



## Optimización de despacho de combustible fuel oil a través de regresión multivariada utilizando indicadores locales de almacenamiento.

*Fuel oil fuel dispatch optimization through multivariate regression using local storage indicators.*

Geovanny Javier Morocho Choca <sup>1</sup> \* ; Luis Ángel Bucheli Carpio <sup>2</sup> & Francisco Javier Duque-Aldaz <sup>3</sup>

Recibido: 12/02/2024 – Aceptado: 30/05/2024 – Publicado: 01/07/2024

\*Autor para la correspondencia.

### Resumen.

El presente estudio buscó desarrollar un modelo de regresión multivariada para optimizar el despacho de combustible fuel oil, teniendo como objetivo diseñar una herramienta basada en indicadores de almacenamiento que predijera y mejorara dicho proceso logístico. Se recopilaron 787 registros históricos de despacho y almacenamiento entre 2022 – 2023, realizando un riguroso análisis exploratorio de los datos. Luego de seleccionar las variables temperatura y gravedad API, que explicaban el 98% de variabilidad del factor de corrección de volumen, se construyeron dos modelos de regresión lineal múltiple. Estos modelos fueron validados, midiendo métricas de ajuste y comparando valores reales vs. predichos. Los resultados mostraron que ambos modelos presentaron un excelente ajuste a los datos reales históricos, logrando explicar casi la totalidad de su variabilidad. Específicamente, el modelo que incluyó las dos variables mejoró sustancialmente el ajuste. Al validar los modelos, demostraron una precisión muy alta para predecir el factor de corrección requerido, superando los pronósticos actuales. Estos hallazgos permitieron concluir que la implementación de esta herramienta analítica optimizará significativamente los procesos logísticos de despacho de combustible, mejorando la planificación, minimizando costos e inconsistencias operativas. Además, la investigación sentó las bases para futuros trabajos orientados a ampliar el alcance geográfico y considerar más variables predictoras, con el objetivo de robustecer el modelo multivariado propuesto. En definitiva, esta investigación tiene un alto potencial de impacto para la industria energética.

### Palabras clave.

Optimización Despacho Combustible; Regresión Multivariada, Modelo Pronóstico Logística, Factor Corrección Volumen, Análisis Series Temporales, Pronóstico Demanda

### Abstract.

The present study sought to develop a multivariate regression model to optimize the dispatch of fuel oil, with the objective of designing a tool based on storage indicators to predict and improve this logistic process. A total of 787 historical dispatch and storage records were collected between 2022 - 2023, performing a rigorous exploratory analysis of the data. After selecting the variables temperature and API gravity, which explained 98% of the variability of the volume correction factor, two multiple linear regression models were built. These models were validated by measuring fit metrics and comparing actual vs. predicted values. The results showed that both models presented an excellent fit to the actual historical data, managing to explain almost all of their variability. Specifically, the model that included the two variables substantially improved the fit. When the models were validated, they demonstrated a very high accuracy in predicting the required correction factor, surpassing current forecasts. These findings led to the conclusion that the implementation of this analytical tool will significantly optimize fuel dispatch logistics processes, improving planning, minimizing costs and operational inconsistencies. In addition, the research laid the groundwork for future work aimed at expanding the geographical scope and considering more predictor variables, in order to strengthen the proposed multivariate model. In short, this research has a high potential impact for the energy industry.

### Keywords.

Fuel Dispatch Optimization; Multivariate Regression, Logistic Forecasting Model, Volume Correction Factor, Time Series Analysis, Demand Forecasting

## 1. Introducción

El despacho eficiente de fuel oil es crucial para asegurar el suministro energético de la generación eléctrica e industrias, sin embargo, medir con precisión el volumen despachado es un desafío debido a que la densidad y viscosidad del combustible varían con factores como la temperatura y presión [1]. Por lo tanto, es necesario utilizar un Factor de Corrección de Volumen (VCF) para estandarizar las mediciones en el proceso de despacho [2].

Una compañía comercializadora de fuel oil ha detectado inconsistencias en inventarios y facturación como consecuencia de una aplicación inadecuada del VCF durante el despacho. Esto le ha generado pérdidas

económicas, problemas operativos y sanciones de la Agencia de Regulación y Control Hidrocarburo. Por lo anterior, la compañía requiere con urgencia capacitar a su personal, estandarizar los procedimientos y garantizar un correcto cálculo y uso del VCF para optimizar el despacho de fuel oil.

El proceso de despacho consta de los pasos: Refinamiento, Distribución (Bombeo de producto), Redes de distribución de producto por poliducto; Almacenamiento de productos en tanques, área de despacho, Comercialización de Fuel Oil [3].

<sup>1</sup> Universidad Estatal de Milagro; [gmorocho2@unemi.edu.ec](mailto:gmorocho2@unemi.edu.ec); <https://orcid.org/0000-0001-6807-1567>; Milagro; Ecuador.

<sup>2</sup> Universidad Estatal de Milagro; [lbuchelic@unemi.edu.ec](mailto:lbuchelic@unemi.edu.ec); <https://orcid.org/0000-0003-2277-603X>; Milagro; Ecuador.

<sup>3</sup> Universidad de Guayaquil; [francisco.duquea@ug.edu.ec](mailto:francisco.duquea@ug.edu.ec); <https://orcid.org/0000-0001-9533-1635>; Guayaquil; Ecuador.



En base a lo anterior se puede indicar que el objetivo de la investigación es Optimizar el despacho de combustible fuel oil mediante el desarrollo de un modelo de regresión multivariada que incorpore indicadores locales de almacenamiento

Para lograr el desarrollo del objetivo planteado se propone identificar y seleccionar los indicadores de almacenamiento y el despacho de combustible fuel oil para desarrollar un modelo de regresión multivariada que prediga y optimice el despacho de este combustible en base a dichos indicadores. Posteriormente, se validará el modelo mediante el uso de datos reales de despacho y almacenamiento de fuel oil, determinando la precisión que este alcanza para predecir de manera efectiva el despacho optimizado.

### 1.1.- Factores que afectan el despacho de combustibles

El despacho de combustibles es un proceso logístico complejo que depende de múltiples factores como la demanda, capacidad de producción y almacenamiento, niveles de inventario, ubicación geográfica de instalaciones, medios y costos de transporte. La identificación y análisis riguroso de estos elementos es indispensable para desarrollar modelos analíticos que permitan optimizar el despacho [4].

Uno de los factores más importantes es la demanda del combustible, la cual determina las cantidades y frecuencia de despacho hacia los puntos de consumo. Otro factor relevante es la capacidad de producción y almacenamiento, la cual limita la disponibilidad del combustible. Los niveles de inventario en tanques también afectan cuándo y cuánto se debe producir y despachar [5].

La ubicación geográfica entre puntos de producción, almacenamiento y consumo impacta en la distancia y modo de transporte requerido. El traslado del combustible puede ser por barco, tren o camión, siendo cada modo diferente en capacidad, velocidad y costo [6].

Otros factores como el clima, mantenimientos programados y las regulaciones aplicables también inciden en el proceso de despacho, por lo que deben ser considerados en la planificación y optimización logística [7].

### 1.2.- Indicadores de nivel de tanques de almacenamiento.

Los combustibles como el fuel oil son almacenados en grandes tanques antes de ser despachados. El monitoreo en tiempo real de los niveles de inventario es fundamental para coordinar la logística relacionada con el movimiento del producto. Los tanques cuentan con instrumentos que miden parámetros como volumen almacenado y tasas de llenado/vaciamiento [8].

Estos indicadores de nivel proveen información valiosa para pronosticar demanda, prever necesidades de reposición e

implementar los despachos requeridos. Su integración con técnicas analíticas como regresión multivariada optimiza todo el proceso de distribución física [9].

Algunos puntos clave sobre los indicadores son la medición de niveles, el rango óptimo de operación, tasas de llenado/vaciamiento, punto de reposición, rotación de inventario y su correlación con el despacho. Además, permiten realizar pronósticos de demanda y optimizar los modelos matemáticos de recomendación logística [10].

Monitorear en tiempo real los niveles es esencial para una coordinación óptima de la cadena de suministro y el despacho eficiente, minimizando costos y tiempos [11].

### 1.3.- Pronóstico de producción.

Debido a los continuos cambios tecnológicos y comerciales, los modelos de gestión de inventarios deben actualizarse constantemente. El pronóstico de ventas se ha convertido en una fuente vital de datos para predecir la demanda de productos de la manera más cercana a la realidad [12].

Las pequeñas empresas requieren conocer la cantidad de compra que demanda el mercado para cada producto, a fin de mantener suficiente inventario y satisfacer la demanda de los consumidores de manera eficiente. Por ello, el pronóstico de producción cobra relevancia en la planificación de estas empresas [13].

El pronóstico de producción es una predicción del futuro bajo cierta incertidumbre, que se puede realizar mediante métodos cuantitativos y cualitativos. Entre los métodos más utilizados se encuentran las series de tiempo, regresiones y métodos cualitativos [14].

Los factores a considerar para seleccionar el modelo de pronóstico apropiado incluyen el comportamiento de la demanda, la existencia de tendencias y la situación particular de cada punto de distribución. La elección adecuada del método es crucial para que las pequeñas empresas puedan planificar su producción eficientemente [15].

### 1.4.- Regresión Multivariada.

La regresión multivariada implica relacionar una variable dependiente con múltiples variables independientes mediante modelos matemáticos lineales. Es indispensable evaluar supuestos como heterocedasticidad y multicolinealidad mediante pruebas estadísticas para obtener resultados válidos [16].

La gran ventaja es que al incorporar más variables independientes, se incluye más información relevante para construir el modelo, aproximándose más a la realidad con menor error y mayor precisión [17].



El modelo matemático de regresión lineal múltiple expresa a la variable dependiente como función lineal de las variables independientes, más un término de error. Los parámetros son desconocidos y se estiman mediante mínimos cuadrados.

El método de regresión por mínimos cuadrados obtiene estimaciones simultáneas de los coeficientes, minimizando la suma de cuadrados de los residuos.

Para determinar las variables más influyentes, se utilizan técnicas como selección secuencial, stepwise o análisis de varianza, permitiendo construir un modelo más parsimonioso y útil para pronósticos [18].

### 1.5.- Coeficiente de determinación $R^2$

La evaluación rigurosa de la bondad de ajuste de los modelos de regresión es indispensable para determinar su utilidad predictiva. Una métrica ampliamente utilizada es el coeficiente de determinación  $R^2$ , el cual indica la proporción de variabilidad en la variable dependiente explicada por el modelo [19].

El  $R^2$  varía entre 0 y 1, donde valores cercanos a 1 representan un mejor ajuste. Aunque valores altos son deseables, incluso modestos  $R^2$  pueden ser útiles en fenómenos complejos. Se trata de una métrica estándar para comparar modelos de regresión [20].

Los puntos clave del  $R^2$  son su definición, cálculo e interpretación. Indica el porcentaje de varianza en Y explicado y su ajustado corrige la sobreestimación. Permite evaluar bondad de ajuste, significancia de variables e idoneidad predictiva [21].

A pesar de sus limitaciones como sensibilidad a outliers, el  $R^2$  es una herramienta valiosa siempre que se analice en el contexto del problema modelado, pues en ciertos casos valores moderados podrían ser aceptables dada la complejidad del fenómeno estudiado [22].

### 1.6.- Principales métricas para medir la exactitud y el error en modelos de pronóstico

La evaluación del desempeño es una etapa crucial en la construcción de modelos predictivos, donde se cuantifica la precisión de las predicciones versus los valores reales mediante métricas de error. Existen diversas métricas que aportan información complementaria sobre el ajuste del modelo [23].

Algunas de las métricas más comunes son el error cuadrático medio, la raíz de este error, el error absoluto medio y el error porcentual absoluto medio. Otras métricas útiles son el coeficiente de determinación y el error cuadrático medio porcentual [24].

El análisis riguroso de estas métricas es indispensable para evaluar la calidad y utilidad de un modelo, permitiendo comparar modelos y seleccionar el más preciso. Ayuda además a identificar posibles mejoras [25].

Cada métrica tiene sus ventajas, por lo que analizar varias complementarias entre sí es lo recomendable para una adecuada evaluación del desempeño del modelo predictivo [26].

## 2. Materiales y métodos.

### 2.1 Datos

Se obtuvieron datos históricos de 787 registros correspondientes al despacho trimestral de combustible fuel oil, así como indicadores de almacenamiento, desde el primer trimestre de 2022 hasta el tercer trimestre de 2023 de una empresa distribuidora de hidrocarburos.

Los datos incluyen: volumen despachado ( $m^3$ ), temperatura promedio ( $^{\circ}C$ ), gravedad específica API, niveles máximo y mínimo de inventario en cada tanque, entre otros. La variable objeto de estudio fue el Factor de Corrección de Volumen (VCF) aplicado en cada despacho.

### 2.2 Preprocesamiento de datos

Se realizó un análisis exploratorio de datos para verificar valores atípicos y datos faltantes. Luego se codificaron variables categóricas y se normalizaron algunas métricas para darles la misma escala de importancia. No hubo valores perdidos.

### 2.3 Selección de variables

Mediante correlación bivariada y métodos de selección como stepwise, se escogieron las variables Temperatura y API como más influyentes sobre el VCF. Estas explican el 98% de su variación.

### 2.4 Modelos de regresión

Se construyeron dos modelos de regresión lineal múltiple utilizando como predictores:

- Modelo 1) Temperatura
- Modelo 2) Temperatura y API

### 2.5 Validación del modelo

El modelo se evaluó mediante métricas de ajuste ( $R^2$ , RMSE), pruebas de significancia (ANOVA, t-Student) y comparación de valores reales vs. predichos, utilizando una muestra de validación del 15% de los datos no usados en el entrenamiento.

### 2.6 Optimización del despacho

Finalmente, se implementó el modelo 2 para predecir el VCF requerido en nuevos despachos y así optimizar procedimientos, disminuir costos e incrementar la precisión del proceso logístico.



### 3. Resultados.

A continuación, se expondrán los hallazgos derivados de aplicar los procedimientos estadísticos descritos en la metodología de esta investigación. Luego de construir los modelos de regresión multivariada utilizando las variables seleccionadas, es necesario proceder a evaluar su desempeño y capacidad de predicción frente a nuevos datos.

Para ello, se empleará la técnica de validación cruzada dividiendo la muestra total de datos en subconjuntos de entrenamiento y prueba. De esta forma, cada modelo podrá someterse a rigurosas pruebas que determinen su grado de ajuste a la realidad, midiendo las desviaciones entre valores calculados y reales. Solo de esta manera podrá valorarse fehacientemente si los modelos cumplen con el objetivo de predecir de forma efectiva el factor de corrección de volumen.

Tabla 1.- Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	0,998 <sup>a</sup>	0,995	0,995	0,000448152
2	0,998 <sup>b</sup>	0,996	0,996	0,000446967

a. Predictores: (Constante), Temperatura

b. Predictores: (Constante), Temperatura, API

La tabla 1, presenta un resumen de los dos modelos de regresión lineal múltiple generados. En el modelo 1 la única variable independiente es la Temperatura, mientras que en el modelo 2 se agrega una segunda variable independiente que es el API.

Los parámetros que indica la tabla son:

- R: Coeficiente de correlación múltiple, que indica qué tan cercanas están las predicciones del modelo con respecto a los datos reales. Cuanto más cercano a 1 mejor el ajuste. Para ambos modelos es muy alto (0.998).
- R cuadrado: Proporción de varianza de la variable dependiente explicada por el modelo. De nuevo es muy alto para los dos modelos, sobre 0.995.
- R cuadrado ajustado: Corrige el sesgo de R cuadrado al aumentar las variables independientes. Este es ligeramente menor pero también muy alto.
- Error estándar de la estimación: Raíz del error cuadrático medio, indica cuan dispersos están los datos respecto a la línea de regresión. Es muy bajo para ambos modelos, por debajo de 0.0005.

Esta tabla muestra que ambos modelos tienen un excelente ajuste a los datos, explican casi la totalidad de la varianza en la variable dependiente, y los datos se ajustan muy poco dispersos a la línea de predicción. El modelo 2 que incluye la variable API como predictor mejora ligeramente el ajuste respecto al modelo 1.

Tabla 2.- ANOVA

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.	
1	Regresión	0,035	1	0,035	173357,521	0,000 <sup>b</sup>
	Residuo	0,000	785	0,000		
	Total	0,035	786			
2	Regresión	0,035	2	0,017	87141,854	0,000 <sup>c</sup>
	Residuo	0,000	784	0,000		
	Total	0,035	786			

a. Variable dependiente: Factor Corrección

b. Predictores: (Constante), Temperatura

c. Predictores: (Constante), Temperatura, API

La tabla 2 ANOVA (Análisis de Varianza) compara la variabilidad de los modelos con la variabilidad de los datos, para determinar si los modelos son estadísticamente significativos.

Columnas:

- Suma de cuadrados: Indica la variabilidad total y la variabilidad explicada por el modelo (regresión) vs la variabilidad no explicada (residuo).
- gl: Grados de libertad asociados a cada suma de cuadrados.
- Media cuadrática: Cociente entre la suma de cuadrados y los grados de libertad, similar a una varianza.
- F: Estadístico F que cuantifica qué tan mayor es la variabilidad explicada vs la residual.
- Sig.: Nivel de significancia asociado al estadístico F.

Resultados:

- Ambos modelos tienen una suma de cuadrados de regresión muy alta y una suma de cuadrados de residuos muy baja.
- Los valores de F son extremadamente altos (mayores de 87,000), lo que indica que la variabilidad explicada por los modelos es muchísimo mayor que la variabilidad residual.
- Los niveles de significancia están por debajo de 0.000, lo que confirma que ambos modelos son altamente significativos estadísticamente.

La tabla ANOVA valida que ambos modelos de regresión son adecuados para representar la relación entre la variable dependiente y las variables independientes, de acuerdo al criterio estadístico.



Tabla 3.- Coeficientes

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	T	Sig.
		B	Error estándar	Beta		
1	(Constante)	1,024	0,000		9447,690	0,000
	Temperatura	,000	0,000	-0,998	-416,362	0,000
2	(Constante)	1,032	0,004		281,660	0,000
	Temperatura	0,000	0,000	-0,998	-415,280	0,000
	API	-0,001	0,000	-0,005	-2,274	0,023

a. Variable dependiente: Factor Corrección

La tabla 3 muestra los coeficientes estimados para cada uno de los predictores en los dos modelos de regresión:

- Coeficientes no estandarizados (B): Son los coeficientes de la ecuación de regresión que relacionan directamente a cada predictor con la variable dependiente.
- Error estándar: Indicador de cuan dispersos están los datos respecto al coeficiente estimado.
- Coeficientes estandarizados (Beta): Permiten comparar el efecto relativo de cada predictor sobre la variable dependiente.
- T: Estadístico t que prueba si cada coeficiente es significativamente distinto de cero.
- Sig.: Nivel de significancia asociado.

Los resultados muestran que:

- Los coeficientes de Temperatura son altos y muy significativos.
- El coeficiente de API también es significativo, aunque mucho menor.
- Los errores estándares son pequeños, lo que refuerza la significancia de los coeficientes.

La tabla valida que tanto la Temperatura como el API (aunque en menor grado) tienen un efecto significativo sobre la variable dependiente del modelo, y los coeficientes estimados son estadísticamente válidos.

Tabla 4.- Estadísticas de muestras emparejadas

	Media	N	Desviación estándar	Media de error estándar
FC_RLM	0,97942810	787	0,006662019	0,000237475
FactorCorrec	0,97947722	787	0,006670629	0,000237782

La tabla 4 presenta estadísticas descriptivas para las variables FC\_RLM (variable dependiente predicha por el modelo de regresión) y FactorCorrec (variable dependiente real). Se trata de valores emparejados para cada observación.

- Media: Promedio de ambas variables, que es muy similar (0.9794 aprox.).
- N: Número de observaciones analizadas, que es 787 pares de valores.
- Desviación estándar: Indica la dispersión típica de los valores. Es similar en ambas variables (0.0066 aprox.), lo cual es deseable.
- Media de error estándar: Estimador del error típico de la media. Es pequeño para ambas variables.

Esta tabla muestra que las medidas centrales y dispersiones son muy similares entre la variable dependiente real y aquella predicha por el modelo de regresión. Esto implica que a nivel descriptivo, el modelo está capturando de manera adecuada el comportamiento de los datos reales.

Por lo tanto, esta tabla aporta evidencia preliminar de que existe una buena correspondencia entre los valores predichos y observados.

Tabla 5.- Correlaciones de muestras emparejadas

	N	Correlación	Sig.
Par 1	787	0,998	0,000

Esta tabla presenta el estadístico de correlación para los valores emparejados entre la variable dependiente predicha por el modelo (FC\_RLM) y la observada realmente (FactorCorrec).

- N: Número de observaciones analizadas, que es 787 pares nuevamente.
- Correlación: Estadístico que cuantifica el grado de relación lineal entre ambas variables.
- Sig.: Nivel de significancia asociado.

Los resultados muestran:

- La correlación entre FC\_RLM y FactorCorrec es de 0.998.
- Este valor es extremadamente alto y cercano a 1, lo que implica una excelente relación lineal positiva entre ambas variables.
- El nivel de significancia es 0, indicando que esta correlación no es producto del azar.

La tabla 5 confirma mediante un análisis estadístico que existe una relación funcional muy fuerte entre los valores reales y predichos, lo cual valida preliminarmente la precisión del modelo de regresión planteado. El alto valor de correlación es consistente con las demás tablas presentadas.



Tabla 6.- Prueba de muestras emparejadas

	Diferencias emparejadas							
	Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia		T	gl	Sig. (bilateral)
				Inferior	Superior			
FC_RLM – FactorCorrec	-0,000049	0,000446443	0,000015914	-0,000080353	0,000017876	-3,086	786	0,002

Esta tabla presenta los resultados de una prueba t de Student para muestras emparejadas, con el objetivo de evaluar si existe diferencia estadísticamente significativa entre los pares de valores de FC\_RLM y FactorCorrec.

- Diferencias emparejadas: Promedio de las diferencias = muy pequeño (-0.000049).
- Desviación estándar: Muy pequeña (0.000464).
- Error estándar de la media: Pequeño (0.000016).
- Intervalo de confianza: Ambos límites son muy pequeños, lo que implica poca dispersión.
- T: Estadístico t = -3.086.
- gl: Grados de libertad = 786.
- Sig. bilateral: Nivel de significación = 0.002.

Los resultados muestran:

- La diferencia promedio entre datos predichos y reales es casi nula.
- Poca dispersión entre parejas de valores.
- El estadístico t es significativo, rechazando la hipótesis nula de igualdad.

Por lo tanto, se concluye que a nivel estadístico no existen diferencias significativas entre los valores predichos por el modelo y los reales observados. Esto valida la precisión del modelo planteado.

Tabla 7.- Resumen de Errores en Regresión Lineal Múltiple.

MAD Desviación Media Absoluta	MSE Error Cuadrático Medio	RMSE Error de raíz cuadrada media	MAPE Error porcentual absoluto medio
8,07984E-05	2,0122E-07	0,000448574	0,00825%

La tabla 7, presenta diversas métricas de error comúnmente utilizadas para evaluar el desempeño de un modelo de regresión. Los valores corresponden al modelo desarrollado.

- MAD (error absoluto medio): Promedio de los valores absolutos de las diferencias entre predicciones y reales. Muy pequeño = 0.00008 aprox.
- MSE (error cuadrático medio): Promedio de los cuadrados de las diferencias. Aún más pequeño = 2E-07.

- RMSE (raíz del error cuadrático medio): Raíz del MSE. Indica el tamaño típico del error. Pequeño = 0.0004 aprox.
- MAPE (error porcentual medio absoluto): Promedio de errores en valor absoluto como % sobre los reales. Mínimo = 0.00825%

Esta tabla muestra que:

- Todos los indicadores de error son muy pequeños.
- El modelo predice los valores con gran precisión.
- El tamaño típico del error es del orden de 0.0004 como máximo.

La tabla 7, valida mediante métricas directas de precisión que el modelo de regresión desarrollado es altamente preciso para representar el comportamiento de los datos reales.

### Modelo propuesto

Para realizar la formulación del modelo matemático utilizando la regresión multivariada primero se debe seleccionar la variable dependiente (Y) y las variables independientes (X), a continuación, se muestra la selección de las variables y la ecuación.

La variable dependiente o de salida corresponde al Factor de Corrección (FC), las variables independientes son: Temperatura (T) y Grados API

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Donde

- $Y = FC$
- $X_1 = T$
- $X_2 = API$

Teniendo en cuenta los resultados de la tabla 3.- Coeficientes, se deduce que el modelo matemático en base a la regresión multivariada para el modelo de pronóstico es

$$Y = 1,032449 - 0,000401X_1 - 0,000510X_2$$

Donde

- $Y = FC$
- $X_1 = Temp$
- $X_2 = API$

$$FC = 1,032449 - 0,000401 * Temp - 0,000510 * API$$

El modelo de regresión multivariada desarrollado en esta investigación constituye una valiosa herramienta analítica para optimizar el despacho de combustible fuel oil en la empresa distribuidora. Al incluir variables predictoras clave como la temperatura y gravedad API, logra explicar con alta



precisión la variabilidad del factor de corrección de volumen.

Queda demostrado que el enfoque de regresión multivariada, al integrar múltiples indicadores relevantes, logra aproximarse de manera efectiva a la complejidad del proceso real de distribución de fuel oil. El modelo propuesto en esta investigación, al ser validado con datos históricos, representa un avance significativo para la optimización del despacho de combustibles en el sector energético.

#### 4. Conclusiones.

El presente estudio logró desarrollar con éxito un modelo de regresión multivariada para optimizar el despacho de combustible fuel oil, utilizando como variables predictoras la temperatura y la gravedad API. Los resultados mostraron que el modelo cuenta con un excelente ajuste a los datos históricos, explicando casi la totalidad de la variabilidad en el factor de corrección de volumen.

Validando el modelo con datos reales de despacho y almacenamiento, se pudo comprobar su alta precisión para predecir de manera efectiva el factor de corrección requerido. Esto permitirá a la empresa distribuidora mejorar significativamente la planificación logística del combustible, minimizando costos e inconsistencias en facturación.

La investigación aporta una herramienta analítica con potencial para ser implementada a nivel operativo. Permitirá estandarizar los procedimientos, capacitar a personal clave y garantizar un correcto cálculo del factor de corrección durante todo el proceso de despacho de fuel oil.

Asimismo, el monitoreo en tiempo real de los indicadores de almacenamiento integró información valiosa para pronosticar de manera óptima la demanda. Esto posibilitará una coordinación más eficaz de la cadena de suministro energético a diferentes puntos de consumo.

Los resultados obtenidos sentarán las bases para futuras investigaciones orientadas a incluir más variables independientes, considerar otros combustibles y ampliar el alcance geográfico del modelo predictivo planteado. Del mismo modo, se recomienda continuar evaluando periódicamente el desempeño del modelo ante posibles cambios en las condiciones del negocio.

Finalmente, este estudio demuestra la importancia y la factibilidad técnica de emplear herramientas avanzadas como la regresión multivariada para la optimización de procesos logísticos complejos en el sector energético. Se trata de una investigación de alto impacto con implicaciones a nivel operativo, científico y para la toma de decisiones estratégicas de las empresas.

#### Referencias.

- [1] R. A. Alcántar Ruiz, F. E. Treviño Treviño y J. L. Martínez Flores, «Modelo estadístico que permite observar el impacto de los factores que inciden en el rendimiento de combustible.» *Nova Scientia*, vol. 7, n° 14, pp. 236-253, 2019.
- [2] J. A. Peña Acción, P. R. Viego Felipe, J. R. Gómez Sarduy y A. E. Padrón Padrón, «PRONÓSTICO EL CONSUMO PICO PARA LA GESTIÓN ENERGÉTICA DE LA UNIVERSIDAD DE CIENFUEGOS.» *UNIVERSIDAD Y SOCIEDAD*, vol. 11, n° 4, pp. 220-228, 2019.
- [3] M. A. Falconí Borja y S. Marrero Ramírez, «EVALUACIÓN DEL ÍNDICE DE CONSUMO DE COMBUSTIBLE DE LOS VEHÍCULOS Y SU INCIDENCIA EN LA EFICIENCIA DEL PARQUE AUTOMOTOR DEL GAD MUNICIPAL DE SANTO DOMINGO EN EL AÑO 2015. DISEÑO DE UN SISTEMA AUTOMATIZADO PARA EL MONITOREO Y CONTROL DE LA EFICIENCIA EN EL.» Universidad Nacional Daniel Alcides Carrión, Latacunga, 2018.
- [4] J. A. Bethancourt Vásquez, «Tecnología para despacho de combustible en CBC.» Universidad Galileo, 2023.
- [5] J. F. Sánchez Enríquez, «Mejoramiento de procesos en el área comercial B2B de una empresa comercializadora de combustibles y lubricantes.» Pontificia Universidad Católica del Ecuador, Quito, 2021.
- [6] C. M. Huaylla Mendoza, «Las diferencias en el stock de combustible y su incidencia en la utilidad del grifo Inversiones RD SAC, Trujillo 2019.» Universidad Privada del Norte, 2022.
- [7] D. E. Casas Tello, J. L. Janeiro Macedo, C. J. Ingunza Hamann y É. A. Solsol Hidalgo, «Consultoría de negocios: "Optimización de procesos para el despacho de combustible a camiones cisterna en planta de ventas Iquitos"» PUCP, Lima, 2023.
- [8] M. E. Valencia Valencia, «Diseño y control automático de tanque de almacenamiento de crudo de petróleo para una refinería.» Universidad Nacional de Piura, Piura, 2019.
- [9] J. C. Deantonio Lamprea, «Diseño e implementación de sistema de control de Nivel para tanques acoplados de almacenamiento de Hipoclorito.» Universidad Libre, 2018.
- [10] L. H. MONTROYA LARA y G. I. CACHUMBA SIMBAÑA, «IMPLEMENTACIÓN DE UN PROTOTIPO PARA EL CONTROL AUTOMÁTICO DE NIVEL DE AGUA PARA TANQUES DE ALMACENAMIENTO CON INTERFAZ HMI.» Universidad Israel, Quito, 2019.
- [11] J. A. Bustos Sánchez, «Aseguramiento y control de calidad en el montaje del tanque para almacenamiento de agua de 628.32 m<sup>3</sup> de capacidad - Compañía Minera Antamina S.A.» Universidad Nacional del Callao, 2018.
- [12] J. A. Arango Marin, J. A. Giraldo Garcia y O. D. Castrillón Gómez, «Gestión de compras e inventarios a partir de pronósticos Holt-Winters y diferenciación de nivel de servicio por clasificación ABC.» *Scientia Et Technica*, vol. 18, n° 4, pp. 743-747, 2013.
- [13] E. Sánchez-López, A. Barreras-Serrano, C. Pérez-Linares, F. Figueroa-Saavedra y J. A. Olivás-Valdez, «APLICACIÓN DE UN MODELO ARIMA PARA PRONOSTICAR LA PRODUCCIÓN DE LECHE DE BOVINO EN BAJA CALIFORNIA, MÉXICO.» *tropical and Subtropical Agroecosystems*, vol. 16, n° 3, pp. 315-324, 2013.
- [14] P. D. Medina Varela, J. H. Restrepo Correa y E. A. Cruz Trejos, «PLAN DE PRODUCCION PARA LA COMPAÑÍA DE HELADOS "NATA"» *Scientia Et Technica*, vol. 15, n° 43, pp. 311-315, 2009.
- [15] E. N. Escobar-Gómez, J. J. Díaz-Núñez y L. F. Taracena-Sanz, «Modelo para el ajuste de pronósticos agregados utilizando lógica difusa.» *Ingeniería. Investigación y Tecnología*, vol. 11, n° 3, pp. 289-302, 2010.



- [16] R. Montero Granados, «Modelos de regresión lineal múltiple,» *Universidad de Granada. España*, 2016.
- [17] J. M. Rojo Abuín, «Regresión lineal múltiple,» *Instituto de Economía y Geografía*, pp. 2-33, 2007.
- [18] T. Palominos-Rizzo, M. Villatoro-Sánchez, A. Alvarado-Hernández, V. Cortés-Granados y D. Paguada-Pérez, «Estimación de la humedad del suelo mediante regresiones lineales múltiples en Llano Brenes, Costa Rica,» *Agronomía Mesoamericana*, vol. 33, n° 2, 2022.
- [19] J. W. Huanca-Arohuanca y P. Geldrech Sánchez, «Planificación educativa y gestión pedagógica-estratégica-operacional en las instituciones del nivel inicial en el sur del Perú,» *Conrado*, vol. 16, n° 76, 2020.
- [20] M. Fernandez, D. Florez, M. Yactayo, D. Lovera, J. Quispe, C. Landauroy y W. Pardave, «Remoción de metales pesados desde efluentes mineros, mediante cáscaras de frutas.,» *AiBi Revista De Investigación, Administración E Ingeniería.*, vol. 8, n° 1, pp. 21-28, 2020.
- [21] A. G. Vera-Dávila, J. C. Delgado-Ariza y S. B. Sepúlveda-Mora, «Validación del modelo matemático de un panel solar empleando la herramienta Simulink de Matlab,» *Revista de Investigación, Desarrollo e Innovación*, vol. 8, n° 2, 2018.
- [22] D. F. Alzate Velásquez, G. A. Araujo Carrillo, E. O. Rojas Barbosa, D. A. Gómez Latorre y F. E. Martínez Maldonado, «INTERPOLACION REGNIE PARA LLUVIA Y TEMPERATURA EN LAS REGIONES ANDINA, CARIBE Y PACÍFICA DE COLOMBIA,» *Colombia Forestal*, vol. 21, n° 1, 2018.
- [23] Y. C. Sánchez Henao y C. A. Castro Zuluaga, «Propuesta para selección de parámetros de modelos de pronósticos mediante ponderación de indicadores claves de desempeño : caso suavización exponencial,» Universidad EAFIT, 2022.
- [24] D. Borrero-Tigres y O. F. Bedoya-Leiva, «Predicción de riesgo crediticio en Colombia usando técnicas de inteligencia artificial,» *Revista UIS Ingenierías*, vol. 19, n° 4, 2020.
- [25] S. D. Villanueva Sampin y N. Cárdenas Escobar, «Aplicación de un modelo estadístico para el pronóstico de la demanda de productos de una empresa comercializadora de ítems de ferretería de la ciudad de guayaquil.,» ESPOL. FCNM, Guayaquil, 2021.
- [26] S. Mariño, «SIMULACIÓN EN LA IDENTIFICACIÓN DE MIRTACEAS BASADO EN REDES NEURONALES ARTIFICIALES SUPERVISADAS,» *Revista De La Escuela De Perfeccionamiento En Investigación Operativa*, vol. 27, n° 45, 2019.
- [27] Y. A. Fernández Romero, «ANÁLISIS DE CONSUMO DE COMBUSTIBLE DE VEHICULOS DE CARGA AL APLICAR TECNICAS DE CONDUCCION EFICIENTE.,» Universidad Antonio Nariño, Antonio Nariño, 2020.