

Survival Analysis Applied to the Return of Containers in a Circular Economy System

Gómez Matias Luciano

¹ Facultad de Ciencias Económicas, Universidad Nacional de Tucumán, Tucumán 4000, Argentina
matiaslucianogomez@hotmail.com

Abstract. Within the framework of the circular economy, the reuse of containers is essential to achieve both sustainability and profitability. Circulo S.A. markets cleaning products using returnable containers, whose return by customers is variable and uncertain. This creates challenges in stock management and forces the company to purchase new containers, impacting operating costs.

This study applies survival analysis to model the return time of containers, using data from the company's internal traceability system (TRAC). Each record was transformed into usage cycles, considering assignments to customers, returns to the warehouse, or losses (censored cycles). Tools such as the Kaplan-Meier curve and the Lognormal distribution were employed.

The results show that 90% of the containers are returned 116 days, according to the Lognormal distribution. However, the Kaplan-Meier curve, which includes censored data, reveals that 45% of the containers had not been returned by day 125. Differences by container type were also observed: 5-liter drums have a higher return rate, while bicarbonate jars are the least likely to be returned.

Based on these findings, it is recommended to establish differentiated penalty timeframes, segment customers according to their return history, and include information on the type of product transported in each container. This study demonstrates the value of survival analysis as a key statistical tool to optimize operational management and strengthen sustainable practices in circular economy businesses.

Keywords: survival analysis, circular economy, customer penalties, sustainability, container return.

Análisis de Supervivencia Aplicado en la Devolución de Envases en un Sistema de Economía Circular

Resumen. En el marco de la economía circular, la reutilización de envases es fundamental para lograr sostenibilidad y rentabilidad. La empresa Circulo S.A. comercializa productos de limpieza mediante envases retornables, y trabaja bajo el marco de la economía circular por lo que afronta grandes dificultades en la gestión de stock de envases que afectan los costos operativos de la misma.

El trabajo aplica análisis de supervivencia para modelar el tiempo de retorno de los envases, utilizando datos del sistema interno de trazabilidad (TRAC). Cada registro fue transformado en ciclos de uso, considerando asignaciones a clientes, devoluciones al depósito o pérdidas (ciclos censurados). Se emplearon herramientas como la curva de Kaplan-Meier y la distribución Lognormal para modelar los tiempos de devolución de cada tipo de envase.

Los resultados muestran que la mayoría de los envases se devuelven antes de los 116 días, según la distribución Lognormal. Sin embargo, la curva de Kaplan-Meier, que incorpora datos censurados, revela que un poco menos de la mitad de los envases aún no han sido devueltos en el día 125. Además, se observan diferencias por tipo de envase, demostrando que existe una correlación en el tipo de envase y su tiempo de devolución.

A partir de estos hallazgos, se recomienda establecer plazos de penalización diferenciados, segmentar clientes según su historial de devolución e incorporar información sobre el producto transportado en cada envase. El estudio demuestra el valor del análisis de supervivencia como herramienta estadística clave para optimizar la gestión operativa y reforzar prácticas sostenibles en empresas circulares.

Palabras clave: análisis de supervivencia, economía circular, penalización de clientes, sustentabilidad, devolución de envases.

1.1 Introducción y Situación Problemática

La contaminación ambiental representa uno de los desafíos más urgentes del siglo XXI. El modelo de producción y consumo lineal, basado en fabricar, usar y desechar, ha incrementado la generación de residuos, particularmente plásticos, que terminan afectando ecosistemas, biodiversidad y salud humana. Según el PNUMA, se generan más de 300 millones de toneladas de desechos plásticos anualmente, una gran parte de los cuales no se gestiona adecuadamente.

En este escenario, la economía circular surge como una alternativa clave. Este enfoque busca cerrar el ciclo de vida de los productos mediante estrategias como la reutilización y el reciclaje, reduciendo la demanda de materias primas y las emisiones de carbono. La Fundación Ellen MacArthur estima que su aplicación podría disminuir las emisiones globales en un 45% para 2050.

Circulo S.A., empresa dedicada a la comercialización de productos de limpieza, adoptó este modelo desde 2020. Su sistema se basa en envases reutilizables que los clientes deben devolver una vez vacíos. En la primera compra, se cobra un cargo adicional por el envase; en las recargas posteriores, solo se paga el contenido. El cargo cubre aproximadamente la mitad del costo real del envase, lo que implica que la empresa asume una parte significativa del costo inicial. Esto refuerza la importancia de fomentar la fidelización, ya que los clientes recurrentes reutilizan envases, reduciendo costos operativos y residuos.

El éxito del modelo depende del cumplimiento del ciclo de retorno de los envases. Sin embargo, el tiempo de devolución es incierto. A diferencia de otros sistemas, como

los utilizados en el sector alimenticio europeo que aplican penalizaciones a los 15 días, en Círculo S.A. este plazo resulta inaplicable: los productos, como bidones de 5 litros, requieren más tiempo para consumirse.

Esta incertidumbre genera consecuencias económicas. La empresa debe reponer envases no devueltos, afectando su rentabilidad. En los casos en que nunca se produce la devolución, se incurre en una pérdida total del activo. Ante esta problemática, es necesario contar con un modelo que permita mejorar la estimación de los tiempos de devolución, incorporar la variabilidad de consumo y orientar políticas como la penalización por no retorno.

1.2 Preguntas y objetivos de investigación

A causa de estos inconvenientes, surgen las siguientes preguntas y objetivos que guían el trabajo:

Preguntas de investigación.

- ¿Cuál es la distribución de probabilidad más adecuada para modelar el tiempo de devolución de los envases?
- ¿Cuál es el tiempo promedio que tardan los clientes en devolver los envases?
- ¿Cuáles son los factores que más influyen en la demora o en la no devolución?

Objetivo general. Modelizar el tiempo de devolución de los envases en Círculo S.A. mediante técnicas de análisis de supervivencia.

Objetivos específicos.

- Determinar la distribución de probabilidad que mejor se ajusta al tiempo de devolución.
- Identificar los factores que influyen en la demora o falta de devolución.
- Estimar el tiempo promedio de retorno de los envases.

1.3 Marco Metodológico

En este trabajo se adopta un enfoque cuantitativo [1] con un diseño longitudinal de panel [2], que permite observar cómo evolucionan los ciclos de devolución de envases a lo largo del tiempo y analizar el comportamiento recurrente de los clientes. La base de datos proviene del sistema de trazabilidad de la empresa Círculo S.A., donde cada envase posee un identificador único que registra sus movimientos de asignación y devolución.

Se estructuraron los datos en ciclos, que reflejan el tránsito del envase desde su entrega al cliente hasta su retorno al depósito. En caso de no retorno, se considera un ciclo censurado. También se trataron situaciones en las que un envase fue reasignado sin

haber sido registrado como devuelto: para mantener la lógica del modelo, se asumió su devolución el mismo día de su nueva asignación.

El análisis se centró en mejorar la estimación de los tiempos hasta la devolución utilizando técnicas de análisis de supervivencia. En este trabajo se empleó la curva de Kaplan-Meier como herramienta principal, la cual permite estimar la probabilidad de que un envase continúe sin devolverse con el paso del tiempo y comparar su comportamiento entre distintos grupos, tales como los tipos de envase.

Se utilizó el lenguaje de programación Python para procesar el gran volumen de datos disponibles, así como para la generación de visualizaciones y la realización de pruebas estadísticas. Se aplicaron validaciones cruzadas y análisis de residuos para asegurar la robustez de los modelos utilizados.

1.4 Marco Teórico

Para comprender el enfoque del presente estudio, se exponen a continuación los principales fundamentos teóricos que lo sustentan. Cada apartado se vincula directamente con los elementos metodológicos y empíricos desarrollados en el trabajo: la economía circular como marco conceptual general; el análisis de supervivencia como herramienta estadística clave para estimar los tiempos de devolución; la teoría de penalización de clientes como base para proponer incentivos adecuados; y las distribuciones estadísticas como soporte técnico para los modelos utilizados

Economía Circular. Según Geissdoerfer [3], la economía circular propone reemplazar el modelo lineal tradicional por uno que mantenga los recursos en uso el mayor tiempo posible. En la industria del plástico, esto implica sistemas de reutilización de envases para minimizar residuos y huella ambiental. En el caso de la empresa Circulo S.A. se propone adoptar este modelo fomentando la devolución de envases para su posterior reutilización. El éxito del sistema requiere que los clientes los devuelvan en un plazo razonable, para lo cual puede ser necesario establecer penalizaciones.

Análisis de supervivencia. El análisis de supervivencia se enfoca en el tiempo hasta que ocurre un evento [4]. En este caso, el evento es la devolución del envase. La función de supervivencia $S(t)$ se define como la probabilidad de que el envase no haya sido devuelto en el tiempo t :

$$S(t) = P(T > t) \quad (1)$$

Se utilizó el estimador no paramétrico de Kaplan-Meier, que estima $S(t)$ considerando datos censurados. Su fórmula:

$$S(t_j) = S(t_{j-1}) \times (n_j - d_j)/n_j \quad (2)$$

donde n_j representa la cantidad de envases en riesgo de ser devueltos en el tiempo t_j , y d_j aquellos efectivamente devueltos en ese momento.

Esta herramienta permite visualizar diferencias entre tipos de envases y estimar el tiempo de devolución según percentiles como la mediana o el percentil 90.

Teoría de Penalización de Clientes. Según Thaler & Sunstein [5] explican que las penalizaciones pueden modificar el comportamiento del cliente si se perciben como justas. Aplicarlas prematuramente puede ser contraproducente si no se ajustan a los patrones reales de consumo. Por eso, es clave determinar un plazo óptimo para su implementación, adaptado a la naturaleza del producto.

Distribuciones Empíricas y Teóricas. Las distribuciones empíricas reflejan los datos observados. Por ejemplo, la distribución de los días hasta la devolución de los envases. En cambio, las distribuciones teóricas (como la lognormal o Weibull) permiten modelar el comportamiento esperado de los datos. Ajustar los datos reales a estas distribuciones ayuda a predecir y entender mejor los tiempos de devolución.

2 Aplicación

2.1 Ajuste del análisis de supervivencia a ciclos repetitivos

En el análisis de supervivencia tradicional, cada unidad de observación presenta un único evento. Sin embargo, en este estudio el evento es la devolución del envase, y cada envase puede completar múltiples ciclos (asignación-devolución). Este patrón genera múltiples observaciones por unidad, lo que exige redefinir el enfoque.

Para resolverlo, se considera cada ciclo como una observación independiente:

1. Cada asignación a cliente inicia un ciclo.
2. Si se registra una devolución (movimiento a “warehouse”), se cierra el ciclo.
3. Si no hay devolución, se trata como dato censurado (evento no observado).

Incluso cuando un envase es asignado a dos clientes consecutivamente, se interpreta que regresó al almacén el mismo día del segundo registro, para cerrar el ciclo.

2.2 Evento de interés y censura

Se define como evento de interés a la devolución del envase al almacén, codificado como “1”. Si no se devuelve dentro del periodo de observación, se considera censurado (“0”). Este tratamiento es esencial para aplicar correctamente los métodos de análisis de supervivencia, ya que permite representar datos incompletos.

2.3 Recolección y estructura de datos

Círculo S.A. implementó el sistema TRAC para registrar movimientos de envases mediante escaneo de códigos QR. Cada registro indica si el envase fue asignado a un cliente (“client”) o regresado al almacén (“warehouse”), permitiendo reconstruir los ciclos de uso.

El flujo de datos de cada venta involucra cinco etapas: registro del pedido (CRM), preparación logística (ERP), asignación del envase (TRAC), entrega, y devolución escaneada (TRAC).

Se recopilan datos como ID del envase, tipo de envase, cliente asignado y fechas de movimiento. Además, se integran datos del ERP y registros manuales de los repartidores.

Tipos de envase identificados:

- Envases tipo 2 y 5: bidones de 5 litros (diferente diseño y durabilidad).
- Envase tipo 3: botellas de 1 litro.
- Envase tipo 4: botes de bicarbonato de sodio.

2.4 Transformación y limpieza de la base de datos

Los registros fueron transformados en ciclos de uso individuales para el análisis. Cada ciclo contiene: ID de envase, tipo de envase, fecha de inicio y fin, tipo de ciclo (completo, censurado o cliente-cliente) y duración en días.

La Tabla 1 muestra un ejemplo representativo de cómo se estructuró la base de datos a partir del sistema de trazabilidad TRAC. Allí puede observarse cómo se clasificaron los ciclos en función de su tipo y duración.

Tabla 1. Base de datos estructurada por ciclos

Asset_id	Tipo_envase	Fecha_asignacion	Fecha_devolucion	Ciclo_completo	Tipo_ciclo	Dias_ciclo
2902	2	30/09/24		0	Censurado	26
2903	2	03/05/24	19/07/24	1	Cliente-warehouse	76
2903	2	19/07/24	22/08/24	1	Cliente-Cliente	34
2904	2	02/05/24	25/06/24	1	Cliente-Cliente	53

Se aplicaron procesos de limpieza que incluyeron:

- Eliminación de valores nulos.
- Detección de duplicados.
- Validación cronológica de los movimientos.

2.5 Análisis inicial de la distribución de probabilidad

Se exploró inicialmente la distribución de los tiempos de duración de los ciclos. En la Figura 1 se muestra un histograma con todos los datos de la base, que permitía observar una forma general similar a una distribución Lognormal, pero con un aumento inesperado e irracional en los últimos días.

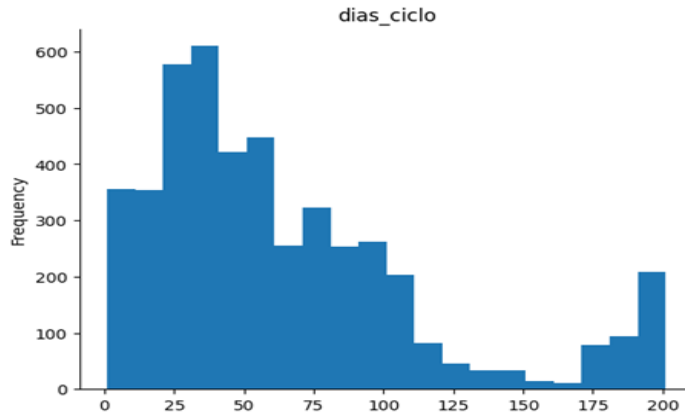


Fig. 1. : Distribución de todos los datos

Este patrón anómalo motivó un nuevo análisis, desagregando la información según el tipo de ciclo. En la **Figura 2** se presentan dos histogramas diferenciados: uno para ciclos completos y otro para datos censurados.

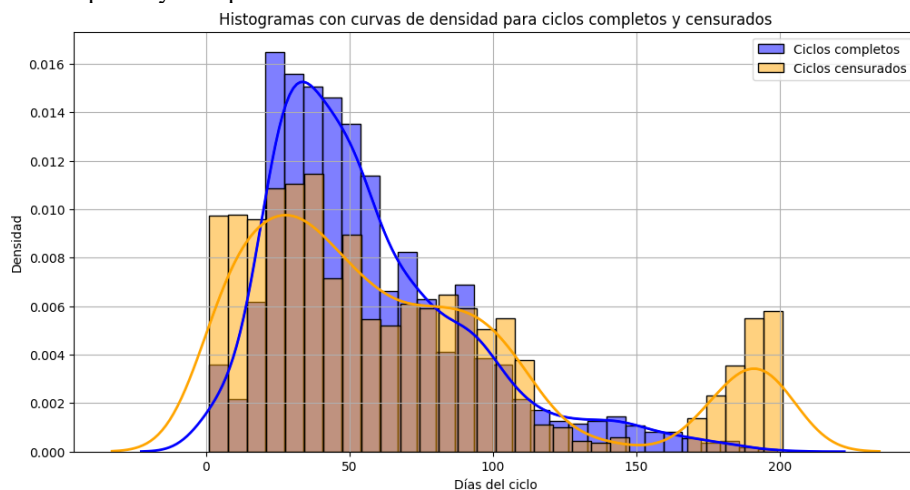


Fig. 2. Histogramas diferenciados por tipo de ciclo

De esta forma podemos ver que el incremento en la frecuencia de los ciclos que demoraban una cantidad de días extremadamente grande son principalmente datos censurados.

Posteriormente, se evaluaron diferentes distribuciones teóricas. La distribución **Log-normal** resultó ser la que mejor se ajustó a los ciclos completos (KS-Statistic = 0.0355, $p = 0.0277$).

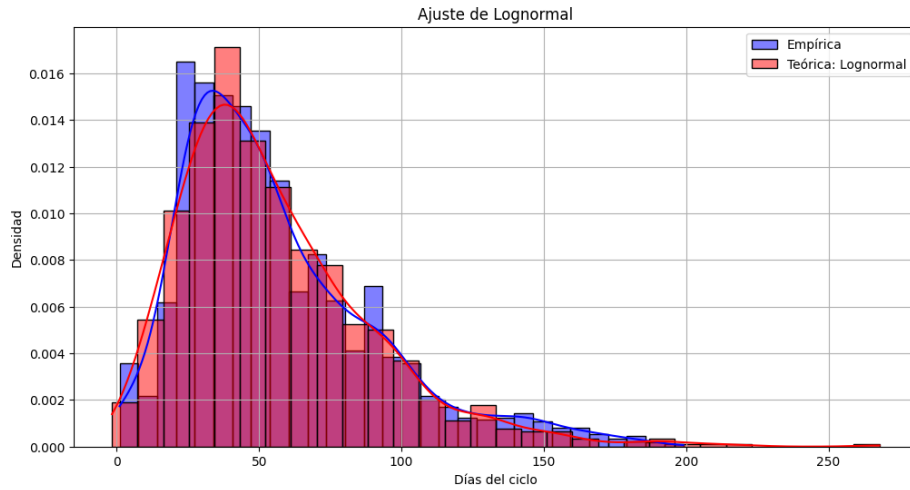


Fig. 3. Ajuste a distribución Lognormal

Parámetros calculados:

- Media logarítmica: $\mu = 3.82$
- Desvío logarítmico: $\sigma = 0.72$
- Mediana: 45.8 días
- Percentil 90: 116.4 días

El percentil 90 indica que el 90% de los ciclos completos se devuelven antes del día 116.

2.6 Análisis de supervivencia

Se implementó el modelo de Kaplan-Meier utilizando el lenguaje Python [6]. Inicialmente se construyó una curva general, que incluye tanto ciclos completos como censurados, permitiendo observar el comportamiento global de todos los envases a lo largo del tiempo. En la Figura 4 se muestra esta curva de supervivencia.

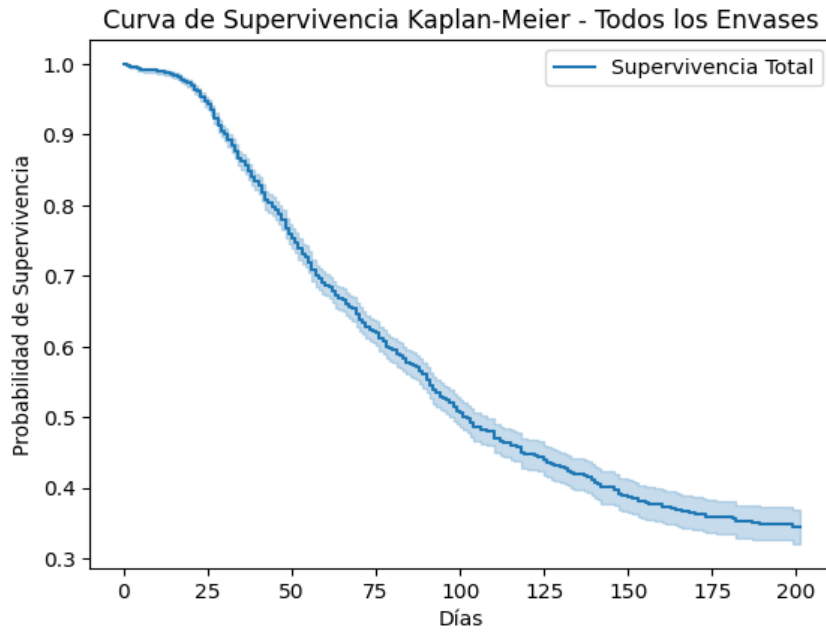


Fig. 4. Curva de Kaplan-Meier con todos los envases

Los valores del eje vertical representan la probabilidad de que un envase aún no haya sido devuelto a determinada cantidad de días. Por ejemplo, al día 100, aproximadamente el 50% de los envases continúan sin ser devueltos; al día 125, este valor es del 45%.

Dado que la curva general no permite identificar diferencias entre tipos de envases, se construyeron curvas específicas para cada tipo. En la Figura 5 se presentan las curvas de Kaplan-Meier desagregadas por tipo de envase.

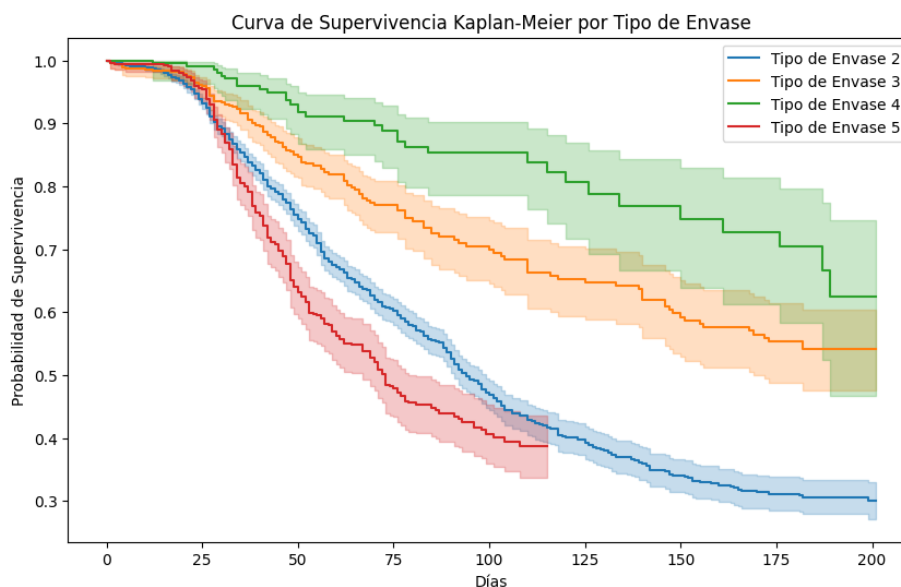


Fig. 5. Kaplan-Meier por tipo de envase

Esta desagregación permite analizar de forma comparativa el comportamiento de cada tipo de envase, mejorando la interpretación de las curvas de Kaplan-Meier..

Interpretación.

- **Curva general:** comportamiento casi lineal, con un 50% de envases sin devolver al día 100 y 45% al día 125.
- **Bidones 5L (Tipos 2 y 5):** mejor tasa de retorno, entre 40-50% sin devolver al día 100.
- **Botellas 1L (Tipo 3):** retorno más lento (70% sin devolver al día 100).
- **Botes bicarbonato (Tipo 4):** peor desempeño, 85% sin devolución al día 100.

Las curvas revelan la necesidad de tratar cada tipo de envase con estrategias diferenciadas.

3 Conclusiones

El análisis desarrollado permitió abordar con profundidad el comportamiento de devolución de envases dentro del modelo de economía circular implementado por Circulo S.A., respondiendo a las preguntas de investigación y alcanzando los objetivos planteados.

A través del análisis de la distribución empírica, se identificó que el 90% de los ciclos completos culminan con la devolución del envase antes del día 116, cuando se

ajusta la curva a una distribución Lognormal. No obstante, la incorporación de los datos censurados mediante el análisis de supervivencia reveló que al día 125 aún el 45% de los envases no habían sido devueltos. Esta diferencia entre ambos enfoques muestra la relevancia de incluir datos censurados, ya que aportan una visión más completa del comportamiento real, incorporando aquellos ciclos que no llegan a cerrarse durante el periodo de observación.

Las curvas de Kaplan-Meier permitieron además comparar grupos según el tipo de envase, evidenciando que los bidones de 5 litros presentan tasas de retorno más favorables, especialmente los de mayor calidad (tipo 5). En contraste, las botellas de 1 litro y, sobre todo, los botes de bicarbonato mostraron tasas de devolución considerablemente más bajas, sugiriendo posibles patrones de uso doméstico o menor compromiso con la devolución.

Asimismo, se concluye que la distribución Lognormal es la que mejor se ajusta a los datos observados en los ciclos completos, mientras que el análisis de supervivencia ofrece un marco robusto para modelar la probabilidad de devolución a lo largo del tiempo, incluyendo los casos incompletos. Este doble enfoque resulta complementario y esencial para el diseño de estrategias más justas y eficaces.

En síntesis, el uso de herramientas estadísticas avanzadas permitió comprender en profundidad los factores que influyen en la devolución de los envases y su variabilidad en función del tipo y del cliente. Estos hallazgos ofrecen una base sólida para optimizar tanto la gestión operativa como las decisiones estratégicas en la empresa, fortaleciendo su compromiso con la sostenibilidad ambiental y la eficiencia económica dentro del modelo de economía circular.

Referencias

1. C. Arana, *Análisis de datos en investigación social*, 1ª ed. Barcelona: Editorial Gedisa, 2015.
2. R. Hernández Sampieri, C. Fernández Collado, and M. D. Baptista Lucio, *Metodología de la investigación*, 6ª ed. México: McGraw-Hill Interamericana, 2014.
3. M. Geissdoerfer, "The Circular Economy – A new sustainability paradigm?," *Journal of Cleaner Production*, 2017.
4. D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, 2nd ed. Hoboken: John Wiley & Sons, 2008.
5. R. H. Thaler and C. R. Sunstein, *Un Pequeño Empujón*. New Haven: Yale University Press, 2008.
6. I. H. Witten, E. Frank, and M. A. Hall, *Minería de datos: Herramientas y técnicas prácticas*, 4ª ed. Burlington: Morgan Kaufmann, 2016.