

Aplicación de técnicas multivariadas en indicadores de productividad procedentes de la simulación de procesos de producción de una empresa láctea

Application of multivariate techniques to productivity indicators derived from the simulation of production processes in a dairy company

Cristian Flores Cadena ^a  and Mairett Rodríguez Balza ^{1,2} 

¹ Universidad Politécnica Estatal del Carchi, Posgrado, Av. Universitaria y Av. Antisana, Tulcán, 040101

² Interoc S.A. Guayas, Ecuador

Recibido: 26/04/2024, Aceptado: 25/07/2025

Autor de correspondencia: Cristina Flores: cristiana.flores@upec.edu.ec

DOI: <https://doi.org/10.53358/ideas.v7i2.1248>



PALABRAS CLAVE

Multivariate Techniques,
MANOVA, PERMANOVA,
Simulación,
Indicadores de productividad,

RESUMEN

En la industria alimentaria, los procesos de producción presentan problemas con el tiempo de procesamiento y la baja productividad, lo que limita la capacidad operativa y competitiva de las empresas del sector. A pesar de contar con tecnologías de manufactura establecidas, no se dispone de modelos estadísticos integrales que permitan analizar y simular estos procesos de manera precisa para identificar oportunidades de mejora. En este contexto, surge la necesidad de modelar y analizar estadísticamente los procesos de producción del yogurt de frutas, mediante técnicas multivariadas y simulaciones, con el fin de optimizar tiempos y aumentar la eficiencia operativa. Para esto se identificaron las operaciones unitarias del proceso de producción de yogurt de frutas, desarrollando una base de datos estructurada con los tiempos de procesamiento de cada proceso. Se validó la integridad de la base de datos, imputando el 3.3% de los datos faltantes mediante el método KNN, y se verificó la ausencia de valores atípicos mediante la Distancia de Mahalanobis. El análisis de los tiempos de ejecución determinó las distribuciones de probabilidad más adecuadas para cada proceso, con una validación mediante la prueba de bondad de ajuste Chi-cuadrado. Se validó el proceso de simulación mediante una prueba de hipótesis T de Student, comparando el proceso real con el simulado, lo que garantizó su representatividad. Posteriormente, se simuló escenarios de mejora usando el software FlexSim y técnicas de Lean Manufacturing (TPM, 5S, Poka Yoke y eliminación de muda), evaluando el impacto en los tiempos de producción y la cantidad de productos terminados. Los resultados fueron analizados con el método PERMANOVA, dado que no se cumplían los supuestos de normalidad, homocedasticidad e independencia para MANOVA. Los análisis revelaron diferencias estadísticamente significativas entre los escenarios de simulación, destacando el escenario 5S como el más eficiente para reducir el tiempo de producción y aumentar el número de productos. El modelo explicó el 44% de la variabilidad en los resultados, subrayando la efectividad de la metodología 5S en la optimización del proceso.

KEYWORDS

PERMANOVA,
MANOVA,
Simulation,
Productivity indicators,
Lean Manufacturing

ABSTRACT

In the food industry, production processes face issues related to processing time and low productivity, which limit the operational and competitive capacity of companies in the sector. Despite having established manufacturing technologies, there is a lack of comprehensive statistical models that allow for accurate analysis and simulation of these processes to identify opportunities for improvement. In this context, it becomes necessary to statistically model and analyze the fruit yogurt production process using multivariate techniques and simulations, in order to optimize processing times and increase operational efficiency. To achieve this, the unit operations involved in the fruit yogurt production process were identified, and a structured database was developed containing the processing times for each stage. The integrity of the database was validated by imputing 3.3% of the missing data using the KNN method, and the absence of outliers was verified using the Mahalanobis Distance. The analysis of processing times identified the most appropriate probability distributions for each process, which were validated using the Chi-square goodness-of-fit test. The simulation model was validated through a Student's t-test by comparing the real and simulated processes, ensuring its representativeness. Subsequently, improvement scenarios were simulated using FlexSim software and Lean Manufacturing techniques (TPM, 5S, Poka Yoke, and waste elimination), evaluating their impact on production times and the number of finished products. The results were analyzed using the PERMANOVA method, as the assumptions of normality, homoscedasticity, and independence required for MANOVA were not met. The analyses revealed statistically significant differences among the simulated scenarios, with the 5S scenario standing out as the most efficient in reducing production time and increasing output. The model explained 44% of the variability in the results, highlighting the effectiveness of the 5S methodology in optimizing the production process.

1. Introducción

En el ámbito de la industria alimentaria, particularmente en la elaboración de productos lácteos, persisten dificultades asociadas a los tiempos prolongados de procesamiento y a la baja eficiencia en la producción. Estas limitaciones afectan la capacidad de respuesta de las empresas frente a la demanda del mercado y reducen su competitividad [1]. Aunque se cuenta con tecnologías industriales estandarizadas, no se aplican de forma generalizada modelos estadísticos avanzados que permitan analizar el desempeño del proceso de manera integral y detectar oportunidades concretas de mejora continua.

En este contexto, optimizar los procesos productivos se convierte en una necesidad estratégica para garantizar la sostenibilidad y el crecimiento de las empresas del sector. Tradicionalmente, las decisiones en las líneas de producción se han basado en la experiencia empírica o en análisis parciales, sin un respaldo estadístico sólido que permita comprender integralmente el comportamiento de los procesos. Esta falta de modelos analíticos y predictivos limita la identificación de cuellos de botella, el control de variabilidad en los tiempos de procesamiento y la implementación efectiva de mejoras. Por tanto, existe una brecha importante en el uso de herramientas estadísticas avanzadas y de simulación para el análisis de los procesos productivos.

Esta investigación cobra relevancia porque aborda dicha brecha mediante el uso de técnicas estadísticas multivariadas, imputación de datos, detección de valores atípicos, análisis de distribuciones de probabilidad y validación de modelos a través de pruebas de hipótesis. Además, incorpora simulaciones con FlexSim y metodologías de mejora continua como Lean Manufacturing, proporcionando una visión integral del sistema productivo.

La validación del modelo simulado frente a datos reales garantiza que los resultados obtenidos no solo tengan un respaldo teórico, sino también una aplicación práctica, lo que permite a los responsables de producción tomar decisiones basadas en evidencia cuantitativa, reduciendo la incertidumbre y mejorando la capacidad de respuesta ante cambios en la demanda o en las condiciones operativas [2]. Además, la investigación aporta valor metodológico al combinar herramientas estadísticas y de ingeniería industrial para evaluar escenarios de mejora, el cual no solo contribuye a la optimización del tiempo y la productividad, sino que también mejora la organización, limpieza y seguridad en el entorno laboral, generando beneficios sostenibles para la cultura organizacional.

Por otro lado, los indicadores de productividad ayudan a medir el progreso y la eficiencia de los procesos, identificando cuellos de botella y áreas de mejora. Proporcionan datos para optimizar recursos, reducir costos y minimizar impactos ambientales. Además, facilitan la comparación con otras empresas del sector, mejorando la competitividad [3]. Los indicadores de productividad deben estar alineados con los objetivos de la empresa, ser cuantificables, fáciles de interpretar y comparables con estándares previos. Su uso adecuado ayuda a mejorar el desempeño, la competitividad y el crecimiento sostenible [4]. En el ámbito empresarial, la productividad no solo significa producir más con menos, sino hacerlo mejor, alcanzando alta eficiencia sin sacrificar calidad. Los indicadores de productividad permiten a las empresas monitorear operaciones, fijar metas alcanzables y comparar su rendimiento con estándares internos o del mercado [4].

La simulación de procesos industriales permite modelar y optimizar sistemas complejos sin intervenir en el sistema real. Es clave en manufactura, logística, energía y salud para mejorar eficiencia, reducir costos y mitigar riesgos. [5]. La simulación permite prever el impacto de cambios en variables clave, probar estrategias antes de su aplicación y optimizar el flujo de materiales, energía e información al identificar cuellos de botella y puntos críticos [6]. Las simulaciones pueden ser discretas, para eventos específicos en logística y producción; continuas, para procesos con cambios constantes como reacciones químicas; o híbridas, combinando ambas para dinámicas mixta [5].

El desarrollo de software especializado, como Arena, FlexSim, AnyLogic y Simul8, ha hecho más accesible la simulación, permitiendo que empresas de diferentes tamaños se beneficien de esta tecnología. Estas herramientas facilitan el modelado de procesos complejos, la realización de análisis detallados y la visualización de resultados mediante gráficos e informes claros. Por tanto, la simulación no solo es un método avanzado de análisis, sino también un recurso fundamental para la toma de decisiones estratégicas en un entorno industrial competitivo y dinámico [6].

Por otro lado, tenemos el análisis multivariado de la varianza (MANOVA, por sus siglas en inglés) este método amplía el análisis de varianza (ANOVA), al permitir evaluar simultáneamente diferencias entre grupos en múltiples variables dependientes. A diferencia del ANOVA, que examina una sola variable dependiente, el MANOVA considera cómo varios resultados se ven afectados por uno o más factores independientes. Esta metodología resulta especialmente valiosa en disciplinas como psicología, biología, ciencias sociales y marketing, donde es necesario comprender patrones complejos y relaciones entre variables [7].

El MANOVA se fundamenta en la premisa de que las variables dependientes están correlacionadas y, por ello, deben analizarse conjuntamente en lugar de hacerlo de forma individual. Esto mejora la potencia estadística y disminuye la probabilidad de errores tipo I (falsos positivos). Por ejemplo, al evaluar el impacto de un programa educativo, el MANOVA podría analizar simultáneamente el rendimiento de los estudiantes en matemáticas, lectura y ciencias, considerando las interacciones entre estas áreas [8].

Entre sus principales ventajas destaca la capacidad de manejar múltiples variables dependientes de manera eficiente, ofreciendo una perspectiva más completa sobre cómo los factores independientes influyen en los resultados. Además, permite identificar patrones subyacentes en los datos, lo que es crucial para la toma de decisiones informadas. No obstante, el MANOVA presenta ciertos desafíos, como cumplir con estrictos supuestos estadísticos, entre ellos la normalidad multivariada, la homogeneidad de las covarianzas y la independencia de las observaciones [7].

El MANOVA es una herramienta poderosa para abordar preguntas de investigación complejas que involucran múltiples variables dependientes. Su capacidad para analizar estas variables de forma conjunta proporciona una visión más integral y detallada de los datos. Sin embargo, requiere un diseño cuidadoso y un análisis riguroso para garantizar resultados válidos y confiables. Es especialmente relevante en estudios donde las variables están interrelacionadas, y un análisis por separado no capturaría toda la complejidad del fenómeno estudiado [8].

Otra de las técnicas multivariantes es el PERMANOVA (Análisis de Varianza Multivariado por Permutación) es un método estadístico no paramétrico que se utiliza para evaluar si existen diferencias significativas entre grupos de datos multivariantes. A diferencia del análisis de varianza multivariable (MANOVA), el PERMANOVA no requiere que los datos sigan una distribución normal ni que las varianzas sean homogéneas. Este método se basa en permutar las observaciones dentro de los grupos y calcular la estadística F observada para comparar la variabilidad entre los grupos con la variabilidad dentro de los mismos, a través de una serie de permutaciones aleatorias. Esto permite determinar si las diferencias entre grupos son más grandes de lo que se esperaría por azar [9].

El uso de PERMANOVA es particularmente valioso en campos como la ecología, la biología y las ciencias sociales, donde los datos multivariantes son comunes y los supuestos de normalidad no siempre se cumplen. Se utiliza frecuentemente para analizar la composición de comunidades biológicas, diferencias en perfiles genéticos o incluso en estudios de percepción del consumidor. En esencia, este método proporciona una forma robusta de evaluar la significancia de las diferencias entre grupos en contextos complejos, donde otras pruebas podrían no ser apropiadas [9].

El objetivo de la presente investigación es modelar y analizar los procesos de producción de una industria alimentaria, enfocándose específicamente en una empresa dedicada a la elaboración de productos lácteos. A través de la simulación, se busca desarrollar escenarios de mejora que permitan identificar y proponer estrategias para optimizar

los indicadores de productividad, considerando aspectos clave como la eficiencia operativa, la reducción de desperdicios y el incremento en la capacidad de respuesta ante la demanda del mercado.

La ausencia de un enfoque analítico riguroso basado en datos impide optimizar los recursos, reducir desperdicios y mejorar la productividad. Asimismo, las iniciativas de mejora como Lean Manufacturing muchas veces se implementan sin una evaluación cuantitativa de su impacto. Por ello, se requiere un estudio que combine el análisis estadístico multivariado con herramientas de simulación y metodologías de gestión de calidad, con el propósito de diseñar y validar propuestas de optimización que respondan a la realidad operativa del proceso productivo

Para la creación de los escenarios de mejora, se emplearon herramientas fundamentales de Lean Manufacturing, como el TPM (Mantenimiento productivo total), 5 S (clasificación, orden, limpieza, estandarización y disciplina), Poka Yoke y Eliminación de Muda (Transporte y Movimiento), las cuales son esenciales para identificar actividades que no aportan valor y para fomentar la mejora continua en los procesos. Estas herramientas permitieron estructurar alternativas de optimización basadas en datos concretos y alineadas con los principios de eficiencia y sostenibilidad.

Además, se utilizó el análisis multivariado de la varianza (MANOVA) y el Análisis de Varianza Multivariado por Permutación (PERMANOVA) como métodos estadísticos para evaluar simultáneamente las diferencias entre grupos en Múltiples variables dependientes, tales como tiempo de producción y número de productos elaborados. Este enfoque proporcionó una visión integral sobre cómo los factores independientes, como cambios en la distribución de recursos o ajustes en los flujos de trabajo, impactan de manera significativa en los resultados globales del sistema.

Esta investigación combina técnicas avanzadas de simulación, principios de Lean Manufacturing y análisis estadístico robusto para ofrecer un enfoque comprensivo y fundamentado en la mejora de los procesos productivos. Este enfoque no solo contribuye a aumentar la competitividad de la empresa láctea considerada en este estudio, sino que también establece una base sólida para futuras investigaciones y aplicaciones en el ámbito industrial.

1.1. Investigaciones relacionadas

El estudio de Vargas (2020) optimizó el proceso de fabricación en una empresa de ladrillos en Santa Marta mediante simulación. Se evaluaron las mejoras con las variables Costo por lote y "Tiempo en el sistema", analizadas con MANOVA. Los resultados indicaron que un modelo tenía mejor desempeño, por lo que se recomendó su implementación. [10].

Duffoo (2024) analizó los indicadores logísticos de la demanda en la industria farmacéutica mediante análisis estadístico multivariado. El estudio correlacional y explicativo evaluó KPI como disponibilidad del producto, precisión del pronóstico y tiempo de entrega. Usando datos de Kaggle, se encontró que las pruebas multivariadas son efectivas para evaluar la demanda, mostrando que su aumento impacta significativamente los KPI y que estos están correlacionados, destacando su importancia en la gestión logística. [11].

Villanueva (2022) aplicó Lean Manufacturing para reducir el trabajo en proceso por escasez de materiales, usando simulación de eventos discretos y herramientas de Industria 4.0 en FlexSim. Se desarrolló un modelo para evaluar la reducción del inventario en proceso mediante una estrategia de suministro justo a tiempo. Los resultados confirmaron la eficacia de FlexSim, aunque no se especificó el tipo de empresa mexicana analizada [12].

En el estudio realizado por Chud, Bedoya y Paredes (2020), aplicó Lean Manufacturing con Value Stream Mapping (VSM) en una cadena de producción del sector del cuero. Se identificaron problemas como subutilización de espacios y tiempos perdidos. A través de la simulación del VSM, se lograron mejoras en productividad y eficiencia, reduciendo el tiempo de ciclo y el lead time. Aunque el estudio se centró en una curtiembre, propone explorar Lean Manufacturing en la producción de lácteos. [13].

Natividad (2019) combinó Lean Manufacturing y simulación estocástica para optimizar el tiempo en la producción de nuggets en una empresa avícola. Usando Crystal Ball y el método de Montecarlo, se identificaron mejoras en el Value Stream Mapping (VSM), reduciendo el tiempo de proceso y aumentando la productividad en 2,29%. La implementación de Kanban permitió disminuir tiempos con alta confianza estadística. [14]. La presente investigación propone emplear el software FlexSim para analizar la aplicación de estas técnicas en una empresa dedicada a la producción de lácteos.

2. Materiales y Métodos

2.1. Recolección de datos en la empresa láctea

La presente investigación se basó en datos recopilados de una empresa alimentaria especializada en la producción de productos lácteos, como yogur de frutas, queso mozzarella, queso fresco, manjar de leche y leche condensada. Dado que el yogur de frutas es el producto de mayor demanda, se seleccionó su línea de producción para el análisis. En este

proceso, se identificaron las etapas involucradas en su elaboración y se registraron los tiempos correspondientes a cada una. Asimismo, se recopilaron datos de 30 lotes de producción.

2.2. Tratamientos de datos faltantes y atípicos

Al utilizar una base de datos como fundamento para investigaciones u otros estudios, es común encontrar la presencia de datos faltantes y atípicos. Estos elementos, si no se tratan adecuadamente, pueden afectar la calidad del análisis y la validez de los resultados. Por esta razón, se realizó un tratamiento exhaustivo de los datos faltantes y atípicos en la base de datos obtenidos. El procesamiento y tratamiento de los datos se llevó a cabo utilizando el software R Studio [15].

Para el tratamiento de datos faltantes, se empleó la técnica de imputación KNN (K-Nearest Neighbors), que permite estimar los valores ausentes basándose en la similitud entre los registros vecinos más cercanos, preservando así la estructura y las relaciones multivariantes de los datos [15]. Por otro lado, para la identificación y tratamiento de datos atípicos, se utilizó la distancia de Mahalanobis, un método estadístico que mide la distancia entre un punto y la media del conjunto de datos en un espacio multivariado, permitiendo detectar observaciones que se desvían significativamente del patrón general. Estas técnicas aseguran que la base de datos final sea confiable y adecuada para los análisis posteriores, minimizando el impacto de posibles sesgos derivados de datos incompletos o anomalías [16].

2.3. Ajuste de la distribución de probabilidad para los tiempos en cada operación unitaria o proceso

Con el objetivo de replicar el comportamiento del proceso real, se llevó a cabo un análisis exhaustivo de los tiempos de ejecución en la línea de producción. Este análisis permitió identificar la distribución de probabilidad que mejor describen los datos recopilados, utilizando el software R Studio como herramienta principal.

Para garantizar la validez del ajuste, se aplicó una prueba de bondad de ajuste basada en el método Chi-cuadrado. Este enfoque estadístico evalúa si los datos observados se desvían significativamente de los valores esperados según la distribución seleccionada. La elección de este método se debe a su capacidad para validar ajustes de distribuciones en muestras de datos, asegurando que el modelo estadístico sea representativo del sistema real.

Este procedimiento no solo asegura la fidelidad del modelo de simulación, sino que también proporciona una base sólida para el análisis y optimización de los procesos productivos

2.4. Construcción y Validación del proceso de simulación

Para la construcción del modelo de simulación, se realizó un análisis detallado del funcionamiento del sistema real, con el objetivo de replicar fielmente los procesos involucrados en la elaboración de yogurt. Este análisis incluyó la identificación de los componentes clave del sistema, como maquinaria, equipos y personal, así como sus interacciones dentro de la línea de producción.

Se utilizaron elementos de la librería del software FlexSim 19.1.1 para modelar el funcionamiento de componentes esenciales, tales como el tanque de recepción, el filtro, el pasteurizador, los tanques de fermentación, el envasador de yogurt, las galletas, los envases, los recipientes y la cámara de refrigeración. Cada uno de estos elementos fue configurado para representar con precisión su comportamiento real, permitiendo al modelo simular de manera integral el flujo de producción.

Además, se programaron las funciones específicas de los equipos en FlexSim, integrando los ajustes de las distribuciones de probabilidad previamente calculadas para los tiempos asociados a cada operación. Estas distribuciones, obtenidas mediante análisis estadístico, proporcionan una base sólida para reproducir la variabilidad propia del sistema real. Como resultado, el modelo de simulación refleja un comportamiento que emula fielmente las condiciones actuales del sistema de producción, ofreciendo una herramienta robusta para el análisis y la optimización del proceso.

2.5. Creación de escenarios de mejoras con la herramienta Lean Manufacturing

Con el propósito de optimizar los indicadores de productividad, se diseñaron y analizaron diversos escenarios de mejora a partir del modelo base. Estos escenarios incorporan principios y herramientas de la filosofía Lean Manufacturing, como Kaizen, las 5S, las 7 Mudras, Six Sigma, TOC (Teoría de las Restricciones), TQM (Gestión de la Calidad Total), TPM (Mantenimiento Productivo Total), SMED (Cambio Rápido de Herramientas),

La integración de estas estrategias permitió identificar y eliminar desperdicios, reducir la variabilidad en los procesos y

optimizar la utilización de recursos humanos, materiales y técnicos en la línea de producción. Cada escenario de mejora fue simulado utilizando el modelo previamente desarrollado, registrando los nuevos indicadores de productividad obtenidos. Estos resultados evidencian el impacto positivo de la implementación de Lean Manufacturing, proporcionando datos concretos para la toma de decisiones y demostrando su eficacia en la búsqueda.

2.6. Supuestos de MANOVA

El MANOVA (Análisis Multivariante de la Varianza) se basa en tres supuestos fundamentales que deben cumplirse para garantizar la validez de los resultados estadísticos:

- **Distribución normal multivariada de las variables dependientes en cada grupo:** El análisis de normalidad múltiple se realizó mediante el Test de Mardia, Test de Henze-Zirkler, Test de Royston, Test de Doornik-Hansen y Test de Energy.
- **Homogeneidad de la matriz de varianza-covarianza:** La Homogeneidad de la matriz de varianza y covarianza se analizó mediante el Test de Box (M-Box).
- **Independencia de las observaciones:** Este supuesto implica que las observaciones individuales no deben influirse entre sí, asegurando que cada medida sea autónoma dentro de los datos analizados.

2.7. Utilización de la técnica multivariante Manova

Además, se utilizó el análisis multivariado de la varianza (MANOVA) como método estadístico para evaluar simultáneamente las diferencias entre grupos en Múltiples variables dependientes, tales como tiempo de producción y número de productos elaborados. Este enfoque proporcionó una visión integral sobre cómo la variable independiente, tal como escenarios de mejora de la línea de producción, con 5 niveles, tales como TPM, 5 S, Poka Yoke 1, Poka Yoke 2 y Eliminación de muda (transporte y movimiento), impactan de manera significativa en los resultados globales del sistema.

3. Resultados y Discusión

3.1. Recolección de datos en la empresa láctea

Se seleccionó la línea de elaboración de yogurt de frutas por ser el producto de mayor demanda en la empresa. En la figura se puede observar la cantidad de leche destinada a este producto en comparación con los demás que se elaboran.

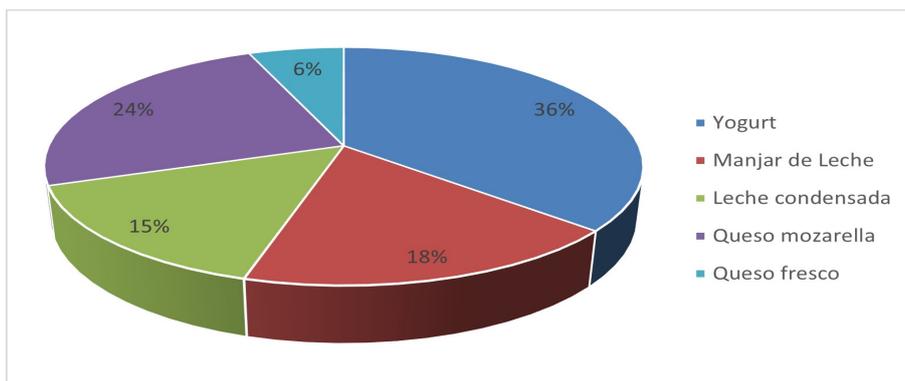


Figura 1: Distribución de la leche utilizada en los diferentes productos elaborados por la empresa

Se identificaron las operaciones unitarias involucradas en el proceso de producción del yogurt de frutas, las cuales incluyen: recepción de la materia prima, enfriamiento, filtración, pasteurizado, enfriamiento, inoculación, agitación, incubación, enfriamiento, agitado, dosificación, envasado y almacenamiento. El diagrama de flujo correspondiente a este proceso se muestra en la Figura 2.

A continuación, se recopiló y organizó una base de datos detallada que incluye los tiempos de procesamiento correspondientes a cada una de las operaciones unitarias identificadas en el proceso de elaboración. Esta información

ANEXO I
Diagrama de bloques del proceso de elaboración del Yogurt

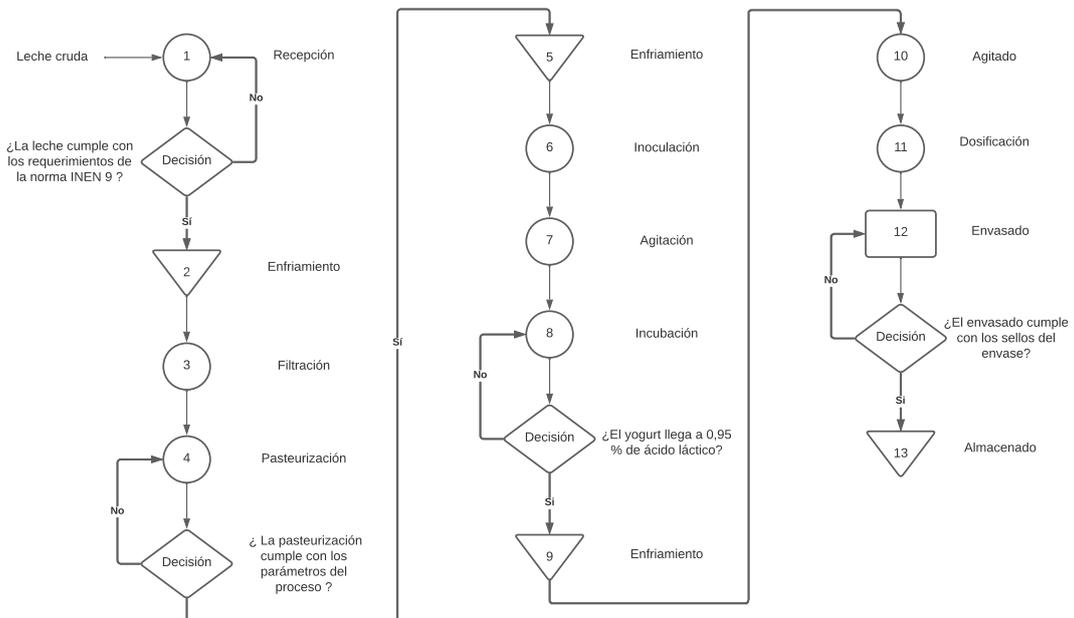
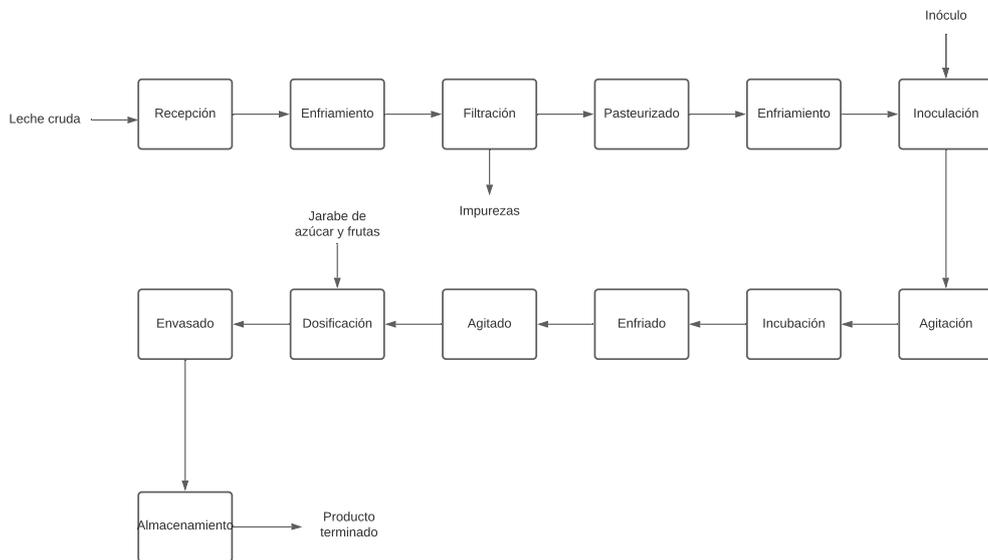


Figura 2: Diagrama de flujo del proceso de elaboración del yogurt

fue registrada de manera sistemática para garantizar su precisión y facilitar el análisis posterior, permitiendo evaluar el desempeño y la eficiencia de cada etapa del proceso.

3.2. Tratamientos de datos faltantes y atípicos

La base de datos utilizada en este estudio constó de 30 observaciones y 13 variables correspondientes a los procesos u operaciones unitarias de la línea de producción. En primera instancia, se identificó la cantidad de datos faltantes, que representaron un 3.3 % del total. Este porcentaje, al ser inferior al 5 %, cumple con el criterio aceptado para proceder con la imputación de datos faltantes sin comprometer la integridad estadística del análisis [17].

Para llevar a cabo dicha imputación, se empleó la técnica KNN (K-Nearest Neighbors), [18] que permite estimar valores faltantes basándose en la similitud con registros cercanos en el espacio multivariado. En total, se imputaron 13 valores ausentes, asegurando la consistencia de la base de datos para los análisis posteriores.

Posteriormente, se realizó un tratamiento exhaustivo de los datos con el objetivo de identificar posibles valores atípicos mediante el cálculo de las Distancias de Mahalanobis. Este procedimiento comenzó estableciendo un puntaje de corte basado en la distribución X^2 (chi-cuadrado) con un nivel de confianza del 99.9 %. Este umbral permitió delimitar el 0.1 % de las observaciones más extremas, que podrían considerarse como atípicas en función de su distancia respecto al centroide multivariado de la muestra [19].

Al aplicar este criterio, se evaluaron todas las observaciones del conjunto de datos. El punto de corte resultante, basado en la distribución X^2 (chi-cuadrado), fue 32.91. Según Lorenzo (2019) los resultados indicaron que ninguna observación superó el valor crítico de dicha distribución. Por lo tanto, no se identificaron valores atípicos y la muestra original, compuesta por 30 observaciones, se mantuvo intacta sin necesidad de excluir ningún dato [20].

Este análisis confirma la homogeneidad de los datos y garantiza que el conjunto de observaciones es adecuado para los análisis estadísticos posteriores sin la influencia de valores atípicos significativos.

3.3. Ajuste de distribución de probabilidad para los tiempos de cada operación unitaria en la elaboración del yogurt

Con el objetivo de simular el proceso real, se analizaron los tiempos de ejecución en la línea de producción y se identificó la distribución de probabilidad que mejor se ajusta a los datos, empleando R Studio [21]. El ajuste se evaluó mediante una prueba de bondad de ajuste utilizando el método Chi-cuadrado [22]. Los resultados se presentan en la Tabla 1, donde se detallan el proceso, la distribución de probabilidad más adecuada y el valor p, empleado para evaluar la hipótesis nula.

Tabla 1: Distribuciones de probabilidad y p-values por proceso

Proceso	Distribución de probabilidad	p-value
Recepción	Uniforme	0,11
Enfriamiento	Log normal	0,47
Filtración	Uniforme	0,25
Pasteurización	Uniforme	0,27
Enfriamiento	Logística	0,78
Inoculación	Log normal	0,24
Agitación	Uniforme	0,22
Incubación	Log normal	0,52
Enfriado	Log logística	0,43
Agitado	Uniforme	0,15
Dosificación	Normal	0,81
Envasado	Logística	0,49
Almacenado	Log logística	0,75

De acuerdo con los datos presentados en la Tabla 1, la prueba de bondad de ajuste arrojó un valor de p-value superior a 0,05 para cada uno de los procesos analizados. Esto indica que no se puede rechazar la hipótesis nula, lo que sugiere que los datos observados se ajustan adecuadamente a las distribuciones de probabilidad propuestas por el programa R Studio. Este resultado garantiza, con un nivel de confianza del 95 %, que las distribuciones seleccionadas son representativas del comportamiento real de los procesos [22].

3.4. Construcción y Validación del proceso de simulación

Se construyó un modelo de simulación de la línea de producción utilizando la información de la Tabla 1, que muestra la distribución de probabilidad ajustada a cada proceso u operación unitaria involucrada en la elaboración del yogurt.

Para validar el modelo simulado, se realizaron 30 simulaciones, siguiendo el Teorema del Límite Central, y se llevó a cabo una prueba de hipótesis para dos muestras independientes mediante la t de Student, comparando los resultados con los del proceso real [23].

En la Tabla 2, se presentan los resultados de la simulación en términos del tiempo de producción y el número de productos obtenidos, junto con los valores correspondientes del proceso real

Tabla 2: Validación del proceso de simulación vs proceso real

No.	Tiempo de producción (min)		Producto Terminado (u)	
	Modelo real	Modelo simulado	Modelo real	Modelo simulado
1	1225	1157	608	603
2	1119	1216	592	594
3	1137	1173	592	595
4	1188	1181	606	598
5	1307	1160	597	604
6	1156	1187	591	609
7	1184	1173	594	599
8	1250	1246	598	608
9	1144	1244	593	603
10	1211	1197	593	593
11	1201	1175	592	607
12	1079	1222	602	597
13	1209	1111	595	609
14	1124	1126	597	609
15	1279	1173	600	591
16	1217	1121	606	605
17	1220	1175	596	590
18	1226	1252	603	592
19	1154	1200	610	602
20	1231	1164	609	598
21	1196	1190	605	593
22	1229	1240	606	610
23	1172	1160	596	594
24	1286	1178	610	601
25	1263	1242	610	606
26	1184	1226	595	593
27	1179	1160	591	605
28	1260	1135	599	601
29	1276	1143	592	598
30	1246	1135	598	592

Luego de realizar la prueba de hipótesis mediante la T de student, se determinó que, no existe evidencia estadísticamente significativa para rechazar la hipótesis nula ($p > 0, 05$). Por lo tanto, la media del tiempo de ciclo del proceso simulado es igual a la media del proceso real con una confianza del 95 %. Además, se determinó que no existe evidencia

estadísticamente significativa en el número de productos terminados entre el proceso real y el proceso simulado ($p > 0,05$).

3.5. Creación de escenarios de mejoras con la técnica Lean Manufacturing

Utilizando el software FlexSim, se desarrollaron escenarios de mejora aplicando técnicas de Lean Manufacturing, tales como TPM (Mantenimiento Productivo Total), 5S (clasificación, orden, limpieza, estandarización y disciplina), Poka Yoke y eliminación de muda (transporte y movimiento).

En cada escenario de mejora, se registraron los tiempos de producción y el número de productos elaborados, que corresponden a las variables dependientes o de respuesta. Para cada escenario, se realizaron 30 simulaciones, siguiendo el Teorema del Límite Central. Los valores obtenidos de las simulaciones fueron sometidos a un análisis para verificar si cumplen con los supuestos requeridos para poder utilizar el análisis multivariado MANOVA.

3.6. Supuestos de MANOVA

Eliminación de datos atípicos

Se realizó un análisis de valores atípicos utilizando la Distancia de Mahalanobis, estableciendo un umbral basado en la distribución X^2 con un nivel de confianza del 99.9%. Este criterio permitió identificar posibles observaciones extremas. Los resultados indicaron que ninguna observación superó el valor crítico (10,82), por lo que no se detectaron valores atípicos. La muestra original de 180 observaciones se mantuvo intacta. Este análisis confirma la homogeneidad de los datos y su idoneidad para los análisis estadísticos posteriores, sin la influencia de valores atípicos significativos.

Distribución normal multivariada de las variables dependientes

En la Tabla 3 se presentan los resultados del análisis de normalidad múltiple según Menoyo (2021). Para evaluar la normalidad múltiple, se pueden emplear diversas pruebas estadísticas, como el Test de Mardia, Test de Henze-Zirkler, Test de Royston, Test de Doornik-Hansen y Test de Energy [24]. A continuación, se muestran los resultados del análisis aplicado a las variables dependientes: tiempo de producción y número de productos.

Tabla 3: Normalidad multivariada de las variables dependientes (Tiempo de producción y número de productos)

Variable dependiente	Test de Mardia	Test de Henze-Zirkler	Test de Royston	Test de Doornik-Hansen	Test de Energy
Tiempo de producción	<0,001	<0,001	<0,001	<0,001	<0,001
Número de productos	<0,001	<0,001	<0,001	<0,001	<0,001

Según los resultados de la Tabla 3, las variables dependientes, no se ajustan a una distribución normal multivariada, con un nivel de significancia del 0,05, por lo tanto, no se cumple el supuesto de Normalidad [24].

Homogeneidad de la matriz de varianza-covarianza:

La homogeneidad de la matriz de varianza y covarianza de los datos se evaluó mediante la prueba de Box (M-Box). El resultado arrojó un valor de $p < 0.001$, lo que indica que existe evidencia estadísticamente significativa para rechazar la hipótesis nula. Por lo tanto, los datos no presentan homogeneidad en la matriz de varianza y covarianza, lo que significa que no se cumple el supuesto de homogeneidad [24].

Independencia de las observaciones:

La independencia de las observaciones se evaluó mediante la prueba de Durbin-Watson, aunque esta prueba es más comúnmente usada en regresiones, o evaluar los residuos del modelo. El resultado arrojó un valor de $p < 0.001$, lo que indica que existe evidencia estadísticamente significativa para rechazar la hipótesis nula. Por lo tanto, los datos no presentan una independencia de las observaciones, lo que significa que no se cumple el supuesto de independencia [24].

Según Rodríguez (2021) si los datos no siguen una distribución normal multivariada, no son homocedasticidad ni independientes, es posible aplicar ciertas transformaciones para intentar cumplir con estos supuestos. En este caso, se llevaron a cabo las siguientes transformaciones: logarítmica, raíz cuadrada, Box-Cox y la función inversa [25]. Después

de cada transformación, se evaluó la normalidad, homocedasticidad e independencia de los datos; sin embargo, ninguna de ellas logró cumplir con los supuestos.

Según Somerfield (2021) una alternativa para analizar los datos cuando no se cumplen los supuestos mencionados es utilizar métodos no paramétricos, como el PERMANOVA, el cual se basa en permutaciones y no requiere el cumplimiento estricto de los supuestos de normalidad y homogeneidad [9].

3.7. Utilización de la técnica multivariante

Se utilizó el método multivariable no paramétrico PERMANOVA, basado en permutaciones, el cual no requiere la asunción de normalidad ni homogeneidad. El análisis paramétrico se realizó mediante la prueba de Bray-Curtis, utilizando 999 permutaciones [26].

Los resultados dieron un valor de $p < 0.001$, indicando evidencia estadísticamente significativa para rechazar la hipótesis nula. Por lo tanto, se concluye que existe una diferencia estadísticamente significativa en las variables dependientes, tiempo de producción y número de productos, entre los diferentes escenarios de simulación utilizando Lean Manufacturing.

Dado que se encontró una diferencia estadísticamente significativa entre los tratamientos evaluados (escenarios de simulación utilizando Lean Manufacturing), se aplicó el método de comparación de rangos múltiples multivariante conocido como prueba post hoc de comparaciones pareadas de PERMANOVA o PERMANOVA Pairwise. Los resultados obtenidos mediante este método se presentan en la siguiente tabla.

Tabla 4: Resultados de la prueba de post hoc de comparaciones pareadas de PERMANOVA

Nº	Comparaciones	Suma de cuadrados	F.Model	R ²	p.value	p.adj usted	sig
1	Real vs TPM	9,39E+02	37.739.26	0.0610925	0.054	0.81	
2	Real vs Poka_Yoke	6,01E+03	201.994.2	0.2583065	0.001	0.01	*
3	Real vs Poka_Yoke_1	5,38E+03	272.754.2	0.3198509	0.001	0.01	*
4	Real vs 5 S	7,38E+03	454.346.4	0.4329594	0.001	0.01	*
5	Real vs Eliminación	5,49E+03	300.666.6	0.3626804	0.001	0.01	*
6	TPM vs Poka_Yoke	2,27E+03	81.392.84	0.1230627	0.004	0.06	
7	TPM vs Poka_Yoke_1	1,85E+03	133.733.0	0.1517156	0.001	0.01	*
8	TPM vs 5 S	1,91E+03	215.809.0	0.2711800	0.001	0.01	*
9	TPM vs Eliminación	1,90E+03	183.597.8	0.2046549	0.001	0.01	*
10	Poka_Yoke vs Poka_Yoke_1	7,33E+01	0.323.500	0.0557447	0.967	1.000	
11	Poka_Yoke vs 5 S	1,87E+02	0.972191	0.0944536	0.366	0.651	
12	Poka_Yoke vs Eliminación	1,62E+02	0.822539	0.0914660	0.387	0.651	
13	Poka_Yoke_1 vs 5 S	1,87E+02	2.443.485	0.1114173	0.051	0.622	
14	Poka_Yoke_1 vs Eliminación	3,66E+01	0.383068	0.0260232	0.814	1.000	
15	5 S vs Eliminación	1,48E+02	2.495.007	0.1178000	0.118	0.9	

Según los resultados de la prueba post hoc presentados en la Tabla 4, se identificaron siete comparaciones con un valor de $p.adjust < 0.05$, lo que indica que dichas comparaciones son estadísticamente significativas. Entre estas siete comparaciones, el escenario que presentó los mejores resultados en cuanto a las variables tiempo de producción y número de productos elaborados fue el escenario de 5S, con un $R^2 = 0.44$. Este valor indica que el 44% de la variabilidad en la variable dependiente es explicada por el modelo o por las variables independientes incluidas en el análisis, mientras que el 56% restante se debe a otros factores no considerados en el modelo o al error aleatorio.

4. Conclusiones

Se logró identificar y documentar detalladamente las operaciones unitarias del proceso de producción de yogurt de frutas, estableciendo una base de datos estructurada con los tiempos de procesamiento de cada etapa. Esta organización sistemática de la información garantiza precisión y facilita su análisis, permitiendo evaluar la eficiencia y desempeño del proceso productivo.

El análisis de la base de datos confirmó su integridad y homogeneidad. Se imputó un 3.3% de datos faltantes con el método KNN, preservando la validez estadística. Además, el análisis de valores atípicos mediante la Distancia de

Mahalanobis y χ^2 no identificó observaciones extremas, manteniendo intacta la muestra de 30 observaciones, lo que garantiza su robustez y representatividad.

El análisis de los tiempos de ejecución permitió identificar la distribución de probabilidad más adecuada para cada proceso en la línea de producción. La prueba de bondad de ajuste Chi-cuadrado confirmó que los datos observados se ajustan correctamente a las distribuciones propuestas, con un nivel de confianza del 95%. Estos resultados aseguran la validez y representatividad de las distribuciones seleccionadas para la simulación del proceso real.

Mediante FlexSim, se simularon escenarios de mejora aplicando técnicas de Lean Manufacturing para optimizar el proceso productivo. Se evaluaron los efectos de TPM, 5S, Poka Yoke y eliminación de muda en los tiempos de producción y la cantidad de productos elaborados. Para garantizar la validez de los resultados, se realizaron 30 simulaciones por escenario, siguiendo el Teorema del Límite Central, lo que permite una representación confiable del comportamiento del sistema.

El análisis de los datos obtenidos en las simulaciones reveló que no se cumplían los supuestos requeridos para aplicar MANOVA. Como alternativa, se optó por el método PERMANOVA, una técnica no paramétrica basada en permutaciones, que permite analizar los datos sin la necesidad de cumplir estrictamente los supuestos de normalidad y homogeneidad, garantizando así la validez del análisis.

El análisis mediante PERMANOVA, basado en permutaciones y utilizando la distancia Bray-Curtis, permitió evaluar diferencias entre los escenarios de simulación sin requerir supuestos de normalidad ni homogeneidad. Los resultados ($p < 0.001$) indicaron una diferencia estadísticamente significativa en las variables dependientes (tiempo de producción y número de productos) al aplicar técnicas de Lean Manufacturing.

La prueba post hoc identificó siete comparaciones estadísticamente significativas ($p_{\text{adjust}} < 0.05$), destacando el escenario 5S como el más eficiente en la reducción del tiempo de producción y el aumento del número de productos elaborados. Con un $R^2 = 0.44$, el modelo explica el 44% de la variabilidad en la variable dependiente, mientras que el 56% restante se atribuye a factores no considerados o al error aleatorio. Estos resultados resaltan la efectividad de la metodología 5S dentro de las estrategias de Lean Manufacturing.

Referencias

- [1] R. Ramirez, “Herramientas y técnicas de mejora de la calidad en la industria de alimentos latinoamericana y su aporte a la competitividad organizacional,” Artículo de grado, Fundación Universidad de América, p. 8, 2022.
- [2] E. Acosta, F. Chiodi, M. Fernandez, and A. F. J. Tcach, “Etapas de validación y verificación en la construcción de modelos para la simulación de procesos industriales,” pp. 1–8, 2016.
- [3] H. Patiño, “Evaluación de la productividad organizacional,” Gestión de indicadores e índices, 2024.
- [4] I. M. L. V. and J. M. A. Pacheco, “Impacto de los indicadores de productividad en la gestión,” *Universidad del Zulia*, vol. 39, no. 101, pp. 567–581, 2022.
- [5] R. M. F. M. V. J. P. N. O. and G. S. E. C. M. R., “Modelación y simulación de procesos en la intensificación de instalaciones en la industria química,” *Revista de Ciencia Tecnología e Innovación Innova*, vol. 2, no. 2, pp. 22–36, 2020.
- [6] J. Bautista, “Evolución de los softwares de simulación para el diseño y construcción en la industria,” *Polo del Conocimiento: Revista científico-profesional*, vol. 5, no. 8, pp. 1332–1343, 2020.
- [7] G. M. A. and E. A. H. Morales, *Análisis multivariante de datos*. México: Comercial Grupo ANAYA, SA, 2021.
- [8] J. B. Vertti, *Análisis multivariado*. México: Universidad Autónoma de Aguascalientes, 2019.
- [9] K. R. C. R. N. G. P. J. Somerfield, “Análisis de similitudes (anosim) para diseños de dos vías utilizando una estadística anosim generalizada, con notas comparativas sobre el análisis de varianza multivariado permutacional (permanova),” *Austral Ecology*, vol. 46, no. 6, pp. 887–1007, 2021.
- [10] J. M. N. P. N. O. J. Vargas, “Aplicación de la técnica multivariada manova a dos variables de control provenientes de tres modelos de simulación estocásticos de un proceso productivo,” *Entre Ciencia y Tecnología*, vol. 14, no. 28, pp. 66–75, 2020.
- [11] J. A. R. M. G. D. Valverde, *Indicadores logísticos en un modelo simulado de la industria farmacéutica*. Huacho: Universidad Nacional José Faustino Sánchez Carrión, 2024.
- [12] Z. J. J. H. M. H. S. and F. Villanueva, “Diseño de un sistema de surtimiento de materiales bajo principios esbeltos y de manufactura 4.0 usando simulación de eventos discretos,” *Pistas Educativas*, vol. 42, no. 138, pp. 1–17, 2022.
- [13] C. V. B. I. and P. A., “Simulación de mejoras en el sistema productivo de una curtiembre basada en el mapeo de su cadena de valor,” *Scientia Et Technica*, vol. 25, no. 3, pp. 394–415, 2020.
- [14] R. O. S. G. J. G. C. J. J. C. P. S. J. F. A. and A. Natividad, “Manufactura esbelta con simulación dinámica estocástica para el incremento de la productividad, línea de nuggets empresa avícola,” *Revista de Investigación Científica*, vol. 5, no. 2, pp. 139–153, 2019.
- [15] L. Freire, “Tratamiento de falta de información en técnicas de minería de datos,” Granada, 2023.
- [16] L. G. D. Monroy, *Estadística multivariada: Inferencia y métodos*. Bogotá: Proceditor Ltda, 2017.
- [17] G. A. R. Duque, “Revisión y perspectivas para la construcción de bases de datos robustas y con datos faltantes: caso aplicado a la información financiera,” *Tecnura*, vol. 27, no. 75, pp. 14–37, 2023.
- [18] L. Y. D. Turco, “Mejora de la predicción en tareas de aprendizaje automático supervisado: imputación de valores perdidos,” Andalucía, España, 2021.
- [19] J. Lorenzo, “Distancias de mahalanobis,” Argentina, 2019.
- [20] —, “Distancia de mahalanobis,” 2019.
- [21] M. A. D. Martínez, R. Z. Cruz, and R. V. R. Salinas, “Simulación con flexsim, una nueva alternativa para la ingeniería hacia la toma de decisiones en la operación de un sistema de múltiples estaciones de prueba,” *Científica*, vol. 22, no. 2, pp. 97–104, 2018.

- [22] A. C. B. Cepeda, “Bondad de ajuste para datos redondeados,” México, 2021.
- [23] K. D. O. S. and J. C. C. Vélez, *Bioestadística aplicada a investigaciones científicas en la salud*. Quito: Ediciones Mawil, 2022.
- [24] M. R. David, G. C. Antonio, and G. L. Eva, *Fundamento de la ciencia de datos*. España: Universidad de Alcalá, 2021.
- [25] A. R. M. Rodríguez, “Métodos alternativos ante la violación de supuestos en diseños de experimentos factoriales,” *Anales Científicos, Universidad Nacional Agraria La Molina*, vol. 82, no. 2, pp. 318–335, 2021.
- [26] P. L. L. B. J. C. C. and B. C. Grollemund, “Las pruebas permanova y el modelado log-normal de poisson revelan cómo se distinguen dos quesos tradicionales mediante tareas de clasificación y verbalización,” *Food Quality and Preference*, vol. 104, 2023.