

**Análisis de regresión y correlación de variables que impactan la tasa de quejas en  
Audifarma S.A**

Diana Milena Alvarez Ramirez

Jhon Fredy Villada Londoño

Asesor

Sandra Patricia Barreto

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnologías e Ingeniería ECBTI

Especialización en Ciencias de Datos y Analítica

2024

## Resumen

En el presente documento el lector puede encontrar un análisis de la correlación entre variables tácticas (satisfacción de usuarios, nivel de servicio, días de inventario y entrega de pendientes) y la tasa de quejas. Utilizando el lenguaje de programación Python, se generaron gráficos de correlación, dispersión y regresión lineal para la predicción de la tasa de quejas. Este enfoque permitió identificar patrones en los datos históricos y plantear diferentes modelos de regresión lineal.

En el proyecto se evaluaron 4 modelos de regresión lineal con diferente cantidad de variables y se interpretó cada uno de los resultados tanto en el conjunto de entrenamiento como en el conjunto de prueba. Al final se muestra un resumen con las medidas de desempeño  $R^2$ , MSE, RMSE y MAE. Obtenidas para cada modelo en el conjunto de entrenamiento, lo cual permitió comparar los modelos y tomar una decisión respecto a cuál era el modelo con el mejor desempeño general.

El proyecto fue especialmente relevante ya que no solo permitió predecir la tasa de quejas mediante un modelo de regresión lineal, sino que también ayudó a comprender la correlación y la importancia relativa de las variables tácticas. Esto será clave para priorizar estrategias organizacionales, reducir las quejas y mejorar la satisfacción de los usuarios, contribuyendo significativamente a la optimización de los procesos internos y la toma de decisiones basada en datos

**Palabras claves:** Regresión lineal, tasa de quejas, modelos predictivos, calidad de servicio, satisfacción del cliente.

## Abstract

In this paper the reader can find an analysis of the correlation between tactical variables (user satisfaction, service level, days of inventory and outstanding delivery) and the complaint rate. Using the Python programming language, correlation, scatter and linear regression graphs were generated for the prediction of the complaint rate. This approach allowed the identification of patterns in the historical data and the development of different linear regression models.

The project evaluated 4 linear regression models with different number of variables and interpreted each of the results in both the training set and the test set. A summary with the  $R^2$ , MSE, RMSE and MAE performance measures is shown at the end. Obtained for each model in the training set, which allowed us to compare the models and make a decision as to which model had the best overall performance.

The project was particularly relevant as it not only allowed predicting the complaint rate using a linear regression model, but also helped to understand the correlation and relative importance of tactical variables. This will be key to prioritize organizational strategies, reduce complaints and improve user satisfaction, contributing significantly to the optimization of internal processes and data-driven decision-making.

**Keywords:** linear regression, complaint rate, predictive models, service quality, customer satisfaction.

## Tabla de Contenido

Justificación .....	10
Objetivos .....	11
Objetivo General .....	11
Objetivos Específicos .....	11
Marco Conceptual y Teórico .....	12
Metodología de Investigación .....	16
Evolución Mensual de la Tasa de Quejas (enero 2020 - abril 2024) .....	16
Categorización de la Tasa de Quejas .....	17
Matriz de Dispersión .....	18
Eliminación de Outliers .....	20
Matriz de Correlación .....	22
Matriz de Dispersión Luego de la Eliminación de Outliers .....	23
Boxplot de Tasa de Quejas .....	25
Boxplot de Nivel de Servicio .....	26
Boxplot de Entrega de Pendientes .....	27
Boxplot de Días de Inventario .....	27
Boxplot Satisfacción de Usuarios .....	28
Participación Nivel Tasa de Quejas .....	29
Análisis de Desempeño de Modelos de Regresión .....	31
Modelo de Regresión Lineal Múltiple (Modelo 4v) .....	31
Evaluación del Modelo 4v .....	33
Modelo de Regresión Lineal Múltiple sin Nivel de Servicio (Modelo 3v) .....	34

Evaluación del Modelo 3v.....	36
Modelo de Regresión Lineal días de Inventario y Nivel de Servicio (Modelo 2v) .....	37
Evaluación del Modelo 2v.....	39
Modelo de Regresión Lineal Simple (RLS).....	40
Evaluación del modelo de RLS .....	41
Resultados de las Métricas para cada Modelo .....	42
Gráfica de Regresión Lineal: Tasa de Quejas vs Nivel de Servicio .....	43
Supuestos del Modelo de Regresión Lineal.....	45
Linealidad.....	45
Independencia de los Errores .....	45
Homocedasticidad .....	45
Normalidad de Errores .....	46
Multicolinealidad .....	47
Conclusiones.....	48
Recomendaciones .....	51
Referencias Bibliográficas .....	53

## Lista de Figuras

<b>Figura 1</b> <i>Evolución Mensual de la Tasa de Quejas (enero de 2020 – abril 2024)</i> .....	17
<b>Figura 2</b> <i>Categorización de la Tasa de Quejas</i> .....	18
<b>Figura 3</b> <i>Tabla de Dispersión (enero 2020 - abril 2024)</i> .....	20
<b>Figura 4</b> <i>Tasa de Quejas antes de la Eliminación de Outliers</i> .....	21
<b>Figura 5</b> <i>Evolución Mensual de la Tasa de Quejas sin Outliers</i> .....	22
<b>Figura 6</b> <i>Matriz de Correlación</i> .....	23
<b>Figura 7</b> <i>Dispersión Después de Eliminación de Outliers</i> .....	25
<b>Figura 8</b> <i>Boxplot de Tasa de Quejas</i> .....	26
<b>Figura 9</b> <i>Boxplot Nivel de Servicios</i> .....	27
<b>Figura 10</b> <i>Diagrama de Caja de Entrega de Pendientes</i> .....	27
<b>Figura 11</b> <i>Boxplot Días de Inventario</i> .....	28
<b>Figura 12</b> <i>Boxplot Satisfacción de Usuario</i> .....	29
<b>Figura 13</b> <i>Participación Nivel de Tasa de Quejas, Gráfico de Pie</i> .....	30
<b>Figura 14</b> <i>Resumen Detallado Ajustado Modelo 4v Utilizando Datos de Entrenamiento</i> .....	33
<b>Figura 15</b> <i>Comparación de Tasas de Quejas Reales y Predichas para el Modelo 4v</i> .....	34
<b>Figura 16</b> <i>Resumen Detallado Ajustado Modelo 3v Utilizando Datos de Entrenamiento</i> .....	36
<b>Figura 17</b> <i>Comparación de Tasas de Quejas Reales y Predichas para el Modelo 3v</i> .....	37
<b>Figura 18</b> <i>Resumen Detallado Ajustado Modelo 2v Utilizando Datos de Entrenamiento</i> .....	39
<b>Figura 19</b> <i>Comparación de Tasas de Quejas Reales y Predichas para el Modelo 2v</i> .....	40
<b>Figura 20</b> <i>Resumen Detallado Ajustado de Modelo RLS Utilizando Datos Entrenamiento</i> .....	41
<b>Figura 21</b> <i>Comparación de Tasas de Quejas Reales y Predichas para Modelo RLS</i> .....	42
<b>Figura 22</b> <i>Resultados de las Métricas de Evaluación para cada uno de los Modelos</i> .....	43

<b>Figura 23</b> <i>Regresión Lineal Tasa de Quejas vs Nivel de Servicio</i> .....	44
<b>Figura 24</b> <i>Supuesto de Homocedasticidad</i> .....	46
<b>Figura 25</b> <i>Supuesto de Normalidad de Errores</i> .....	47

## Introducción

Nuestra propuesta es desarrollar un proyecto aplicado como opción de grado, basado en el modelo de gestión organizacional. El objetivo principal es estudiar la correlación entre las variables independientes: satisfacción de usuarios, nivel de servicio, días de inventario y entrega de pendientes llamadas en adelante variables tácticas y la variable objetivo: tasa de quejas. Para ello, utilizaremos herramientas como Python y sus librerías especializadas para generar gráficos de correlación, dispersión y regresión lineal. Además, se desarrollará un modelo de regresión que permita predecir la tasa de quejas.

El modelo de regresión lineal es una técnica estadística que permite identificar patrones en los datos históricos para describir y predecir la relación entre las variables tácticas y la dependiente. Este enfoque se basa en la construcción de una fórmula estadística que no solo explica el comportamiento pasado de las variables, sino que también permite realizar predicciones confiables a partir de nuevos valores de las variables tácticas. Para garantizar la validez de las predicciones, el modelo debe cumplir con ciertos supuestos estadísticos relacionados con la regresión lineal, asegurando así su fiabilidad y aplicabilidad.

Entre los supuestos más relevantes se encuentran la linealidad, la independencia de los errores, que evita correlaciones entre los errores residuales; la homocedasticidad, que asegura que los errores tengan una varianza constante; la normalidad de los errores, necesaria para validar las pruebas estadísticas; y la ausencia de multicolinealidad, que garantiza que las variables tácticas no estén altamente correlacionadas entre sí. Estos supuestos serán verificados utilizando análisis gráficos y métricas estadísticas en Python.

Actualmente, la organización no dispone de un modelo que permita predecir la tasa de quejas a partir de las variables independientes mencionadas. Tampoco existe claridad sobre la

magnitud del impacto que cada una tiene en la variación de la tasa de quejas. Este proyecto es especialmente relevante, ya que no solo permitirá predecir la tasa de quejas mediante un modelo robusto, sino que también ayudará a comprender la correlación y la importancia relativa de las variables tácticas. Esto será clave para priorizar estrategias organizacionales, reducir las quejas y mejorar la satisfacción de los usuarios, contribuyendo significativamente a la optimización de los procesos internos y la toma de decisiones basada en datos.

### **Justificación**

Clasificar cada una de las variables tácticas según su impacto en la tasa de quejas, para identificar cuáles tienen mayor influencia y orientar las estrategias de mejora.

La gestión organizacional es una práctica integrada en los diferentes aspectos de la empresa. A través de la tasa de quejas, se pueden evaluar perspectivas del Balance ScoreCard, como la de clientes y procesos internos. Con el uso de nuevas herramientas como Python, es posible realizar análisis de correlación y desarrollar modelos predictivos. Estas herramientas permiten analizar la relación entre las variables tácticas y la tasa de quejas, facilitando la identificación de los factores clave y la creación de estrategias efectivas para mejorar este indicador.

## **Objetivos**

### **Objetivo General**

Desarrollar un análisis de correlación de variables y un modelo de regresión lineal para la predicción de resultados para la tasa de quejas en Audifarma S.A

### **Objetivos Específicos**

Realizar un análisis de correlación entre las variables tácticas y la tasa de quejas para la identificación de relaciones significativas entre ellas.

Definir el mejor modelo de regresión lineal a partir de los datos proporcionados para la predicción de la tasa de quejas en Audifarma S.A.

Validar los supuestos de la regresión lineal para el modelo seleccionado para la predicción de la tasa de quejas en Audifarma S.A.

## Marco Conceptual y Teórico

El análisis de la tasa de quejas y su relación con variables tácticas clave es fundamental para la mejora continua en la gestión organizacional. Este proyecto, mediante el uso de métodos estadísticos avanzados y herramientas tecnológicas, busca desarrollar un modelo que no solo permita predecir esta métrica, sino que también proporcione un marco claro para la toma de decisiones basadas en datos, asegurando así la competitividad y sostenibilidad a largo plazo.

En la gestión moderna de las organizaciones, la satisfacción del cliente es un pilar esencial para mantener la competitividad. Un indicador crucial en este contexto es la tasa de quejas, que refleja expectativas no cumplidas por parte de los clientes y se relaciona directamente con la percepción de la calidad del servicio. Comprender los factores que influyen en esta métrica permite implementar estrategias más efectivas para mejorar la calidad del servicio y fortalecer la relación con los clientes. En un mundo globalizado, las organizaciones enfrentan el reto constante de superar las expectativas de los clientes mediante la mejora continua de sus procesos.

El análisis de las quejas ha evolucionado hacia enfoques más estructurados y eficientes, considerando que el comportamiento de estas representa las respuestas de los clientes frente a experiencias de insatisfacción en servicios o productos (Moliner, Berenguer y Gil, 2008). En este contexto, herramientas estadísticas como la correlación y la regresión lineal son fundamentales para explorar y modelar las relaciones entre variables relevantes, como la tasa de quejas y factores como la satisfacción del usuario, los días de inventario, el nivel de servicio y las entregas pendientes. Estas herramientas no solo permiten identificar las variables más influyentes, sino también facilitan la mejora continua en la gestión de la experiencia del cliente (Fornell y Wernerfelt, 1987).

En la actualidad, el uso de herramientas estadísticas avanzadas, combinado con tecnologías como Python y sus librerías especializadas, potencia la capacidad de las organizaciones para construir modelos predictivos precisos. Estos modelos contribuyen no solo al entendimiento teórico, sino también al diseño de estrategias prácticas para mejorar la satisfacción del cliente y reducir la tasa de quejas (Engel y Sedlmeier, 2011). En un entorno altamente competitivo, esta combinación de teoría y práctica es esencial para mantener la relevancia y el éxito organizacional (Gea, 2014).

La regresión lineal es una técnica estadística que modela la relación entre una variable dependiente y una o más variables independientes. En su forma más simple, esta técnica asume una relación lineal y busca identificar la mejor línea recta que se ajuste a los datos observados. Más allá de su sencillez, la regresión lineal es una herramienta poderosa aplicada en diversos campos, como el análisis de datos en la gestión organizacional (Montgomery, 2006).

La correlación lineal y la regresión lineal simple son métodos fundamentales para estudiar la relación entre dos variables. Mientras que la correlación cuantifica la fuerza y dirección de la relación sin implicar causalidad, la regresión busca construir un modelo que permita predecir el valor de una variable (dependiente) a partir de otra (independiente). En estudios de correlación, las variables se observan y analizan sin control, mientras que, en la regresión, una variable se modela en función de otra, permitiendo realizar predicciones. En este sentido, los estudios de correlación suelen preceder al desarrollo de modelos de regresión, estableciendo primero la relación entre las variables antes de construir un modelo predictivo más preciso (Amat, 2016).

Un modelo de regresión lineal permite comprender y predecir la relación entre una variable dependiente (como la tasa de quejas) y una o más variables independientes (como el

nivel de servicio o los días de inventario). Este modelo se describe mediante la ecuación:  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \epsilon$ . El objetivo de este modelo es aprender patrones a partir de datos históricos y realizar predicciones basadas en nuevos valores de las variables independientes. Esto permite a las organizaciones anticiparse a problemas y optimizar sus procesos.

Para garantizar la validez de un modelo de regresión lineal, se deben cumplir ciertos supuestos estadísticos como la Linealidad, Independencia de los errores, Homocedasticidad (varianza constante), Normalidad de los errores (distribución normal) y Ausencia de multicolinealidad (no deben estar altamente correlacionadas) (Montgomery, 2006). Si estos supuestos no se cumplen, el modelo puede generar predicciones poco confiables o resultados sesgados. Por ello, es fundamental verificar su cumplimiento mediante gráficos de dispersión, análisis de residuos y métricas.

Dentro de este proyecto aplicado, es importante citar investigaciones previas que han abordado las variables de tasa de quejas, satisfacción al usuario, correlación y regresión lineal. Estos estudios previos han proporcionado un marco de referencia establecido sobre el cual se basa esta investigación, permitiendo una comparación y contextualización fundamentada de los resultados obtenidos.

El estudio “Determinantes del comportamiento de queja y su importancia en la segmentación de clientes insatisfechos” explora cómo la intensidad de la insatisfacción y otras variables discriminan segmentos de consumidores. Utilizando la metodología CHAID, se identifican dos tipos de respuestas: quejas directas y respuestas privadas. Los resultados muestran que los consumidores muy insatisfechos con baja probabilidad de éxito en sus quejas tienden a optar por respuestas privadas. Además, quienes se quejan desarrollan más respuestas

privadas que aquellos que no lo hacen. Estos hallazgos destacan la importancia de variables como la intensidad de la insatisfacción y la probabilidad de éxito en la queja (Moliner, 2004).

El proyecto “Efecto del marketing relacional en la lealtad de los huéspedes de hoteles de corta estancia” utiliza un enfoque cuantitativo y diseño ex post facto. Basado en estudios previos, emplea métodos estadísticos como la correlación y la regresión lineal para analizar la relación entre variables y predecir resultados. La investigación muestra que estrategias de marketing relacional, centradas en construir confianza y mejorar la experiencia del cliente, son clave para aumentar la lealtad. Aunque la comunicación, el compromiso y la resolución de conflictos también son importantes, su impacto es menor. Se recomienda fortalecer estas áreas y explorar contextos culturales adicionales (Quintana, 2020).

El estudio “Análisis de correlación entre las peticiones, quejas, reclamos y sugerencias (PQRS) y factores asociados a la prestación de servicios de salud en el subsistema de la policía nacional” investiga la relación entre las PQRS y los servicios de salud en el SSPN durante 2018 y 2019. Se observó un incremento significativo en las PQRS, reflejando una creciente preocupación por la calidad y accesibilidad de estos servicios. Este análisis cuantitativo busca proporcionar insights para mejorar la calidad y eficiencia de la atención en salud dentro de la institución policial (Ovalle, 2019).

## **Metodología de Investigación**

En este capítulo, se recopilan aspectos clave en la exploración de la base de datos proporcionada por Audifarma S.A. para las variables tácticas y la tasa de quejas. Se incluyen líneas de tiempo con la evolución mensual de la tasa de quejas, la identificación y descripción de variables, gráficos de cajas, matrices de correlación, gráficos de dispersión y los diferentes modelos de regresión lineal, así como los supuestos del modelo escogido, todos desarrollados en Python. Esto facilita la identificación visual del comportamiento de los datos. Finalmente, se presentan los análisis, hallazgos y conclusiones sobre los resultados obtenidos en este ejercicio.

### **Evolución Mensual de la Tasa de Quejas (enero 2020 - abril 2024)**

A continuación, presentamos un gráfico de línea que muestra la evolución mensual de la tasa de quejas, con datos desde enero de 2020 hasta abril de 2024, abarcando un total de 52 registros. En este gráfico se observa que la tasa de quejas presenta fluctuaciones a lo largo del tiempo, con un notable incremento a partir de mediados de 2023, alcanzando su punto más alto en abril de 2024.

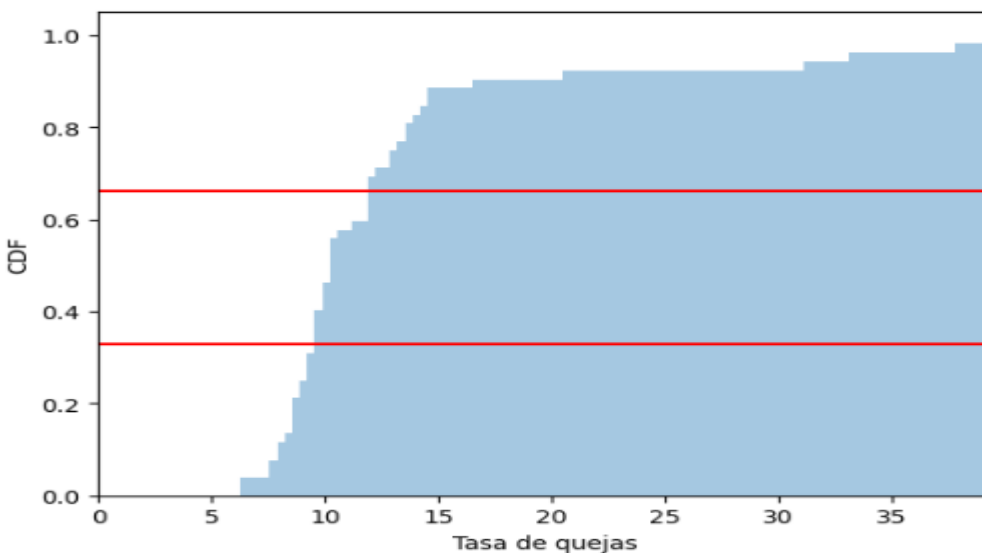
## Figura 1

*Evolución Mensual de la Tasa de Quejas (enero de 2020 – abril 2024)*



## Categorización de la Tasa de Quejas

La tasa de quejas es una variable que, cuanto más cercana a 0 esté, es mejor para la organización. Para la categorización de esta variable se eligen tres categorías: bajo, medio y alto. El objetivo interno de la organización es que la tasa de quejas sea de un dígito. No obstante, se considera que una tasa de quejas de hasta 13 es tolerable, aunque genera diferentes alertas. Un valor superior se convierte en un foco de atención y requiere la implementación de planes de acción para reducirla. Esta clasificación sirve como base para definir los rangos de las categorías. Además, se utiliza la herramienta de distribución acumulativa de probabilidad para identificar rangos de acuerdo a la distribución de los datos.

**Figura 2***Categorización de la Tasa de Quejas***Matriz de Dispersión**

La matriz de gráficos de dispersión organiza todas las combinaciones de gráficas posibles entre las cinco variables tácticas y la variable objetivo: Nivel de servicio, Pendientes, Días de inventario, Satisfacción usuarios y Tasa de quejas. A continuación, se presenta el análisis de las dispersiones de las variables con todos los datos incluyendo el incremento en la tasa de quejas de los últimos meses de análisis.

Tasa de quejas vs nivel de servicio: El gráfico de dispersión muestra una relación inversa fuerte entre la Tasa de quejas y el Nivel de servicio. A medida que el Nivel de servicio aumenta, la Tasa de quejas disminuye. Estas variables presentan la correlación más fuerte en el análisis del conjunto de datos con un coeficiente de correlación de  $-0.9395$ . Esto sugiere un comportamiento lineal con una pendiente negativa, lo cual tiene sentido ya que un alto nivel de servicio debería resultar en una menor tasa de quejas.

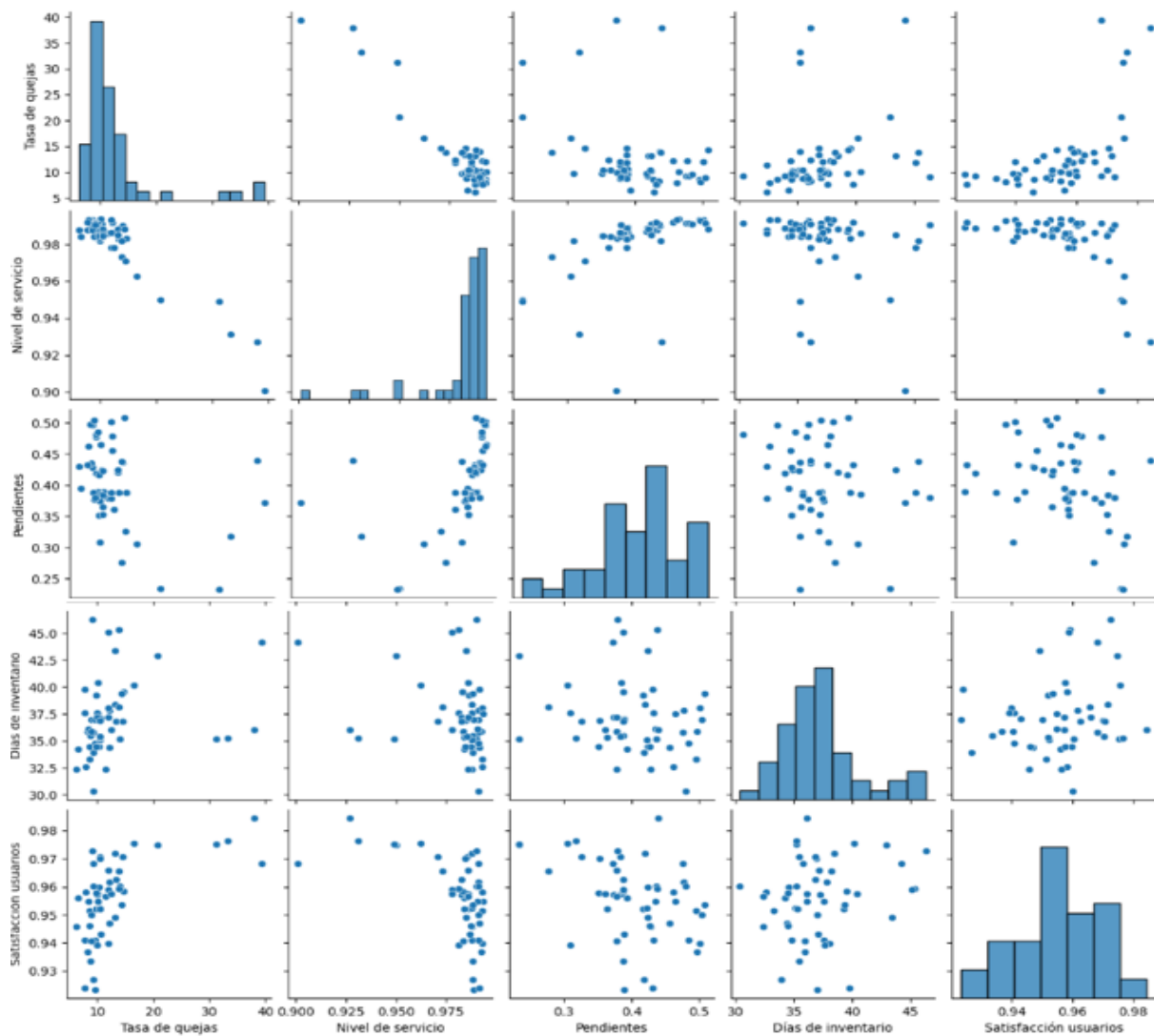
Tasa de quejas vs pendientes: El gráfico de dispersión muestra una relación débil entre la Tasa de quejas y los Pendientes. La correlación es de  $-0.37$ , lo que indica una relación inversa débil. No obstante, esto es coherente con la interpretación de que, si el nivel de pendientes disminuye, las quejas podrían aumentar, ya que los pendientes miden el cumplimiento de la promesa de entrega de medicamentos.

Tasa de quejas vs días de Inventario: El gráfico de dispersión muestra una relación muy débil entre la Tasa de quejas y los Días de inventario. La correlación es de  $0.23$ , lo cual es bajo y no coherente. Se esperaría que una mayor cantidad de días de inventario resultara en menos quejas, por lo que el signo del resultado debería ser negativo para tener coherencia.

Tasa de quejas vs satisfacción usuarios: El gráfico de dispersión muestra una relación moderada entre la Tasa de quejas y la Satisfacción de los usuarios. La correlación es de  $0.57$ , indicando una relación directamente proporcional. En la práctica, un mayor número de quejas debería reflejar una mayor insatisfacción de los usuarios, lo que sugiere que esta correlación no es coherente con la realidad esperada.

### Figura 3

*Tabla de Dispersión (enero 2020 - abril 2024)*



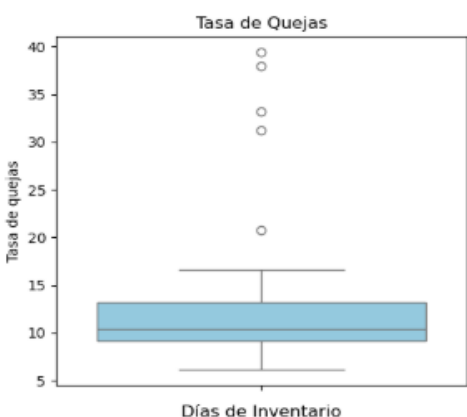
### Eliminación de Outliers

A continuación, presentamos el gráfico boxplot para la identificación de outliers, el cual muestra que hay 5 datos que no se deberían tener en cuenta, específicamente los 4 primeros meses del año 2024 y diciembre de 2023. Se toma la decisión de eliminar solo los últimos 4 registros y conservar el dato de diciembre de 2023, considerando que la organización dispuso de

mayores canales para que los usuarios manifestaran sus inconformidades con el servicio (entre ellos, la página web y la aplicación móvil, además de las líneas telefónicas de atención al usuario existentes). Además, teniendo en cuenta que las condiciones del sistema de salud se vieron afectadas por el desabastecimiento de algunos medicamentos, se consideró probable un escenario de incremento significativo en las quejas, lo cual se reflejaría en un inicio de tendencia alcista con el dato de diciembre de 2023.

#### Figura 4

*Tasa de Quejas antes de la Eliminación de Outliers*



La siguiente gráfica muestra una base final de 48 registros, en la cual se sigue evidenciando la tendencia creciente en los últimos meses de análisis para la tasa de quejas en Audifarma.

## Figura 5

### *Evolución Mensual de la Tasa de Quejas sin Outliers*



## Matriz de Correlación

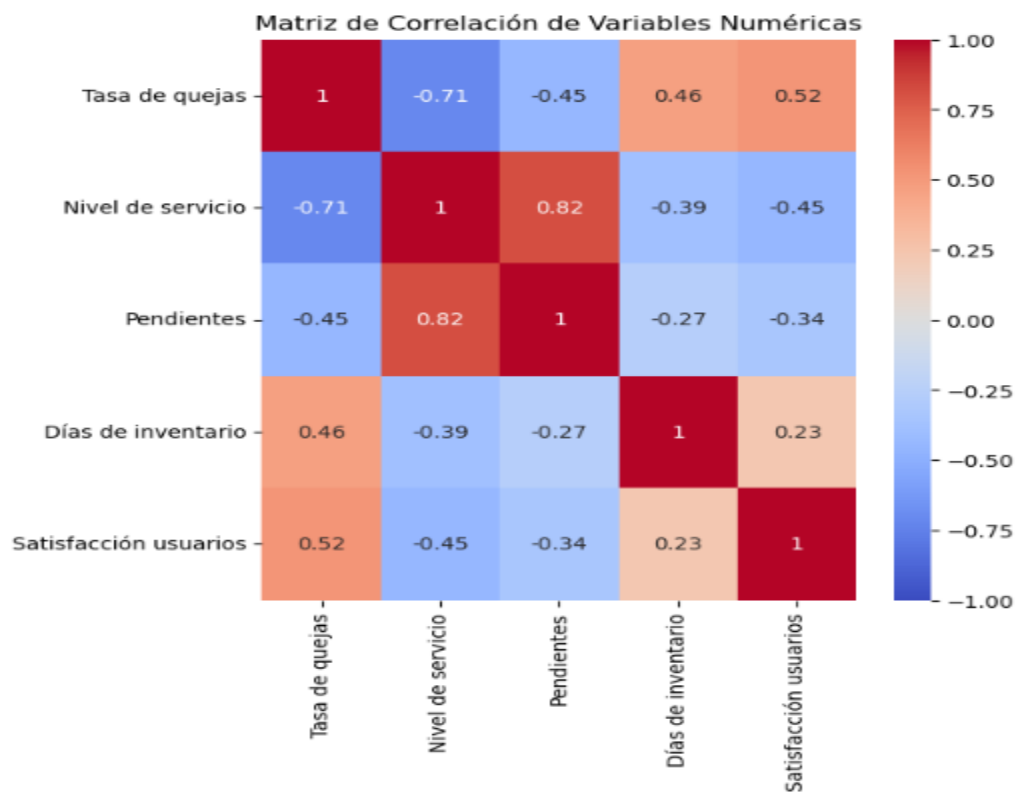
Este mapa es relevante porque permite identificar relaciones significativas entre las variables, lo cual puede ser útil para la toma de decisiones y la mejora de procesos.

Tasa de quejas vs nivel de servicio: Correlación negativa fuerte (-0.71), indicando que, a mayor nivel de servicio, menor tasa de quejas.

Tasa de quejas vs satisfacción de usuarios: Correlación positiva moderada (0.52), sugiriendo que, a mayor tasa de quejas, mayor satisfacción, aunque esto puede parecer contradictorio.

Tasa de quejas vs pendientes: Correlación positiva fuerte (-0.45), mostrando que, a mayor nivel de servicio, más pendientes.

Tasa de quejas vs días de inventario: Correlaciones bajas con las demás variables, siendo la más alta con la Tasa de quejas (0.46).

**Figura 6***Matriz de Correlación***Matriz de Dispersión Luego de la Eliminación de Outliers**

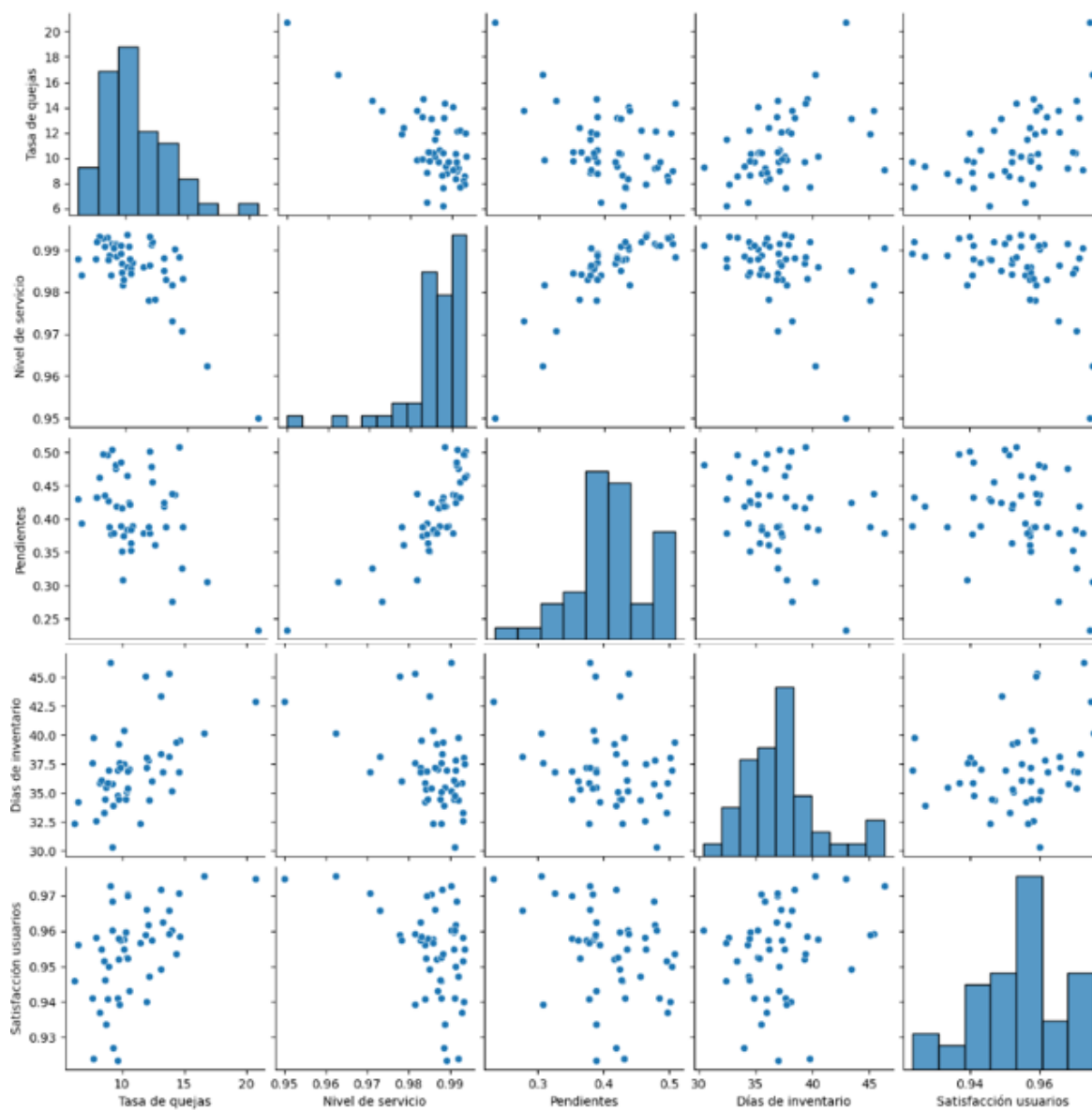
Luego de la eliminación de los outliers se puede observar cómo se continúa percibiendo un comportamiento lineal entre las variables tasa de quejas y nivel de servicio con una pendiente negativa que indica que a menor nivel de servicio mayor serán las quejas que se reciban y con una correlación importante de -0.71 que si bien bajó respecto a la medición con todo el universo de datos (-0.94) sigue siendo la variable con el resultado más significativo.

Pendientes vs tasa de quejas también tiene una correlación negativa, aunque en este caso es mucho menor con -0.45. Este resultado es coherente porque los pendientes miden el porcentaje de cumplimiento de la promesa de entrega a los usuarios cuando van a la farmacia y

no se les entrega el medicamento en el primer acercamiento. La interpretación es que, a mayor cantidad de cumplimiento de los pendientes, menor debería ser la tasa de quejas. No obstante, el resultado de la correlación no muestra representatividad.

A pesar de la eliminación de los outliers se siguen presentando incoherencias respecto a los resultados de las demás variables. Satisfacción de usuarios vs tasa de quejas con una correlación de 0.52 no tiene sentido en la práctica dado que no es posible que a mayor cantidad de quejas los usuarios estén más satisfechos. Días de inventario vs tasa de quejas con un resultado de 0.46 tampoco es coherente que entre más días de inventario tenga mayores serán las cantidades de quejas especialmente porque los días de inventario son una medida de la cantidad de medicamentos que tengo para satisfacer la demanda.

De acuerdo con lo anterior se procede a trabajar solo con la variable nivel de servicio dado que es la que muestra un comportamiento lineal con la tasa de quejas y es la que tiene el nivel de correlación más importante en el ejercicio con las 4 variables.

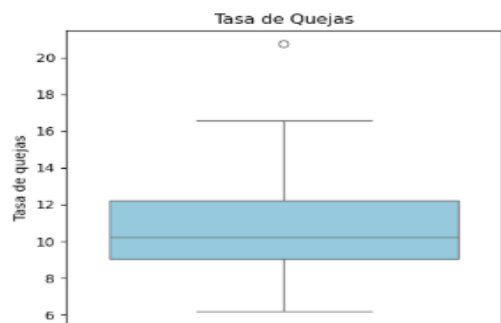
**Figura 7***Dispersión Después de Eliminación de Outliers****Boxplot de Tasa de Quejas***

De la distribución de los datos de tasa de quejas se puede decir que están en un rango que está entre el 6 y el 16 aproximadamente. Se identifica un dato atípico que está sobre un nivel de tasa de quejas de 20, pero la mayor cantidad de datos está distribuida en un rango entre 9 y 12 de

tasa de quejas esto es positivo para la organización teniendo en cuenta que tal como se mencionó al comienzo, se considera aceptable un nivel de tasa de quejas inferior a 13 en este caso aproximadamente el 75% de los datos (los tres primeros cuartiles) cumplen con esa premisa.

### Figura 8

*Boxplot de Tasa de Quejas*

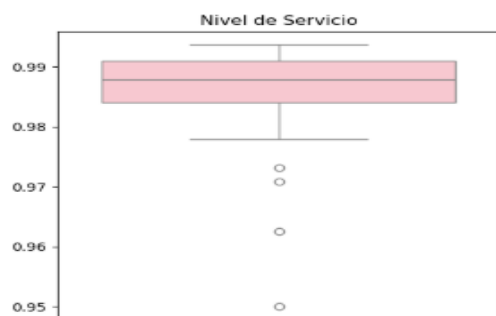


### *Boxplot de Nivel de Servicio*

De la gráfica de cajas del nivel de servicio se puede decir que la distribución está en datos altos de nivel de servicio (mayores a 97.8% aproximadamente) lo cual se habla con que la mayoría de los datos de tasa de quejas están en niveles aceptables por la organización. Lo anterior gracias a que entre mejor nivel de servicio se tenga en la empresa los usuarios estarán más satisfechos y por tanto no se quejarán del servicio recibido. También se observan 4 valores atípicos los cuales están relacionados con los valores más altos de tasa de quejas.

## Figura 9

### *Boxplot Nivel de Servicios*

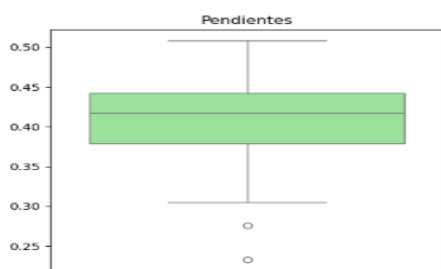


### *Boxplot de Entrega de Pendientes*

Este diagrama muestra un rango entre 30% y 50% con una concentración de los datos entre el 37% y el 45% de los datos al gráfico identifica dos outliers al rededor del 25%. Esta es una variable que tiene mucho por mejorar en la organización teniendo en cuenta que lo ideal es tener una oportunidad del 100% pero el máximo dato alcanzado ronda el 50%.

## Figura 10

### *Diagrama de Caja de Entrega de Pendientes*



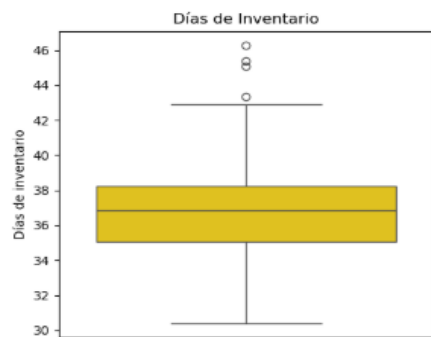
### *Boxplot de Días de Inventario*

Días de inventario muestra un rango que va de 30 a 42 días, con 4 outliers que llegan hasta los 46 días. La centralidad de los datos se encuentra entre los 34.5 y 38 días. Es importante mencionar que la organización debe optimizar los inventarios, ya que no es productivo mantener

muchos días de inventario con dinero inmovilizado en productos de forma ineficiente. Al mismo tiempo, no se deben tener pocos días de inventario, lo cual podría afectar el servicio o no contar con suficientes reservas ante una contingencia por la posible no entrega de algún proveedor.

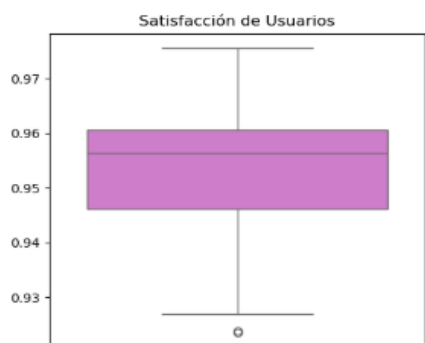
### Figura 11

*Boxplot Días de Inventario*



### *Boxplot Satisfacción de Usuarios*

Finalmente, de satisfacción de usuarios se puede decir que se mueve en valores muy buenos de satisfacción a nivel general con el 75% de los datos por encima del 94.7% de nivel de satisfacción y un 25% que está por encima del 96% de satisfacción. Si bien hay niveles altos de quejas y hay bajos niveles de cumplimiento en la oportunidad, los usuarios valoran otros aspectos del servicio como lo son la atención y la empatía por parte de los colaboradores de la empresa, así como también las condiciones de los establecimientos donde son atendidos y demás variables que en conjunto son tenidas en cuenta al momento de dar una calificación.

**Figura 12***Boxplot Satisfacción de Usuario***Participación Nivel Tasa de Quejas**

El gráfico de pie muestra la distribución porcentual de las categorías creadas para la variable nivel de tasa de quejas. El nivel de tasa de quejas denominado “Medio” agrupa el 52.1% de los datos obtenidos de las quejas, es decir, el 52.1% de los datos se encuentra entre 9 y 12.5. Es importante recordar que la meta es mantener la tasa de quejas lo más baja posible, y solo el 25% de los datos se encuentra en la categoría ideal denominada “Bajo” (con niveles de quejas entre 0 y 9). Además, el 77.1% de los datos cumplen con la premisa de la organización de estar por debajo de 13. Por otra parte, el 22.9% de los datos se encuentran en la categoría “Alto”, siendo este el grupo con menor participación. No obstante, la organización deberá generar estrategias para disminuir la participación de esta categoría y aumentar la del nivel “Bajo”.

**Figura 13**

*Participación Nivel de Tasa de Quejas, Gráfico de Pie*



### **Análisis de Desempeño de Modelos de Regresión**

A continuación, se presentarán y explicarán diversos modelos de regresión con el objetivo de determinar cuál es el más adecuado para desarrollar en el proyecto aplicado. Los modelos de regresión son herramientas estadísticas fundamentales que permiten analizar y predecir la relación entre variables. A través de este análisis, se busca identificar el modelo que mejor se ajuste a los datos y proporcione las predicciones más precisas. Se evaluarán diferentes enfoques, considerando tanto la simplicidad del modelo como su capacidad predictiva, para seleccionar el modelo óptimo que contribuirá al éxito del proyecto. Se hace una partición de los datos en 70% (33 datos) entrenamiento y 30% (15 datos) prueba.

#### **Modelo de Regresión Lineal Múltiple (Modelo 4v)**

En el primero modelo evaluado se toman todas las variables tácticas (Nivel de servicio, pendientes, días de inventario y satisfacción usuarios) para explicar la variable dependiente tasa de quejas.

En este modelo de regresión lineal múltiple, se evidencia que hay una explicación del 53.1% de la variabilidad de la variable tasa de quejas en el conjunto de entrenamiento por parte de las variables en mención. Aunque este no es un gran resultado, logra explicar una parte considerable de la variabilidad. No obstante, en el valor de  $R^2$  ajustado hay una disminución, pasando a un 46.5%, lo que sugiere que el modelo está siendo penalizado por incluir variables que no aportan significativamente.

De los p-valores de cada una de las variables se puede decir que Nivel de servicio (se toma a pesar de estar en el límite) y Días de inventario son las únicas variables que son significativas al ser menores que el valor de significancia de referencia de 0.05. Cuando se observa el valor del coeficiente de cada una de las dos variables mencionadas la magnitud de

nivel de servicio es la que más aporta. A pesar de tener significancia en el modelo hay que mencionar que en la práctica días de inventario debería tener un signo negativo.

Por otra parte, los p-valores altos (mayores a 0.05) para Omnibus y Jarque Bera (0.904 y 0.985) sugieren que los residuos del modelo siguen una distribución normal, lo cual es positivo ya que da cumplimiento a uno de los supuestos de la regresión lineal conocido como Normalidad de los errores. Siguiendo con los supuestos de la regresión lineal, se tiene la prueba Durbin-Watson la cual establece que valores cercanos a 2 muestran independencia de errores, para este caso el resultado fue de 1.65, siendo un valor relativamente cercano a 2 parece que este supuesto se cumple en este modelo.

**Figura 14***Resumen Detallado Ajustado Modelo 4v Utilizando Datos de Entrenamiento*

OLS Regression Results						
Dep. Variable:	Tasa de quejas	R-squared:	0.531			
Model:	OLS	Adj. R-squared:	0.465			
Method:	Least Squares	F-statistic:	7.941			
Date:	Tue, 03 Dec 2024	Prob (F-statistic):	0.000207			
Time:	11:27:55	Log-Likelihood:	-64.084			
No. Observations:	33	AIC:	138.2			
Df Residuals:	28	BIC:	145.6			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	143.7934	111.755	1.287	0.209	-85.126	372.713
Nivel de servicio	-195.6184	98.775	-1.980	0.058	-397.949	6.712
Pendientes	12.2454	10.061	1.217	0.234	-8.363	32.854
Días de inventario	0.2762	0.110	2.521	0.018	0.052	0.501
Satisfacción usuarios	46.7895	32.382	1.445	0.160	-19.542	113.121
Omnibus:	0.203	Durbin-Watson:	1.630			
Prob(Omnibus):	0.904	Jarque-Bera (JB):	0.031			
Skew:	0.069	Prob(JB):	0.985			
Kurtosis:	2.941	Cond. No.	1.74e+04			

Teniendo en cuenta que el conjunto de entrenamiento es el mismo en todos los modelos, los resultados de R2 en este segmento mostraron una tendencia bajista en la medida que se fueron quitando variables.

***Evaluación del Modelo 4v***

En el conjunto de prueba se logra tener un R2 de 68% lo cual es un mejor resultado en comparación con el conjunto de entrenamiento, esto se puede interpretar como subajuste del modelo dado que no logra comprender la variabilidad en el conjunto de entrenamiento, pero logra generalizar mejor. Una posible causa sería la cantidad de datos con los que se cuenta.

La diferencia media de 1.5 entre los datos reales y los predichos reflejado por el error medio absoluto muestra que en caso de usar este modelo Audifarma debería tener en cuenta ese tamaño de flexibilización para los datos que se prediga.

- El R2 es: 0.6811363345082038
- El error cuadrático medio es: 3.1712734854090616
- El RMSE es: 1.7808069758985845
- El error medio absoluto es: 1.5383649974733276

### Figura 15

#### Comparación de Tasas de Quejas Reales y Predichas para el Modelo 4v

Nº	Nivel de servicio	Pendientes	Días de inventario	Satisfacción usuarios	\	Nº	Tasa_de_quejas_real	Tasa_Predicha	Error
34	0.984475	0.352523	36.880000	0.969970	34	10.440000	11.098749	-0.658749	
24	0.991939	0.431889	39.770000	0.923883	24	7.680000	9.252278	-1.572278	
39	0.982887	0.388273	36.820000	0.962463	39	13.287097	11.479463	1.807634	
37	0.990378	0.379171	46.290000	0.972647	37	9.078519	12.994525	-3.916006	
3	0.987827	0.432829	37.599277	0.940924	3	7.640000	10.265953	-2.625953	
4	0.990842	0.424460	34.524406	0.959788	4	10.310000	9.607082	0.702918	
32	0.985916	0.378102	32.380000	0.956520	32	11.460000	9.258006	2.201994	
27	0.987975	0.422278	35.140000	0.952498	27	10.410000	9.970270	0.439730	
45	0.973125	0.276200	38.181405	0.965700	45	13.769970	12.544129	1.225841	
28	0.982975	0.374385	37.250000	0.957337	28	9.880000	11.171045	-1.291045	
29	0.983987	0.377278	35.910000	0.940658	29	8.860000	9.858030	-0.998030	
48	0.950052	0.233210	42.922862	0.974800	48	20.736182	18.266399	2.469783	
22	0.987751	0.427680	34.450000	0.946147	22	8.640000	9.592464	-0.952464	
35	0.986289	0.378669	37.190000	0.965991	35	12.010000	10.963595	1.046405	
20	0.989106	0.389513	36.960000	0.923495	20	9.660000	8.493354	1.166646	

*Nota.* Se muestra el error de predicción para cada observación en el conjunto de prueba.

#### Modelo de Regresión Lineal Múltiple sin Nivel de Servicio (Modelo 3v)

En este segundo modelo se tiene en cuenta todas las variables independientes a excepción del Nivel de servicio.

Aquí se evidencia que hay una explicación de 46,6% de la variabilidad de la tasa de quejas en el conjunto de entrenamiento por parte de las variables seleccionadas. Este resultado de R2 es moderado bajo y muestra al igual que en el modelo anterior que en el conjunto de entrenamiento hay subajuste y al tener menos variables la explicación de la variable objetivo es menor.

De los p-valores de cada una de la variable se puede decir que Días de inventario es la única variable que es significativa ya que con un resultado de p-valor de 0.001 es menor que el valor de significancia de referencia de 0.05. Sin embargo al observar el coeficiente no es significativamente diferente de 0 además aporta positivamente, incoherente en la práctica.

Por otra parte, los p-valores altos (mayores a 0.05) para Omnibus y Jarque Bera (0.45 y 0.58) sugieren que los residuos del modelo siguen una distribución normal, lo cual es positivo ya que da cumplimiento al supuesto de Normalidad de los errores. Siguiendo con los supuestos de la regresión lineal, se tiene la prueba Durbin-Watson la cual establece que valores cercanos a 2 muestran independencia de errores, para este caso el resultado fue de 1.487, siendo un valor relativamente lejano a 2 parece que este supuesto no se cumple en este modelo.

**Figura 16***Resumen Detallado Ajustado Modelo 3v Utilizando Datos de Entrenamiento*

OLS Regression Results						
=====						
Dep. Variable:	Tasa de quejas	R-squared:		0.466		
Model:	OLS	Adj. R-squared:		0.411		
Method:	Least Squares	F-statistic:		8.431		
Date:	Tue, 03 Dec 2024	Prob (F-statistic):		0.000350		
Time:	11:27:56	Log-Likelihood:		-66.247		
No. Observations:	33	AIC:		140.5		
Df Residuals:	29	BIC:		146.5		
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	-70.1178	30.093	-2.330	0.027	-131.664	-8.571
Pendientes	-3.9664	6.137	-0.646	0.523	-16.517	8.584
Días de inventario	0.3715	0.103	3.598	0.001	0.160	0.583
Satisfacción usuarios	72.2275	31.187	2.316	0.028	8.443	136.012
=====						
Omnibus:	1.572	Durbin-Watson:		1.487		
Prob(Omnibus):	0.456	Jarque-Bera (JB):		1.090		
Skew:	0.139	Prob(JB):		0.580		
Kurtosis:	2.155	Cond. No.		4.80e+03		
=====						

*Evaluación del Modelo 3v*

En el conjunto de prueba se logra tener un R2 de 32,6% lo cual es un resultado más bajo que el obtenido en el conjunto de entrenamiento, es decir que este modelo en ninguno de los dos segmentos logra capturar la variabilidad de la tasa de quejas de forma significativa.

- El R2 es: 0.3261782221630488
- El error cuadrático medio es: 6.7015259786646855
- El RMSE es: 2.588730572822264
- El error medio absoluto es: 1.8331910779027845

**Figura 17***Comparación de Tasas de Quejas Reales y Predichas para el Modelo 3v*

Pendientes	Días de inventario	Satisfacción usuarios	\	Tasa_de_quejas_real	Tasa_Predicha	Error
N°			N°			
34	0.352523	36.880000	0.969970	34	10.440000	12.242845 -1.802845
24	0.431889	39.770000	0.923883	24	7.680000	9.672848 -1.992848
39	0.388273	36.820000	0.962463	39	13.287097	11.536543 1.750554
37	0.379171	46.290000	0.972647	37	9.078519	15.826177 -6.747658
3	0.432829	37.599277	0.940924	3	7.640000	10.093576 -2.453576
4	0.424460	34.524406	0.959788	4	10.310000	10.346976 -0.036976
32	0.378102	32.380000	0.956520	32	11.460000	9.498257 1.961743
27	0.422278	35.140000	0.952498	27	10.410000	10.057789 0.352211
45	0.276200	38.181405	0.965700	45	13.769970	12.720594 1.049376
28	0.374385	37.250000	0.957337	28	9.880000	11.381141 -1.501141
29	0.377278	35.910000	0.940658	29	8.860000	9.667138 -0.807138
48	0.233210	42.922862	0.974800	48	20.736182	15.309755 5.426426
22	0.427680	34.450000	0.946147	22	8.640000	9.321310 -0.681310
35	0.378669	37.190000	0.965991	35	12.010000	11.966858 0.043142
20	0.389513	36.960000	0.923495	20	9.660000	8.769080 0.890920

*Nota.* Se muestra el error de predicción para cada observación en el conjunto de prueba.

**Modelo de Regresión Lineal días de Inventario y Nivel de Servicio (Modelo 2v)**

Tomando como base lo obtenido en el Modelo 4v donde a nivel de coeficientes de variables mostró que las variables más significativas para predecir la tasa de quejas en Audifarma eran Días de inventario y Nivel de servicio se procede con el planteamiento del modelo con estas dos variables.

Aquí se evidencia que hay una explicación del 45.6% de la variabilidad de la variable tasa de quejas en el conjunto de entrenamiento por parte de las variables seleccionadas, lo cual es un resultado moderado bajo y el R2 ajustado disminuye un poco pasando a 42%, esto sugiere que el modelo está siendo penalizado por incluir variables que no aportan significativamente.

En los p-valores se puede observar que las dos variables son significativas ya que sus resultados son menores que el valor de significancia de referencia de 0.05. Sin embargo al

observar el coeficiente, Días de inventario no es significativamente diferente de 0 además continúa aportando positivamente, incoherente en la práctica.

Por otra parte, los p-valores altos (mayores a 0.05) para Omnibus y Jarque Bera (0.84 y 0.83) sugieren que los residuos del modelo siguen una distribución normal, lo cual es positivo ya que da cumplimiento a uno de los supuestos de la regresión lineal conocido como Normalidad de los errores. Siguiendo con los supuestos de la regresión lineal, se tiene la prueba Durbin-Watson la cual establece que valores cercanos a 2 muestran independencia de errores, para este caso el resultado fue de 1.407, siendo un valor relativamente lejano a 2 parece que este supuesto no se cumple en este modelo.

**Figura 18***Resumen Detallado Ajustado Modelo 2v Utilizando Datos de Entrenamiento*

OLS Regression Results						
Dep. Variable:	Tasa de quejas	R-squared:	0.456			
Model:	OLS	Adj. R-squared:	0.420			
Method:	Least Squares	F-statistic:	12.57			
Date:	Tue, 03 Dec 2024	Prob (F-statistic):	0.000108			
Time:	11:27:56	Log-Likelihood:	-66.550			
No. Observations:	33	AIC:	139.1			
Df Residuals:	30	BIC:	143.6			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	134.6351	55.994	2.404	0.023	20.279	248.991
Nivel de servicio	-137.3500	54.908	-2.501	0.018	-249.487	-25.213
Días de inventario	0.3163	0.110	2.888	0.007	0.093	0.540
Omnibus:	0.326	Durbin-Watson:	1.407			
Prob(Omnibus):	0.849	Jarque-Bera (JB):	0.373			
Skew:	0.211	Prob(JB):	0.830			
Kurtosis:	2.696	Cond. No.	8.76e+03			

*Evaluación del Modelo 2v*

Al aplicar las métricas de evaluación para el Modelo 2v en el conjunto de prueba hay una mejora en el resultado obteniendo un 55,1% de R2 lo que sugiere que el modelo se desempeña mejor con resultados desconocidos que con los datos que sí conoce.

- El R2 es: 0.5511914666471796
- El error cuadrático medio es: 4.463646240947282
- El RMSE es: 2.112734304390233
- El error medio absoluto es: 1.791663637619802

## Figura 19

### Comparación de Tasas de Quejas Reales y Predichas para el Modelo 2v

Nº	Nivel de servicio	Días de inventario	Tasa_de_quejas_real	Tasa_Predicha	Nº	Error
34	0.984475	36.880000	10.440000	11.081215	34	-0.641215
24	0.991939	39.770000	7.680000	10.970034	24	-3.290034
39	0.982887	36.820000	13.287097	11.280430	39	2.006667
37	0.990378	46.290000	9.078519	13.246506	37	-4.167987
3	0.987827	37.599277	7.640000	10.848252	3	-3.208252
4	0.990842	34.524406	10.310000	9.461682	4	0.848318
32	0.985916	32.380000	11.460000	9.460143	32	1.999857
27	0.987975	35.140000	10.410000	10.050259	27	0.359741
45	0.973125	38.181405	13.769970	13.051793	45	0.718177
28	0.982975	37.250000	9.880000	11.404290	28	-1.524290
29	0.983987	35.910000	8.860000	10.841535	29	-1.981535
48	0.950052	42.922862	20.736182	17.720343	48	3.015838
22	0.987751	34.450000	8.640000	9.862775	22	-1.222775
35	0.986289	37.190000	12.010000	10.930179	35	1.079821
20	0.989106	36.960000	9.660000	10.470446	20	-0.810446

### Modelo de Regresión Lineal Simple (RLS)

Se plantea este modelo teniendo en cuenta que la variable Nivel de servicio es la variable que más correlación muestra con la tasa de quejas de acuerdo con los datos proporcionados. Además, en el modelo 2v fue la variable con mayor significancia por su p-valor y por su coeficiente.

En el conjunto de entrenamiento, al tener solo una variable, el resultado del R2 disminuye aún más, llegando al 30.5%, lo cual es un resultado bajo. A pesar de ello, el resultado del p-valor y del coeficiente muestra que la variable tiene una significancia y un aporte importante en la explicación de la tasa de quejas.

Por otra parte, los p-valores altos (mayores a 0.05) para Omnibus y Jarque Bera (0.93 y 0.99) sugieren que los residuos del modelo siguen una distribución normal, lo cual es positivo ya que da cumplimiento a uno de los supuestos de la regresión lineal conocido como Normalidad de los errores. Siguiendo con los supuestos de la regresión lineal, se tiene la prueba Durbin-Watson

la cual establece que valores cercanos a 2 muestran independencia de errores, para este caso el resultado fue de 1.42, siendo un valor relativamente lejano a 2 parece que este supuesto se cumple en este modelo. Sin embargo se más adelante se muestra pruebas para validarlo.

## Figura 20

### *Resumen Detallado Ajustado de Modelo RLS Utilizando Datos Entrenamiento*

OLS Regression Results						
Dep. Variable:	Tasa de quejas	R-squared:	0.305			
Model:	OLS	Adj. R-squared:	0.282			
Method:	Least Squares	F-statistic:	13.59			
Date:	Tue, 03 Dec 2024	Prob (F-statistic):	0.000867			
Time:	11:27:57	Log-Likelihood:	-70.598			
No. Observations:	33	AIC:	145.2			
Df Residuals:	31	BIC:	148.2			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	212.2081	54.640	3.884	0.001	100.770	323.646
Nivel de servicio	-204.1396	55.383	-3.686	0.001	-317.094	-91.185
Omnibus:	0.132	Durbin-Watson:	1.426			
Prob(Omnibus):	0.936	Jarque-Bera (JB):	0.013			
Skew:	0.013	Prob(JB):	0.994			
Kurtosis:	2.908	Cond. No.	296.			

### *Evaluación del modelo de RLS*

Al aplicar las métricas de evaluación para el Modelo de Regresión Lineal Simple en el conjunto de prueba hay una mejora significativa del R2 pasando de un 30,5% a un 74.49% con esto se puede hacer un planteamiento de subajuste del modelo, dado que no responde muy bien a los valores conocidos, de tal forma que parece no entender su variabilidad pero en el conjunto de prueba se desempeña mucho mejor obteniendo buenas predicciones y por lo tanto un valor moderado alto para este indicador, por lo tanto es un modelo que se puede usar para predecir la tasa de quejas.

- El R2 es: 0.7449269544213621
- El error cuadrático medio es: 2.536840894174827
- El RMSE es: 1.592746337046432
- El error medio absoluto es: 1.3305385248612411

## Figura 21

### *Comparación de Tasas de Quejas Reales y Predichas para Modelo RLS*

Nº	Nivel de servicio	Tasa_de_quejas_real	Tasa_Predicha	Error
34	0.984475	10.440000	11.237731	-0.797731
24	0.991939	7.680000	9.714040	-2.034040
39	0.982887	13.287097	11.562022	1.725075
37	0.990378	9.078519	10.032770	-0.954251
3	0.987827	7.640000	10.553388	-2.913388
4	0.990842	10.310000	9.937910	0.372090
32	0.985916	11.460000	10.943601	0.516399
27	0.987975	10.410000	10.523336	-0.113336
45	0.973125	13.769970	13.554823	0.215147
28	0.982975	9.880000	11.543991	-1.663991
29	0.983987	8.860000	11.337450	-2.477450
48	0.950052	20.736182	18.264842	2.471340
22	0.987751	8.640000	10.569019	-1.929019
35	0.986289	12.010000	10.867535	1.142465
20	0.989106	9.660000	10.292357	-0.632357

## Resultados de las Métricas para cada Modelo

En la tabla resumen se observan los resultados de las métricas de evaluación para cada uno de los modelos en el conjunto de prueba, la conclusión es que el modelo Regresión Lineal Simple en términos generales (R2, significancia, coherencia y simplicidad) es el modelo que mejores resultados obtuvo luego del análisis de los datos proporcionados por la Audifarma y por tanto debería ser el modelo que la empresa utilice para la predicción de la tasa de quejas.

## Figura 22

*Resultados de las Métricas de Evaluación para cada uno de los Modelos*

	Modelo	R2	MSE	RMSE	MAE
0	Regresión Lineal Múltiple 4V	0.681136	3.171273	1.780807	1.538365
1	Regresión Lineal Múltiple 3V	0.326178	6.701526	2.588731	1.833191
2	Regresión Lineal Múltiple 2V	0.551191	4.463646	2.112734	1.791664
3	Regresión Lineal Simple	0.744927	2.536841	1.592746	1.330539

El modelo 4V a pesar de tener el mejor resultado de R2 en el conjunto de entrenamiento, tiene un R2 menor en el conjunto de prueba y además incluye variables que no aportan significativamente el modelo según sus resultados de p-valor, la magnitud y los signos de los coeficientes. Caso similar ocurre con el modelo 3V y 2V sus resultados de R2 son mejores en que el de RLS contienen variables no significativas y con coeficientes incoherentes.

### Gráfica de Regresión Lineal: Tasa de Quejas vs Nivel de Servicio

A continuación, se presenta gráfica de regresión lineal que relaciona la tasa de quejas con el nivel de servicio. Los datos de entrenamiento están representados por puntos azules, mientras que los datos de prueba se muestran en puntos verdes. La línea roja representa la línea de regresión calculada a partir de estos datos. La ecuación de la línea de regresión es:

$$\text{Tasa de quejas} = -204.1396x + 212.2081$$

$$\text{Valor de } R^2 = 0.7449$$

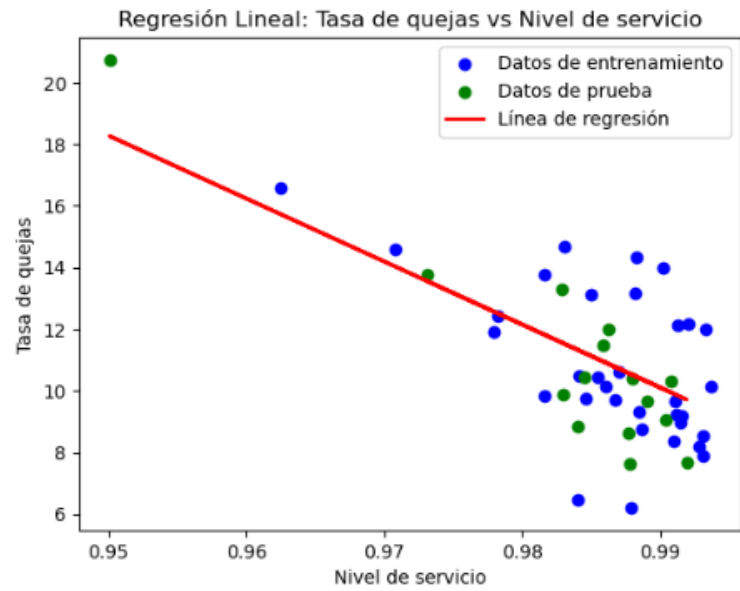
Esta visualización es relevante porque permite observar cómo el nivel de servicio impacta en la tasa de quejas, proporcionando una herramienta basada en datos históricos para mejorar la calidad del servicio.

**Figura 23**

*Regresión Lineal Tasa de Quejas vs Nivel de Servicio*

$$\text{Tasa de quejas} = -204.1396x + 212.2081$$

$$R^2 = 0.7449$$



## **Supuestos del Modelo de Regresión Lineal**

En este capítulo, se presentan y explican los supuestos fundamentales del modelo de regresión lineal, los cuales son esenciales para garantizar la validez y precisión del modelo.

### **Linealidad**

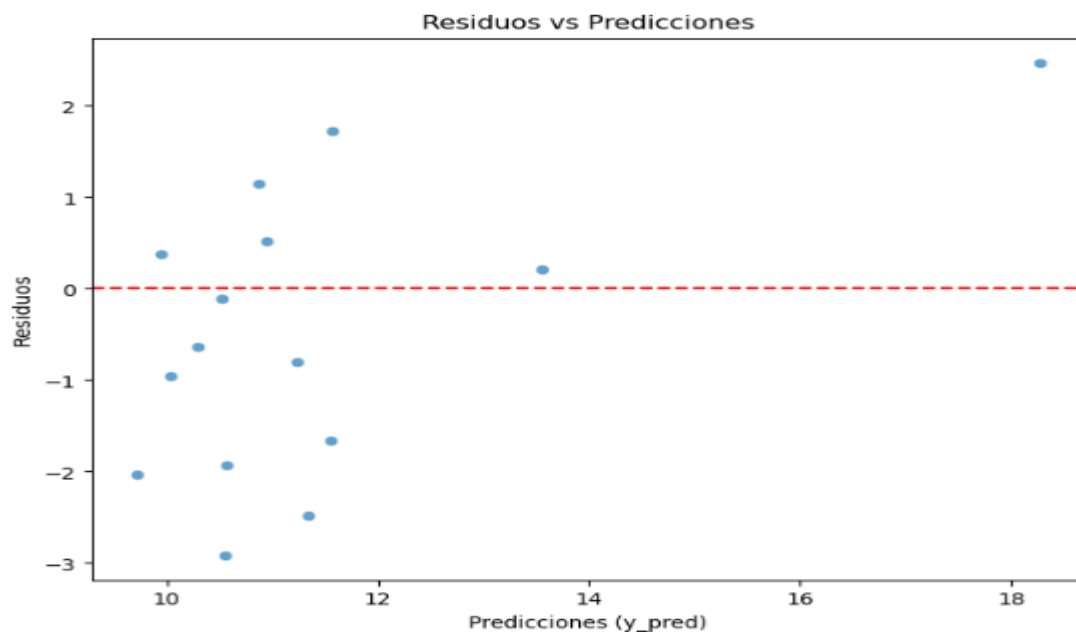
La mayoría de los puntos se encuentran cerca de la recta, lo que indica que existe una relación lineal entre las variables

### **Independencia de los Errores**

Para identificar la independencia de los errores, se utiliza la prueba de Durbin-Watson. Valores cercanos a dos indican independencia; en este caso, el resultado es 2.5, lo que sugiere independencia de los errores. Estadístico de Durbin-Watson: 2.5918488415119962.

### **Homocedasticidad**

Se observa aleatoriedad en los residuos alrededor del cero, sin ningún patrón aparente, lo que indica homocedasticidad.

**Figura 24***Supuesto de Homocedasticidad*

De igual forma se aplica la prueba Breusch-Pagan para obtener un valor objetivo, en este caso el P-Valor es de 0.27 teniendo en cuenta que la hipótesis nula afirma que los residuos son homocedásticos (la varianza es constante) y que el resultado del P-Valor es superior al nivel de significancia de 0.05 no hay evidencia estadística suficiente para rechazar la hipótesis nula. Por lo tanto, se puede concluir que hay homocedasticidad en los residuos.

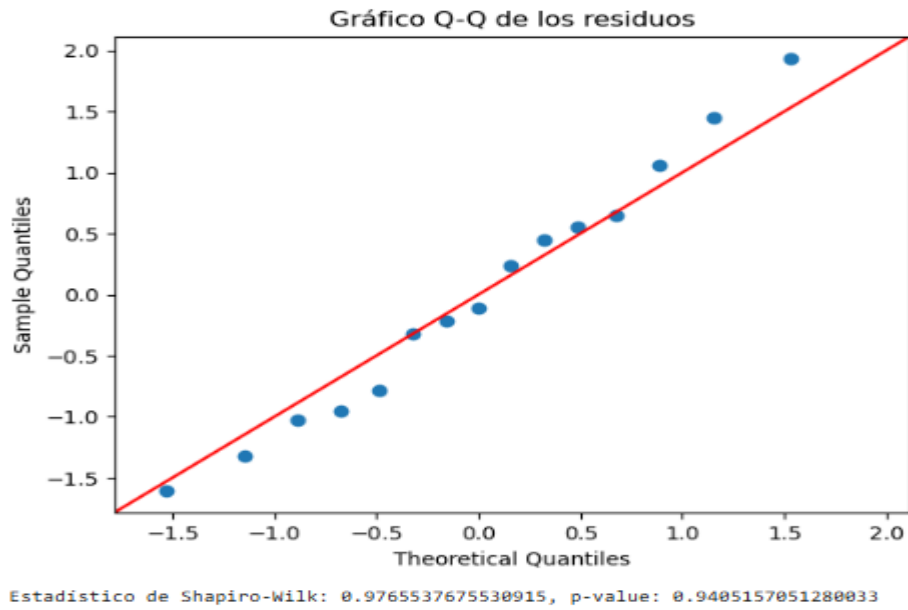
```
{'LM Statistic': 1.2115356696112378, 'p-value': 0.27102819958801755, 'F-Statistic': 1.1422565506612894, 'F p-value': 0.30462040274790375}
```

### **Normalidad de Errores**

Validada con la prueba Shapiro-Wilk, donde el P-Valor es 0.94. Este valor, al ser mayor que 0.05, permite aceptar la hipótesis nula de que los residuos siguen una distribución normal.

**Figura 25**

*Supuesto de Normalidad de Errores*

**Multicolinealidad**

En este caso, no aplica dado que solo se tiene una variable independiente.

## Conclusiones

Del análisis de correlación de las variables tácticas (Nivel de Servicio, Pendientes, Días de Inventario y Satisfacción Usuarios) con la tasa de queja se puede concluir que la variable Nivel de Servicio muestra la mejor correlación con la tasa de quejas con un valor  $-0.71$  además muestra coherencia práctica dado que en la medida que Audifarma garantice mejores niveles de servicio podrá entregar una mayor cantidad de medicamentos a sus usuarios y por lo tanto se esperaría que al probabilidad de obtener quejas por la no entrega de productos disminuya. Algo similar sucede con La variable Pendientes, ya que muestra una correlación práctica coherente a pesar de su valor moderado-bajo de  $-0.45$ , la coherencia práctica se da porque la oportunidad mide el porcentaje de cumplimiento de la promesa de entrega a los usuarios cuando van a la farmacia y no se les entrega el medicamento en el primer acercamiento, en ese momento Audifarma se compromete en entregar los medicamentos pendientes en 48 horas. la lectura es que a mayor cantidad de cumplimiento en los pendientes menor debería ser el nivel de tasa de quejas dado que un usuario que recibe su medicamento cuando se le prometió no tendría razones para quejarse.

Caso contrario sucede con las variables Días de Inventario y Satisfacción Usuarios, las cuales muestran resultados de correlación moderados-bajos, pero con incoherencia práctica. En el caso de la satisfacción los datos muestran una correlación directamente proporcional con la tasa de quejas, pero en la práctica no se obtienen mayores niveles de quejas porque los usuarios están más satisfechos, aquí los resultados se ven un poco alterados porque si bien los niveles de quejas han venido creciendo la satisfacción se ha mantenido en valores muy buenos (esto se puede observar en las gráficas de cajas), esto se da porque los usuarios valoran otras cosas del servicio

como lo son la atención empática de los trabajadores, las instalaciones donde son atendidos, la entrega de información clara, la facilidad para acceder a los servicios, entre otras cosas.

Finalmente, Días de Inventario muestra una correlación moderada-baja positiva que en la práctica tampoco tiene sentido porque mayores días de inventario significa tener mayor cantidad de producto disponible en inventario para satisfacer la demanda

Durante el desarrollo se identificó la necesidad de implementar cuatro modelos para posteriormente evaluarlos y definir el mejor de acuerdo a sus resultados en las métricas de desempeño definidas ( $R^2$ , MSE, RMSE y MAE). El primer modelo (Modelo Regresión Lineal Múltiple 4v) contempló las cuatro variables tácticas para la predicción de la tasa de quejas, el segundo modelo (Modelo Regresión Lineal Múltiple 3v) tuvo en cuenta todas las variables excepto la variable Nivel de servicio. Aquí se pretendía observar el desempeño del modelo sin la variable con mayor correlación. El tercer modelo (Modelo Regresión Lineal Múltiple 2v) tuvo en cuenta las variables Días de Inventario y Nivel de servicio, las cuales fueron seleccionadas por sus resultados de significancia en el primer modelo y finalmente el cuarto modelo (Modelo Regresión Lineal Simple) solo tuvo en cuenta la variable Nivel de Servicios por su alto nivel de correlación. Una vez definidos entrenados los modelos se hace la comparación de resultados en las métricas de desempeño mencionadas y los mejores resultados los obtuvo el cuarto modelo, por lo tanto, este es el mejor modelo obtenido de acuerdo con los datos proporcionados por Audifarma S.A para la predicción de la tasa de quejas.

Se seleccionó el cuarto modelo debido a su parsimonia, un principio que sugiere que, entre modelos con un rendimiento similar, se debe seleccionar el modelo más simple. Este modelo es el que tiene el mejor  $R^2$  en pruebas, (aunque en entrenamiento no obtuvo el mejor  $R^2$ ). Para mejorar esta métrica, se podrían utilizar técnicas como la validación cruzada (cross-

validation), que permite evaluar la capacidad del modelo para generalizar a nuevos datos y asegurar su robustez. Los resultados de las métricas de desempeño mostraron que el cuarto modelo no solo es más sencillo, sino que también ofrece una mayor precisión y generalización en los datos de prueba, lo que lo convierte en la opción más adecuada para la predicción de la tasa de quejas en Audifarma S.A.

Una vez seleccionado el modelo de Regresión Lineal Simple, se procedió con la validación de los supuestos de la regresión lineal, obteniendo los siguientes resultados: La linealidad (supuesto 1), la independencia de los errores (supuesto 2), la homocedasticidad (supuesto 3), la normalidad de errores (supuesto 4) y la multicolinealidad (supuesto 5).

Estos resultados proporcionan robustez a las conclusiones obtenidas, ya que el modelo cumple con todos los supuestos de la regresión lineal. En conclusión, el modelo de Regresión Lineal Simple es adecuado y tiene completa aplicabilidad para la predicción de la tasa de quejas en Audifarma S.A., garantizando la validez y fiabilidad de los resultados obtenidos.

## Recomendaciones

A partir del análisis y los resultados obtenidos en este estudio, se han identificado varias recomendaciones que pueden contribuir a mejorar la gestión y el desempeño de la empresa. Estas recomendaciones no solo buscan optimizar los procesos actuales, sino también preparar a la organización para enfrentar posibles escenarios futuros. A continuación, se presentan las recomendaciones específicas basadas en los hallazgos del estudio.

Se recomienda que Audifarma pueda continuar trabajando en el fortalecimiento del modelo alimentándolo con nuevos datos históricos con el fin de dotarlo de una mayor capacidad de entender la variabilidad de la tasa de quejas. Este fortalecimiento se puede dar incluyendo los datos de los meses transcurridos durante el año 2024 o podría considerarse la opción de trabajarlo con datos semanales y no mensuales como se trabajó en este proyecto de grado. Otra idea sería poder en futuras investigaciones aplicar Cross validation para aumentar el tamaño del conjunto de entrenamiento y maximizar de esta forma el conjunto de los datos disponibles.

Considerar otras variables diferentes a las tácticas que se trabajaron en este proyecto podría dar pie a una nueva investigación que permita encontrar nuevos hallazgos que le entreguen información relevante a la empresa para la toma de decisiones basada en datos respecto a la tasa de quejas.

Se recomienda poder identificar lo sucedido en mayo de 2020 y enero de 2024 donde para que la tasa de quejas haya tenido cambios importantes en un solo mes.

Aunque el cuarto modelo de Regresión Lineal Simple fue seleccionado como el mejor modelo para la predicción de la tasa de quejas en Audifarma S.A (a partir de los datos de para análisis que la empresa proporcionó), es importante no descartar los otros modelos evaluados. Los modelos de Regresión Lineal Múltiple (4v, 3v y 2v) pueden ser útiles en escenarios atípicos

o excepcionales que la empresa pueda enfrentar en el futuro. Estos modelos pueden proporcionar información valiosa cuando se presentan cambios en las variables tácticas o cuando se requiere un análisis más detallado de múltiples factores. Por lo tanto, se recomienda mantener estos modelos como opciones viables para futuras evaluaciones y combinaciones con nuevas técnicas de análisis.

Los modelos no seleccionados deben ser considerados como herramientas adicionales que pueden generar alertas tempranas en la empresa. En situaciones donde las condiciones operativas cambien o se presenten nuevos desafíos, estos modelos pueden ofrecer perspectivas diferentes y complementarias al modelo principal. Evaluar estos modelos con nuevas técnicas o combinaciones puede mejorar la capacidad de Audifarma S.A. para adaptarse a diversas circunstancias y optimizar la toma de decisiones; inclusive para futuras investigaciones de la tasa de quejas sería interesante que Audifarma tuviera en cuenta modelos de series de tiempos porque facilita la interpretación de los puntos de inflexión que se presentaron a finales del 2023 y comienzos del 2024, lo cual podría darle un panorama diferente y nuevos modelos para predecir la tasa de quejas

Se recomienda que la empresa siga poniendo sus esfuerzos en la identificación de variables que impactan la tasa de quejas con el fin de tomar decisiones que le permitan tener niveles cada vez más bajos y usuarios cada vez más satisfechos para que se logre la continuidad del negocio en el tiempo.

### Referencias Bibliográficas

- Amat Rodrigo, J. (2016). *Correlación lineal y regresión lineal simple*. *Ciencia de Datos*, 5(2), 45-58. [https://cienciadedatos.net/documentos/24\\_correlacion\\_y\\_regresion\\_lineal](https://cienciadedatos.net/documentos/24_correlacion_y_regresion_lineal)
- Ccorisapra-Quintana, J., Farfán-Inca-Roca, M. C., Ramirez-Palomino, A., Gutierrez-Muñoz, K. D., y Villegas-Casaverde, M. (2020). *Impacto del marketing relacional en la lealtad de los huéspedes de hoteles de corta estancia*. *Revista de Investigación Valdizana*, <https://revistas.unheval.edu.pe/index.php/riv/article/download/1556/1492?inline=1>
- Engel, J., y Sedlmeier, P. (2011). *Correlation and regression in the training of teachers*. In C. Batanero, G. F. Burrill, & C. Reading, *Teaching statistics in school mathematics: Challenges for teaching and teacher education. A joint ICMI/IASE study* (pp. 247-258). Dordrecht: Springer
- Fornell, C., y Wernerfelt, B. (1987). *Defensive marketing strategy by customer complaint management: A theoretical analysis*. *Journal of Marketing Research*, 24(4), 337-346. <https://web.mit.edu/bwerner/www/papers/DefensiveMarketingStrategybyCustomerComplaintManagement-ATheoreticalAnalysis.pdf>
- Gea, M. (2014). *La correlación y regresión en bachillerato: Análisis de libros de texto y del conocimiento de los futuros profesores* (Tesis de doctorado, Universidad de Granada). <https://digibug.ugr.es/bitstream/handle/10481/34257/24031550.pdf?sequence=1>
- Maute, M. F., y Forrester, W. R. (1993). *The structure and determinants of consumer complaint intentions and behavior*. *Journal of Economic Psychology*, 14(June), 219-247. <https://www.sciencedirect.com/journal/journal-of-economic-psychology/vol/14/issue/2>.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introducción al Análisis de Regresión Lineal* (3ra ed.). México: Compañía Editorial Continental. Academia.edu.

[https://www.academia.edu/42811449/Introduccion\\_al\\_Analisis\\_de\\_Regresion\\_Lineal\\_Tercera\\_Edicion\\_Montgomery\\_Peck\\_Vining](https://www.academia.edu/42811449/Introduccion_al_Analisis_de_Regresion_Lineal_Tercera_Edicion_Montgomery_Peck_Vining)

Moliner Velázquez, B. (2004). *La formación de la satisfacción/insatisfacción del consumidor y del comportamiento de queja: Aplicación al ámbito de los restaurantes* (Tesis doctoral). Universitat de València. Departament de Direcció d'Empreses. Juan Jose Renau Piqueras. <https://www.tdx.cat/handle/10803/9662>

Moliner Velázquez, B., Gil Saura, I., Berenguer Contri, G., y Fuentes Blasco, M. (2008). *Determinantes del comportamiento de queja y su importancia en la segmentación de clientes insatisfechos. Cuadernos de Economía y Dirección de la Empresa* [Archivo pdf]. <https://www.redalyc.org/pdf/807/80703605.pdf>

Ovalle Guarín, L. F., & Parra López, P. C. (2019). *Análisis de correlación entre las peticiones, quejas, reclamos y sugerencias (PQRS) y factores asociados a la prestación de servicios de salud en el subsistema de la Policía Nacional*. Universidad Santo Tomás, División de Ciencias Económicas y Administrativas, Facultad de Economía, Especialización en Auditoría de Salud, Sede Bogotá. <https://repository.usta.edu.co/bitstream/handle/11634/21324/2019paulaparra.pdf?sequence=9&isAllowed=y>

Reglamento Estudiantil Universidad Nacional Abierta y a Distancia UNAD. (2013). *Capítulo 8. De las opciones de grado. En Acuerdo 0029 del 13 de diciembre de 2013* (pp. 27-32). [https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2013/COA\\_C\\_ACUE\\_029\\_20131213.pdf](https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2013/COA_C_ACUE_029_20131213.pdf)

Stephens, N., & Gwinner, K. P. (1998). *Why don't some people complain? A cognitive-emotive process model of consumer complaint behavior. Journal of the Academy of Marketing Science, 26(3), 172-189.*

Universidad Nacional Abierta y a Distancia UNAD. (s.f). *Líneas de Investigación*. Consultado el 22 de abril del 2024. <https://academia.unad.edu.co/investigacion-y-productividad-ecbti/lineas>