por el cliente a través de las características organolépticas, tanto el sabor, como el color, olor, textura, entre otros (Rosero, 2016). Según un estudio publicado en la revista Food Science and Technology International (Ebrahimi, 2021), la leche semidescremada se fabrica mediante el proceso de centrifugación que parte de la leche entera para lograr separar la grasa y obtener una leche con menor contenido de grasa, varía entre 1,5% y el 2,5%. Dicho

proceso de elaboración puede realizarse en una sola etapa o en varias, para lograr obtener un mayor grado de descremado.

Según un estudio publicado por la revista *Journal of Dairy Science* (Chandan & Kilara, 2015), la leche semidescremada tiene variaciones en propiedades físicas con respecto a la leche entera, por ejemplo, tiene una menor viscosidad debido a su menor contenido de grasa a diferencia de la leche entera. Además, también se menciona que tiene una menor capacidad de retención de aire y una menor capacidad de formación de espuma que la leche entera.

La revista *LWT – Food Science and Technology* (Jung, Lee, & Kim, 2019) publicó un estudio acerca del proceso de elaboración de la leche semidescremada para la estabilidad química. Dicho estudio agregó un beneficio para el almacenamiento del producto, generando durabilidad sin afectar los nutrientes que componen la leche.

## **2.2. Control de procesos y calidad**

Lácteos San Antonio brinda un servicio de calidad, partiendo desde la selección del producto crudo hasta el correcto empaquetado del producto terminado. El Departamento de Gestión de Calidad se encarga principalmente del control de la inocuidad en el proceso de la leche, la cadena láctea debe ser controlada y llevada de manera correcta a través de las Buenas Prácticas de Manufactura (BPM), cumpliendo además con la manipulación higiénica de los productos y el control de los estándares de calidad.

Un producto inocuo para el consumo debe tener las características organolépticas correctas, esto además se debe mantener bajo control para cumplir con un proceso de calidad adecuado, pues no debe poseer residuos o sedimentos que favorezcan a un sabor y color no deseado de la leche; esta a su vez debe contener un nivel de bacterias bajo y mantener la composición de variables físicas químicas apto para el consumo, esto sin hacer uso de sustancias como antibióticos (FAO, 2023).

Es importante controlar durante todo el proceso la calidad de la leche, sometiendo a pruebas físico-químicas, de composición, organolépticas, entre otras. El área de calidad tiene la responsabilidad de realizar dichas pruebas al producto terminado después de haber culminado todo el proceso y concluido el tiempo de liberación (Lácteos San Antonio, 2022). La organización cumple con los requisitos de calidad a través de la implementación de la

norma ISO 9000, la misma que describe a la calidad como un conjunto de características que cumplen con los requisitos planteados con el fin de tener la capacidad para la satisfacción del cliente (ISO 9000, 2015). A su vez, Lácteos San Antonio enfoca su Sistema de Gestión de Calidad (SGC), basado en la norma ISO 9001, esta dirige las actividades mediante la identificación de los objetivos, determinando los procesos y medios necesarios para obtener los fines que se desean en la empresa (ISO 9000, 2015). La planificación de producción, el desarrollo del proceso y la obtención del producto terminado son controlados y direccionados gracias al Software SAP, dicho programa informático integra la gestión empresarial, con el fin de automatizar las diferentes áreas de producción y administración de la empresa, permite la visualización referente de los datos correspondientes de la planta de producción de la leche (Tusa, 2022).

En cuanto a la calidad nutricional, un estudio publicado en la revista *Nutrients* (Crichton, Murphy, & Bryan, 2016), indica que al consumir la leche semidescremada se completa la dieta de ingesta diaria recomendada, esto gracias a que la leche contiene los nutrientes esenciales que son controlados por medio de diversos análisis sensoriales que incluyen el color, olor y sabor; análisis físico-químico y análisis sanitario. La identificación de factores que alteran la calidad de la leche puede generar posibles soluciones para las inconformidades, involucrando nuevas herramientas que reduzcan la pérdida de materia prima y cumplan con los estándares normativos que satisfagan las necesidades del consumidor (Carrisoza, 2022).

El seguimiento y control de los procesos para los productos alimenticios conforma una de las etapas de elaboración más importantes para garantizar la seguridad alimentaria, la calidad del producto y la inocuidad de los alimentos (FAO, 2023). Hoy en día la calidad cumple un papel fundamental para garantizar la seguridad alimentaria y la satisfacción del consumidor, la mejora continua posee diversas técnicas y herramientas para cumplir con los requisitos deseados de cada industria, por lo que se emplean estas para generar un sistema de gestión de calidad (SGC) conforme a las necesidades de proceso de producción del producto (Cannon, 2013).

Una de las herramientas que se utilizan comúnmente en las industrias son los sistemas de control alimentario. En un estudio publicado por el *Journal of Food Processing and Preservation* (Salih & Alghamdi, 2021), se utilizó un enfoque de diseño experimental con el objetivo de cumplir con la inocuidad alimentaria durante la elaboración del queso halloumi. Además de obtener una mejora considerable en tiempos y costos, también se realizó una



mejora significativa en la calidad del queso, garantizando las buenas prácticas de manufactura y mejorando los análisis de calidad.

El uso de las tecnologías avanzadas ha contribuido también en la optimización de la calidad de la leche semidescremada. El artículo de Nongonierma, FitzGerald, Brodkorb, & Hogan (2019), utiliza la tecnología de alimentos, la misma que aplica la ciencia físico-química para la conservación y mejora de la calidad de la leche, a través del análisis de propiedades organolépticas y la composición de este lácteo, generando un resultado significativo para el proceso de control de calidad y contribuyendo a la satisfacción del consumidor. Para la mejora de la calidad de la leche durante un estudio (Pérez, Montilla, & López, 2020), se evaluó la influencia de la adición de diferentes aditivos en la leche semidescremada, resultó una mejora significativa en la textura, formación de espuma en la leche, sin llegar a afectar o generar un desbalance en el pH y en la acidez.

Así también el análisis de las variables físico-químicas de la leche se ha vuelto crucial para determinar la calidad y composición de la leche. Además, el análisis de estas variables permite identificar problemas de contaminación durante el proceso de producción, con lo cual las industrias optan por aplicar las Buenas Prácticas de Manufactura (BPM), así como también las medidas correctivas y la mejora continua para la calidad de la leche. En este sentido, estudios como el de (Sievanen, Huppertz, & Kelly, 2018) destacan la importancia del análisis de variables fisicoquímicas como la grasa, acidez, proteínas, pH y lactosa en la evaluación de la calidad de la leche y en la identificación de posibles contaminantes que afecta a la inocuidad del producto.

### **2.3. Estadística en el control de procesos**

Las técnicas de inferencia estadística son ampliamente utilizadas para la evaluación de la calidad de la leche, así como también en factores que pueden llegar a afectar la salud y seguridad alimentaria de los consumidores (Ross, 2021).

Como primer punto es importante recalcar que la calidad de la leche se puede diagnosticar con la ayuda de diferentes parámetros, como la composición química, higiene o propiedades físicas (Cedeño, y otros, 2015). Dichos parámetros son analizados para conocer la variabilidad de los factores físico-químicos y saber cómo estos afectan a la inocuidad del producto, esto a través de las técnicas de inferencia estadística (FAO, 2019).

Una de las técnicas más utilizadas en la evaluación de la calidad de la leche es el análisis de varianza ANOVA, este corresponde al estudio de variables que dependen de distintos factores, permitiendo reconocer irregularidades, obteniendo comparaciones logísticas y significativas que proporcionan sentido a las hipótesis planteadas de acuerdo al caso de estudio (Ambuludi, Jumbo, Fernández, & Vargas, 2017). La relación entre variables usualmente se reporta mediante gráficos, a través de estos, las variables analizadas presentarán los resultados que generan el análisis para la selección de las relaciones significativas de los factores estudiados (Instituto Valenciano de Investigaciones Agrarias, 2015).

Un estudio realizado por Hassan, El-Nouty y Attia (2017), ha utilizado ANOVA para analizar y determinar la variabilidad en la composición de la leche en diferentes regiones geográficas, permitiendo a su vez diferenciar la composición y los efectos que causa a los parámetros físico químicos como el pH, acidez, grasa y lactosa. Así también, esta herramienta ha sido utilizada para evaluar el efecto de la alimentación en la producción de leche de vaca (Leduc, Souchet, Gele, Provost, & Botinaud, 2021), permitiendo estudiar el balance energético negativo y demostrar la significancia de los cambios en las variables físico químicas.

Una vez obtenida la prueba ANOVA se pueden obtener diferentes tipos de graficas que facilitan la interpretación por medio de la visualización de los resultados obtenidos. Por ejemplo: la gráfica de revisiduos vs valores ajustados, esta indica un diagrama de dispersión para identificar no linealidad, desigualdad en varianzas y valores atípicos (Penn State, 2023). La gráfica de residuos vs apalancamiento se define como un diagrama de diagnóstico el cual permite identificar las observaciones atípicas de un modelo de regresión, por lo tanto, posibilita diferenciar las muestras que generan dispersión en el análisis de varianza (STATOLOGY, 2021).

Otro de los gráficos más comunes es el diagrama de cajas y bigotes, o conocido también como boxplot, este muestra la distribución de los datos de cada grupo analizado y permite comparar la mediana, el rango intercuartílico y los valores extremos entre las diferentes agrupaciones. Los boxplot son útiles para identificar grupos que tienen medias similares pero diferentes varianzas, o también grupos que tienen medias diferentes y una notable superposición de los datos (Myers & Montgomery, 2017).

El análisis Post Hoc es una técnica estadística que se utiliza para comparar las medias de los grupos después de que se hayan encontrado diferencias significativas en la prueba ANOVA, es decir, el análisis Post Hoc determina en qué grupos se difiere estadísticamente (Ortiz, Gómez, Herrera, & Brea, 2017). Añaña, Anjos y Pereira (2017) mencionan en su estudio que existen diferentes técnicas de análisis para el Post Hoc, entre ellas la prueba de Tukey, Scheffé, Bonferroni, Dunnett, Duncan, entre otras. La elección de estas técnicas para el modelo dependerá de los objetivos del estudio y el tipo de muestreo que se realice.

Un estudio realizado por Chaves (2021) utilizó ANOVA y Tukey para analizar la calidad de la leche de vaca en diferentes regiones de Brasil. Otro estudio realizado por Kechichian (2020) hizo uso de la prueba ANOVA y Scheffé para examinar la calidad de la leche cruda y la pasteurizada producida en una industria de lácteos en Argentina. Ambos estudios han presentado resultados importantes para la mejora del producto, así también han generado nuevos procedimientos de acuerdo al sistema de gestión de calidad con el fin de obtener un producto apto para el consumo humano y a su vez la satisfacción de su mercado.

Así entonces, las técnicas de inferencia estadística hoy en día son ampliamente utilizadas en la evaluación de la calidad de la leche y en la identificación de diversos factores que pueden afectar a la composición del producto lácteo.

### **2.3.1. Regresión logística**

Una técnica de inferencia estadística utilizada para reconocer el comportamiento del conjunto de datos y el efecto simultáneo de diferentes variables es el análisis multivariado o la estadística multivariante (Palacio, Apodaca, & Crisci, 2020). En el contexto de la calidad de la leche, se ha utilizado el análisis multivariado a través de la regresión logística, con el fin de evaluar la relación entre higiene y composición de la leche (Kumar, Roy, & Gupta, 2018), así como también para analizar el efecto de las prácticas de manejo en la calidad de la leche (Ghaly, Mahmoud, & El-Zairy, 2019).

Otra herramienta es la regresión logística, utilizada para modelar una probabilidad de un suceso binario, indicando si dicho evento puede darse o no (Rodríguez Mejía, 2022). De acuerdo a un artículo publicado en la revista *Statistical Science* (Hosmer, Lemeshow, & Sturdivant, 2013), la regresión logística tiene diversos campos para ser aplicado, algunos de estos son en medicina, economía, biología, ingeniería industrial y en ciencia social. Además, este artículo menciona que esta técnica se utiliza para crear modelos de

probabilidad de eventos, siendo ejemplos la presencia o ausencia de enfermedades, la supervivencia o mortalidad, y la ocurrencia o no de un evento.

Las técnicas de modelo matemático también son utilizadas para el seguimiento y control de procesos de alimentos. Un estudio realizado recientemente (Gadrich, 2022) desarrolló un modelo de regresión logística simple para predecir el nivel de calidad de la carne de cerdo y conocer si cumplía las especificaciones de acuerdo a sus requisitos. Las variables predictoras utilizadas para el modelo fueron la edad del cerdo, peso y raza, demostrando que el modelo cumplía y se adaptaba a la predicción de la calidad de la carne de cerdo.

De acuerdo a su formulación matemática, la regresión logística se basa en la función logística, la misma que relaciona la probabilidad de un evento con una combinación lineal de variables explicativas. Un estudio realizado en la revista *Annals of Internal Medicine* (Harrell, 2015) indica que la regresión logística puede presentar mejoras en sus modelos a través de la interacción y valores polinómicos, haciendo uso de técnicas de regularización y evaluación con herramientas afines a la formulación dada.

La capacidad para realizar un análisis predictivo ha sido verificada en un sinnúmero de estudios y casos, un estudio publicado en la revista *Journal of Clinical Oncology* (Steyerberg, Wang, van der Helm, Lee, & Wei, 2010) evaluó un modelo predictivo a través de la regresión logística para el cáncer de mama y este obtuvo que el modelo planteado fue capaz de predecir la probabilidad de recurrencia del cáncer hasta con una precisión del 70%.

Para la creación de los modelos de regresión logística se puede implementar diversos paquetes de software estadísticos, tales como RStudio, SAS y SPSS, además, existen programas determinados únicamente para la regresión logística, como es el Stata (Fuentes, Rivera, Pinos, & Zurita, 2012). La regresión logística interviene en diversos estudios de técnicas estadísticas, este determina una estimación para la predicción de acuerdo a la correlación que existe entre las variables dependientes o independientes (Melis, 2019).

Este modelo permite obtener respuestas discretas, teniendo como partida una función exponencial con un mejor ajuste en la curva, siendo aplicable en casos reales de estudio y permitiendo interpretar el modelo de predicción (Peláez, 2016). Uno de los principales objetivos de la regresión logística es configurar la correlación de las variables y determinar

la influencia de estas de acuerdo a la probabilidad de ocurrencia dependiendo de un acontecimiento específico (Berlanga & Vilà Baños, 2014).

### 2.3.2. Regresión logística simple

La regresión logística simple estima la probabilidad de una variable cualitativa binaria categórica en función de una variable cuantitativa, esta permite modelar el logaritmo de la probabilidad de que una variable pertenezca o no a un grupo. Además de obtener una relación significativa, esta calcula la probabilidad de que una variable dependiente corresponda a una de las categorías binarias tomando en consideración el valor que adquiera la variable independiente (Rodrigo, 2016).

El artículo de regresión logística simple y múltiple (Rodrigo, 2016), proporciona la codificación y fórmulas del modelo Logit, indicando lo siguiente: el modelo estima la probabilidad de que la variable dependiente sea igual a 1, es decir,  $Y=1$ , por lo tanto, se pretende encontrar la probabilidad de ocurra cierto suceso. A continuación, se presenta un desglose para la obtención de la fórmula Logit.

Este modelo Logit utiliza la distribución logística estándar acumulativa F.

Dicha clasificación está en función del threshold (límite) que se indique, en donde  $x$  será los valores de su predictor. Por consiguiente, se obtiene la *Ecuación 1*:

$$\Pr(Y = 1 | X_1, X_2, \dots, X_K) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K) \quad (1)$$

La *Ecuación 1* indica que el modelo logit estima la probabilidad de que “Y” sea 1 teniendo las variables  $x_1, x_2$  hasta  $X_K$  y que está representado por la distribución F, expresándolo de la siguiente manera en la *Ecuación 2*:

$$\Pr(Y = k | X = x) = \frac{1}{1 + e^{-(B_0 + B_1 X)}} = \quad (2)$$

Se puede interpretar  $\Pr(Y = k | X = x)$  como la probabilidad de que la variable cualitativa Y tome el valor de K.

Para la regresión logística se hace uso de la función logística, proporcionada en la *Ecuación 3*:

$$p(X) = \frac{e^{B_0 + B_1 X}}{1 + e^{B_0 + B_1 X}} \quad (3)$$

La gráfica de la función logística se presenta en una curva en forma de s, por lo que la variable Y puede comprenderse entre los valores de [0, 1], teniendo la siguiente reestructuración de la *Ecuación 3* en la *Ecuación 4*:

$$e^{B_0+B_1X} = \frac{p(X)}{1-p(X)} \quad (4)$$

Donde  $\frac{p(X)}{1-p(X)}$  corresponde a los odds, los cuales representan la posibilidad de que suceda un acontecimiento de un evento de interés haciendo uso de probabilidades, por lo tanto, se tiene un cociente entre el número de sucesos del evento y de no eventos.

La revista médica de Chile (Cerde, Vera, & Rada, 2013), indica en uno de sus artículos que una ratio puede tomar un valor entre 0 y 1, indicando una baja probabilidad de éxito hasta un valor infinito, con la certeza de que tiene una alta probabilidad de éxito, dejando en claro cuanto más probable es la aceptación que el rechazo.

Continuando con la ecuación, se obtiene una función lineal en la *Ecuación 5*,

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1X \quad (5)$$

$\log\left(\frac{p(X)}{1-p(X)}\right)$  es conocido como logaritmo de odds o logit.

La regresión logística simple es un método que se ha desarrollado en diversas áreas técnicas. En un reciente estudio (Chen, y otros, 2021), se han comparado varios rendimientos de modelos de regresión logísticas para evaluar la predicción de la mortalidad en pacientes hospitalizados por el COVID-19. Dicho estudio encontró que el modelo de regresión logística simple obtuvo un buen nivel de precisión y sensibilidad, con una tasa de acierto del 81%.

Un estudio publicado en el Journal of Education Computing Research (Wang, Chen, & Li, 2020), utilizó un modelo de regresión logística simple para predecir la probabilidad de que los estudiantes se retiren de sus estudios universitarios. El modelo obtuvo como resultado una herramienta efectiva para determinar e identificar a los estudiantes que presentaban mayor riesgo de abandonar sus carreras.

Dentro de lo que conlleva el análisis de la calidad en un producto alimenticio se ha utilizado a la regresión logística como principal herramienta estadística. El estudio publicado por el *Journal of Food Quality* (Zhu & Guo, 2020), emplea un modelo de regresión simple para evaluar la calidad de los granos de café, los resultados obtenidos muestran que el modelo permitió clasificar con precisión los granos de alta calidad.

Para la industria de alimentos la regresión logística ha resultado conveniente para el progreso de ciertas investigaciones, en un estudio publicado por la revista científica *Talanta* (Sardá, Sánchez, & Pons, 2021), se utilizaron sensores para medir la concentración de ácido acético en el vinagre. Se desarrolló modelo de regresión logística simple a partir de los datos recolectados del sensor y se obtuvo como resultado que el modelo tiene la capacidad de detectar con precisión la presencia de ácido acético en el vinagre.

En un estudio reciente se ha demostrado que la regresión logística simple puede ser utilizado también en la evaluación de la calidad de los alimentos, (Khajeh, Mirjalili, & Ale-Ebrahim, 2020) utilizaron este modelo para predecir la calidad del aceite de oliva virgen extra. Se analizaron diversas variables físico químicas del aceite de oliva, como la acidez, índice de peróxidos y la densidad, siendo estas variables las predictoras para la creación del modelo de regresión logística simple. Los resultados indicaron que el modelo creado era capaz de predecir con precisión la calidad del aceite de oliva, con una tasa de acierto del 89%. Dicho estudio demuestra como la regresión logística simple tiene la capacidad de funcionar como una herramienta de evaluación para la calidad de los alimentos.

La regresión logística simple también puede ser utilizada para la predicción de la devolución de productos en diferentes áreas de comercio. Un nuevo estudio publicado en el *Journal of Business Research* (Xiey & Luio, 2021), hizo uso de esta técnica estadística para identificar los factores que mayormente influyen en la probabilidad de devolución de productos en línea de un comercio electrónico. Se consideraron como variables predictoras, el precio del producto, la calidad del producto y la puntuación del vendedor. Se obtuvo como resultado que el modelo era capaz de predecir con precisión la probabilidad de devolución de productos, con una tasa de acierto del 75%.

El identificar las variables físico químicas claves para este caso de estudio definirá el desarrollo del análisis predictivo, las mismas serán analizadas a través de la regresión logística, el cual posibilita definir y describir la relación que existen dichas variables, tanto

dependientes como independientes, permitiendo interpretar las variables significativas para el modelo de predicción.

### **2.3.3. Modelo de predicción**

La industria de lácteos ha optado cada vez más por los modelos de predicción para mejorar la calidad y eficiencia en la producción. Ortega (2023) menciona que estos modelos utilizan técnicas estadísticas y matemáticas para predecir variables de interés, como la composición de la leche, la calidad del queso y la vida útil de los productos lácteos.

Un modelo de predicción utilizado en la industria de lácteos es el modelo de regresión, este se emplea con el interés basado en una o más variables predictoras. Durante el estudio realizado por Golshan Tafti (2021), se utilizó un modelo de regresión para predecir la composición de la leche de vaca utilizando variables como la edad, el periodo de lactancia y el rendimiento lechero.

Así también se han utilizado los modelos de redes neuronales para la interpretación de patrones complejos en los datos y predicción de variables de interés con alta precisión. En el estudio realizado por Jha y otros autores (2021), se utilizó un modelo de redes neuronales para predecir la calidad del queso Cheddar, haciendo uso de variables físico químicas como la temperatura de pasteurización y el pH. El resultado fue un modelo capaz de predecir los parámetros adecuados del queso para mantener estable su composición y mejorar su durabilidad y calidad del producto.

En resumen, se sabe que los modelos de predicción tienen una gran variabilidad de usos en la industria de alimentos, estos permiten mejorar la calidad y la eficiencia en la producción, siendo comúnmente utilizados para predecir variables de interés (Calvo-Brenes, 2013). Por lo tanto, se permite adaptar esta técnica en este caso de estudio, con el fin de obtener una herramienta para el control de calidad de la leche semidescremada.

### **2.3.4. Evaluación del modelo de predicción**

La evaluación de un modelo de predicción es un proceso fundamental para que el aprendizaje automático genere una técnica adecuada que cumpla con los requisitos y parámetros establecidos durante la programación (Aldave, Grassi, & Mantilla, 2023).



Para los modelos de regresión lineal, el coeficiente de determinación  $R^2$  se utiliza para determinar la proporción de varianza en la variable dependiente, la misma que se relaciona con las variables predictoras. Este a su vez explica el modelo, hasta con un valor máximo de 1, permitiendo estimar la capacidad o nivel del alcance que lo defina mejor, es decir, el modelo con mayor valor en el estadístico  $R^2$  será la mejor opción para la construcción del modelo de predicción (Smith & McKenna, 2014).

El pseudo  $R^2$  de McFadden se muestra en la *Ecuación 6*:

$$R_{MF}^2 = 1 - \frac{LL(PROPUESTO)}{LL(NULO)} \tag{6}$$

Donde,

LL(Propuesto)= log-verosimilitud del modelo ajustado (con todas las covariables).

LL(Nulo)= log-verosimilitud del modelo sólo con el intercepto.

Además de conocer los estimados estadísticos para la elección del mejor modelo de regresión, se establecen técnicas o herramientas para la evaluación determinística. Una de ellas es la matriz de confusión, esta detalla el rendimiento del modelo en términos verdaderos positivos, falsos negativos, falsos positivos y verdaderos negativos, como se puede observar en la *Ilustración 3* (López, Avi, & Fernández, 2018).

<b>Valores Predicción</b>	Verdaderos Positivos (VP)	Falsos Positivos (FP)
	Falsos Negativos (FN)	Verdaderos Negativos (VN)
	<b>Valores Reales</b>	

La matriz de confusión es útil para evaluar modelos de clasificación binaria, así como también para la evaluación de modelos de aprendizaje automáticos en diferentes áreas e industrias, como en diagnósticos médicos, detección de fraude o predicciones de variables cualitativas y cuantitativas (Kim & Koh, 2018).

Según (Barrios, 2019) los resultados que se obtienen a partir de la matriz de confusión son:

**Exactitud (Accuracy):** Hace referencia a lo cerca que se encuentra el resultado de una medición del valor verdadero.

**Precisión (Precision):** Indica la dispersión del conjunto de datos obtenidos a partir de mediciones repetidas de una variable. Es importante saber que, entre menor sea dispersión mayor será la precisión.

**Sesgo (Inaccuracy):** Se representa por ser la diferencia entre el valor medio y el valor real de la variable medida; este valor hace referencia al concepto de exactitud.

**Sensibilidad (Recall o Sensitivity):** Es una proporción de los casos positivos que fueron seleccionados e identificados correctamente por el algoritmo.

**Especificidad (Specificity):** Este indica los casos negativos que fueron clasificados correctamente por el algoritmo.

En la *Ilustración 4* se pueden observar las ecuaciones correspondientes de cada parámetro que se obtiene en los resultados de la matriz de confusión. Estas ecuaciones permiten generar una evaluación precisa del algoritmo utilizado en el modelo.

Exactitud	Precisión	Sensibilidad	Especificidad
$\frac{VP + VN}{VP + FP + FN + VN}$	$\frac{VP}{VP + FP}$	$\frac{VP}{VP + FN}$	$\frac{VN}{VN + FP}$

**2.3.5. Herramientas de control de calidad**

Un proceso crítico para asegurar la calidad de los productos y garantizar la satisfacción del cliente es la implementación de las herramientas de control de calidad (Sales, 2013). Se

han desarrollado y mejorado estas herramientas y técnicas con el fin de mejorar los procedimientos y agilizar los controles en diferentes industrias (Solano, 2014).

Una de las herramientas que permite identificar las principales causas de los defectos en la producción de lácteos es el diagrama Pareto, este se caracteriza por ser una de las gráficas analíticas con mejores resultados para la evaluación de los problemas reales, gracias a su principio 80/20, en donde este se interpreta que el 80% de las pérdidas económicas son originadas por el 20% de los productos (Sales, 2013).

Un estudio realizado por Jafari (2016), reveló que los principales factores que contribuyen a los defectos en la producción de productos lácteos son la contaminación microbiana, la falta de higiene y la mala calidad de las materias primas. La aplicación del diagrama Pareto permitió la identificación rápida y efectiva de las causas más importantes de los problemas en la producción, lo que facilitó la implementación de medidas correctivas y preventivas para mejorar la calidad del producto.

La aplicación de otras herramientas como las cartas de control, permiten generar control y monitoreo durante los procesos. Una de ellas es la carta X-bar y R, esta se utiliza para controlar la media y la variabilidad de un proceso. La carta P se utiliza para monitorear la proporción de defectos en una muestra y la carta C se utiliza para supervisar el número de defectos en una muestra (Montgomery D. , 2017).

Un caso de estudio que utilizó la carta de control para dar seguimiento a la calidad de la leche fue realizado por Jaffery (2018) en una planta procesadora de lácteos en Pakistán. En este estudio se aplicó la carta de control X-bar y R para monitorear la calidad de la leche cruda en términos de su contenido de grasa. Los datos fueron recolectados diariamente durante un periodo de tres meses y se analizaron utilizando el software Minitab. Los resultados mostraron que el proceso de producción era estable y que la variabilidad en el contenido de grasa de la leche permitió a la empresa detectar rápidamente cualquier desviación en la calidad de la leche y tomar medidas preventivas para evitar la producción de lotes no conformes.

Para lo que implica la evaluación del modelo de predicción haciendo uso de una herramienta de control de calidad, un estudio realizado por Kourou y otros autores (2015), se propone el uso de una carta de control para evaluar el rendimiento de los modelos de

aprendizaje automático en un modelo de predicción. La carta de control permite monitorear el rendimiento del modelo de pronóstico, detectando cualquier desviación en el modelo.

### 3. Hipótesis y/o preguntas de investigación

- Las variables físico químicas del producto devuelto de leche semidescremada UHT polietileno 1L están correlacionadas e influyen en el rechazo del mismo.
- El uso de regresión logística para las variables físico químicas significativas permite generar un modelo de predicción de devolución del producto, el cual genera un recálculo de las líneas de control de una herramienta de gestión de calidad para el producto leche semidescremada UHT polietileno 1L en la empresa “Lácteos San Antonio”.

### 4. Objetivos

#### 4.1. Objetivo General

Diseñar un modelo de predicción para las variables físico químicas que generan devolución del producto leche semidescremada UHT polietileno 1L, a partir del uso de técnicas de regresión logística para el Área de Gestión de Calidad en el sector lácteo.

#### 4.2. Objetivos específicos

- Identificar las principales causas de devolución del producto leche semidescremada UHT polietileno 1L, y cuáles son sujetas a predicción.
- Establecer la correlación entre las variables físico-químicas significativas del producto lácteo, por medio de técnicas de inferencia estadística, para definir los factores que determinen el modelo de predicción.
- Obtener un modelo de predicción para las variables físico químicas que generan devolución del producto, a partir del uso de técnicas de regresión logística para el área de calidad en el sector lácteo.
- Evaluar el modelo de predicción obtenido en una herramienta de gestión de calidad.

5. Descripción de la Metodología

5.1. Análisis del proceso de devoluciones dentro del Área de Gestión de Calidad de la empresa Lácteos San Antonio.

El Área de Gestión de Calidad cuenta con una base de datos conocida como Matriz de Recepción de Quejas, la cual se dedica al registro exclusivo para el proceso de devoluciones. Es obtenida a través del contacto directo con el cliente, en donde se brindan datos que corresponden al pedido que se ha realizado y su respectivo proceso de devolución. Las etapas que determinan el proceso de devolución se enmarcan en la *Ilustración 5*.

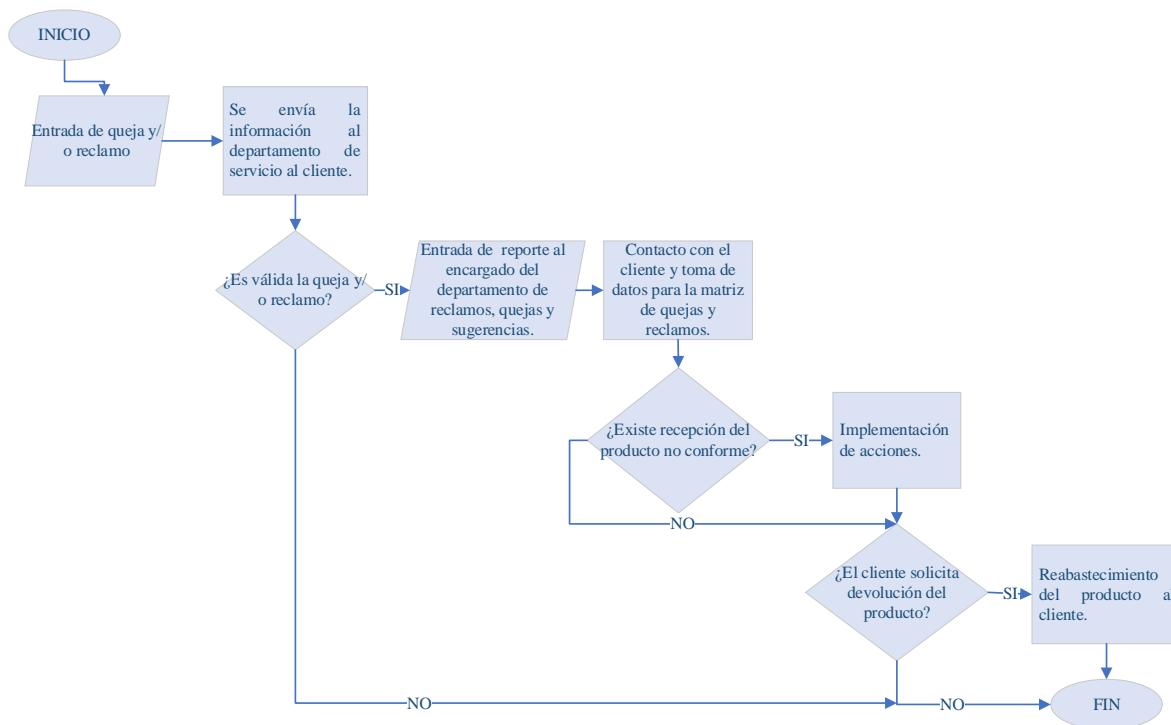


Ilustración 5 Diagrama de flujo para el proceso de recepción de devoluciones. Elaborado por: Paula Lanchimba P.

El uso la matriz de quejas y reclamos para la creación del modelo consistió en el análisis del código del producto, el nombre completo del producto, el lote, la descripción de la devolución y la cantidad de unidades compradas, *Tabla A1* en Anexo I en la página 76. A partir de estas referencias, se realizó la definición de la muestra de estudio, para lo cual fue importante identificar el número de rechazos que se generan por producto y su baja económica para la empresa.

## 5.2. Definición de la muestra de estudio

En primer lugar, se definió el producto con el que se realizó el modelo de predicción, el cual presentó mayor cantidad de devoluciones y rechazo por parte de los consumidores. Se realizó a partir de un estudio no probabilístico intencional, en donde la Matriz de Recepción de Quejas fue el principal objeto de estudio para la obtención de dicho producto. Así también, para validar la selección del producto, se realizó un diagrama Pareto a partir de las pérdidas generadas.

En un inicio se hizo el análisis económico de las devoluciones por producto que la empresa Lácteos San Antonio ha obtenido como registro, los mismos fueron brindados por el Área de Gestión de Calidad, este se elaboró para el periodo mayo – agosto 2022, de acuerdo a la base que fue proporcionada. La información que incluye es el producto, código correspondiente, la cantidad devuelta y las pérdidas financieras, detallado en la *Tabla A3* en Anexo I en la página 77.

## 5.3. Identificación de las principales causas de devolución del producto leche semidescremada UHT polietileno 1L, y cuáles son sujetas a predicción

Una vez elegido el producto leche semidescremada UHT polietileno 1L, se analizaron las causas de devolución, esto a partir de la Matriz de Recepción de Quejas. El código, el producto, el lote, la descripción y las unidades compradas son algunos de los datos que se encuentran en dicha matriz, *Tabla A1* en Anexo I en la página 76, por lo que se realizó una segmentación de acuerdo al producto que fue seleccionado.

Seguidamente, se seleccionaron las causas de rechazo con mayor registro en la matriz. A partir de dichas descripciones se procedió a identificar las tres principales causalidades de devolución con el uso de la *Tabla A2* en Anexo I, de acuerdo a la cantidad en las que estas son mencionadas por los consumidores. Esta simplificación de características de rechazo permitió a su vez identificar las variables físico químicas que se relacionaron directamente con las devoluciones atendidas.

Para la identificación de las variables físico-químicas del producto leche semidescremada UHT polietileno 1L, se usaron los datos que se obtienen a través del software SAP, estos surgen a partir de dos puntos críticos que en el Área de Gestión de Calidad realiza el seguimiento, considerándolas como relevantes para el producto.

Las variables físico-químicas del producto terminado se enlistan a continuación:

- ✓ pH
- ✓ Temperatura
- ✓ % Grasa
- ✓ Acidez °Dornic
- ✓ Crioscopía
- ✓ % Agua
- ✓ % Lactosa
- ✓ Características organolépticas

#### **5.4. Recolección de datos**

Una vez reconocidas las principales causas de rechazo se procedió a la recolección de datos que fueron utilizados para el análisis de varianza ANOVA. Los productos se agruparon en tres conjuntos para la recolección de datos:

- Primer grupo: Devueltos, son aquellos productos que tras su distribución y comercialización el cliente se encuentra desconforme y realiza la queja del producto.
- Segundo grupo: No devueltos, corresponde al grupo de lotes o productos terminado que se encuentran en planta, antes del envasado.
- Tercer grupo: Son los productos comprados que se utilizan para el estudio como grupo de control.

En la *Ilustración 6* se puede apreciar un VSM que indica el proceso de devolución del producto lácteo, en este diagrama se identifican los productos no devueltos como antes y los productos devueltos como después.

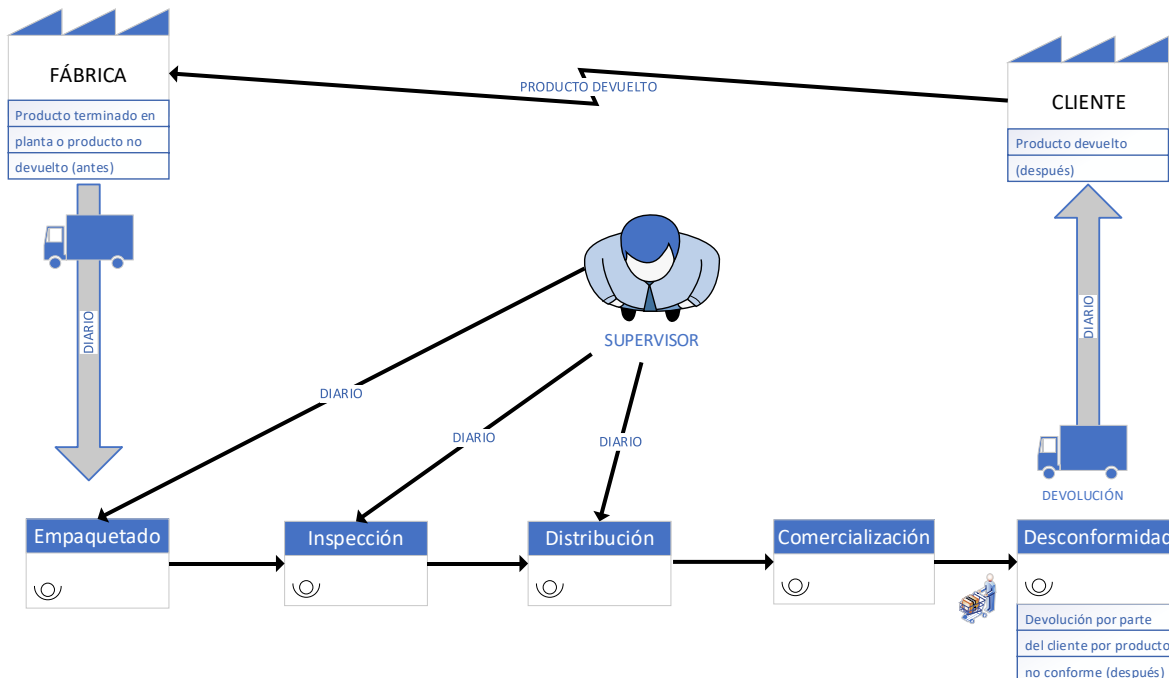


Ilustración 6 Diagrama VSM del proceso de devolución. Elaborado por: Paula Lanchimba P.

Para la recolección de datos se realizó en primer lugar la compra del producto y se realizaron los estudios bromatológicos en unidades que no hayan presentado problemas de devolución, para conocer sus características físico-químicas y analizar la desviación estándar.

Seguidamente, se definió el grupo de los productos devueltos, en donde se eligieron muestras de leche para realizar el correspondiente análisis de componentes, *Ilustración 22* e *Ilustración 23* en Anexo I, tomando en cuenta su fecha de elaboración, fecha de vencimiento y número de lote correspondiente, *Tabla A4* en Anexo I en la página 80.

Las muestras de las devoluciones del producto se obtuvieron directamente en la empresa “Lácteos San Antonio”, para lo cual el departamento de Gestión de Calidad autorizó el análisis bromatológico, estudio completo para conocer las condiciones organolépticas y alteraciones químicas. Los análisis bromatológicos se realizaron en los laboratorios de la Universidad de Cuenca, Universidad del Azuay y Laboratorios MSV.

Para finalizar, los datos del conjunto de producto no devuelto, se utilizó la base de datos de Producto Terminado (PT), que recoge el software SAP, haciendo relación de lotes y fecha de elaboración, permitiendo tener una misma referencia con el grupo de los devueltos.



### 5.5. Definición de los factores que determinan el modelo de predicción mediante técnicas de inferencia estadística

Se realizó en primer lugar el análisis de varianza de un solo factor ANOVA en RStudio, haciendo un análisis comparativo entre los productos devueltos, no devueltos y comprados. Se utilizaron para este análisis los tipos de defectos que proporcionaron mayor devolución, así también los datos de productos comprados y producto terminado (PT); de los cuales se obtuvieron los respectivos valores de pH, acidez y crioscopía, *Tabla A5* en Anexo II en la página 81.

Para este análisis se consideró lo siguiente:

$$H_0: \mu A = \mu B$$

$$H_1: \mu A \neq \mu B$$

Es decir,

Hipótesis nula 0: Las medias son iguales entre las variables independientes, valores de pH, acidez y crioscopía de los productos no devueltos; con respecto a las variables dependientes, valores de pH, acidez y crioscopía de los productos devueltos.

Hipótesis alterna 1: Una media no es igual entre las variables independientes, valores de pH, acidez y crioscopía de los productos no devueltos; con respecto a las variables dependientes, valores de pH, acidez y crioscopía de los productos devueltos.

El análisis de varianza tuvo como finalidad la comparación entre las medias de las diferentes categorías con respecto al pH, acidez y crioscopía, en donde los datos seleccionados para el estudio pertenecen a las observaciones independientes definidas por las condiciones del producto. Por lo tanto, los resultados indicaron las diferencias significativas entre cada relación, proporcionando los factores que determinen el modelo de predicción.

El análisis de varianza se completó con un post hoc en los resultados que indicaban significancia estadística para determinar las diferencias entre grupos e identificar las diferencias entre las muestras independientes del pH, acidez y crioscopía.

## 5.6. Obtención de un modelo de predicción mediante el uso de regresión logística

El modelo de regresión logística se empleó con el fin de encontrar las variables significativas para la predicción de la devolución del producto, para ello se utilizó el programa RStudio. Además, se empleó la información del producto leche, *Tabla A6* en Anexo III en la página 85, se encuentra segmentada por antes, correspondiente a los productos no devueltos que se encuentran en planta como producto terminado y por después, productos devueltos tras la distribución, así también se incluyen los valores de pH y acidez.

Los productos categorizados por antes y después corresponden a las leches que no han sido devueltas y las que han presentado rechazo, correspondientes a un mismo lote, es por ello que se buscó obtener una diferencia significativa que indique la causalidad del rechazo del producto a través de la regresión logística.

Para la creación del modelo de regresión logística, se empezó con la nueva base de datos organizada y se procedió a subirla al software de RStudio, *Tabla A6* en Anexo III en la página 85. Se obtuvo un análisis univariado, con el fin de obtener una retro inspección de cada una de las variables estudiadas por separado.

El modelo matemático es el resultado del modelo de regresión logística, esta ecuación permitió predecir y relacionar las probabilidades para que la leche semidescremada 1L polietileno sea o no devuelta de acuerdo a la variable predictora.

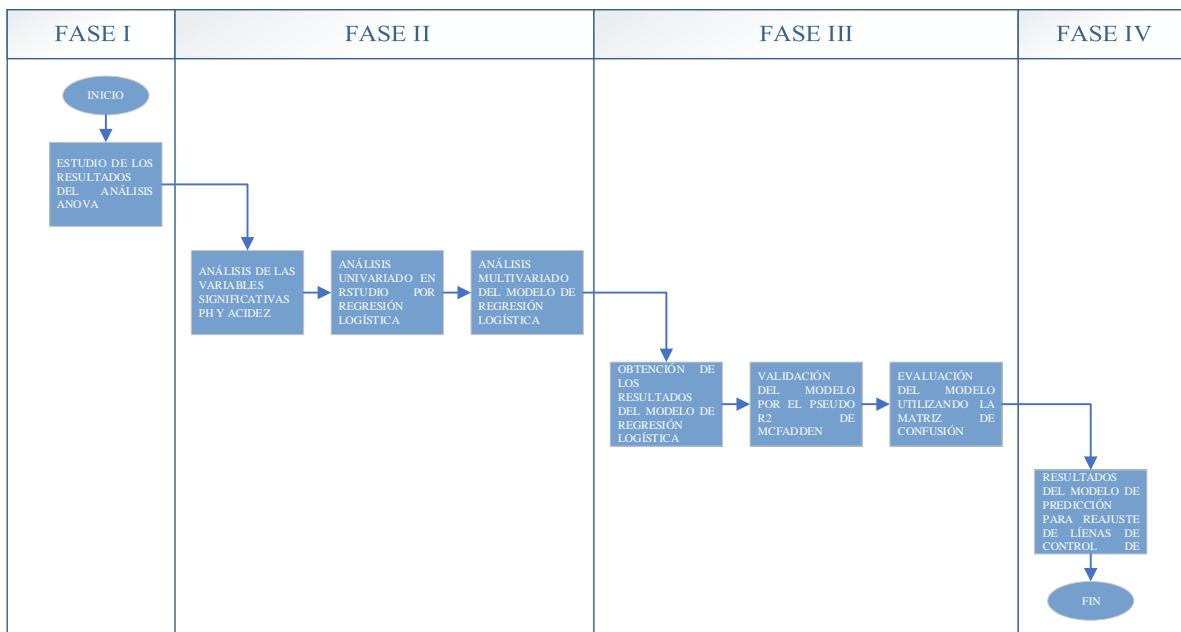
Para la evaluación del modelo en primer lugar se aplicó el método de McFadden, el mismo que consiste en la obtención del valor pseudo  $R^2$ . El pseudo  $R^2$  de McFadden cuantifica la mejora respecto al modelo nulo y la varianza explicada, pero no la correlación entre los valores predichos y los observados. En otros términos, el valor oscila entre 0 y 1, interpretando cuanto explica el modelo los datos, siendo los valores más cercanos a 1, un modelo adecuado para predecir las probabilidades del estudio.

De igual manera, el uso del paquete *CARET (classification and regression training)* para el entrenamiento y prueba de los datos para evaluar el modelo, fue importante para el proceso de creación del modelo predictivo, pues esta función genera un análisis exploratorio a partir de un entrenamiento con el 80% de los datos y un 20% para el test del modelo.

A partir de este paquete se obtuvo la matriz de confusión, la misma que proporcionó los resultados del modelo, es decir, funciona como una herramienta que permite visualizar como se resuelve el algoritmo del modelo. La matriz indicó en una columna el número de predicciones de cada categoría, y en cada fila se representan las instancias en la categoría real, en otras palabras, permitió identificar cuando el modelo genera aciertos y errores cuando se da el entrenamiento de los datos procesados.

La matriz de confusión originó cuatro métricas diferentes, la exactitud, la precisión, la sensibilidad y la especificidad. Por lo tanto, de acuerdo a los resultados obtenidos del modelo, se brindó un resumen analítico de estas métricas, de acuerdo a la devolución de la leche semidescremada en relación su variable predictora.

La *Ilustración 7* representa a través de un diagrama las fases en las que se procedió a realizar el modelo de regresión logística en RStudio, así como también la validación y evaluación del modelo de predicción.



*Ilustración 7 Diagrama de flujo por fases del modelo de predicción. Elaborado por: Paula Lanchimba P.*

### 5.7. Aplicación de la herramienta de gestión de calidad para el reajuste de las líneas de control del producto lácteo.

Se realizaron cartas de control I-MR como herramientas para la gestión de calidad para la leche semidescremada UHT polietileno 1L, esto con la ayuda del desarrollador Minitab, permitiendo conocer los criterios significativos del producto lácteo, tales como los límites de control superior e inferior y valor de  $\bar{x}$  media ( $\bar{X}$ ) de la variable físico química significativa para el modelo de predicción. La gráfica I-MR indicó el valor individual de la variable físico química de cada observación de la base de datos y representó visualmente a las desviaciones estándar de las muestras con respecto a la  $\bar{X}$ , logrando determinar el comportamiento de la variable físico-química de la leche.

La primera carta de control se realizó en base a los datos de la variable físico química significativa del producto terminado, en donde se utilizan como LCI y LCS los valores internos de la empresa, *Tabla A7* en Anexo III en la página 92.

Se estableció un nuevo rango para los límites de control inferior y superior a través de la herramienta de calidad, los valores que se sobresalen del límite de control inferior (LCI) serán los que se deben descartar para cumplir con la meta de devoluciones, la misma que debe ser menor al 0.5% de acuerdo al Departamento de Gestión de Calidad y esto a su vez permitirá hacer válido el modelo.

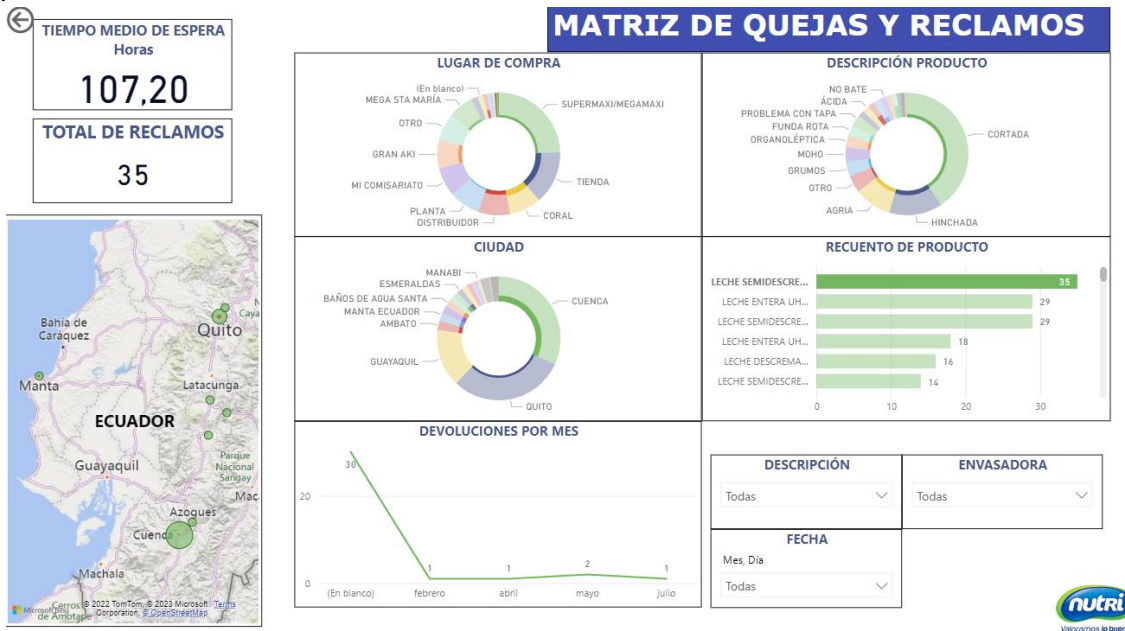
Para la obtención del nuevo LCI de la variable físico química del producto lácteo se reemplazaron los valores que se incluyen dentro del rango actual que maneja la empresa en la ecuación resultante del modelo, logrando verificar cuales son los valores que presentan las probabilidades de devolución más altas y generando un nuevo rango establecido para la variable significativa.

Para establecer el rango se eliminaron los valores que presentaban mayor riesgo de probabilidad de devolución, *Tabla A7* en Anexo III en la página 92, logrando diferenciar en la segunda carta de control, los valores causantes de que exista una cantidad notable de datos que se encontraban fuera del límite de control inferior (LCI).

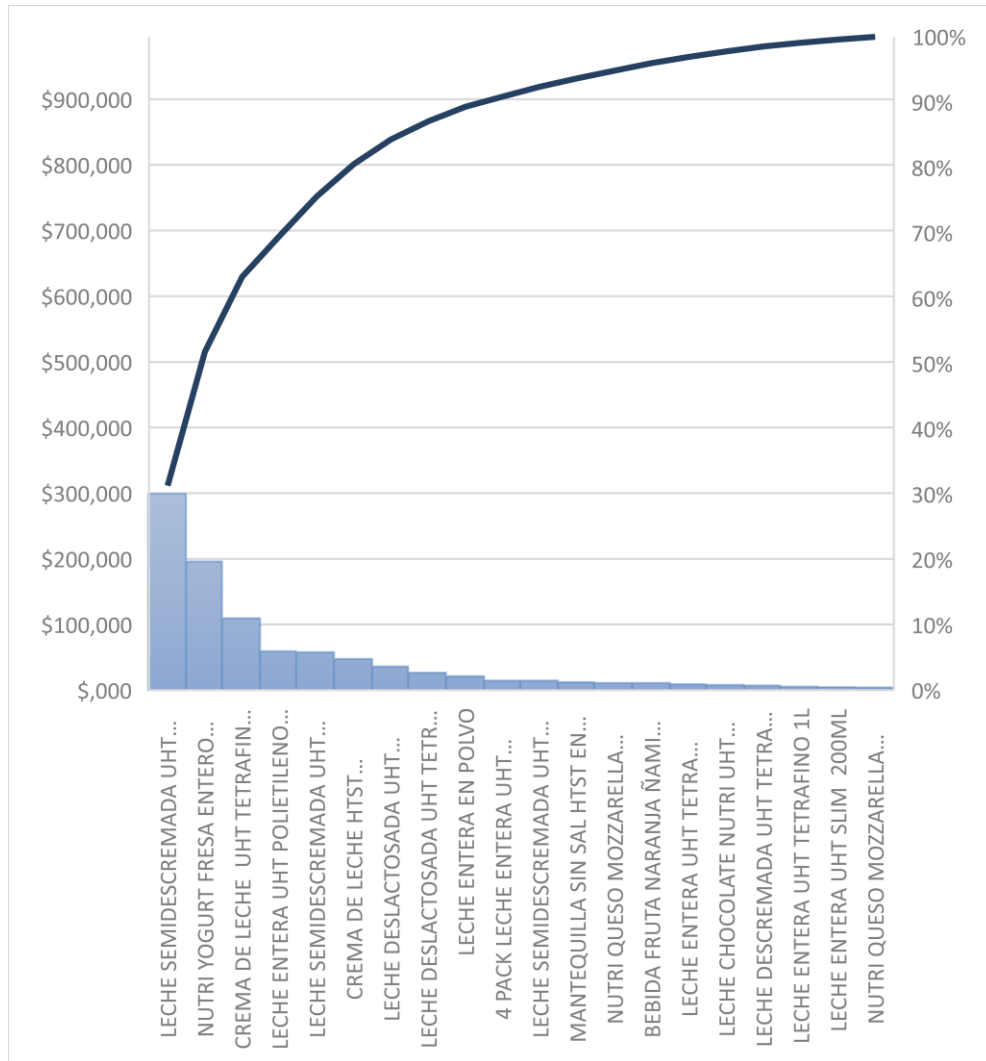
6. Resultados y Discusión

6.1. Definición de la muestra de estudio

Los Indicadores de la Matriz de Quejas y Reclamos, *Ilustración 8*, mostraron información de un cuatrimestre del año 2022, la Leche Semidescremada UHT Polietileno 1L registró un alto índice en las devoluciones con respecto a otros productos lácteos elaborados por la empresa.



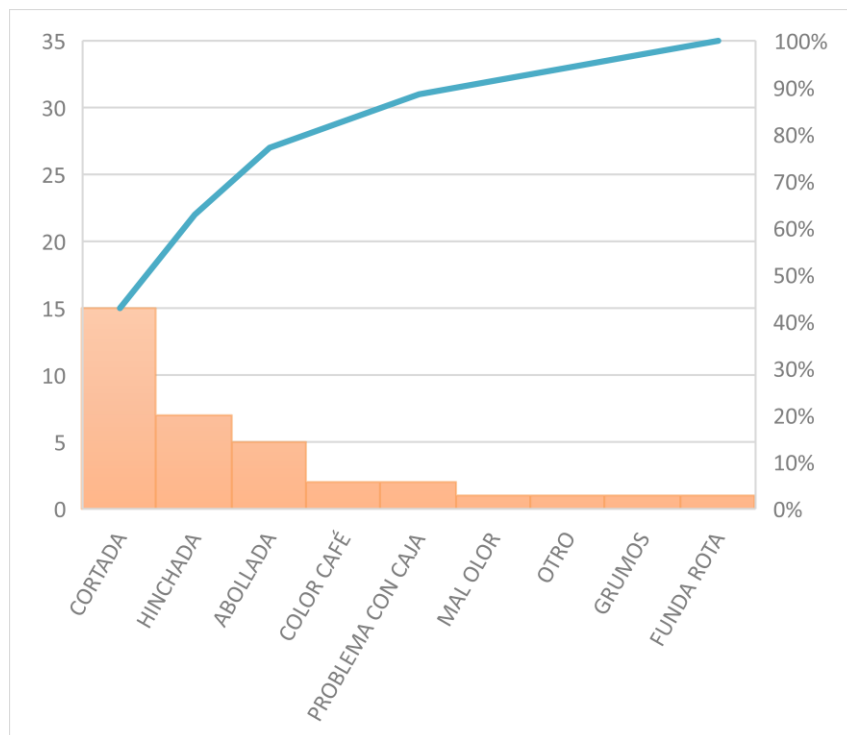
En la *Ilustración 9* se muestra el Diagrama Pareto, este indica que el 80% de los productos devueltos en los que se incluye la leche semidescremada UHT polietileno 1L, Nutri yogurt fresa entero HTST 180G, crema de leche UHT tetrafino 1/2L, leche entera UHT polietileno 1L, leche semidescremada UHT tetrafino 1L y crema de leche HTST pasteurizada 10L, son los que representa mayor pérdida financiera para la compañía, y el 20% indica en menor cantidad las pérdidas económicas por devoluciones.



En otras palabras, se puede decir que, la leche semidescremada UHT polietileno 1L es el producto con mayor número de devoluciones que tiene la empresa, además de tener pérdidas económicas representativas para Lácteos San Antonio. Estos resultados no difieren de otros estudios que han demostrado la existencia de desperdicios y devoluciones de los productos lácteos UHT, donde se han generado hasta un 7,72% de devoluciones trimestrales de la leche UHT en relación a los demás productos lácteos (Granja, 2010).

**6.2. Principales causas de devolución del producto leche semidescremada UHT polietileno 1L**

Una vez seleccionado el producto se realizó un diagrama Pareto para determinar las principales causas por las que el consumidor rechaza el producto. Se empleó la descripción por devolución y las unidades compradas, *Tabla A2* en Anexo I. Con base en el porcentaje de devolución representado en la *Ilustración 10* se determina que las tres causas principales de devolución son leches cortadas, hinchadas y abolladas.



Las variables físico químicas que la empresa considera críticas de acuerdo a las causas identificadas son el pH, acidez °D y crioscopía. Lo mismo sucede al comparar los resultados con estudios que se han examinado la calidad de la leche, en donde, el producto presenta problemas en su mayoría por acidificación, aumentos de alcalinidad y alteración de las propiedades físicas de la leche (González, 2013). Por lo tanto, se determina que las principales variables que afectan a las condiciones de la leche son el pH, Acidez y Crioscopía.

### 6.3. Resultados del análisis bromatológico

A continuación, se presentan los resultados de los análisis bromatológicos de laboratorio para los productos devueltos, no devueltos y comprados:

- **Datos de los análisis bromatológicos de los productos devueltos**

En la *Tabla 1* se puede visualizar el grupo al que corresponde la muestra, en este caso son los productos devueltos o catalogados como después, debido a que las muestras que se eligieron fueron de los productos que se encontraban en distribución y el cliente realizó la devolución del mismo. La tabla se encuentra segmentada por subgrupos, abollada, hinchada y cortada, detallando el nombre del producto, el código de lote y los valores correspondientes al pH, acidez y crioscopía.



Tabla 1 Estudio bromatológico de las variables físico-químicas de los productos devueltos. Fuente: Elaboración propia.

DEVUELTOS (DESPUÉS)	ABOLLADA				
	PRODUCTO	LOTE	PH	ACIDEZ	CRIOSCOPIA
	LECHE SEMIDESCREMADA POLIETILENO 1L	L2213NO	6.1	0.2	-0.563
		L2217NO	6.71	0.156	-0.523
		L2230NO	6.67	0.168	-0.567
		L2230NO	6.69	0.154	-0.524
		L2234NO	6.68	0.171	-0.521
		L2237NO	6.65	0.169	-0.522
		L2234NO	6.72	0.173	-0.526
		L2241NO	6.66	0.171	-0.564
HINCHADA					
PRODUCTO	LOTE	PH	ACIDEZ	CRIOSCOPIA	
LECHE SEMIDESCREMADA POLIETILENO 1L	L2248NO	5	0.9	-0.46	
	L2248NO	4.68	0.772	-0.523	
	L2250NO	5.1	0.892	-0.517	
	L2253NO	4.76	0.901	-0.483	
	L2253NO	4.93	0.872	-0.511	
	L2255NO	5.23	0.901	-0.504	
	L2243NO	4.92	0.893	-0.499	
	L2237NO	5	0.766	-0.487	
CORTADA					
PRODUCTO	LOTE	PH	ACIDEZ	CRIOSCOPIA	
LECHE SEMIDESCREMADA POLIETILENO 1L	L2217NO	4.5	0.8	-0.455	
	L2255NO	4.5	0.767	-0.521	
	L2255NO	4.61	0.792	-0.524	
	L2241NO	4.68	0.801	-0.492	
	L2260LM	4.4	0.794	-0.518	
	L2234NO	4.71	0.756	-0.522	
	L2243NO	4.56	0.8	-0.498	
	L2250NO	4.67	0.796	-0.467	

- **Datos de los análisis bromatológicos de los productos no devueltos**

En la *Tabla 2* se pueden visualizar los resultados del grupo de muestras de los productos no devueltos. La tabla indica el nombre del producto, los valores resultantes del análisis del

pH, acidez y crioscopía y el lote codificado de las mismas muestras de los productos que fueron devueltos.

Tabla 2 Estudio bromatológico de las variables físico-químicas de los productos no devueltos. Fuente: Elaboración propia.

NO DEVUELTOS (ANTES)					
	PRODUCTO	LOTE	PH	ACIDEZ TOTAL	CRIOSCOPIA
	LECHE SEMIDESCREMADA POLIETILENO 1L	L2213NO	6.66	0.135	-0.546
		L2217NO	6.56	0.135	-0.546
		L2230NO	6.63	0.137	-0.545
		L2230NO	6.7	0.131	-0.544
		L2234NO	6.65	0.1351	-0.543
		L2237NO	6.62	0.1393	-0.543
		L2234NO	6.66	0.1373	-0.544
		L2241NO	6.62	0.1373	-0.54
PRODUCTO	LOTE	PH	ACIDEZ TOTAL	CRIOSCOPIA	
LECHE SEMIDESCREMADA POLIETILENO 1L	L2248NO	6.63	0.1391	-0.54	
	L2248NO	6.67	0.136	-0.543	
	L2250NO	6.69	0.133	-0.542	
	L2253NO	6.62	0.138	-0.547	
	L2253NO	6.66	0.136	-0.545	
	L2255NO	6.63	0.1373	-0.545	
	L2243NO	6.66	0.1363	-0.545	
	L2237NO	6.6	0.1393	-0.545	
PRODUCTO	LOTE	PH	ACIDEZ TOTAL	CRIOSCOPIA	
LECHE SEMIDESCREMADA POLIETILENO 1L	L2217NO	6.68	0.133	-0.542	
	L2255NO	6.65	0.138	-0.543	
	L2255NO	6.69	0.135	-0.546	
	L2241NO	6.62	0.138	-0.541	
	L2260LM	6.66	0.1381	-0.537	
	L2234NO	6.64	0.1363	-0.539	
	L2243NO	6.69	0.1343	-0.543	
	L2250NO	6.63	0.1373	-0.54	

- **Datos de los análisis bromatológicos del grupo de control**

En la *Tabla 3* se encuentran los resultados de los análisis de laboratorio de los productos lácteos comprados en percha, esta tabla proporciona el nombre del producto analizado, el número de lote, y los valores de pH, acidez y crioscopía.

*Tabla 3 Estudio bromatológico de las variables físico-químicas de los productos comprados. Fuente: Elaboración propia.*

PRODUCTO	LOTES COMPRADOS			
LECHE SEMIDESCREMADA POLIETILENO 1L	LOTE	PH	ACIDEZ DORNIC	CRIOSCOPIA
	L2300NO	6.6	0.17	-0.545
	L2286NO	6.7	0.17	-0.546
	L2304NO	6.71	0.17	-0.543

Se puede observar que los parámetros de pH para el grupo de leches devueltas disminuyen cuando la causa de devolución es por producto hinchado y cortado, afectando a la seguridad alimentaria y sus propiedades organolépticas, lo que podría causar problemas de salud al consumidor, sin embargo, para los productos no devueltos y comprados el pH se encuentra dentro del valor de la norma interna de la empresa. De acuerdo al estudio de (Alcón, 2015), los valores de pH de la leche cortada concuerdan con los valores obtenidos por los análisis realizados, siendo estos menores a 5,8; generando un aumento de bacterias que actúan en la liberación de ácidos en el estómago de los consumidores, siendo perjudicial la ingesta del producto para su salud.

En el caso de la Acidez Dornic, se observa que, el grupo de los productos no devueltos presenta valores de acidez dentro de los límites de control manejados por la empresa (0,13 – 0,144). Para el grupo de productos devueltos, existe una elevación de los valores de la acidez de hasta 0,9, disminuyendo el valor nutricional por desestabilización de las proteínas. Estos resultados no son diferentes de los estudios que han demostrado que la acidez de la leche genera textura arenosa y áspera al producto cuando su valor es mayor a 0,26 (Díaz, y otros, 2018).

**6.4. Definición de los factores que determinan el modelo de predicción por medio de técnicas de inferencia estadística**

**Análisis ANOVA en RStudio**

El análisis de varianza ANOVA se utilizó como herramienta estadística para comparar las medias de los grupos de muestras entre pH, acidez y crioscopía, obteniendo como primer resultado del análisis de varianza la comparación de medias entre las categorías devuelto y no devuelto en relación a la variable pH. A partir de la programación en RStudio, los resultados de la *Tabla 4* indicaron un valor de  $p < 2 \times 10^{-16}$  que es menor al nivel de significancia alfa ( $\alpha$ ) de 0.05, por lo tanto, se rechaza la hipótesis nula, concluyendo que existe diferencia estadísticamente significativa entre el pH de las muestras devueltas con las no devueltas. Se reconoce al pH como una de las variables predictoras estadísticamente significativas.

*Tabla 4 Resultados ANOVA, pH vs Categoría. Elaborado por: Paula Lanchimba P.*

	Grados de libertad	Suma de cuadrados	Media	Valor crítico para F	Probabilidad
<b>Categoría</b>	4	26.52	6.63	271.6	$< 2 \times 10^{-16}$ ***
<b>Residuales</b>	25	0.61	0.024		

Se realizó la prueba de post hoc con la función `pairwise.t.test()` y los resultados de la *Tabla 5* demostraron que los valores menores a la probabilidad p de 0.05, obtenidos de la prueba estadística entre la variable independiente y la variable dependiente, indicaron que las dos variables están relacionadas, es decir, existe una correlación estadística entre las categorías de devolución de la leche con respecto al pH. Se concluye que no todas las medias de los grupos de muestra entre devueltos y no devueltos son iguales, por lo tanto, se identificaron las categorías que fueron significativas para el modelo de predicción.

*Tabla 5 Resultados prueba post hoc de la variable pH. Elaborado por: Paula Lanchimba P.*

	Abollado	Comprado	Cortado	Hinchado
<b>Comprado</b>	1	-	-	-
<b>Cortado</b>	$< 2 \times 10^{-16}$	$7.1 \times 10^{-16}$	-	-
<b>Hinchado</b>	$< 2 \times 10^{-16}$	$5.2 \times 10^{-14}$	0.00026	-
<b>PT</b>	1	1	$8.4 \times 10^{-16}$	$6.2 \times 10^{-14}$

Los pares que resultaron significativos se obtuvieron por medio de las variables independientes (categoría por devolución) sobre la variable dependiente (pH). La diferencia estadística se generó debido a que los pares entre categorías y pH presentaron medias diferentes. Así también se interpreta que entre comprado y abollado no existen diferencias significativas en el pH, esto se debe a que el pH de los productos comprados no se encontraba afectado.

Por otro lado, se obtuvieron las correlaciones significativas para la acidez vs categoría, *Tabla 6*. Los resultados indican que la categoría tiene un valor p de  $< 2 \times 10^{-16}$ , por lo tanto, la acidez representa una variable predictora al igual que el pH.

*Tabla 6 Resultados ANOVA, Acidez vs Categoría. Elaborado por: Paula Lanchimba P.*

	Grados de libertad	Suma de cuadrados	Media	Valor crítico para F	Probabilidad
<b>Categoría</b>	4	0.0030427	0.0007607	164.9	$< 2 \times 10^{-16}$
<b>Residuales</b>	25	0.0001154	0.0000046		

\*\*\*

En la *Tabla 7* se pueden visualizar los resultados de la prueba de post hoc para conocer las diferencias significativas, se obtiene lo siguiente:

*Tabla 7 Resultados de prueba post hoc de la variable acidez. Elaborado por: Paula Lanchimba P.*

	Abollado	Comprado	Cortado	Hinchado
<b>Comprado</b>	$< 2 \times 10^{-16}$	-	-	-
<b>Cortado</b>	1	$< 2 \times 10^{-16}$	-	-
<b>Hinchado</b>	1	$< 2 \times 10^{-16}$	1	-
<b>PT</b>	1	$3.1 \times 10^{-15}$	1	1

Para la identificación de las diferencias significativas, resultó que la variable dependiente acidez obtuvo diferentes medias entre los grupos, por lo que se aprueba la hipótesis alterna, que menciona que al menos una media no es igual entre las categorías.

Los valores de la acidez del producto abollado y del producto comprado presentan variabilidad, por lo que el análisis de varianza y el post hoc generan esta correlación significativa, así mismo sucede para las siguientes correlaciones resultantes de la variable acidez. Para cada una de estas correlaciones se demuestra por la prueba post hoc que la variable física-química acidez si es significativa para la creación del modelo de predicción.

Para finalizar con el análisis ANOVA se realiza la prueba para la variable de crioscopía. La *Tabla 8* obtenida indica que la categoría tiene un valor p de 0.0871, por lo tanto, al ser un valor mayor que p se demuestra que no existe suficiente evidencia para rechazar la hipótesis de que las medias de la población son todas iguales, en otros términos, la crioscopía no presenta diferencias significativas y por ende no es una variable predictora para la creación del modelo de predicción. Las diferencias entre las medias no son estadísticamente significativas, por lo tanto, no se realiza la prueba post hoc.

*Tabla 8 Resultados ANOVA, Crioscopía vs Categoría. Elaborado por: Paula Lanchimba P.*

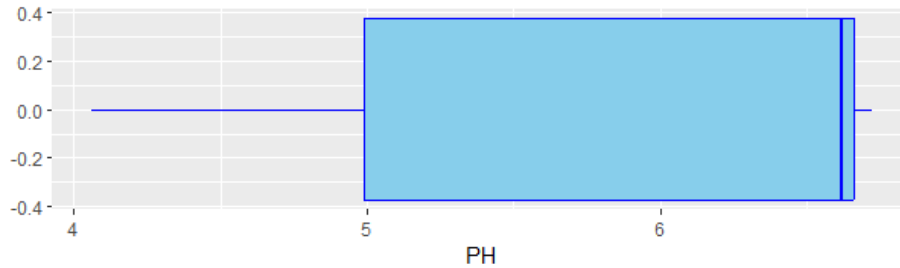
	Grados de libertad	Suma de cuadrados	Media	Valor crítico para F	Probabilidad
<b>Categoría</b>	4	$4.488 \times 10^{-05}$	$1.122 \times 10^{-05}$	2.298	0.0871
<b>Residuales</b>	25	$1.221 \times 10^{-04}$	$4.883 \times 10^{-06}$		

## 6.5. Creación del Modelo de Predicción para las variables físico químicas que generan devolución del producto en RStudio

### 6.5.1. Modelo de Regresión Logística Lineal

En primer lugar, para la creación del modelo de predicción se generó a través del modelo de regresión logística y se estudiaron los datos de ingreso de manera gráfica, permitiendo conocer su distribución estadística.

El diagrama de caja de la *Ilustración 11* indica que la mediana de los datos del pH es 6.63, este valor permite distinguir la distribución de datos que el pH ha presentado durante los análisis de laboratorio del producto terminado, demostrando que existe una amplia segregación del rango límite. Así también, el bigote derecho va desde 6.66 hasta 6.720, mientras que el bigote izquierdo corresponde al valor mínimo de 4.40, siendo este el valor de pH más bajo que fue tomado en las muestras. Demostrando que el valor de pH presenta la tendencia de ser menor a los límites de control.



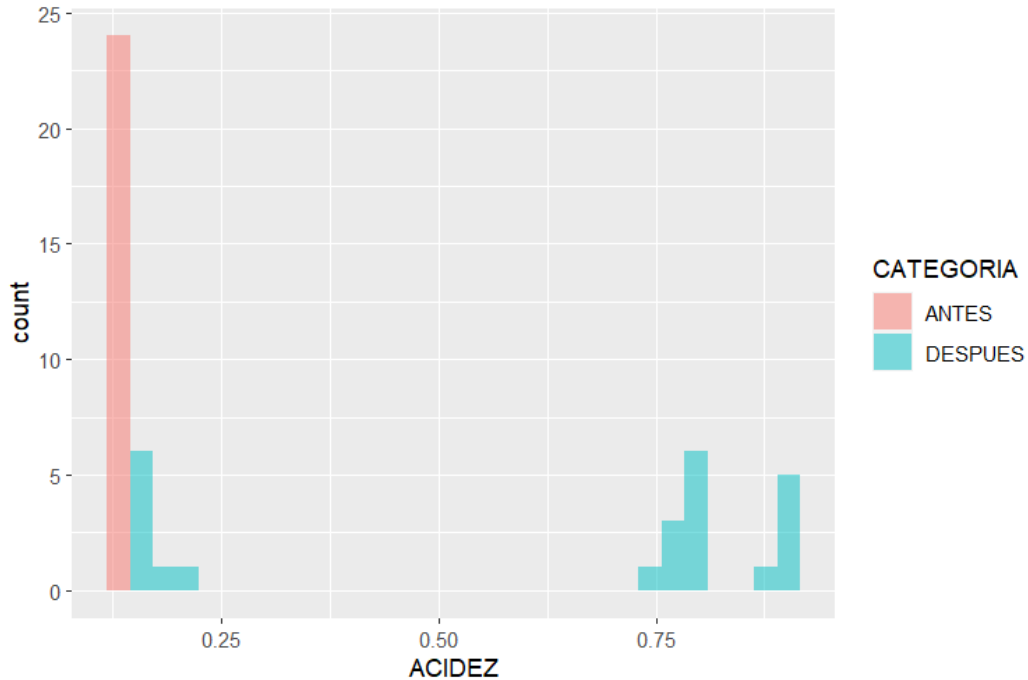
Para el análisis del diagrama de caja de la variable acidez, *Ilustración 12*, se puede apreciar una distribución muy diferente a la del pH, en donde su bigote izquierdo va desde el 0.1310 hasta el 0.1368, la agrupación es muy corta, su bigote derecho va desde 0.777 a 0.9010, lo que demuestra que la acidez presenta una elevación del rango establecido por el Departamento de Gestión de Calidad. La mediana corresponde a un valor de 0.1467, en donde, el 50% que se visualiza con la línea marcada indica la existencia de una concentración muy alta de los datos que van desde el 50 al 75% de la distribución, designando a la acidez como una variable inestable de acuerdo a la condición que se encuentre el producto lácteo.



La gráfica de análisis multivariado, *Ilustración 13*, es una representación visual los productos terminados en planta y los devueltos con respecto a la variable del pH. Se puede apreciar que existen valores de pH menores a 5, mientras que también se aprecia un efecto simultáneo entre el producto no devuelto y el producto devuelto. Existe un rango diferenciable de los valores de pH entre las leches que han sido devueltas y leches que no han sido devueltas. Se determina entonces que la mayoría de productos que han sido devueltos se debe a la variación en el pH.







Se ha realizado además un análisis multivariado numérico vs numérico, es decir, pH vs acidez. Esta técnica estadística generó un contraste univariado, que se centra en una sola variable y permitió analizar como las variables se relacionan entre sí. Además, al ser utilizado para la construcción del modelo de predicción, permitió encontrar la correlación de las asociaciones numéricas, de los grupos de muestras, *Tabla 9*, para la identificación de patrones o tendencias en los datos.

*Tabla 9 Análisis Multivariado, pH vs Acidez. Elaborado por: Paula Lanchimba P.*

ANÁLISIS MULTIVARIADO NUMÉRICO VS NUMÉRICO
<b>CORRELACIÓN PH VS ACIDEZ</b>
-0,9709858

Al obtener este valor se indica que existe una correlación negativa casi perfecta, la cual refleja que el pH y la acidez dependen considerablemente el uno del otro, y esto es correcto debido a que como se había mencionado, el pH es una medición relativa de la acidez. Seguidamente, se obtuvieron los resultados de desviación estadística que generó la programación de la regresión logística:

La *Tabla 10* proporciona los valores de desviación del pH en las muestras de leche devueltas y no devueltas, el extremo mínimo, el cuartil inferior, la mediana, el cuartil superior y los valores de mínimo y máximo.

*Tabla 10 Resultados de desviación. Elaborado por: Paula Lanchimba P.*

Min	1Q	Mediana	3Q	Max
-0,75799	-0,56816	-0,45337	0,00433	2,22354

Los resultados de los valores de la distribución del pH, representados en la *Tabla 11*, indican el valor estimado, tanto del intercepto como del pH, estos valores son los que se utilizaron para generar la ecuación de probabilidad de devolución, la misma que permitió conocer el rango de probabilidades en que el producto puede llegar a ser rechazado por el consumidor. El error estándar, se interpreta como una referencia proporcional de falsos resultados cuando son valores elevados, en este caso, el error estándar de pH no es elevado, por lo tanto, el predictor incrementa la probabilidad de rechazar el producto y este sirvió como contribuyente para el modelo de predicción. El estadístico z o probabilidad (>z) al ser diferente de cero, demostró que la variable predictora generó una ayuda significativa para poder llegar a definir la variable físico-química que causa las devoluciones del producto lácteo.

*Tabla 11 Resultados de los valores de la distribución del pH. Elaborado por: Paula Lanchimba P.*

	Estimado	Error estándar	Valor de z	Probabilidad (>z)	
<b>Intercepto</b>	51.529	21.184	2.432	0.0150	*
<b>pH</b>	-8.023	3.203	-2.505	0.0123	*

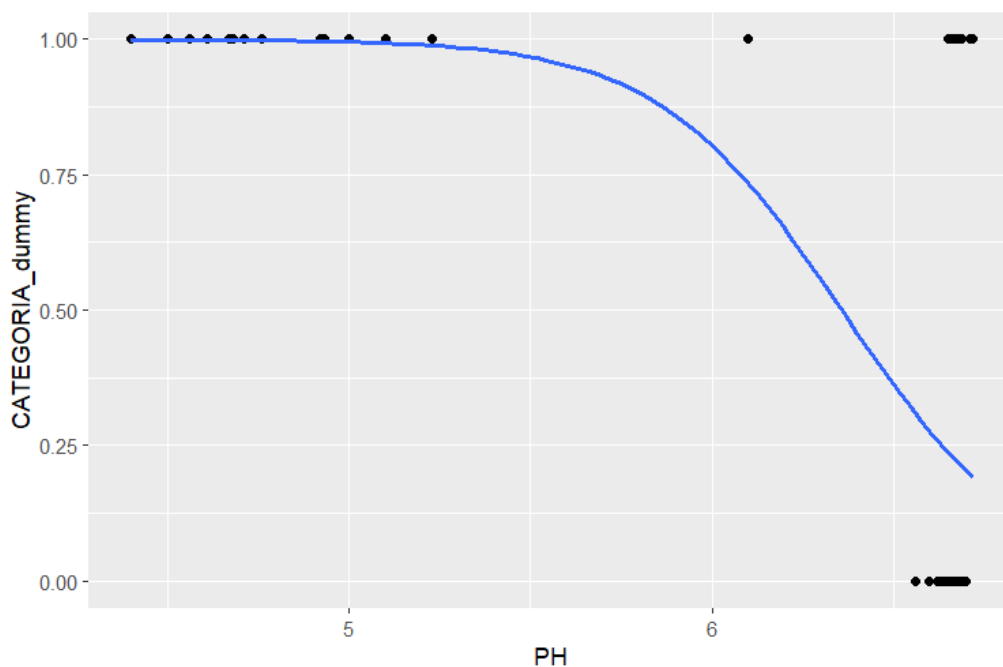
La ecuación resultante del modelo es:

$$\text{logit (Probabilidad de devolución)} = 51.529 - 8.023 * \text{pH} \tag{7}$$

El modelo de regresión obtiene como resultado la significancia estadística tanto del intercepto como de la variable pH. Se entiende que aproximadamente un 90% presenta una probabilidad de que el producto sea devuelto por la variabilidad del pH. En otros términos, el código de significado \* representa que es el pH es representativo para el modelo de regresión.

El logaritmo de odds indica que la probabilidad de ocurrencia de que leche sea devuelta está negativamente relacionada con el pH de la leche (coeficiente parcial = -8.023), en otros términos, el pH al tener una medida menor, mayor será la cantidad de devoluciones del producto, siendo significativa esta relación ( $p\text{-value} = 0.0123$ ) para un nivel de confianza del 10%.

Se determina entonces que la probabilidad para que sean devueltas las leches es más alta cuando se tiene un pH bajo, mientras que para un pH elevado se tiene menor probabilidad de que estas sean devueltas. Sin embargo, en la *Ilustración 15* se puede diferenciar que existen leches devueltas que corresponden al grupo de las leches abolladas que mantienen un pH dentro de los parámetros. Por consiguiente, se estableció que las leches cortadas e hinchadas son devueltas por variaciones del pH. Se debe recordar que 1 corresponde a devuelto o después y 0 a antes o no devuelto.



Para finalizar con el modelo de regresión logística se realizó por el ajuste del modelo nulo para pseudo  $R^2$  de McFadden, *Tabla 12*. El valor que da como resultado 0.6335013 en el Pseudo estadístico  $R^2$  de McFadden, este coeficiente indica que el modelo de regresión logística es fiable para obtener las probabilidades de devolución de la leche semidescremada a través del modelo logístico.

Tabla 12 Resultados del Pseudo Estadístico  $R^2$  de McFadden. Elaborado por: Paula Lanchimba P.

llh	llhNull	G2	McFadden	r2ML	r2CU
-23,6235221	-64,45733113	81,667579	0,6335013	0,5844474	0,779293

El modelo de regresión logística determinó que la predicción es fiable para determinar las probabilidades de devolución de la leche causadas por la variable físico química, incrementando la posibilidad de mejora del control de calidad. La validación del modelo se realizó con el método de McFadden, el mismo que obtuvo un valor de 0,63, considerándolo un valor estadístico aceptable. Así, sucede lo mismo con el estudio de (Gázquez & Pérez, 2017), en donde se considera que el coeficiente de McFadden mayor a 0,4 es un indicador de un ajuste satisfactorio del modelo.

### 6.5.2. Evaluación del modelo de regresión logística

Para la evaluación del modelo se obtuvo la matriz de confusión, *Tabla 13*. La matriz de confusión permitió analizar los resultados del modelo de regresión logística, supervisando y entrenando al modelo para poder conocer como trabajó el algoritmo a través del aprendizaje supervisado, logrando generar una programación confiable para la predicción de la variable físico-química.

Tabla 13 Resultados de la matriz de confusión. Elaborado por: Paula Lanchimba P.

Matriz de confusión		Estimado por el modelo	
		No devuelto (0)	Devuelto (1)
Real	No devuelto (0)	A: (VP)= 9	B: (FP)= 0
	Devuelto (1)	C: (FN)= 1	D: (VN)= 7

Los resultados de la matriz de confusión son representados en la *Tabla 13*, la diagonal principal A igual a 9 y D igual a 7, corresponden con los valores estimados de forma correcta por el modelo, siendo estos los verdaderos negativos (VN) y los verdaderos positivos (VP). La diagonal B igual a 0 y C igual a 1, representan los casos en los que el modelo ha presentado error durante su prueba de algoritmo, falsos positivos (FP) y falsos negativos (FN), respectivamente.

Entonces:

- ✓ Verdaderos positivos (VP): Hay 9 muestras de leche que no son devueltas y se clasificaron como no devueltas.

- ✓ Verdaderos negativos (VN): Hay 7 muestras de leche que son devueltas y fueron clasificadas como devueltas.
- ✓ Falsos positivos (FP): Hay 0 muestras de leche no devueltas que han sido clasificadas incorrectamente como devueltas.
- ✓ Falsos negativos (FN): Hay 1 muestra de leche devuelta y fue clasificada de manera incorrecta como no devuelta.

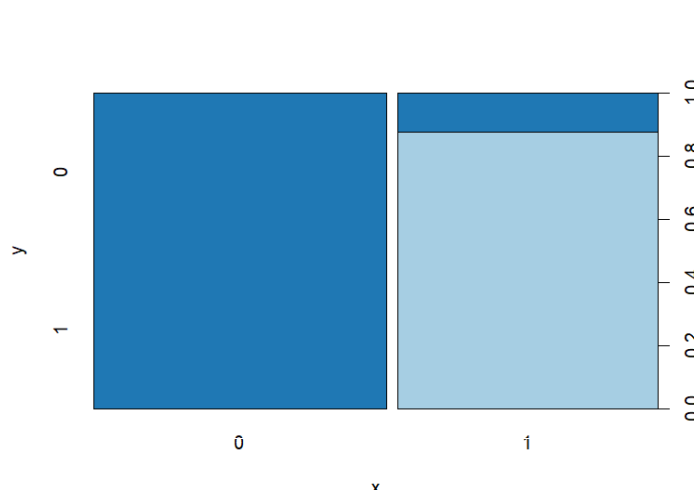
En resumen, la proporción del número de predicciones que son correctas en la matriz presentan un conjunto de datos equilibrado, por lo tanto, estos resultados indican que el desarrollo del algoritmo tiene un desempeño correcto para el modelo de predicción.

La *Tabla 14* presenta los resultados conseguidos y estos demuestran que el modelo es más específico que sensible, es decir, que para predecir la devolución de la leche semidescremada polietileno 1L, se evitan los falsos positivos. Por lo que, para las devoluciones de este producto, gracias al modelo de predicción, no se obtendrán leches rechazadas con un valor de pH que se encuentre dentro del límite establecido, en otros términos, no existirán leches devueltas que tengan el pH adecuado, permitiendo disminuir la cantidad de devoluciones y reducir costos que impliquen la logística de un reproceso o de reabastecimiento del producto al consumidor.

*Tabla 14 Resultados de las métricas. Elaborado por: Paula Lanchimba P.*

Accuracy	Precision	Recall	Specify
0,94117647	1	0,875	1
94,1%	100,0%	88,9%	100,0%

De igual manera, se representan los resultados de las métricas de la matriz de confusión en la *Ilustración 16*. Tanto las leches devueltas como las no devueltas presentan una



predicción correcta, se observa que la sensibilidad o recall tiene un 88,9%, siendo una proporción alta de casos positivos (devueltos) que fueron correctamente identificados por el algoritmo. Así mismo, la precisión y la especificidad de 100% en ambas métricas, proporcionan que los casos negativos (no devueltos) han sido clasificados correctamente. La evaluación del modelo a partir de la matriz de confusión, determinó que el algoritmo de regresión logística obtuvo una precisión del 100% y una exactitud del 94,1%, indicando que el modelo obtuvo la cantidad de predicciones positivas correctas y un alto porcentaje de casos positivos detectados. Esto se podría relacionar con lo que se plantea en el estudio de (Borja & Monleon, 2020), donde indican que la precisión y exactitud evalúan la efectividad general del algoritmo, sin embargo, los autores destacan la desventaja de estas métricas al mencionar que producen menor cantidad de valores discriminatorios para modelos de regresión logística de multivariantes.

Por otro lado, la sensibilidad y la especificidad con valores de 88,9% y 100%, respectivamente, demostraron que el modelo tiene la capacidad de poder detectar e identificar de manera correcta el grupo de leches devueltas y no devueltas. Estos resultados son distintos de otros artículos, que han demostrado que la sensibilidad y la especificidad deben cumplir con el 100% para cada clase para que el modelo sea perfecto y proporcione las predicciones óptimas para el estudio (Alemán, 2021).

#### **6.6. Aplicación de los resultados del modelo de predicción para el recálculo de las líneas de control de una herramienta de gestión de calidad.**

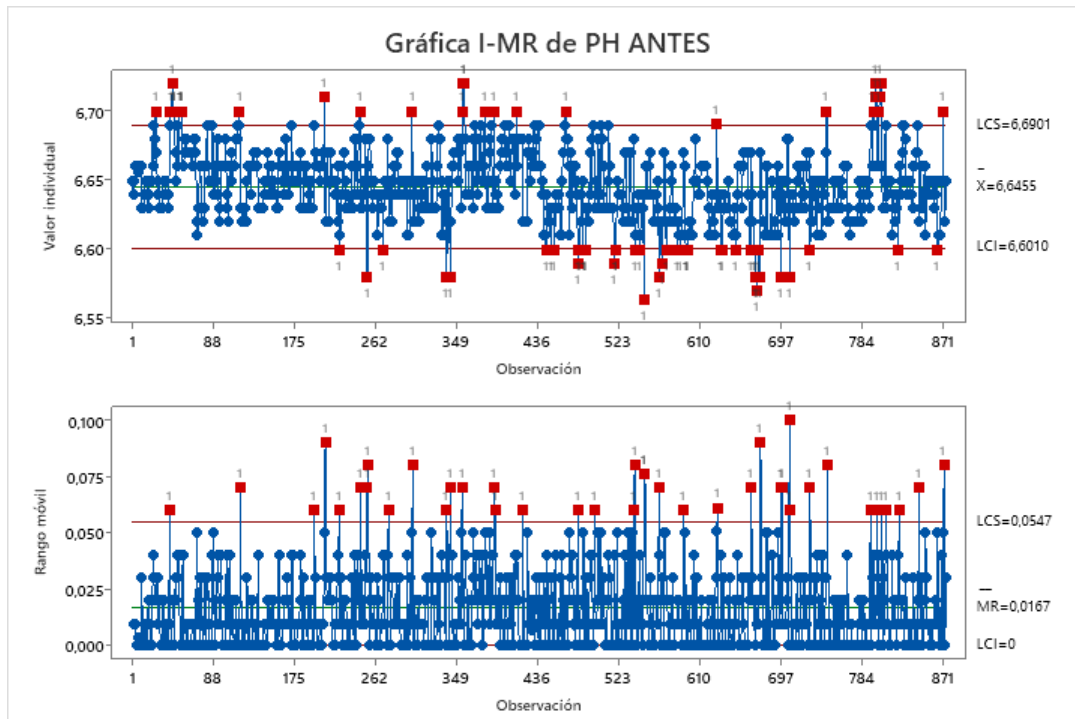
La herramienta de gestión de calidad que se utilizó fueron las cartas de control I-MR, estas se generaron con base a la tabla de datos proporcionada por la empresa, en donde se puede observar el lote de creación y el valor del pH antes y después, *Tabla A7* en Anexo III. A partir de ello se realizaron dos cartas de control, la primera correspondiente al rango de pH actual (antes) que maneja la empresa y la segunda, de acuerdo al rango propuesto por el modelo (después).

En primer lugar, se realizó un análisis a la tabla de valores de productos en proceso y envasado correspondiente a la empresa Lácteos San Antonio. En la *Ilustración 17* se puede visualizar los límites control en tanque y envasado de la variable físico-química pH de la leche semidescremada UHT.

Producto	pH			
	Tanque		Envasado	
	min	max	min	max
Leche Semidescremada*	6,75	6,85	6,55	6,80

Estos valores se rigen con el rango de pH establecido por el Departamento de Gestión de Calidad. Los datos que fueron utilizados para la elaboración de las cartas de control fueron obtenidos de la base central de datos de producto terminado, receptados en el año 2022.

La gráfica I-MR del pH, *Ilustración 18*, con el rango permitido de 6,55 a 6,80, indica los parámetros de Límite de Control Superior (LCS) con un valor de 6.69, y el Límite de Control Inferior (LCI) es 6.60; dando un valor medio de 6.64. A partir de la observación se demuestra que existen causas asignables por lo que se genera una desviación estándar de la línea central, lo que genera a su vez inestabilidad en el proceso de control para la variable del pH, indicando una gran cantidad de valores fuera del LCI y LCS.



*Ilustración 18 Carta de Control pH, rango actual. Elaborado por: Paula Lanchimba P.*

Esto a su vez, permite demostrar que los valores que se encuentran fuera del LCI tienen mayor probabilidad de que los lotes correspondientes sean devueltos o presenten rechazo por el consumidor, por lo tanto, se procedió a realizar la carta de control con respecto a los nuevos límites de control para el pH.

En primer lugar, se obtuvo a través del modelo de regresión logística la ecuación que permitió conocer las probabilidades de devolución de acuerdo a los valores especificados por la empresa, es decir, conocer e identificar desde que valor de pH presentaba mayor probabilidad de rechazo, aplicando la fórmula desde el 6,55 al 6,80.

El modelo además indica, que entre menor sea el valor de pH mayor será la probabilidad de devolución, por lo que se enfocó principalmente en la intervención del límite de control inferior, empezando por 6,55.

Entonces,

El modelo logístico resultante fue:

$$\text{logit (Probabilidad de devolución)} = 51.529 - 8.023 * pH \quad (8)$$

Al despejar la probabilidad de devolución, se obtiene la siguiente ecuación:

$$\text{Probabilidad de devolución} = \frac{e^{51.529-8.023*pH}}{1-e^{51.529-8.023*pH}} \quad (9)$$

Los resultados obtenidos con la ecuación de la regresión logística se indican en la *Tabla 15*. El modelo acierta con las probabilidades de devolución más altas con los valores de pH menores.



Tabla 15 Probabilidades de devolución. Elaborado por: Paula Lanchimba P.

pH	Probabilidad de devolución
6,55	56,25%
6,56	49,76%
6,57	44,22%
6,58	39,47%
6,59	35,35%
6,6	31,76%
6,61	28,61%
6,62	25,83%
6,63	23,38%
6,64	21,19%
6,65	19,24%
6,66	17,50%
6,67	15,94%
6,68	14,53%
6,69	13,26%
6,7	12,11%
6,71	11,08%
6,72	10,14%
6,73	9,28%
6,74	8,51%
6,75	7,80%
6,76	7,15%
6,77	6,57%
6,78	6,03%
6,79	5,54%
6,8	5,09%

Por lo tanto, se procedió a obtener el nuevo rango, para ello se eliminaron los valores de pH desde 6,55 a 6,59, dado a que estos presentan una probabilidad de devolución elevada del producto desde 35,35% a 56,25%, respectivamente. El nuevo rango de pH fue elegido para cumplir con la meta de devoluciones, siendo esta menor a 0,5. Obteniendo los nuevos límites de control de 6,60 a 6,80. El límite de control superior (LCS), fue también parte del análisis, sin embargo, el modelo no presentó probabilidades considerables para la devolución cuando el valor de pH era cercano a 6,80, por lo que no hubo modificación de este límite para la evaluación del modelo.

El objetivo de obtener un nuevo límite de control para los valores de la variable físico-química fue reducir la cantidad de devoluciones del producto lácteo, a través de un ajuste que debe ser leve para que no existan mayores cambios en composición y estructura físico-química del producto, permitiendo cumplir con la meta de devoluciones, por lo tanto, en la *Tabla 16*, se demuestra que eliminando los valores de 6,55 a 6,59 se reduce la probabilidad de devolución a 0,25 %.

*Tabla 16* Indicador de la meta de devoluciones de la Leche Semidescremada UHT polietileno 1L. Fuente: (Lácteos San Antonio, 2023). Elaborado por: Paula Lanchimba P.

PRODUCCIÓN MENSUAL 2023	Mes			Total
	enero	febrero	marzo	
LECHE SEMIDESCREMADA UHT POLIETILENO 1L	862444	860532	666996	2389972
DEVOLUCIONES (LÍMITE DE 6.6 A 6.8)	1741	2581	1602	5924
META DE DEVOLUCIONES (<0.5)	0,20%	0,30%	0,24%	0,25%

La gráfica I-MR del pH, *Ilustración 19*, expone una disminución de los valores que se encontraban fuera del LCI, permitiendo demostrar que el modelo de predicción resuelve de manera correcta el algoritmo, reduciendo una cantidad notoria de productos devueltos y proporcionando una mejora para el proceso de control de calidad, dando de baja a los lotes de pH que tienen altas probabilidades de ser devueltos, en otros términos, para el Departamento de Gestión de Calidad, además de cumplir con la meta de devoluciones, como se demuestra en la *Tabla 16*, se reduciría todo lo que implica el reproceso del producto devuelto, tal como la materia prima, suministros, costos y procedimientos.

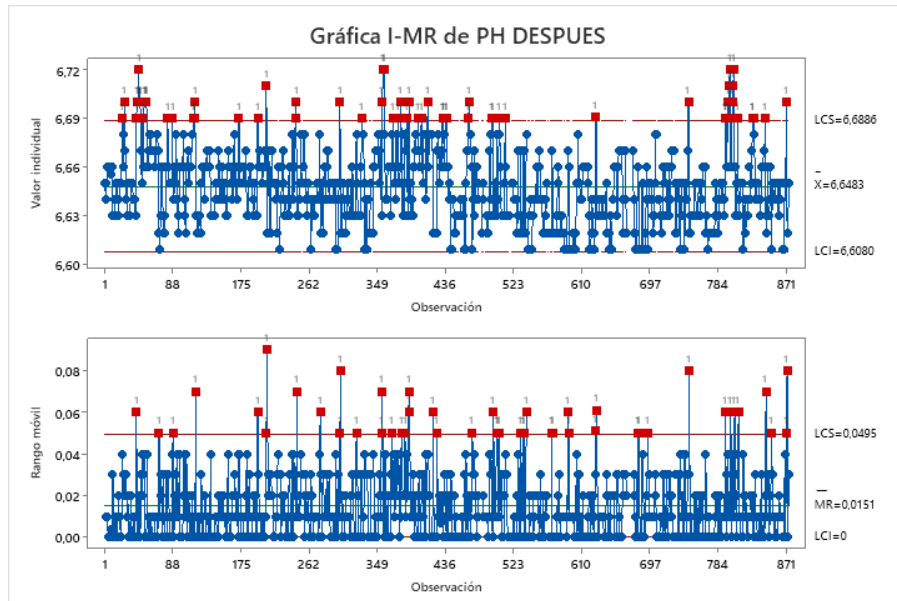


Ilustración 19 Carta de Control pH, rango propuesto. Elaborado por: Paula Lanchimba P.

La evaluación del modelo de predicción mediante las cartas de control, demostró que un ajuste en el límite inferior del valor de pH para la Leche Semidescremada UHT polietileno 1L, favorece a la disminución de devoluciones del producto a la empresa Lácteos San Antonio, siendo este nuevo rango de 6,60 a 6,80. En relación a ciertos estudios, la presencia de lecturas de pH inferiores a 6,60 presentan resultados inestables, planteando problemas de aumento de grasa y alterando la calidad de la leche, por lo que, el valor de pH de una leche fresca suele encontrarse entre 6,60 y 6,80 (López, Barriga, Jara, & Ruz, 2015). Por consiguiente, se estableció que el ajuste del límite de control del pH aporta a la mejora de la calidad de las propiedades de la leche y a la reducción de costos que refiere el proceso de devoluciones.

## 7. Conclusiones

En el presente estudio se analizó de manera detallada la Matriz de Recepción de Quejas y el informe económico de pérdidas del Departamento de Gestión de Calidad de la empresa Lácteos San Antonio. A través del análisis del Diagrama Pareto, se determinó a la Leche Semidescremada UHT polietileno 1L como el producto lácteo que presentó mayor cantidad de devolución y pérdidas económicas para la empresa. Este estudio se complementó con la identificación de las principales causas de devolución de este producto, donde se encontró que las descripciones de devoluciones más frecuentes correspondieron a cortada, hinchada y abollada con el 43%, 20% y 14%, respectivamente. Por lo que, el estudio se centró en las variables físico químicas de estos productos.

La investigación se centró en las variables de pH, Acidez y Crioscopía. Se realizaron análisis bromatológicos de los productos devueltos y no devueltos, y se obtuvieron los valores de pH, Acidez y Crioscopía. Mediante el análisis de varianza ANOVA de las variables físico-químicas, se definieron los factores que determinaron el modelo de predicción. Se concluyó que la variable Crioscopía no resultó ser significativa, a diferencia de las variables físico-químicas, pH y Acidez, que fueron variables influyentes en el proceso del modelo de predicción. De igual manera, se determinó que el pH no afecta a la devolución por abolladura.

Se obtuvo un modelo de predicción usando la regresión logística y se estableció que el pH fue la variable que determina la devolución del producto. Se analizó la ecuación resultante del modelo y se concluyó que los valores de pH que van desde 6,55 a 6,59 generan las probabilidades de devolución más elevadas, 56,25% y 35,35, respectivamente. Además, para este modelo se adicionó la evaluación de la regresión logística por medio de la Matriz de Confusión, en donde se obtuvo una sensibilidad del 88,9%, una precisión de 100% y especificidad de 100%, determinando que los casos fueron correctamente identificados. Se estableció que el modelo realiza una predicción correcta para las leches devueltas y no devueltas con respecto al pH.

Para este estudio se analizaron los límites de control del pH que son establecidos por el Departamento de Gestión de Calidad para la evaluación del modelo de predicción. Mediante la comparación de las Cartas de Control con el límite que maneja la empresa (6,55 a 6,80) y el nuevo límite adquirido por el análisis de las probabilidades de devolución (6,60 a 6,80),

se estableció que existe una reducción en la cantidad de devoluciones del producto causadas por el pH. Por lo tanto, se demostró que reducir el límite inferior de control de pH puede disminuir el porcentaje de devolución del producto de 0,5% a 0,25%, logrando cumplir con la meta de devoluciones que se maneja internamente en la empresa.

La construcción del modelo de predicción ha analizado la posibilidad de disminuir la cantidad de producto devuelto. Esto permite a la empresa tomar medidas preventivas antes de que el producto empiece su distribución y llegue al mercado. De tal manera, se recomienda a la empresa que continúe invirtiendo en la implementación y desarrollo del modelo de predicción como una herramienta para mejorar la calidad del producto y disminuir la cantidad de devoluciones. Así mismo, la empresa debe desarrollar un plan de muestreo para los productos abollados, a fin de identificar la causa de la devolución. Además, la empresa debe considerar no realizar un análisis para la crioscopía en los puntos críticos del producto terminado, debido a que este resulta innecesario para una de las últimas etapas del producto lácteo. En resumen, la ejecución del modelo de predicción es una solución prometedora para optimizar la calidad del producto y reducir los costos asociados a las devoluciones, lo que puede contribuir significativamente al éxito y a la rentabilidad de la empresa.

## Referencias

- Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.
- Alcón, S. (2015). *Aplicación del sistema de control de calidad en química industrial J.Montes Bolivia S.R.L.* La Paz: Universidad Mayor de San Andrés.
- Aldave, H. B., Grassi, J., & Mantilla, H. (2023). Application of a spatial risk model of the crystalline spider mite (*Oligonychus* sp) to avocado crop damage using remote sensing. *Revista Facultad Nacional de Agronomía Medellín*, 76-78.
- Alemán, S. (2021). Análisis de sentimientos para Twitter con Vader y TextBlob. *Revista ODIGOS: Ciencias de la Ingeniería y Ciencias Exactas*.
- Ambuludi, J., Jumbo, N., Fernández, P., & Vargas, J. (2017). Control de calidad de leche cruda en la parroquia Zumbi, provincia de Zamora Chinchipe. . *Revista del colegio de médicos veterinarios del estado de Lara*.
- Ariza, E. (2017). *Evaluación de la estabilidad proteica en la leche y su relación con la calidad microbiológica en los municipios de Tocancipá, Tabio y Zipaquirá*. La Salle: Universidad de La Salle.
- Ayora Ortega, M. E. (2023). *Predicción in silico de los tiempos de retención de antibióticos identificados en muestras de leche mediante UHPLC/ESI Q-Orbitrap*. Cuenca: Universidad del Azuay.
- Barrios, J. (2019, Julio 26). *Big Data*. Retrieved from La matriz de confusión y sus métricas: <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- Berlanga, V., & Vilà Baños, R. (2014). Cómo obtener un modelo de regresión logística binaria con SPSS. *Revista d'Innovació i Recerca en Educació*.
- Bonilla, A. (2019). *Evaluación de las transferencias realizadas por el gobierno central al sector lechero ecuatoriano durante el periodo 2012-2017*. Quito: Pontifica Universidad Católica del Ecuador.
- BOOSTEDML. (2019, Marzo 28). *Articles on Statistics and Machine Learning for Healthcare*. Retrieved from The QQ Plot in Linear Regression: <https://boostedml.com/2019/03/linear-regression-plots-how-to-read-a-qq-plot.html>
- Boquero, M., & Paredes, A. (2010). La Agroindustria ecuatoriana: un sector importante que requiere de una ley que promueva su desarrollo. *LA GRANJA*, 44-46.
- Borja, R., & Monleon, A. (2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Revista Iberica de Sistemas e Tecnologías de Informacao*, 184-196.

- Calvo-Brenes, G. (2013). *Modelo de predicción de la calidad del agua en ríos basado en índices e indicadores del recurso hídrico y el entorno socio ambiental*.
- Cannon, G. (2013). *Food and health: the experts agree. An analysis of one hundred authoritative scientific reports on food, nutrition and public health published throughout the world in thirty years*. Londres: Consumers' Association.
- Carrisoza, I. (2022, Septiembre 20). *Ganadería*. Retrieved from Principales pruebas de calidad de leche, factores que la afectan y cómo corregirlos: <https://www.ganaderia.com/destacado/principales-pruebas-de-calidad-de-leche-factores-que-la-afectan-y-como-corregirlos>
- Cedeño, C., Castañeda, H., Švarc-Gaji, J., Guerra, J., Velázquez, V., & Wolter, W. (2015). *Producción y calidad de la leche*. Sinaloa: Universidad Autónoma de Sinaloa.
- Cerda, J., Vera, C., & Rada, G. (2013). Odds ratio: Theoretical and practical issues. *Revista Médica de Chile. MEDICINA BASADA EN EVIDENCIAS*, 141.
- Chambers, J. M., & Cleveland, W. S. (2017). *Graphical methods for data analysis*. Wadsworth.
- Chandan, R. C., & Kilara, A. (2015). *Dairy ingredients for food processing*. John Wiley & Sons.
- Chaves, M. A. (2021). Variability of milk quality parameters from different regions in Brazil. *Journal of Dairy Science*, 641-649.
- Chen, X., Liu, J., Wang, D., Sun, L., Liu, M., & Zhang, X. (2021). Comparison of logistic regression models for predicting mortality in hospitalized patients with COVID-19. *Journal of Medical Virology*, 93- 96.
- Contexto ganadero. (2022, Enero 06). ¿Sabe qué es el punto crioscópico de la leche y para qué sirve? *Contexto Ganadero*.
- Crichton, G. E., Murphy, K. J., & Bryan, J. (2016). Dairy intake and cognitive health in middle-aged South Australians. *Nutrients*, 8-12.
- da Silva Añaña, E., Dos Anjos, F. A., & de Lima Pereira, M. (2017). Segmentación post-hoc del mercado turístico en los destinos del litoral brasileño: análisis según los valores personales de los turistas. *Estudios y perspectivas en turismo*, 662-677.
- De Marchi, V., Ramanzin, M., Bianchi, L., Rocchia, F., Rosati, A., & Segato, S. (2019). ISO 9001:2015 standard and PDO cheese quality: A case study in Italian dairy sector. *Food Control*, 29-34.
- Díaz, J., Yannelli, L., Vargas, L., Vega, P., Gasull, A., & Paccusse, S. (2018). Acidez Dornic en Leche Humana Donada. *Revista ICU. Investigación, Ciencia y Universidad*.

- Ebrahimi, M. S. (2021). Effect of different centrifugation speeds on some physicochemical and sensory properties of semiskimmed milk. *Food Science & Technology International*.
- Falcón, A., & Serpa, G. (2021). Acerca de los métodos teóricos y empíricos de investigación: significación para la investigación educativa. *Revista Conrado*, 22-31.
- FAO. (2011). *Portal Lácteo*. Organización de las Naciones Unidas .
- FAO. (2019). *Codex 2019: El año de la inocuidad de los alimentos*. Food & Agriculture Org.
- FAO. (2023). *Organización de las Naciones Unidas*. Retrieved from Inocuidad y calidad de los alimentos: <https://www.fao.org/food-safety/food-control-systems/es/>
- FAO. (2023). *Portal Lácteo*. Retrieved from Calidad y evaluación: <https://www.fao.org/dairy-production-products/products/calidad-y-evaluacion/es/>
- Faraway, J. J. (2014). *Linear models with R*. CRC press.
- Fuentes, A., Rivera, N., Pinos, R., & Zurita, G. (2012). *Software estadístico para regresión. El caso de regresión logística y regresión Poisson* . Doctoral dissertation.
- Gadrich, T. (2022). *Ordinal Analysis of Variation of Sensory Responses in Combination with Multinomial Ordered Logistic Regression vs. Chemical Composition: A Case Study of the Quality of a Sausage from Different Producers*. Karmiel: Braude College of Engineering, Department of Industrial Engineering and Management.
- García, C. (2014). Grasa y proteína de la leche de vaca. *DIALNET*, 85-105.
- Gázquez, J., & Pérez, M. (2017). *Consideración de la heterogeneidad en el comportamiento de elección del consumidor a través de modelos logit: enfoque paramétrico vs.semiparamétrico*. Almería: Universidad de Almería. Departamento de Dirección y Gestión de Empresas. Facultad de Ciencias Económicas y Empresariales.
- Ghaly, A. E., Mahmoud, N. S., & El-Zairy, M. R. (2019). Impact of management practices on milk quality of buffaloes in Egypt using multivariate analysis. *Environmental Science and Pollution Research*, 2285-2292.
- Golshan Tafti, A., Toriki-Harchegani, M., & Mirzaei-Alamouti, H. (2021). Modeling and predicting milk composition of dairy cows using the multivariable regression method. . *Journal of Dairy Science*, 69-79.
- González, M. (2013). *Estudio del punto crioscopico de leche cruda bovina, en dos pisos altitudinales y dos épocas del año, Ecuador 2012*. Quito: Universidad Politécnica Salesiana Sede Quito.



- Granja, E. (2010). "IMPLEMENTACIÓN DE UN SISTEMA DE CONTROL DE CALIDAD DE LOS MATERIALES DE ENVASES Y DETERMINACIÓN DEL DESPERDICIO, RENDIMIENTO EN LAS LÍNEAS DE PRODUCCIÓN: UHT (TBA) Y (APP) EN LA PLANTA "LECOCEM – PARMALAT". Riobamba: ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO .
- Harrell, F. E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. *Annals of Internal Medicine* .
- Hassan, A. A., El-Nouty, F. D., & Attia, Y. A. (2017). Geographical variation in milk composition of Egyptian buffalo in relation to animal breed, lactation stage and season. . *Journal of Agriculture and Environmental Sciences*, 47-54.
- Hernández, I. T., Berruecos, J. M., & Z., H. M. (2011). Análisis bromatológico de alimentos empleados como ingredientes en nutrición animal. *Revista Mexicana de ciencia pecuarias*.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación: Selección de la muestra*. México : McGraw-Hill.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
- Hosmer, J., Lemeshow, D. W., & Sturdivant, R. X. (2013). Applied logistic regression. *Statistical Science*.
- INEC. (2020). *La producción de leche en Ecuador*. Quito: Instituto Nacional de Estadísticas y Censos.
- INEN. (2009). *Leche Pasteurizada. Requisitos*. Quito: Instituto Ecuatoriano de Normalización.
- Instituto Valenciano de Investigaciones Agrarias. (2015). Análisis de la varianza. In IVIA, *Introducción al análisis de varianza* (pp. 1-37). Valencia: IVIA.
- Ionita, E. (2022). La producción de leche en Ecuador. *Veterinario Digital*, 1.
- ISO 9000. (2015). *Sistemas de gestión de la calidad: Fundamentos y vocabulario*. Suiza: Plexusmex, S.C.
- Jafari, S., Mirzaei, H., Ahmadi, H., & Mozafari, H. (2016). Using Pareto chart for identifying and prioritizing of factors affecting quality of dairy products: a case study in milk and yogurt production factory. *Journal of Food Quality and Hazards Control*, 104-108.
- Jaffery, S., Javed, M., Ahmad, Z., Ali, N., & Farooq, U. (2018). Implementation of Statistical Process Control to Monitor Fat Content of Raw Milk in Dairy Industry of Pakistan. . *Journal of Food Quality*.

- Jekel, J. F., Katz, D. L., Elmore, J. G., & Wild, D. M. (2014). *Epidemiology, biostatistics, and preventive medicine*. Elsevier Health Sciences.
- Jha, R., Yadav, B. S., Kaur, M., & Kumar, A. (2021). Prediction of Cheddar cheese quality attributes using artificial neural network. *Journal of Dairy Science*, 75-84.
- Jung, J. Y., Lee, N. K., & Kim, Y. (2019). The effects of hydrocolloids on physical and rheological properties of reduced-fat milk. *LWT-Food*.
- Kechichian, V. (2020). Physicochemical characteristics of raw and pasteurized milk produced by a dairy company in Argentina. *Journal of Dairy Science*, 403-413.
- Khajeh, M., Mirjalili, M. A., & Ale-Ebrahim, M. (2020). Predicting extra virgin olive oil quality using logistic regression. *Journal of Food Science and Technology*, 57-60.
- Kim, J., & Koh, Y. (2018). A comparison of cross-validation methods for the estimation of prediction error in multivariate calibration. . *Journal of Chemometrics*, 32-41.
- Kourou, K., Exarchos, T. P., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. . *Computational and structural biotechnology journal*, 8-17.
- Krzywinski, M., & Altman, N. (2014). Points of significance: visualizing samples with box plots. *Nature methods*, 119-120.
- Kumar, S., Roy, B., & Gupta, M. K. (2018). Effect of hygiene on milk quality of dairy cows in and around Mathura, India. . *International Journal of Livestock Research*, 24-28.
- Lácteos San Antonio. (2022). *Matriz de Costos*. Cuenca: Lácteos San Antonio, Área de Gestión de Calidad.
- Lácteos San Antonio. (2022). *Matriz de Recepción de Quejas*. Cuenca.
- Lácteos San Antonio. (2022). *Producto en terminado 2022*. Cuenca: Lácteos San Antonio.
- Lácteos San Antonio. (2022). *Tabla de Valores fisico-químicos de producto en proceso y envasado*. Cuenca: Lácteos San Antonio.
- Lácteos San Antonio. (2023). *Producción mensual Leche Semidescremada UHT polietileno 1L*. Cuenca: Lácteos San Antonio.
- Leduc, A., Souchet, S., Gele, M., Provost, F. L., & Botinaud, M. (2021). Efecto de la restricción de alimentos en la producción de leche de vacas lecheras: una revisión. *Journal of Animal Science*, 1-10.
- López, Á., Barriga, D., Jara, J., & Ruz, J. (2015). *Determinaciones Analíticas en Leche*. Córdoba: Instituto de Investigación y Formación Agraria y Pesquera.

- López, F. J., Avi, J. R., & Fernández, M. V. (2018). Control estricto de matrices de confusión por medio de distribuciones multinomiales. *Geofocus: Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 21-27.
- MAG. (2020). *Resumen Ejecutivo de los Diagnósticos Territoriales del Sector Agrario*. Ministerio de Agricultura y Ganadería . Quito: Coordinación General de Planificación y Gestión Estratégica. .
- Mateu, E., & Casal, J. (2013). Tamaño de la muestra. *Epidem. Med. Prev*, 8-14.
- Melis, J. (2019). *El modelo de regresión log-binomial: una alternativa al modelo de regresión logística en estudios de cohortes y transversales*. Bachelor's thesis, Universitat Politècnica de Catalunya.
- Menard, S. (2016). *Applied logistic regression analysis*. Sage.
- Montgomery, D. (2017). *Introduction to Statistical Quality Control*. John Wiley & Sons.
- Montgomery, D. C. (2017). *Design and analysis of experiments*. John Wiley & Sons.
- Myers, R. H., & Montgomery, D. C. (2017). *Response surface methodology: Process and product optimization using designed experiments (4th ed.)*. John Wiley & Sons.
- Navas, S. (2016). *INFLUENCIA DE LA COMPOSICIÓN DE LA LECHE SOBRE EL RENDIMIENTO QUESERO*. Santa Ana de Coro: Universidad Nacional Experimental Francisco de Miranda .
- Nongonierma, A. B., FitzGerald, R. J., Brodtkorb, A., & Hogan, S. A. (2019). Improving the amino acid profile of semi-skimmed bovine milk through membrane fractionation. . *International Dairy Journal*, 37-44.
- NTE INEN 201:2009. (2009). *LECHE LARGA VIDA REQUISITOS*. Quito: NORMA TÉCNICA ECUATORIANA.
- Ortiz, A., Gómez, S., Herrera, O., & Brea, O. (2017). Inclusión del yogurt artesanal de leche de búfala en el pienso de gallinas ponedoras Isa Brown y su efecto en la producción y calidad del huevo. *Revista Ciencia y Agricultura*.
- Palacio, F. X., Apodaca, M. J., & Crisci, J. V. (2020). *Análisis multivariado para datos biológicos: teoría y su aplicación utilizando el lenguaje R*.
- Peláez, I. M. (2016). Modelos de regresión: lineal simple y regresión logística. *Revista Seden*, 195-214.
- Penn State. (2023, Febrero 15). *Eberly College of Science*. Retrieved from Applied Regression Analysis:  
<https://online.stat.psu.edu/stat462/node/117/#:~:text=When%20conducting%20a%20residual%20analysis,unequal%20error%20variances%2C%20and%20outliers>

- Pérez, A., Montilla, A., & López, R. (2020). Effect of calcium chloride, casein and lactose addition on the quality parameters of semi-skimmed milk. *Foods*, 9-16.
- Rainusso, P. M. (2008). *Metodología para la aplicación del Diseño de Experimentos (DoE) en la industria*. San Sebastián: Servicio de Publicaciones de la Universidad de Navarra ISBN 84-8081-153-6.
- Ríos, Z. (2022, junio 25). Retrieved from Acidez y Grados Dornic: <https://es.scribd.com/document/400767620/Acidez-y-Grados-Dornic>
- Rodrigo, J. A. (2016). *Regresión logística simple y múltiple*. CC BY 4.0.
- Rodríguez Mejía, M. S. (2022). *Detección de ovulación por medio de resistencia eléctrica del moco vaginal "Batersuarg" en vacas productoras de leche en la Hacienda "La Ovejería"*. Provincia del Carchi: UPEC.
- Romero, P. A., Leyvas, G., Castillo, J. G., & Santos, A. (2009). Evaluación de la calidad sanitaria de quesos crema tropical mexicano de la región de Tonalá, Chiapas. *Revista mexicana de ingeniería química*, 111-119.
- Rosero, C. (2016). *ESTANDARIZACIÓN DE LA CALIDAD DEL QUESO FRESCO A TRAVÉS DEL DISEÑO EXPERIMENTAL*. Ibarra: Universidad Técnica del Norte.
- Ross, S. (2021). *Introduction to Probability and Statistics for Engineers and Scientists. Sixth Edition*. London: Elsevier Inc.
- Sales, M. (2013). *Diagrama de Pareto*. EALDE.
- Salih, S. M., & Alghamdi, M. A. (2021). Optimization of halloumi cheese production process by response surface methodology. *Journal of Food Processing and Preservation*, 45-49.
- Sardá, L., Sánchez, A. J., & Pons, A. (2021). Detection of acetic acid in vinegar by combining metal oxide gas sensors and logistic regression. *Talanta*, 224.
- Sievanen, K., Huppertz, T., & Kelly, A. (2018). Influence of added calcium chloride on the heat stability of unconcentrated bovine milk. *International Journal of Dairy Technology*, 151-155.
- Smith, T. J., & McKenna, C. M. (2014). *A Comparison of Logistic Regression Pseudo R2 Indices*. Illinois: Northern Illinois University.
- Solano, L. (2014). *Diagrama de Pareto*. Retrieved from <http://www.quees.info/diagrama-de-pareto.html>.
- STATOLOGY. (2021, Septiembre 7). *Statics, Simplified*. Retrieved from What is a Residuals vs. Leverage Plot? (Definition & Example): <https://www.statology.org/residuals-vs-leverage-plot/>

- Stevens, J. P. (2019). *Applied multivariate statistics for the social sciences*. . Routledge.
- Steyerberg, E. W., Wang, Y., van der Helm, E. W., Lee, K. L., & Wei, L. J. (2010). Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. *Journal of Clinical Oncology*, 28-15.
- Superintendencia de Control del Poder de Mercado. (2017). *Informe del sector lácteo en Ecuador*. Quito: Intendencia de Abogacía de la Competencia.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. . Pearson Education.
- Torres, M. B. (2022). *Evaluación del bicarbonato de sodio en la estabilidad de leche de vaca para elaboración de manjar Tipo I*. Ibarra: Universidad Técnica del Norte.
- Tusa, F. (2022, Junio 13). SAP. (P. Lanchimba, Interviewer)
- Wang, X., Chen, H., & Li, M. (2020). Predicting the probability of college students dropping out using logistic regression. *Journal of Educational Computing Research*, 58-65.
- Xiey, Z., & Luio, Q. (2021). Predicting the probability of product returns in online shopping using logistic regression. *Journal of Business Research*, 741-747.
- Zhu, C., & Guo, W. (2020). Application of near-infrared spectroscopy combined with logistic regression analysis for the quality evaluation of coffee beans. *Journal of Food Quality*, 1-8.

## Anexos

## Anexo I

La *Tabla A1* representa un extracto o resumen de la Matriz de Recepción de Quejas, el mismo que es controlado y ejecutado por el Departamento de Gestión de Calidad para recibir las devoluciones de los productos lácteos que no son conformes por parte de los consumidores, este incluye el código del producto, el nombre del producto, el lote, la descripción de la devolución de producto y la cantidad comprada por el cliente.

*Tabla A 1 Matriz de Recepción de Quejas. Fuente: (Lácteos San Antonio, 2022)*

CÓDIGO	PRODUCTO	LOTE	DESCRIPCIÓN	UNIDAD COMPRADA
60000055	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	L8066	CORTADA	12
60000002	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	L8186	HINCHADA	12
60000311	LECHE ENTERA UHT POLIETILENO 1L	L8193	MOHO	2
60000039	LECHE ENTERA UHT TETRAFINO 1L	L8197	CORTADA	12
60000050	CREMA DE LECHE UHT TETRAFINO 100ML	NO HAY	CORTADA	12
60000002		L8193	CORTADA	6
60000035	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	L8215	MAL OLOR	12
60000034	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	I8224	OTRO	1
60000053	LECHE ENTERA UHT POLIETILENO 1L	I8219	CORTADA	12
60000035	CREMA DE LECHE UHT TETRAFINO 1/2LT	L8253	AGRIA	12
60000051	CREMA DE LECHE HTST PASTEURIZADA 10LT	L8164	CORTADA	12
60000039	NUTRI YOGURT FRESA ENTERO HTST 180G	L8240	CORTADA	12
60000035	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	L8241	PROBLEMA CON TAPA	12
60000038	NUTRI YOGURT FRESA ENTERO HTST 180G	L8239	AGRIA	6
60000052	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	L8275	HINCHADA	12
60000052	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	L8244	CORTADA	12
60000038	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	L8224	HINCHADA	12
60000051	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	L8271	CORTADA	12
60000030	NUTRI YOGURT DURAZNO MIX ENTERO HTST170G	NO HAY	AGRIA	1
60000053	BEBIDA FRUTA NARANJA ÑAMI UHT SQUARE 1L	L8327	CORTADA	4
60000035	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	L8356	CADUCADA	10
60000039	LECHE SEMIDESCREMADA UHT SQUARE 1L	L8225	CORTADA	24
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	L9015	CORTADA	3
60000049	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	L8224	CORTADA	12
60000044	LECHE DESLACTOSADA UHT TETRAFINO 900ML	L98296	CORTADA	12
60000051	CREMA DE LECHE UHT TETRAFINO 1/2LT	L8271	CORTADA	12
60000045	CREMA DE LECHE UHT TETRAFINO 1/2LT	L8300	CORTADA	12
60000057	CREMA DE LECHE UHT TETRAFINO 1/2LT	L9024	HINCHADA	12
60000039	LECHE ENTERA EN POLVO	L8323	CORTADA	12

La *Tabla A2* es un resumen de la Matriz de Recepción de Quejas, obtenido del periodo junio-agosto 2022, en esta tabla se encuentra la descripción o motivo de la devolución del producto y la cantidad correspondiente a cada descripción. Esta información se utilizó para determinar las principales causas por rechazo del producto.

*Tabla A 2 Resumen cantidad devuelta durante el periodo junio - agosto 2022. Fuente: Elaboración propia.*

DESCRIPCIÓN	CANTIDAD
MAL OLOR	1
CORTADA	15
AGRIA	5
COLOR CAFÉ	2
HINCHADA	7
PROBLEMA CON CAJA	2
OTRO	1
GRUMOS	1
FUNDA ROTA	1

La *Tabla A3* proporciona la información del registro económico durante el periodo mayo-agosto 2022 en pérdidas ocasionadas por las devoluciones de los productos lácteos por parte de los clientes hacia la empresa Lácteos San Antonio. Esta tabla posee el código del producto, el nombre del producto, pérdidas financieras por devolución y la cantidad devuelta a la empresa.

*Tabla A 3 Pérdidas financieras durante periodo mayo-agosto. Fuente: (Lácteos San Antonio, 2022)*

CÓDIGO	PRODUCTO	PÉRDIDAS FINANCIERAS	CANTIDAD DEVUELTA
60000002	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$4.44	12
60000039	LECHE ENTERA UHT POLIETILENO 1L	\$6.30	12
60000002	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$5.38	1
60000039	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$0.00	6
60000039	LECHE ENTERA UHT POLIETILENO 1L	\$6.30	6
60000035	CREMA DE LECHE UHT TETRAFINO 1/2LT	\$9.80	14
60000030	CREMA DE LECHE HTST PASTEURIZADA 10LT	\$47.58	6
60000000	NUTRI YOGURT FRESA ENTERO HTST 180G	\$196.00	1
60000215	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$11.04	24
60000215	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$30.56	7
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$1.68	2
60000039	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000223	NUTRI QUESO MOZZARELLA NUTRIFOOD 250g	\$11.04	24



60000281	NUTRI QUESO MOZZARELLA NUTRIFOOD 500G	\$4.20	3
60000280	LECHE ENTERA UHT POLIETILENO 1L	\$5.60	2
60000048	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$5.58	12
60000039	BEBIDA FRUTA NARANJA ÑAMI UHT SQUARE 1L	\$11.04	24
60000094	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$5.11	24
60000045	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000002	LECHE DESLACTOSADA UHT TETRAFINO 900ML	\$12.60	2
60000052	CREMA DE LECHE UHT TETRAFINO 1/2LT	\$4.60	2
60000030	CREMA DE LECHE UHT TETRAFINO 1/2LT	\$95.16	96
60000030	LECHE ENTERA EN POLVO	\$12.60	12
60000304	CREMA DE LECHE UHT TETRAFINO 1/2LT	\$0.00	1
60000038	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$13.44	36
60000055	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000311	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$3.14	2
60000050	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000035	LECHE ENTERA EN POLVO	\$8.76	12
60000053	LECHE ENTERA UHT POLIETILENO 1L	\$6.84	12
60000039	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$20.16	24
60000052	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$5.04	6
60000038	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$0.84	1
60000045	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$5.04	6
60000057	LECHE CHOCOLATE NUTRI UHT SQUARE 1 LT	\$8.16	7
60000138	LECHE DESLACTOSADA UHT TETRAFINO 900ML	\$16.13	4
60000052	LECHE DESLACTOSADA UHT TETRAFINO 900ML	\$7.32	10
60000052	LECHE DESLACTOSADA UHT TETRA SQUARE 1L	\$7.32	6
60000041	LECHE DESLACTOSADA UHT TETRA SQUARE 1L	\$14.64	12
60000041	4 PACK LECHE ENTERA UHT POLIETILENO 1L	\$14.64	12
60000120	LECHE ENTERA UHT TETRA SQUARE 1L	\$9.12	1
60000044	MANTEQUILLA SIN SAL HTST EN BLOQUE 7LB	\$12.24	4
60000004	LECHE ENTERA UHT POLIETILENO 1L	\$10.55	4
60000035	LECHE ENTERA UHT POLIETILENO 1L	\$9.12	12
60000035	LECHE ENTERA UHT POLIETILENO 1L	\$14.64	12
60000035	LECHE ENTERA UHT SLIM 200ML	\$4.56	4
60000036	LECHE DESLACTOSADA UHT TETRA SQUARE 1L	\$4.56	3
60000041	LECHE SEMIDESCREMADA UHT SQUARE 1L	\$14.64	12
60000045	LECHE DESCREMADA UHT TETRA SQUARE 1L	\$7.20	3
60000043	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$6.42	3
60000138	LECHE ENTERA UHT TETRAFINO 1L	\$5.38	1
60000038	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$6.72	6
60000002	LECHE SEMIDESCREMADA UHT TETRAFINO 1L	\$9.72	12
60000039	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.80	12
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12



60000035	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000044	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$3.36	4
60000044	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$20.16	24
60000048	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000055	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$2.52	3
60000057	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$1.68	2
60000042	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$3.36	4
60000043	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$11.76	14
60000180	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000035	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$1.68	2
60000038	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$3.36	4
60000002	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000043	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000042	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000045	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$2.52	3
60000035	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
60000185	LECHE SEMIDESCREMADA UHT POLIETILENO 1L	\$10.08	12
		<b>\$995.25</b>	

La *Ilustración 20* y la *Ilustración 21* son fotografías de las muestras de la leche semidescremada UHT polietileno 1L que han sido devueltas a la empresa, las mismas fueron enviadas a laboratorios para su análisis bromatológico.



La *Tabla A4* corresponde a los datos generales tomados de las leches devueltas que fueron utilizadas como muestra para los análisis de laboratorio. Se presenta el lote, la causa de devolución, la fecha de elaboración y fecha de vencimiento.

*Tabla A 4 Información general de las leches que fueron tomadas para muestra de las devoluciones. Elaborado por: Paula Lanchimba P.*

LOTE	CAUSAS	FECHA DE ELABORACIÓN	FECHA DE VENCIMIENTO
L2213NO	ABOLLADA	01/08/2022	28/10/2022
L2217NO	ABOLLADA	05/08/2022	01/11/2022
L2230NO	ABOLLADA	18/08/2022	14/11/2022
L2230NO	ABOLLADA	18/08/2022	14/11/2022
L2234NO	ABOLLADA	22/08/2022	18/11/2022
L2237NO	ABOLLADA	25/08/2022	21/11/2022
L2234NO	ABOLLADA	22/08/2022	18/11/2022
L2241NO	ABOLLADA	30/08/2022	26/11/2022
L2248NO	HINCHADA	05/09/2022	02/12/2022
L2248NO	HINCHADA	05/09/2022	02/12/2022
L2250NO	HINCHADA	07/09/2022	04/12/2022
L2253NO	HINCHADA	10/09/2022	07/12/2022
L2253NO	HINCHADA	10/09/2022	07/12/2022
L2255NO	HINCHADA	12/09/2022	09/12/2022
L2243NO	HINCHADA	31/08/2022	27/11/2022
L2237NO	HINCHADA	25/08/2022	21/11/2022
L2217NO	CORTADA	05/08/2022	01/11/2022
L2255NO	CORTADA	01/08/2022	28/10/2022
L2255NO	CORTADA	12/09/2022	09/12/2022
L2241NO	CORTADA	30/08/2022	26/11/2022
L2260LM	CORTADA	17/09/2022	14/12/2022
L2234NO	CORTADA	22/08/2022	18/11/2022
L2243NO	CORTADA	31/08/2022	27/11/2022
L2250NO	CORTADA	07/09/2022	04/12/2022

## Anexo II

La *Tabla A5* es la base de datos generada para el análisis de varianza ANOVA realizado en el software de programación RStudio. Los argumentos utilizados son la categoría del producto y los valores de pH, acidez y crioscopía, obtenidos de los resultados de los análisis bromatológicos en laboratorio.

*Tabla A 5 Base de datos para RStudio, prueba ANOVA. Fuente: Elaboración propia.*

<b>Categoría</b>	<b>pH</b>	<b>ACIDEZ</b>	<b>CRIOSCOPIA</b>
Comprado	6,6	0,17	-0,545
Comprado	6,7	0,17	-0,546
Comprado	6,71	0,17	-0,543
PT	6,6	0,1393	-0,545
PT	6,64	0,1393	-0,545
PT	6,69	0,1343	-0,543
Abollado	6,1	0,135	-0,546
Abollado	6,71	0,135	-0,546
Abollado	6,67	0,137	-0,545
Abollado	6,69	0,131	-0,544
Abollado	6,68	0,1351	-0,543
Abollado	6,65	0,1393	-0,543
Abollado	6,72	0,1373	-0,544
Abollado	6,66	0,1373	-0,54
Hinchado	5	0,1391	-0,54
Hinchado	4,68	0,136	-0,543
Hinchado	5,1	0,133	-0,542
Hinchado	4,76	0,138	-0,547
Hinchado	4,93	0,136	-0,545
Hinchado	5,23	0,1373	-0,545
Hinchado	4,92	0,1363	-0,545
Hinchado	5	0,1393	-0,545
Cortado	4,5	0,133	-0,542
Cortado	4,5	0,138	-0,543
Cortado	4,61	0,135	-0,546
Cortado	4,68	0,138	-0,541
Cortado	4,4	0,1381	-0,537
Cortado	4,71	0,1363	-0,539
Cortado	4,56	0,1343	-0,543
Cortado	4,67	0,1373	-0,54

### Código de programación del análisis de varianza ANOVA

Como primer argumento se tuvo la importación de la base de datos, con la función *str* se indica en la consola todos los valores de la tabla base.

Así mismo, la función *aov()* se utilizó para obtener un resumen del modelo de análisis de varianza, el mismo que indica los valores de F y Pr (>F), dichas cifras resultantes permitieron conocer la significancia de las variables físico químicas.

La función *summary* expone de manera general una tabla resumen de la media, mediana, valor crítico para F, suma de cuadrados, promedio de los cuadrados, F y la probabilidad. Además, agrega los códigos de significancia, los mismos que determinan si la variable presenta o no una diferencia significativa estadística.

Se agregó también un *plot()* de cada variable, estos indicaron por medio de graficas los valores atípicos que afectan a la normalidad del modelo y a la varianza, dichos valores son reconocidos por puntos que se rotulan numéricamente con respecto a la tabla base.

La prueba de post hoc se realizó a partir de la función *pairwise.t.test()*, la misma que se generó para que la prueba de post hoc de una cola proporcione las medias que difieren del modelo y a su vez identificar la correlación entre las variables físico químicas que determinando los factores claves para la creación del modelo de predicción.

```
> library(readxl)
> BASE_DATOS <- read_excel("C:/Users/Pau1a/Desktop/INFORMACION TESIS/base
/BASE_DATOS.xlsx")
> str(BASE_DATOS)
tibble [30 × 4] (S3: tbl_df/tbl/data.frame)
 $ Categoría   : chr [1:30] "Comprado" "Comprado" "Comprado" "PT" ...
 $ pH          : num [1:30] 6.6 6.7 6.71 6.6 6.64 6.69 6.1 6.71 6.67 6.69
 ...
 $ ACIDEZ.TOTAL: num [1:30] 0.17 0.17 0.17 0.139 0.139 ...
 $ CRIOSCOPIA  : num [1:30] -0.545 -0.546 -0.543 -0.545 -0.545 -0.543 -0.
546 -0.546 -0.545 -0.544 ...
> anova.pH <- aov(BASE_DATOS$pH ~ BASE_DATOS$Categoría)
> summary(anova.pH)
              Df Sum Sq Mean Sq F value Pr(>F)
BASE_DATOS$Categoría  4  26.52   6.630   271.6 <2e-16 ***
Residuals             25   0.61   0.024
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> plot(anova.pH)
Hit <Return> to see next plot:

# diferencias significativas
pairwise.t.test(BASE_DATOS$pH, BASE_DATOS$Categoría)
```

```
#Post hoc de una cola
```

```
pairwise.t.test(BASE_DATOS$pH, BASE_DATOS$Categoría, alternative="less")
```

Pairwise comparisons using t tests with pooled SD

data: BASE\_DATOS\$pH and BASE\_DATOS\$Categoría

	Abollado	Comprado	Cortado	Hinchado
Comprado	1	-	-	-
Cortado	< 2e-16	3.6e-16	-	-
Hinchado	< 2e-16	3.0e-14	1	-
PT	1	1	1	1

P value adjustment method: holm

```
> pairwise.t.test(BASE_DATOS$pH, BASE_DATOS$Categoría, alternative="greater")
```

Pairwise comparisons using t tests with pooled SD

data: BASE\_DATOS\$pH and BASE\_DATOS\$Categoría

	Abollado	Comprado	Cortado	Hinchado
Comprado	1.00000	-	-	-
Cortado	1.00000	1.00000	-	-
Hinchado	1.00000	1.00000	0.00026	-
PT	1.00000	1.00000	6.0e-16	5.6e-14

P value adjustment method: holm

```
> #ANOVA para la siguiente variable
```

```
> anova.AT<-aov(BASE_DATOS$ACIDEZ.TOTAL ~ BASE_DATOS$Categoría)
```

```
> summary(anova.AT)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
BASE_DATOS\$Categoría	4	0.0030427	0.0007607	164.9	<2e-16 ***
Residuals	25	0.0001154	0.0000046		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> plot(anova.AT)
```

```
Hit <Return> to see next plot:
```

```
#post hoc para saber diferencias significativas
```

```
pairwise.t.test(BASE_DATOS$ACIDEZ.TOTAL, BASE_DATOS$Categoría)
```

Pairwise comparisons using t tests with pooled SD

data: BASE\_DATOS\$ACIDEZ.TOTAL and BASE\_DATOS\$Categoría

	Abollado	Comprado	Cortado	Hinchado
Comprado	< 2e-16	-	-	-
Cortado	1	< 2e-16	-	-
Hinchado	1	< 2e-16	1	-
PT	1	3.1e-15	1	1

P value adjustment method: holm

```
#Abollado y comprado
```

```
#comprado con cortado, con hinchado y con PT
```

```
> pairwise.t.test(BASE_DATOS$ACIDEZ.TOTAL, BASE_DATOS$Categoría, alternative = "less")
```

Pairwise comparisons using t tests with pooled SD

data: BASE\_DATOS\$ACIDEZ.TOTAL and BASE\_DATOS\$Categoría

	Abollado	Comprado	Cortado	Hinchado
Comprado	1	-	-	-
Cortado	1	< 2e-16	-	-
Hinchado	1	< 2e-16	1	-
PT	1	1.8e-15	1	1

P value adjustment method: holm

```
> pairwise.t.test(BASE_DATOS$ACIDEZ.TOTAL, BASE_DATOS$Categoría, alternative = "greater")
```

Pairwise comparisons using t tests with pooled SD

data: BASE\_DATOS\$ACIDEZ.TOTAL and BASE\_DATOS\$Categoría

	Abollado	Comprado	Cortado	Hinchado
Comprado	<2e-16	-	-	-
Cortado	1	1	-	-
Hinchado	1	1	1	-
PT	1	1	1	1

P value adjustment method: holm

```
> #ANOVA
```

```
> anova.C<- aov(BASE_DATOS$CRIOSCOPIA ~ BASE_DATOS$Categoría)
```

```
> summary(anova.C)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
BASE_DATOS\$Categoría	4	4.488e-05	1.122e-05	2.298	0.0871
Residuals	25	1.221e-04	4.883e-06		

```
----
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#no existen diferencias entre ningún par, por lo tanto, no se hace post hoc

```
> plot(anova.C)
```

```
Hit <Return> to see next plot:
```

```
Hit <Return> to see next plot:
```

```
Hit <Return> to see next plot:
```

```
Hit <Return> to see next plot:
```

```
#REPRESENTACION GRAFICA MODELO BOX PLOT
```

```
> library(ggplot2)
```

```
> #PH
```

```
> ggplot(data = BASE_DATOS, aes(x = Categoría, y = pH, color = Categoría)) +
```

```
+ geom_boxplot() +
```

```
+ theme_bw()
```

```
> #ACIDEZ
```

```
> ggplot(data = BASE_DATOS, aes(x = Categoría, y = ACIDEZ.TOTAL, color = Categoría)) +
```

```
+ geom_boxplot() +
```

```
+ theme_bw()
```

```
> #CRIOSCOPIA
```

```
> ggplot(data = BASE_DATOS, aes(x = Categoría, y = CRIOSCOPIA, color = Categoría)) +
```

```
+ geom_boxplot() +
```

```
+ theme_bw()
```

## Anexo III

La *Tabla A6* segmenta la categoría de las leches en antes (no devueltas) y después (devueltas), así mismo, se indican los valores de pH y acidez correspondientes a cada muestra.

*Tabla A 6 Base de datos para Regresión Logística en RStudio. Fuente: Elaboración propia.*

CATEGORIA	PH	ACIDEZ
DESPUÉS	6,1	0,2
DESPUÉS	6,71	0,156
DESPUÉS	6,67	0,168
DESPUÉS	6,69	0,154
DESPUÉS	6,68	0,171
DESPUÉS	6,65	0,169
DESPUÉS	6,72	0,173
DESPUÉS	6,66	0,171
DESPUÉS	5	0,9
DESPUÉS	4,68	0,772
DESPUÉS	5,1	0,892
DESPUÉS	4,76	0,901
DESPUÉS	4,93	0,872
DESPUÉS	5,23	0,901
DESPUÉS	4,92	0,893
DESPUÉS	5	0,766
DESPUÉS	4,5	0,8
DESPUÉS	4,5	0,767
DESPUÉS	4,61	0,792
DESPUÉS	4,68	0,801
DESPUÉS	4,4	0,794
DESPUÉS	4,71	0,756
DESPUÉS	4,56	0,8
DESPUÉS	4,67	0,796
ANTES	6,66	0,135
ANTES	6,56	0,135
ANTES	6,63	0,137
ANTES	6,7	0,131
ANTES	6,65	0,1351
ANTES	6,62	0,1393
ANTES	6,66	0,1373
ANTES	6,62	0,1373
ANTES	6,63	0,1391
ANTES	6,67	0,136
ANTES	6,69	0,133

ANTES	6,62	0,138
ANTES	6,66	0,136
ANTES	6,63	0,1373
ANTES	6,66	0,1363
ANTES	6,6	0,1393
ANTES	6,68	0,133
ANTES	6,65	0,138
ANTES	6,69	0,135
ANTES	6,62	0,138
ANTES	6,66	0,1381
ANTES	6,64	0,1363
ANTES	6,69	0,1343
ANTES	6,63	0,1373

### Modelo de regresión logística

La realización del modelo de regresión logística se realizó en el programa RStudio, se ingresó la base de datos, y seguidamente se procedió a estructurar el algoritmo, haciendo uso de funciones estadísticas y permitiendo generar como resultado la ecuación de probabilidades para la devolución del producto leche semidescremada UHT polietileno 1L.

```
#Establecemos la carpeta en la que trabajamos
> setwd("C:/Users/Paula/Desktop/INFORMACION TESIS/REGRESION LOGISTICA")
> library(readxl)
> lacteos <- read_xlsx('BASE_REGRESION.xlsx')
> #Análisis exploratorio ####
> #Inspeccionamos la estructura
> str(lacteos)
tibble [93 × 3] (S3: tbl_df/tbl/data.frame)
 $ CATEGORIA: chr [1:93] "DESPUES" "DESPUES" "DESPUES" "DESPUES" ...
 $ PH       : num [1:93] 6.1 6.71 6.67 6.69 6.68 6.65 6.72 6.66 5 4.68 ...
 $ ACIDEZ   : num [1:93] 0.2 0.156 0.168 0.154 0.171 0.169 0.173 0.171 0.9 0.772 ...
> #Resumen estadístico de las variables
> summary(lacteos)
  CATEGORIA          PH          ACIDEZ
Length:93          Min.   :4.06    Min.   :0.1310
Class :character   1st Qu.:4.99    1st Qu.:0.1368
Mode  :character   Median :6.62    Median :0.1467
                Mean   :5.96    Mean   :0.3716
                3rd Qu.:6.66    3rd Qu.:0.7770
                Max.   :6.72    Max.   :0.9010
                NA's   :45
> #Observamos las distribuciones de los dos grupos
> library(ggplot2)
> #Análisis univariado
> ggplot(lacteos, aes(x = CATEGORIA)) +
+   geom_bar()
> ggplot(lacteos, aes(x = PH)) +
+   geom_boxplot()
> ggplot(lacteos, aes(x = ACIDEZ)) +
+   geom_boxplot()
Warning message:
```



```

Removed 45 rows containing non-finite values ( `stat_boxplot()` ).
> #Análisis multivariado
> #CATEGORÍA VS PH
> ggplot(lacteos, aes(x = PH, fill = CATEGORIA)) +
+   geom_histogram(alpha=0.5, position="identity")
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
> ggplot(lacteos, aes(x = PH)) +
+   geom_boxplot() +
+   facet_wrap(vars(CATEGORIA), nrow = 2)
> #CATEGORÍA VS ACIDEZ
> ggplot(lacteos, aes(x = ACIDEZ, fill = CATEGORIA)) +
+   geom_histogram(alpha=0.5, position="identity")
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
Warning message:
Removed 45 rows containing non-finite values ( `stat_bin()` ).
> ggplot(lacteos, aes(x = ACIDEZ)) +
+   geom_boxplot() +
+   facet_wrap(vars(CATEGORIA), nrow = 2)
Warning message:
Removed 45 rows containing non-finite values ( `stat_boxplot()` ).
> #PH VS ACIDEZ
> cor(lacteos$PH, lacteos$ACIDEZ)
[1] NA
> #Modelado ####
> #Creo una dummy variable a partir de la categoría
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> lacteos <- lacteos %>%
+   mutate(CATEGORIA_dummy = if_else(CATEGORIA == "DESPUES", 1 , 0))
> #Modelo 1
> log.model1 <- glm(CATEGORIA_dummy ~ PH, data = lacteos, family = 'binomial')
> #Observamos el resumen del modelo
> summary(log.model1)

Call:
glm(formula = CATEGORIA_dummy ~ PH, family = "binomial", data = lacteos)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.75799 -0.56816 -0.45337  0.00433  2.22354

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  51.529     21.184   2.432  0.0150 *
PH           -8.023     3.203  -2.505  0.0123 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 128.915  on 92  degrees of freedom
Residual deviance:  47.247  on 91  degrees of freedom
AIC: 51.247

```

Number of Fisher Scoring iterations: 9

```
> #Observamos el detalle de los coeficientes
> summary(log.model1)$coefficients
              Estimate Std. Error  z value  Pr(>|z|)
(Intercept) 51.528744  21.184356  2.432396 0.01499930
PH           -8.022711   3.202892 -2.504833 0.01225091
> #Estimar ODDS Ratio Modelo 1
> exp(log.model1$coefficients)
(Intercept)          PH
2.391382e+22 3.279298e-04
> #Intervalos de Confianza de ODDS ratio
> exp(confint(log.model1))
waiting for profiling to be done...
              2.5 %          97.5 %
(Intercept) 3.114981e+09 1.879752e+48
PH           4.081977e-08 2.946835e-02
There were 12 warnings (use warnings() to see them)
> #Observamos el modelo
> ggplot(lacteos, aes(PH, CATEGORIA_dummy)) +
+   geom_point() +
+   geom_smooth(
+     method = "glm",
+     se = FALSE,
+     method.args = list(family = binomial)
+   )
`geom_smooth()` using formula = 'y ~ x'
> #Evaluación del modelo
> #McFadden's Pseudo R2
> library(pscl)
Classes and Methods for R developed in the
Political Science Computational Laboratory
Department of Political Science
Stanford University
Simon Jackman
hurdle and zeroinfl functions by Achim Zeileis
> pR2(log.model1)
fitting null model for pseudo-r2
      llh      llhNull      G2      McFadden      r2ML      r2CU
-23.6235221 -64.4573113  81.6675785  0.6335013  0.5844474  0.7792932
```

### Evaluación del modelo de predicción

La evaluación del modelo de predicción consistió en la programación en el software de RStudio, haciendo uso del paquete CARET, el mismo que permitió entrenar al modelo y supervisar el aprendizaje del algoritmo, obteniendo la matriz de confusión y de tal manera explicar la fiabilidad del modelo para predecir.

```
#EVALUACIÓN DEL MODELO PAQUETE CARET
> setwd("C:/Users/Paula/Desktop/INFORMACION TESIS/REGRESION LOGISTICA")
> library(readxl)
> data <- read_xlsx('BASE_PH.xlsx')
> str(data)
tibble [93 × 2] (S3: tbl_df/tbl/data.frame)
 $ CATEGORIA: num [1:93] 1 1 1 1 1 1 1 1 1 1 ...
 $ PH       : num [1:93] 6.1 6.71 6.67 6.69 6.68 6.65 6.72 6.66 5 4.68 ...
> #Introducimos la librería CARET
> library(mlr3)
> library(caret)
```

```

> library(lattice)
> #Se utiliza el 80% de los datos para entrenar el modelo y el 20% para la prueba
> trainIndex <- createDataPartition(data$CATEGORIA, p=0.8, list=F)
> train <- data[trainIndex,]
> test <- data[-trainIndex,]
> #Se especifica el tipo de validación, número de repeticiones, para el control
> #del entrenamiento
> fitControl <- trainControl( method = "cv", number = 2, savePredictions = 'final',
classProbs = FALSE)
> #Se declaran las variables predictoras
> predictors <-c("PH")
> outcomeName<-'CATEGORIA'
> model1 <- as.data.frame(train[,predictors], train[,outcomeName], method='lm', tr
Control = fitControl)
> summary(model1)
      PH
Min.   :4.060
1st Qu.:4.995
Median :6.620
Mean   :5.987
3rd Qu.:6.660
Max.   :6.720
> reg1 <- glm(CATEGORIA ~ PH, data=data, family = "binomial")
> summary(reg1)

```

```

Call:
glm(formula = CATEGORIA ~ PH, family = "binomial", data = data)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.75799 -0.56816 -0.45337  0.00433  2.22354

```

```

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   51.529     21.184   2.432  0.0150 *
PH             -8.023      3.203  -2.505  0.0123 *
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

(Dispersion parameter for binomial family taken to be 1)

```

```

Null deviance: 128.915 on 92 degrees of freedom
Residual deviance: 47.247 on 91 degrees of freedom
AIC: 51.247

```

```

Number of Fisher Scoring iterations: 9

```

```

> exp(coef(reg1))
      (Intercept)           PH
2.391382e+22 3.279298e-04
> #El estadístico sigue una distribución chi-cuadrado con grados de libertad
> #equivalentes a la diferencia de grados de libertad de los dos modelos
> anova(reg1, test = "Chisq")
Analysis of Deviance Table

```

```

Model: binomial, link: logit

```

```

Response: CATEGORIA

```

```

Terms added sequentially (first to last)

```

```

      Df Deviance Resid. Df Resid. Dev Pr(>Chi)

```

```

NULL          92    128.915
PH            1    81.668    91    47.247 < 2.2e-16 ***
----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #Convertir probabilidad en clasificación (EVALUACIÓN)
> predicciones <- ifelse(test = reg1$fitted.values > 0.5, yes = 1, no = 0)
> t1 <- table(data$PH, predicciones)
> t1
      predicciones
      0  1
4.06  0  1
4.22  0  1
4.4   0  1
4.5   0  2
4.56  0  1
4.61  0  1
4.65  0  1
4.67  0  1
4.68  0  2
4.71  0  1
4.76  0  2
4.84  0  2
4.88  0  2
4.9   0  1
4.92  0  1
4.93  0  1
4.98  0  2
4.99  0  1
5     0  2
5.02  0  2
5.1   0  1
5.21  0  1
5.23  0  1
5.33  0  1
5.37  0  1
5.39  0  1
5.83  0  1
6.1   0  1
6.2   0  1
6.22  0  1
6.3   0  1
6.56  1  0
6.6   1  0
6.61  1  0
6.62  8  0
6.63  7  0
6.64  6  0
6.65  5  0
6.66 13  0
6.67  2  0
6.68  3  0
6.69  4  0
6.7   1  0
6.71  1  0
6.72  1  0
> #Paquete CARET
> trainIndex1 <- createDataPartition(data$PH, p=0.8, list=F)
> train1 <- data[trainIndex1,]
> test1 <- data[-trainIndex1,]
> predictors1<-c("PH")
> outcomeName1<-'CATEGORIA'
> library(dplyr)
> train1$CATEGORIA<- as.factor(train1$CATEGORIA)

```

```
> model2<- train(CATEGORIA ~ PH, data= train1, method= 'glm', trControl = fit
Control)
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
> #model2 <-train(train1[,predictors1], train1[,outcomeName1], method='glm',
trControl = fitControl)
> summary(model2)
```

```
Call:
NULL
```

```
Deviance Residuals:
```

```
      Min       1Q   Median       3Q      Max
-0.72972 -0.57347 -0.24507  0.01316  2.17077
```

```
Coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   42.653     21.743   1.962  0.0498 *
PH             -6.683      3.286  -2.034  0.0420 *
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 105.358 on 75 degrees of freedom
Residual deviance: 38.558 on 74 degrees of freedom
AIC: 42.558
```

```
Number of Fisher Scoring iterations: 9
```

```
> test1$CATEGORIA <- as.factor(test1$CATEGORIA)
> test1$predicho <- predict(model2, test1[,predictors1])
> plot(test1$CATEGORIA, test1$predicho, col=palette("Paired"))
> confusionMatrix(test1$CATEGORIA, test1$predicho)
```

```
Confusion Matrix and Statistics
```

```

      Reference
Prediction 0 1
0 9 0
1 1 7
```

```
Accuracy : 0.9412
95% CI : (0.7131, 0.9985)
No Information Rate : 0.5882
P-Value [Acc > NIR] : 0.001559
```

```
Kappa : 0.8811
```

```
Mcnemar's Test P-Value : 1.000000
```

```
Sensitivity : 0.9000
Specificity : 1.0000
Pos Pred Value : 1.0000
Neg Pred Value : 0.8750
Prevalence : 0.5882
Detection Rate : 0.5294
Detection Prevalence : 0.5294
Balanced Accuracy : 0.9500
```

```
'Positive' Class : 0
```

### Evaluación del modelo de predicción por Minitab

La *Tabla A7* indica la base de datos resumida que se utilizó para la creación de las cartas de control de la variable predictor física-química para la evaluación del modelo de predicción realizado en Minitab. Esta tabla proporciona la fecha de creación del lote, el pH antes y después del cambio en los valores de los límites de control de la variable físico-química predictor que se realizaron para disminuir las devoluciones del producto.

*Tabla A 7 Base de datos resumida de producto terminado. Fuente: (Lácteos San Antonio, 2022)*

FECHA	PH ANTES	PH DESPUES
3/1/2022	6,65	6,65
3/1/2022	6,64	6,64
3/1/2022	6,65	6,65
3/1/2022	6,66	6,66
6/1/2022	6,66	6,66
6/1/2022	6,66	6,66
6/1/2022	6,66	6,66
6/1/2022	6,656	6,656
6/1/2022	6,66	6,66
6/1/2022	6,66	6,66
10/1/2022	6,63	6,63
10/1/2022	6,63	6,63
10/1/2022	6,63	6,63
10/1/2022	6,64	6,64
10/1/2022	6,65	6,65
10/1/2022	6,65	6,65
10/1/2022	6,65	6,65
10/1/2022	6,65	6,65
10/1/2022	6,63	6,63
10/1/2022	6,64	6,64
10/1/2022	6,65	6,65
10/1/2022	6,65	6,65
10/1/2022	6,65	6,65
13/1/2022	6,69	6,69
13/1/2022	6,66	6,66
13/1/2022	6,68	6,68
13/1/2022	6,7	6,7
13/1/2022	6,67	6,67
17/1/2022	6,64	6,64
17/1/2022	6,65	6,65
17/1/2022	6,63	6,63
17/1/2022	6,63	6,63

17/1/2022	6,63	6,63
17/1/2022	6,63	6,63
17/1/2022	6,65	6,65
17/1/2022	6,65	6,65
17/1/2022	6,65	6,65
17/1/2022	6,64	6,64
17/1/2022	6,64	6,64
17/1/2022	6,64	6,64
17/1/2022	6,63	6,63
20/1/2022	6,69	6,69

**CONVENIO DE CONFIDENCIALIDAD**SIG  
QM-H1.11/1.0 03-23**CONVENIO DE CONFIDENCIALIDAD**

Para la celebración del presente convenio comparecen por una parte, la compañía LÁCTEOS SAN ANTONIO C.A, representada legalmente por su Gerente General el Señor Luis David Moncayo Salavarría, con documento de identidad número 0101138162, con domicilio en las calles Carlos Tosi y Cornelio Vintimilla – Parque Industrial de la ciudad de Cuenca, República del Ecuador (en adelante NUTRI); y por otra parte, la señorita Paula Karina Lanchimba Parra, con documento de identidad número 0107336752, con domicilio en la ciudad de Cuenca (en adelante tesista). Los comparecientes, son de nacionalidad ecuatoriana, capaces para contratar y obligarse en las calidades en las que comparecen, quienes, de forma libre y voluntaria, deciden suscribir el presente contrato de confidencialidad, de conformidad con las siguientes cláusulas:

**Cláusula Primera: Antecedentes**

Lácteos San Antonio C.A. es una Empresa constituida en Ecuador cuya actividad principal es la producción y venta al por mayor de Lácteos y sus derivados, así como néctares, avenas y bebidas de frutas.

Dentro del giro de su negocio, NUTRI brinda la oportunidad de realizar trabajos de titulación a la comunidad educativa.

Los administradores y empleados de la empresa con motivo de la relación que surge entre NUTRI y el tesista, con la finalidad de proteger la propiedad intelectual, producción y comercialización de los productos, ha convenido en el manejo adecuado y confidencial de la información técnica, secretos productivos, comerciales, de investigación y/o desarrollo, así como, los datos financieros y de costos que genera.

Por lo señalado, se han comprometido a guardar reserva acerca de:

- No revelar, publicar, ni utilizar ninguna Información Confidencial entregada durante el proceso de ejecución del servicio, y a no revelar a terceros ningún informe, conclusiones ni recomendaciones producto de resultados finales del servicio.

Página 1 de 4





## CONVENIO DE CONFIDENCIALIDAD

SIG  
QM-11.11/1.0 03-23

- Mantener la confidencialidad de cualquier informe, análisis, información técnica, comercial, financiera, corporativa o cualquier otro documento preparado por las partes durante el proceso, de la misma manera en la que acostumbra a proteger su propia información confidencial.
- Tomar tantas medidas como sean necesarias, a fin de evitar revelar la Información confidencial a otros, ya sea a través de sus dependientes, proveedores, clientes y cualquier otra persona natural o jurídica, involucrada en la realización del trabajo encomendado.
- De ser necesario, para la efectiva consecución de los trabajos encomendados, que alguna de las partes revele alguna Información confidencial a terceros o a algunos de sus empleados, deberán con anterioridad asegurarse que esa Información confidencial sea guardada por dichas personas bajo este mismo mecanismo de confidencialidad, haciéndose responsables por cualquier incumplimiento de privacidad de estos terceros.

### Cláusula Segunda: Confidencialidad

Por las razones expresadas en la cláusula precedente, la señorita Paula Karina Lanchimba Parra, en su calidad de tesista, se compromete a mantener la más estricta y absoluta reserva y confidencialidad sobre todos los datos y asuntos relativos a los procesos y actividades del negocio, quedando expresamente impedido, tanto él como sus ayudantes y demás personal que labore a sus órdenes, de comunicar, divulgar, bajo cualquier forma, la información que hubieren obtenido en el proceso de ejecución del servicio.

En consecuencia, acuerdan las partes, que toda la información y documentación, recibida directamente de la empresa o de terceros, y que tenga relación con el negocio, deberán ser considerados estrictamente confidenciales, comprometiéndose, en consecuencia, cada una de las partes recíprocamente a lo siguiente:

1. Mantener en custodia toda información que se obtenga de forma física, digital o virtual; del negocio, esta supone que se guardarán en las oficinas y bases de datos, a la cual solamente tendrán acceso las personas que autorizadas por la empresa para intervenir en el proceso y las que formen parte del equipo de trabajo de la empresa que brinda el servicio.

Página 2 de 4

**CONVENIO DE CONFIDENCIALIDAD**SIG  
QM-11.11/1.0 03-23

2. Cumplir el presente convenio de buena fe, de manera tal que han de entenderse incorporadas al mismo, todas las obligaciones que derivan de su naturaleza.

**Cláusula Tercera: Duración**

El presente convenio durará de manera indefinida desde la suscripción del presente convenio, es decir que los datos e información del negocio no podrá revelarse aun cuando termine la relación contractual.

**Cláusula Cuarta**

Las partes acuerdan que, ante cualquier incumplimiento a los términos del presente convenio de confidencialidad, la parte incumplidora será responsable por todos los daños y perjuicios tanto morales como económicos que se ocasionen, para lo cual se someterán a todas las acciones que en derecho correspondan. En consecuencia, se deja expresamente establecido que podrá pedirse la indemnización de daños y perjuicios, en los términos que establece el Código Civil y, demás acciones que la ley señale.

**Cláusula Quinta: Incumplimiento**

El incumplimiento de cualquiera de los términos del presente convenio dará lugar a una sanción de cien salarios básicos unificados y/o a la terminación de la relación contractual con NUTRI.

**Cláusula Sexta: Resolución De Conflictos- Jurisdicción Y Competencia**

Para todos los efectos legales derivados de este contrato, las partes fijan su domicilio en la ciudad de Cuenca. Toda dificultad o conflictos que pudieren suscitarse, se someterán a un proceso de Arbitraje en el centro de Arbitraje y Mediación de la Cámara de Producción del Azuay, de acuerdo con las reglas de la Ley de Arbitraje y Mediación y el Reglamento de dicho Centro.

**Cláusula Séptima: Aceptación.**

Las partes aceptan el contenido de todas y cada una de las cláusulas del presente contrato y en consecuencia se comprometen a cumplirlas en toda su extensión.

Para constancia de lo anterior, las partes firman este contrato en dos ejemplares del mismo tenor y valor, en la ciudad de Cuenca, al día 17 del mes de marzo del 2023.



## CONVENIO DE CONFIDENCIALIDAD

SIG  
QM-11.11/1.0 03-23

Sr. Luis David Moncayo S.  
Lácteos San Antonio C.A.

Srta. Paula Karina Lanchimba Parra  
Tesisista