



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO  
ESCUELA SUPERIOR CIUDAD SAHAGÚN  
LICENCIATURA EN INGENIERÍA INDUSTRIAL

**TESIS:**

Ingeniería de confiabilidad: marco metodológico para el modelado estadístico de tiempos de falla en forma de curva de bañera

**PARA OBTENER EL TÍTULO DE  
LICENCIADO EN INGENIERÍA INDUSTRIAL**

**PRESENTA:**

Aidee Domínguez Islas

**DIRECTOR:**

Dr. Isidro Jesús González Hernández

**CO-DIRECTOR:**

Dr. Luis Carlos Méndez González

**COMITÉ TUTORIAL:**

Dra. Francisca Santana Robles

Mtra. Claudia García Pérez

Ciudad Sahagún, Hidalgo, México, febrero 2026



Universidad Autónoma del Estado de Hidalgo  
Escuela Superior de Ciudad Sahagún  
Campus Sahagún

CLII-006-26  
Asunto: Autorización de impresión

P.D.L.I.I. C. Aidee Domínguez Islas  
Presente

Por este conducto le comunico que el jurado que le fue asignado a su trabajo de "TESIS" con el nombre: **"Ingeniería de confiabilidad: marco metodológico para el modelado estadístico de tiempos de falla en forma de curva de bañera"**, y que después de revisarlo en reunión de sinodales han decidido autorizar la digitalización del mismo, hechas las correcciones que fueron acordadas.

A continuación se anotan las firmas de conformidad de los integrantes del jurado:

|               |                                     |
|---------------|-------------------------------------|
| Presidente    | Dra. Francisca Santana Robles       |
| Primer vocal  | Dr. Isidro Jesús González Hernández |
| Segundo vocal | Dr. Luis Carlos Méndez González     |
| Secretario    | Mtra. Claudia García Pérez          |

Sin otro particular, me es grato reiterar a usted la seguridad de mi atenta consideración y respeto.

Atentamente  
"Amor, Orden y Progreso"  
Fray Bernardino de Sahagún, Hgo., a 04 de febrero de 2026

Dra. Francisca Santana Robles  
Coordinadora de la Licenciatura en Ingeniería Industrial

c.c.p.- Expediente  
FSR

"Amor, Orden y Progreso"

Carretera Otumba - Cd. Sahagún No. 7, Colonia  
Legaspi, Zona Industrial, Ciudad Sahagún, Hidalgo,  
México C.P. 43998  
Teléfono: 52 (771)7172000 Ext. 50201  
[essahagun@uaeh.edu.mx](mailto:essahagun@uaeh.edu.mx)



2025



[uaeh.edu.mx](http://uaeh.edu.mx)

## Dedicatoria

Dedico esta tesis, con todo mi amor y gratitud, a mi mamá, quien ha sido mi mayor ejemplo de fortaleza, sacrificio y perseverancia.

Gracias por tu apoyo incondicional, por tu paciencia y por nunca soltar mi mano.

Sobre todo, gracias por siempre creer en mí incluso cuando yo misma dudo de mis capacidades.

Eres y siempre serás mi principal motivación para seguir adelante. Este logro también es tuyo, mamá.

## Agradecimientos

Expreso mi más sincero agradecimiento a todas las personas que de alguna forma contribuyeron a la realización de esta tesis.

De manera muy especial, agradezco al Dr. Isidro Jesús González Hernández por su acompañamiento constante, orientación académica y apoyo durante todo el proceso de titulación. Su guía, confianza y compromiso con mi formación fueron fundamentales tanto para el desarrollo de este trabajo como para mi crecimiento profesional.

Agradezco a la Universidad Autónoma del Estado de Hidalgo y a los docentes que formaron parte de mi preparación académica, por los conocimientos transmitidos y por fomentar en mí el pensamiento crítico y profesional.

A mis compañeros, por el apoyo, la motivación y los momentos compartidos a lo largo de esta etapa.

Finalmente, agradezco profundamente a mi familia, por su amor, paciencia y respaldo incondicional, sin los cuales este logro no habría sido posible.

## Tabla de contenido

|                                                                                                                         |            |
|-------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Resumen .....</b>                                                                                                    | <b>9</b>   |
| <b>Abstract .....</b>                                                                                                   | <b>10</b>  |
| <b>Capítulo I. Planteamiento del problema .....</b>                                                                     | <b>11</b>  |
| 1.1 Introducción .....                                                                                                  | 11         |
| 1.2 Antecedentes .....                                                                                                  | 13         |
| 1.3 Planteamiento del problema .....                                                                                    | 16         |
| 1.4 Justificación .....                                                                                                 | 19         |
| 1.5 Objetivos del proyecto .....                                                                                        | 21         |
| 1.6 Hipótesis .....                                                                                                     | 22         |
| 1.7 Alcances y limitaciones .....                                                                                       | 24         |
| 1.8 Descripción general de la estructura de la tesis.....                                                               | 26         |
| <b>Capítulo II. Marco teórico .....</b>                                                                                 | <b>29</b>  |
| 2.1 Introducción .....                                                                                                  | 29         |
| 2.2 Estado del arte .....                                                                                               | 30         |
| 2.3 Fundamentos teóricos.....                                                                                           | 34         |
| <b>2.4 Marco conceptual.....</b>                                                                                        | <b>67</b>  |
| <b>Capítulo III. Marco metodológico para el análisis de tiempos de falla.....</b>                                       | <b>73</b>  |
| 3.1 Selección de la metodología de la tesis.....                                                                        | 73         |
| 3.2 Propuesta del procedimiento metodológico para la modelación de<br>tiempos de falla en forma de curva de bañera..... | 76         |
| <b>Capítulo IV. Resultados y análisis de los datos .....</b>                                                            | <b>103</b> |
| 4.1 Introducción .....                                                                                                  | 103        |
| 4.2 Resultados de la Fase 1: Preparación y comprensión de los datos .....                                               | 104        |

|                                                                                                                    |            |
|--------------------------------------------------------------------------------------------------------------------|------------|
| 4.3 Resultados de la Fase 2: Selección preliminar del modelo probabilístico .....                                  | 109        |
| 4.4 Resultados de la Fase 3: Ajuste paramétrico del modelo mediante Estimación de Máxima Verosimilitud (MLE) ..... | 112        |
| 4.5 Resultados de la Fase 4: Evaluación de la calidad del modelo .....                                             | 120        |
| 4.6 Resultados de la Fase 5. Interpretación de los resultados en el contexto de ingeniería de confiabilidad .....  | 132        |
| 4.7 Resultados de la Fase 6. Documentación y comunicación técnica .....                                            | 140        |
| <b>Capítulo V. Discusión</b> .....                                                                                 | <b>145</b> |
| 5.1 Interpretación de los principales hallazgos .....                                                              | 145        |
| <b>Conclusiones</b> .....                                                                                          | <b>149</b> |
| <b>Referencias bibliográficas</b> .....                                                                            | <b>152</b> |
| <b>Anexo: Código en R</b> .....                                                                                    | <b>161</b> |

## Índice de Tablas

|                                                                                                 |     |
|-------------------------------------------------------------------------------------------------|-----|
| <b>Tabla 1.</b> Comparación de las distribuciones utilizadas para modelar curvas de bañera..... | 33  |
| <b>Tabla 2.</b> Conjunto de datos históricos.....                                               | 36  |
| <b>Tabla 3.</b> Relación entre las funciones $Ft$ , $ft$ , $Rt$ y $ht$ . ....                   | 43  |
| <b>Tabla 4.</b> Estadísticas descriptivas de los tiempos de falla .....                         | 105 |
| <b>Tabla 5.</b> Diagnóstico de forma de los tiempos de falla .....                              | 110 |
| <b>Tabla 6.</b> Clasificación preliminar del tipo de función de falla .....                     | 112 |
| <b>Tabla 7.</b> Estimación de parámetros de distribuciones clásicas .....                       | 113 |
| <b>Tabla 8.</b> Estimación de parámetros de la distribución Beta–Weibull.....                   | 114 |
| <b>Tabla 9.</b> Métricas de confiabilidad de los modelos .....                                  | 117 |
| <b>Tabla 10.</b> Comparación de los criterios de información.....                               | 121 |
| <b>Tabla 11.</b> Pruebas de bondad de ajuste.....                                               | 122 |
| <b>Tabla 12.</b> Identificación de las fases de la curva de bañera .....                        | 135 |
| <b>Tabla 13.</b> Métricas de confiabilidad estimadas a partir del modelo Beta-Weibull .....     | 137 |
| <b>Tabla 14.</b> Implicaciones operativas del modelo Beta–Weibull para el mantenimiento.....    | 140 |
| <b>Tabla 15.</b> Estructura sugerida para un reporte técnico de confiabilidad .....             | 141 |
| <b>Tabla 16.</b> Visualizaciones recomendadas y su propósito técnico .....                      | 142 |
| <b>Tabla 17.</b> Relación entre métricas de confiabilidad y decisiones operativas ..            | 144 |

## Índice de Figuras

|                                                                                                                                                                              |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1.</b> La variable de estado y el tiempo hasta la falla de un producto. ....                                                                                       | 35 |
| <b>Figura 2.</b> Histograma y curva de supervivencia del conjunto de datos (Tabla 2).<br>.....                                                                               | 37 |
| <b>Figura 3.</b> Función de Densidad de Probabilidad, $f_t$ para el tiempo hasta la falla $T$ .....                                                                          | 39 |
| <b>Figura 4.</b> La FDA $F_t$ y FDP $f_t$ . ....                                                                                                                             | 40 |
| <b>Figura 5.</b> Cálculo integral de la probabilidad de fallar dentro del intervalo $t_1, t_2 = 5.0, 5.7$ . ....                                                             | 41 |
| <b>Figura 6.</b> Función de supervivencia $R_t$ .....                                                                                                                        | 41 |
| <b>Figura 7.</b> Curva de bañera empírica. ....                                                                                                                              | 45 |
| <b>Figura 8.</b> Curva de bañera.....                                                                                                                                        | 45 |
| <b>Figura 9.</b> Vida residual de un ítem que continúa funcionando en el tiempo $x$ . .                                                                                      | 47 |
| <b>Figura 10.</b> FDP $f_t$ (línea continua) y FDA $F_t$ (línea discontinua) para la distribución exponencial $\lambda = 0.4$ . ....                                         | 49 |
| <b>Figura 11.</b> La función de tasa de falla de la mezcla de dos distribuciones exponenciales $\lambda_1 = 1, \lambda_2 = 3$ , y $p = 0.4$ . ....                           | 57 |
| <b>Figura 12.</b> Función de densidad de probabilidad de la distribución de Weibull para valores seleccionados del parámetro de forma $\alpha\theta = 1$ .....               | 59 |
| <b>Figura 13.</b> Función de tasa de falla de la distribución de Weibull, con $\theta = 1$ y cuatro valores diferentes del parámetro de forma $\alpha$ . ....                | 60 |
| <b>Figura 14.</b> El factor de proporcionalidad del MTTF como función de $\alpha$ . ....                                                                                     | 62 |
| <b>Figura 15.</b> La función de vida residual media escalada $g_t = MRL_x MTTF$ para la distribución de Weibull con parámetros $\alpha = 2.25$ y $\theta = 8760$ horas. .... | 64 |
| <b>Figura 16.</b> MTTFs como función de $n$ , el número de componentes independientes e idénticos en una estructura en serie (Ejemplo 4).....                                | 66 |
| <b>Figura 17.</b> Curva de bañera para tasa de fallas (González-Hernández et al., 2025). ....                                                                                | 69 |

|                                                                                           |     |
|-------------------------------------------------------------------------------------------|-----|
| <b>Figura 18.</b> Histograma de tiempos de falla. ....                                    | 106 |
| <b>Figura 19.</b> Diagrama de caja de los tiempos de falla. ....                          | 107 |
| <b>Figura 20.</b> Curva de supervivencia empírica $Rt$ mediante Kaplan–Meier. ....        | 107 |
| <b>Figura 21.</b> Estimación no paramétrica de la función de falla $ht$ . ....            | 108 |
| <b>Figura 22.</b> Comparación de supervivencia $Rt$ teórica vs $Rtempírica$ . ....        | 123 |
| <b>Figura 23.</b> Comparación de la función de falla $ht$ teórica vs. $htempírica$ . .... | 124 |
| <b>Figura 24.</b> Histograma de tiempos y funciones de densidad ajustadas. ....           | 126 |
| <b>Figura 25.</b> QQ-plot: modelos clásicos y Beta–Weibull. ....                          | 128 |
| <b>Figura 26.</b> PP-plot: modelos clásicos y Beta–Weibull. ....                          | 129 |
| <b>Figura 27.</b> Función de falla: comparación empírica vs. Beta–Weibull. ....           | 133 |
| <b>Figura 28.</b> $MRLt$ bajo el modelo Beta–Weibull. ....                                | 138 |

## Resumen

La presente tesis tuvo como objetivo desarrollar un procedimiento metodológico para aplicar la ingeniería de confiabilidad al análisis de productos, dispositivos, equipos o sistemas industriales, mediante la modelación de tiempos de falla con comportamiento en forma de curva de bañera, utilizando distribuciones de probabilidad clásicas y avanzadas. La investigación abordó un problema de carácter académico identificado en la formación del ingeniero industrial: la ausencia de una metodología aplicada, clara y reproducible que permita seleccionar, ajustar e interpretar de manera adecuada modelos probabilísticos capaces de representar fallas con comportamiento no monótono.

El estudio se fundamentó en datos de referencia ampliamente utilizados en la literatura de confiabilidad y combinó análisis exploratorio, estimación no paramétrica de la función de riesgo y ajuste de modelos paramétricos mediante máxima verosimilitud. Se compararon distribuciones tradicionales y extendidas con base en criterios estadísticos y de coherencia ingenieril, evaluando métricas como el tiempo medio hasta la falla, la vida media residual y la probabilidad de supervivencia.

Los resultados evidenciaron que los modelos probabilísticos flexibles ofrecen una representación más adecuada de las fases de mortalidad temprana, vida útil estable y desgaste acelerado, superando las limitaciones de la distribución Weibull estándar. Asimismo, se demostró que la correcta selección del modelo influye directamente en la interpretación del comportamiento de falla y en la generación de información útil para la toma de decisiones en confiabilidad.

La principal aportación del trabajo consiste en la sistematización de un procedimiento metodológico reproducible, documentado mediante scripts en el software R, que fortalece el vínculo entre la estadística aplicada y la ingeniería de confiabilidad en el ámbito académico de la Ingeniería Industrial.

**Palabras clave:** ingeniería de confiabilidad, curva de bañera, tiempos de falla, distribuciones de probabilidad, análisis de supervivencia.

## Abstract

The objective of this thesis was to develop a methodological procedure for applying reliability engineering to the analysis of industrial products, devices, equipment, or systems by modeling failure times with bathtub curve behavior using classical and advanced probability distributions. The research addressed an academic problem identified in industrial engineering education: the absence of a clear and reproducible applied methodology that allows for the selection, adjustment, and interpretation of probabilistic models capable of representing failures with non-monotonic behavior.

The study was based on reference data widely used in reliability literature and combined exploratory analysis, nonparametric risk function estimation, and parametric model fitting using maximum likelihood. Traditional and extended distributions were compared based on statistical and engineering consistency criteria, evaluating metrics such as mean time to failure, mean residual life, and survival probability.

The results showed that flexible probabilistic models offer a more accurate representation of the early mortality, stable lifetime, and accelerated wear phases, overcoming the limitations of the standard Weibull distribution. It was also demonstrated that the correct selection of the model directly influences the interpretation of failure behavior and the generation of useful information for reliability decision-making.

The main contribution of this work consists of the systematization of a reproducible methodological procedure, documented using R scripts, which strengthens the link between applied statistics and reliability engineering in the academic field of industrial engineering.

**Keywords:** reliability engineering, bathtub curve, failure times, probability distributions, survival analysis.

# Capítulo I. Planteamiento del problema

## 1.1 Introducción

La Ingeniería Industrial es una disciplina orientada al diseño, análisis, mejora y optimización de sistemas productivos y de servicios, con el propósito de maximizar la eficiencia, la productividad y la calidad organizacional. En los últimos años, el entorno industrial ha experimentado una transformación profunda impulsada por la automatización, la digitalización y la integración de tecnologías asociadas a la Industria 4.0. Estos cambios han incrementado la complejidad operativa de los sistemas, lo que a su vez ha generado nuevos retos vinculados con la continuidad de los procesos, la gestión estratégica de activos y la necesidad de disponer de metodologías confiables para anticipar, modelar y controlar fallas en equipos críticos, dispositivos o productos finales (Bracke & Pieper, 2025). En este contexto, la capacidad para comprender y predecir comportamientos de falla se ha convertido en un elemento esencial para garantizar la estabilidad y competitividad de las operaciones industriales.

La Ingeniería de Confiabilidad ha adquirido una relevancia creciente al proporcionar métodos cuantitativos que permiten analizar el comportamiento de los componentes a lo largo del tiempo, estimar la probabilidad de ocurrencia de fallas, identificar patrones de deterioro y diseñar estrategias de mantenimiento fundamentadas en evidencia estadística. Según Wu et al. (2023), la confiabilidad constituye un pilar central en la gestión moderna de activos, al reducir interrupciones no programadas, mitigar pérdidas económicas asociadas al tiempo muerto y mejorar la disponibilidad técnica de los sistemas. Diversos estudios reportan que la implementación adecuada de técnicas de confiabilidad puede reducir los costos de mantenimiento entre un 20 % y un 30 %, además de incrementar la disponibilidad de equipos en proporciones superiores al 15 % (Nor, Pedapati & Muhammad, 2021; Elsahhar et al., 2025), lo que evidencia su impacto económico y operativo.

Dentro de este campo, el análisis de los tiempos falla constituye una de las herramientas fundamentales para comprender los mecanismos de deterioro de los equipos, productos o dispositivos. Uno de los comportamientos más estudiados en la literatura es la llamada “curva de bañera”, que describe la

evolución típica del riesgo de falla a través de tres fases: mortalidad temprana, vida útil estable y desgaste. Este patrón permite relacionar distintos mecanismos físicos de falla con etapas específicas del ciclo de vida y constituye una base indispensable para diseñar políticas de mantenimiento preventivo y predictivo (Lai, Xie & Murthy, 2001; Du & Sun, 2022). La modelación precisa de la curva de bañera es esencial para anticipar la transición entre fases, evitar fallas catastróficas e intervenir en momentos oportunos.

Más allá de su dimensión técnica, la confiabilidad tiene implicaciones directas en la seguridad, la sostenibilidad y la competitividad de las organizaciones. Fallas inesperadas en sistemas críticos pueden producir daños materiales, interrupciones prolongadas en la cadena de suministro, riesgos para la integridad del personal y afectaciones a la productividad. En industrias como la manufacturera, minera, energética y aeronáutica se ha documentado que las políticas de mantenimiento basadas en datos (particularmente aquellas sustentadas en modelos probabilísticos de falla) aumentan la seguridad operativa y reducen la variabilidad en los tiempos de intervención (Giorgio, Piscopo & Pulcini, 2023; Rahman et al., 2024).

Desde la perspectiva académica, la Ingeniería de Confiabilidad constituye un puente entre la estadística aplicada, la analítica de datos, la ingeniería de mantenimiento y la gestión de operaciones. Su estudio integra conceptos como modelos de supervivencia, análisis de riesgo, criterios de selección de modelos, estimación de parámetros mediante el método de máxima verosimilitud y métodos no paramétricos para caracterizar la función de falla (hazard<sup>1</sup> function). Estas herramientas permiten transformar datos empíricos en información para la toma de decisiones tácticas y estratégicas. La literatura reciente enfatiza que la capacidad de un ingeniero para modelar distribuciones de falla, interpretar parámetros y comunicar resultados cuantitativos es una competencia clave en entornos productivos altamente tecnificados (Giorgio et al., 2023; Liu et al., 2024).

No obstante, a pesar de la importancia de estos enfoques, diversos diagnósticos educativos señalan que los egresados de Ingeniería Industrial

---

<sup>1</sup> A lo largo de la tesis se estará utilizando la palabra “hazard” para referirnos al concepto de función de falla

suelen presentar limitaciones para aplicar de manera práctica los conceptos de confiabilidad en escenarios reales. Aunque cursan asignaturas de probabilidad, estadística y análisis de datos, dichos contenidos rara vez se vinculan explícitamente con el análisis real de fallas, la interpretación de la función de falla o la selección de modelos adecuados para comportamientos no monótonos. Esta desconexión entre teoría y práctica dificulta que los futuros profesionales utilicen de forma efectiva las herramientas estadísticas disponibles.

En el contexto institucional de la Universidad Autónoma del Estado de Hidalgo (UAEH), cuya misión educativa enfatiza la formación de profesionales capaces de resolver problemas complejos mediante herramientas científicas y tecnológicas, se vuelve indispensable fortalecer las competencias relacionadas con la confiabilidad y la modelación de fallas. En este marco, el presente proyecto de tesis se orienta al estudio de analizar fallas en forma de curva de bañera y a la modelación de tiempos de falla mediante distribuciones de probabilidad adecuadas, con el propósito de proponer un procedimiento metodológico reproducible que facilite la selección del modelo, la estimación de parámetros, la evaluación del ajuste y la interpretación de los resultados. De esta manera, esta tesis pretende responder tanto a las necesidades del sector industrial como a los principios institucionales de calidad educativa, pertinencia social y formación analítica del ingeniero industrial.

## **1.2 Antecedentes**

La Ingeniería de Confiabilidad se ha consolidado como una de las áreas esenciales dentro de la Ingeniería Industrial debido a su impacto directo en la continuidad operativa, la gestión estratégica de activos y la reducción de riesgos asociados a fallas inesperadas. Su objetivo principal consiste en analizar y modelar el comportamiento de los equipos productivos, dispositivos o productos finales a lo largo del tiempo, permitiendo anticipar fallas, estimar la vida útil de los componentes o productos y diseñar estrategias de mantenimiento basadas en evidencia cuantitativa. En un contexto industrial marcado por la automatización, la digitalización y la transición a la Industria 4.0, estas capacidades se han vuelto cruciales para asegurar la productividad, la eficiencia y la disponibilidad de los sistemas (Bracke & Pieper, 2025; Wu et al., 2023).

Históricamente, el campo de la confiabilidad surgió en sectores donde la predicción y prevención de fallas eran críticas, como la industria militar, la aviación y los sistemas eléctricos. Los primeros enfoques se apoyaban en inspecciones, pruebas de vida acelerada y herramientas estadísticas básicas. Conforme los sistemas industriales incrementaron su complejidad, fue necesario desarrollar modelos probabilísticos más robustos que permitieran interpretar con precisión la evolución del riesgo de falla. Este proceso dio lugar a la modelación mediante distribuciones de probabilidad y funciones matemáticas para describir la función de falla, sentando las bases modernas de la Ingeniería de Confiabilidad (Lai, Xie & Murthy, 2001).

Entre los conceptos centrales del área destaca la curva de bañera, modelo conceptual que describe la variación del riesgo de falla a lo largo del ciclo de vida de un componente o producto en tres etapas: mortalidad infantil, vida útil estable y desgaste. La fase inicial se asocia con fallas tempranas debidas a defectos de fabricación o instalación; la fase intermedia representa un periodo de riesgo aproximadamente constante; y la fase final refleja un incremento del riesgo por envejecimiento y deterioro acumulado (Lai et al., 2001). La correcta modelación de estas fases resulta fundamental para implementar políticas efectivas de mantenimiento preventivo y predictivo, reduciendo costos operativos y mejorando la disponibilidad de los sistemas (Du & Sun, 2022).

En este contexto, la representación estadística de los tiempos hasta la falla mediante distribuciones de probabilidad se ha vuelto una práctica estándar. Por ejemplo, la distribución Weibull ha sido ampliamente utilizada por su flexibilidad y por la interpretación relativamente sencilla de sus parámetros (Wang, Sha, Gu & Xu, 2014). Sin embargo, múltiples trabajos han demostrado que la distribución Weibull clásica presenta limitaciones para representar comportamientos no monótonos del riesgo (falla), como los asociados a la curva de bañera. Esta limitación ha impulsado el desarrollo de modelos más sofisticados, entre los que se incluyen la Weibull modificada, las distribuciones Power-X, la Weibull-Pareto y las distribuciones definidas en intervalos finitos, que ofrecen un ajuste más adecuado en escenarios con cambios múltiples de tendencia en la función de falla (Almetwally et al., 2025; Jiang et al., 2023; Wang, Yu & Li, 2015).

Paralelamente, la literatura reciente ha explorado enfoques estocásticos para modelar procesos de degradación complejos. Entre ellos destacan los procesos tipo Wiener con tasa de falla en forma de bañera, que integran explícitamente la variabilidad de las condiciones de operación y permiten inferir la proximidad al umbral de falla (Giorgio, Piscopo & Pulcini, 2023). En el contexto de la Industria 4.0, han surgido también modelos híbridos que combinan distribuciones de probabilidad con algoritmos de aprendizaje automático, los cuales han mostrado mejoras en la predicción del Tiempo Promedio Hasta la Falla (MTTF<sup>2</sup>) y en la identificación de patrones de deterioro cuando se cuenta con datos provenientes de sensores y sistemas ciberfísicos (Wu et al., 2023).

Desde la perspectiva operativa, metodologías como el Mantenimiento Basado en Condición (CBM, por sus siglas en inglés), el Mantenimiento Centrado en Confiabilidad (RCM, por sus siglas en inglés) y el Mantenimiento Productivo Total (TPM, por sus siglas en inglés) se han consolidado como marcos de referencia para reducir el riesgo de fallas inesperadas y maximizar la disponibilidad de los equipos. No obstante, su eficacia depende en gran medida de la calidad del análisis estadístico subyacente y de la capacidad del personal para interpretar correctamente el comportamiento del riesgo (Giorgio et al., 2023). Diversos estudios coinciden en que la falta de estandarización en la modelación estadística de fallas sigue siendo un problema recurrente, especialmente cuando los sistemas exhiben comportamientos tipo curva de bañera.

En México, universidades como la UNAM, el IPN y la UAEH han promovido investigaciones orientadas a fortalecer el uso de modelos estadísticos en la gestión de confiabilidad y en la industria de transformación (Martínez-Reyes, 2021; Ramírez-Cortés, 2020). Sin embargo, la adopción en la industria, particularmente en pequeñas y medianas empresas, sigue enfrentando obstáculos relacionados con la disponibilidad de datos de calidad, la capacitación insuficiente del personal técnico y la ausencia de metodologías accesibles que conecten la teoría estadística con aplicaciones reales (Hernández-Santiago, 2019). Esta situación refleja una brecha entre la formación

---

<sup>2</sup> Tiempo Promedio Hasta la Falla (Mean Time To Failure (MTTF)). Este termino se utilizará a lo largo de la tesis.

académica y las necesidades del entorno productivo, con impacto directo en la competitividad y sostenibilidad operativa.

Este desafío se agrava al considerar los requerimientos de normativas internacionales como ISO 9001:2015, ISO 31000:2018, ISO 14224:2016 e ISO 55000:2014, que demandan la implementación de sistemas de gestión basados en evidencia y análisis riguroso del riesgo. Si bien las Normas Oficiales Mexicanas relacionadas con seguridad industrial no abordan explícitamente la confiabilidad, sí exigen condiciones de operación seguras y controladas, lo cual implica la necesidad de evaluar y gestionar adecuadamente el comportamiento de las fallas. En este contexto, el dominio de herramientas modernas de análisis de confiabilidad se convierte en una competencia clave para los ingenieros industriales.

El análisis de los antecedentes permite identificar un punto común: existe un cuerpo teórico amplio sobre la modelación de fallas y sobre distribuciones avanzadas capaces de representar comportamientos complejos, pero su aplicación práctica sigue siendo limitada. En particular, la literatura coincide en señalar la necesidad de contar con metodologías estructuradas que faciliten la selección, ajuste y comparación de modelos para datos con comportamiento tipo curva de bañera, y que además sean compatibles con las capacidades formativas de los estudiantes de Ingeniería Industrial. Esta brecha entre teoría y práctica sirve de base para el planteamiento del problema que aborda la presente tesis.

### **1.3 Planteamiento del problema**

En los sistemas industriales contemporáneos, caracterizados por altos niveles de automatización, integración digital y exigencias operativas cada vez más estrictas, la gestión de la confiabilidad se ha consolidado como un componente estratégico para asegurar la continuidad del negocio o las operaciones, la eficiencia de los procesos y la competitividad de las organizaciones. A pesar del desarrollo de modelos estadísticos avanzados y de la creciente disponibilidad de datos operativos, la industria continúa enfrentando dificultades para seleccionar, ajustar e interpretar adecuadamente las distribuciones de probabilidad que describen el comportamiento real de los equipos, dispositivos o productos,

especialmente cuando los datos presentan patrones de riesgo no monótonos, propios de la curva de bañera (Lai, Xie & Murthy, 2001; Wang, Yu & Li, 2015).

En la práctica industrial, persiste un uso predominante de la distribución Weibull estándar como modelo casi universal para el análisis de fallas. Aunque esta distribución ha demostrado ser versátil en escenarios con riesgo creciente o decreciente, presenta limitaciones importantes cuando los datos exhiben dos cambios de tendencia, multimodalidad o fases claramente diferenciadas de mortalidad temprana, vida útil estable y desgaste acelerado (Wang, Sha, Gu & Xu, 2014). Como consecuencia, los modelos basados exclusivamente en la Weibull clásica no capturan con precisión la estructura temporal del riesgo, lo que conduce a estimaciones imprecisas de indicadores como el tiempo medio hasta la falla, la tasa de falla instantánea o la probabilidad de falla en intervalos específicos (Jiang et al., 2023).

Esta situación tiene repercusiones directas en la operación industrial. Estudios en sectores intensivos en activos han documentado discrepancias significativas entre el comportamiento real de los equipos y las predicciones de los modelos tradicionales. En turbinas eólicas envejecidas, por ejemplo, se observa un incremento abrupto de las tasas de fallo al final de la vida útil que no es reproducido adecuadamente por modelos convencionales, lo cual compromete la planificación del mantenimiento y las decisiones sobre extensión de vida útil (Elsahhar, Rahman & van Bussel, 2025). De manera similar, en sistemas multiestado sometidos a cargas variables, la coexistencia de múltiples modos de falla rompe el supuesto de estacionariedad de los modelos clásicos, reduciendo su capacidad predictiva y la confiabilidad de los indicadores utilizados en la toma de decisiones (Varshney, Bajaj, Choudhary, Pushkarna & Zaitsev, 2025).

El problema se intensifica al analizar bases de datos reales empleadas en mantenimiento predictivo y analítica industrial. Repositorios como SCANIA Component X presentan desbalanceo severo, censura, ruido, variaciones temporales y condiciones operativas heterogéneas, lo que dificulta la estimación estable de parámetros mediante el método de máxima verosimilitud y complica la comparación justa entre modelos (Kharazian, Lindgren, Magnússon, Steinert & Reyna, 2025). Sin procedimientos metodológicos claros, las decisiones de

mantenimiento pueden basarse en modelos estadísticamente débiles o físicamente poco plausibles.

A nivel metodológico, la literatura ha respondido a estas limitaciones mediante el desarrollo de distribuciones extendidas y modelos más flexibles, como las Weibull modificadas (incluyendo las Improved Modified Weibull), las distribuciones Power-X, las variantes Weibull–Pareto y las distribuciones definidas en intervalos finitos (Jiang et al., 2023; Almetwally et al., 2025; Wang, Yu & Li, 2015). Asimismo, se han propuesto modelos basados en procesos de degradación, como el proceso de Wiener con tasa de riesgo en forma de bañera, que integran la variabilidad ambiental y operacional de manera más realista (Giorgio, Piscopo & Pulcini, 2023). Sin embargo, la adopción de estos enfoques en la industria sigue siendo limitada porque no existen guías comparativas estandarizadas que indiquen cómo seleccionar, ajustar y evaluar estos modelos de acuerdo con criterios estadísticos y operativos.

En paralelo, se manifiesta un problema significativo en la formación del ingeniero industrial. Aunque los estudiantes cursan asignaturas de probabilidad, estadística, análisis de datos y mantenimiento industrial, estos contenidos suelen impartirse de manera fragmentada y sin conexión explícita con el análisis real de fallas (Hernández-Santiago, 2019; Martínez-Reyes, 2021; Ramírez-Cortés, 2020; Giorgio, Piscopo & Pulcini, 2023). Como resultado, los egresados presentan dificultades para:

- Seleccionar distribuciones adecuadas para modelar comportamientos no monótonos del riesgo.
- Estimar parámetros utilizando métodos robustos.
- Comparar modelos mediante métricas como AIC (Criterio de Información de Akaike), BIC (Criterio de Información Bayesiano) o pruebas K–S (Kolmogórov–Smirnov).
- Interpretar los resultados en términos de confiabilidad, disponibilidad y costo del ciclo de vida.
- Comunicar análisis de manera profesional, transparente y reproducible.

Ante esta brecha formativa, muchos análisis de confiabilidad realizados en la práctica profesional se llevan a cabo de forma intuitiva o se limitan al uso de la distribución Weibull clásica, incluso cuando existen evidencias claras de comportamientos tipo curva de bañera, multimodalidad o deterioro no lineal. Esto deriva en diagnósticos incompletos, intervalos de inspección mal diseñados y decisiones de reemplazo que no optimizan la disponibilidad ni el costo total de propiedad del activo.

En síntesis, se identifica una brecha metodológica, académica y profesional: no existe un marco metodológico estructurado, claro y reproducible que permita a estudiantes y egresados de Ingeniería Industrial modelar tiempos de falla con comportamiento tipo curva de bañera utilizando distribuciones de probabilidad avanzadas, criterios objetivos de comparación y procedimientos de interpretación en ingeniería. Esta ausencia limita la formación del ingeniero industrial y restringe la adopción de estrategias de mantenimiento basadas en datos por parte del sector productivo.

Por lo tanto, se plantea como problema central de esta tesis la necesidad de desarrollar un procedimiento metodológico claro, didáctico y reproducible que permita a los futuros ingenieros industriales modelar tiempos de falla con comportamiento tipo curva de bañera mediante distribuciones de probabilidad adecuadas, facilitando la selección del modelo, la estimación de parámetros, la comparación objetiva entre alternativas y la interpretación técnica de los resultados en términos de confiabilidad, disponibilidad y toma de decisiones de mantenimiento.

#### **1.4 Justificación**

La confiabilidad industrial constituye un eje fundamental en entornos productivos caracterizados por altos niveles de automatización, digitalización e integración tecnológica. En estos contextos, la continuidad operativa de equipos, dispositivos o productos se ha convertido en requisitos indispensables para la competitividad y sostenibilidad de las organizaciones. Diversos estudios han documentado que las fallas inesperadas en sistemas críticos generan pérdidas económicas significativas, alcanzando valores de decenas de miles de dólares por minuto de inactividad en sectores como el automotriz, energético y aeronáutico (Varshney et al., 2025; Elsayhar, Rahman & van Bussel, 2025). Bajo

estas condiciones, el desarrollo y la aplicación de modelos estadísticos rigurosos para predecir el comportamiento de falla adquieren una relevancia estratégica para la gestión de activos, la planeación del mantenimiento y la optimización del ciclo de vida de los productos o equipos (Wu et al., 2023; Nor, Pedapati & Muhammad, 2021).

En el ámbito de la Ingeniería de Confiabilidad, la curva de bañera se reconoce como un modelo conceptual central debido a su capacidad para representar las fases de mortalidad temprana, vida útil estable y desgaste acelerado a lo largo del ciclo de vida de un componente (Lai, Xie & Murthy, 2001; Du & Sun, 2022). No obstante, la práctica industrial continúa basándose predominantemente en el uso de la distribución Weibull clásica, a pesar de sus limitaciones ampliamente documentadas para modelar tasas de riesgo no monótonas o estructuras de falla con múltiples cambios de tendencia (Wang, Sha, Gu & Xu, 2014; Wang, Yu & Li, 2015). Aunque existen alternativas metodológicas avanzadas como las distribuciones: Power-X, Weibull-Pareto, Improved Modified Weibull y Beta-Weibull que han demostrado un mejor desempeño en la representación de comportamientos complejos del riesgo (Jiang et al., 2023; Almetwally et al., 2025), su adopción en la industria permanece restringida. Esta situación conduce a la utilización de modelos incompletos que comprometen la precisión de las estimaciones del tiempo medio hasta la falla, afectan la programación del mantenimiento y limitan la eficiencia del ciclo de vida de los activos o productos.

En el contexto nacional, estas limitaciones se acentúan debido a la escasa adopción de herramientas estadísticas avanzadas en pequeñas y medianas empresas manufactureras. Investigaciones recientes señalan que una proporción considerable de la pérdida de productividad está asociada a fallas no detectadas oportunamente y a decisiones de mantenimiento basadas en criterios subjetivos (Martínez-Reyes, 2021; Hernández-Santiago, 2019). Adicionalmente, la ausencia de metodologías accesibles para seleccionar, ajustar y comparar modelos probabilísticos dificulta el cumplimiento de estándares internacionales de gestión de activos, tales como ISO 55000, ISO 14224 e ISO 9001, los cuales exigen prácticas fundamentadas en datos y análisis de riesgo.

Desde la perspectiva académica, la justificación del presente estudio se vincula con la necesidad de fortalecer la formación metodológica y analítica del

ingeniero industrial. Si bien los planes de estudio incluyen asignaturas relacionadas con estadística, análisis de datos e ingeniería del mantenimiento, estos contenidos se imparten generalmente de forma fragmentada, lo cual dificulta su integración en procedimientos prácticos para la modelación de fallas con distribuciones avanzadas. Esta brecha formativa limita la capacidad de los egresados para aplicar técnicas de máxima verosimilitud, comparar modelos mediante criterios tales como AIC, BIC o pruebas K–S, e interpretar los resultados en términos de indicadores de confiabilidad y toma de decisiones (Giorgio, Piscopo & Pulcini, 2023; Wu et al., 2023).

Ante este panorama, la presente tesis se justifica por su contribución metodológica y educativa. Proporciona un procedimiento estructurado, didáctico y reproducible para modelar datos de falla con comportamiento tipo curva de bañera, integrando la selección de distribuciones apropiadas, la estimación de parámetros, la evaluación estadística del ajuste y la interpretación ingenieril de los resultados. Este enfoque fortalece competencias esenciales como el análisis estadístico aplicado, la comprensión de la función de riesgo y la capacidad de comunicar hallazgos técnicos de manera profesional (Ramírez-Cortés, 2020; Ramaki et al., 2025; El-Sagheer & Ramadan, 2025).

Finalmente, la pertinencia práctica del estudio radica en que el procedimiento propuesto constituye un marco aplicable a entornos reales, al facilitar la transición hacia estrategias de mantenimiento preventivo y predictivo basadas en métricas como el MTTF, la función de falla y la vida residual. De esta manera, el trabajo se alinea con los objetivos formativos de la Universidad Autónoma del Estado de Hidalgo (UAEH) y contribuye a la formación de ingenieros capaces de enfrentar los desafíos de la industria contemporánea mediante decisiones fundamentadas en el análisis estadístico riguroso.

## **1.5 Objetivos del proyecto**

### **Objetivo general**

Desarrollar un marco metodológico, didáctico y reproducible para modelar tiempos de falla en forma de curva de bañera utilizando distribuciones de probabilidad avanzadas, tomando como caso de estudio el conjunto de datos de Aarset (1987), con el fin de mejorar la precisión del análisis de confiabilidad y orientar decisiones de mantenimiento que incrementen la disponibilidad y

eficiencia de los sistemas industriales.

### **Objetivos específicos**

- Caracterizar el comportamiento estadístico de los tiempos de falla del conjunto de datos de Aarset (1987) mediante un análisis exploratorio que identifique su variabilidad, posibles censuras y las fases asociadas al patrón de curva de bañera, empleando para ello la estimación no paramétrica de la función de supervivencia y el diagnóstico preliminar de la función de falla.
- Evaluar el ajuste de diversas distribuciones de probabilidad candidatas mediante criterios estadísticos (AIC, BIC, KS y análisis del hazard), con el fin de determinar su capacidad para representar adecuadamente el comportamiento no monótono observado en los datos y seleccionar el modelo óptimo para la curva de bañera.
- Desarrollar un procedimiento metodológico que establezca las fases de análisis, la estimación de parámetros, la comparación de modelos y la interpretación ingenieril de métricas como MTTF, función de falla, probabilidad de falla y vida residual, con el propósito de generar una guía aplicada, reproducible y útil para apoyar la toma de decisiones en mantenimiento y gestión de activos.

### **1.6 Hipótesis**

El presente estudio parte del supuesto de que los datos de falla que exhiben comportamientos no monótonos en su tasa de riesgo (como el conjunto de datos clásico de Aarset (1987)) requieren modelos probabilísticos más flexibles que la distribución Weibull estándar para representar adecuadamente las fases de mortalidad temprana, vida útil estable y desgaste acelerado asociadas al patrón de curva de bañera. A partir de este fundamento teórico y metodológico, esta tesis plantea que la utilización de distribuciones extendidas permite mejorar el ajuste estadístico, representar con mayor fidelidad la estructura temporal de las fallas y, en consecuencia, fortalecer la interpretación ingenieril en el análisis de confiabilidad.

La literatura especializada señala que modelos avanzados como Weibull modificada, Power-X, Beta-Weibull o Weibull-Pareto incrementan su capacidad

para capturar cambios en la curvatura de la función de falla, manejar datos con censura y representar comportamientos complejos que la Weibull clásica no puede reproducir debido a su forma restrictiva (Lai, Xie & Murthy, 2001; Wang, Yu & Li, 2015; Jiang et al., 2023). Asimismo, estudios aplicados en sistemas energéticos, manufactura avanzada y transporte han mostrado mejoras en la precisión del tiempo medio hasta la falla y en la identificación de las fases de deterioro cuando se emplean modelos con mayor flexibilidad paramétrica (Elsahhar, Rahman & van Bussel, 2025; Varshney et al., 2025).

Aunque esta tesis utiliza un conjunto de datos de referencia y no una base industrial, el propósito analítico radica en comprobar si estos modelos avanzados presentan un desempeño superior bajo un escenario académico controlado, permitiendo establecer lineamientos metodológicos extrapolables a futuros estudios o aplicaciones profesionales. Con base en lo anterior, se formulan las siguientes hipótesis:

#### **Hipótesis nula ( $H_0$ )**

No existe diferencia estadísticamente significativa en el desempeño del modelo de confiabilidad al comparar la Weibull estándar con distribuciones de probabilidad avanzadas. Por lo tanto, la utilización de modelos extendidos no mejora la capacidad para representar el comportamiento tipo curva de bañera ni incrementa la precisión en estimaciones como MTTF, hazard o probabilidad de falla.

#### **Hipótesis alternativa ( $H_1$ )**

Existe una diferencia estadísticamente significativa en el desempeño del modelo de confiabilidad cuando se emplean distribuciones avanzadas. En consecuencia, estos modelos sí mejoran la representación del patrón de curva de bañera y permiten obtener estimaciones más precisas de indicadores como MTTF, tasa de falla y métricas derivadas de la función de supervivencia.

#### **Variables del estudio**

##### **Variable independiente ( $X$ ): Tipo de modelo probabilístico aplicado**

Corresponde a la categoría del modelo utilizado en el ajuste:

- Weibull estándar (modelo tradicional).
- Distribuciones avanzadas evaluadas: Beta–Weibull, Power–X, Weibull modificada, Weibull–Pareto u otros modelos flexibles.

Esta variable permite contrastar si el nivel de flexibilidad paramétrica influye en la calidad del ajuste y en la interpretación de la función de falla.

### **Variable dependiente (Y): Desempeño del modelo de confiabilidad**

Se evalúa mediante indicadores cuantitativos derivados del ajuste y de la interpretación estadística:

- Ajuste estadístico (AIC, BIC, K–S).
- Precisión en el MTTF.
- Representación de la función de falla (forma, monotonidad y coherencia física).
- Probabilidad de falla  $F(t)$  y supervivencia  $R(t)$ .
- Estabilidad de los parámetros estimados.

La relación entre ambas variables permite comprobar si los modelos avanzados aportan mejoras significativas frente al modelo tradicional.

## **1.7 Alcances y limitaciones**

### **Alcances**

Este proyecto de tesis se desarrolla en el ámbito académico y metodológico de la Ingeniería de Confiabilidad, con el propósito de diseñar, documentar y validar un procedimiento estructurado para modelar tiempos de falla con comportamiento tipo curva de bañera mediante distribuciones de probabilidad avanzadas. A diferencia de investigaciones aplicadas en entornos industriales, el alcance de esta tesis se circunscribe al análisis detallado del conjunto de datos clásico de Aarset (1987), ampliamente utilizado en la literatura para evaluar modelos de riesgo no monótono y validar metodologías de modelación estadística (Lai, Xie & Murthy, 2001; Jiang et al., 2023). Esto permite trabajar con

un escenario controlado, replicable y teóricamente fundamentado, adecuado para fines formativos y metodológicos.

El estudio abarca el análisis exploratorio del comportamiento de los tiempos de falla; la estimación no paramétrica de la función de supervivencia y de la función de falla; la comparación formal entre modelos tradicionales (como Weibull estándar) y distribuciones avanzadas (como Beta–Weibull, Weibull modificada o Power–X); y la evaluación de criterios de bondad de ajuste como AIC, BIC y K–S. Asimismo, se incluyen métricas de confiabilidad derivadas del modelo óptimo, tales como el tiempo medio hasta la falla, la probabilidad de supervivencia condicional, la función de falla  $h(t)$  y la vida residual media, con el fin de demostrar la aplicabilidad del procedimiento propuesto.

El alcance metodológico se centra exclusivamente en herramientas estadísticas implementadas en el software de código abierto “R”, empleando librerías especializadas en análisis de supervivencia, funciones de densidad, estimación por máxima verosimilitud y diagnóstico visual mediante curvas de supervivencia, hazard empírico y gráficos QQ y PP. Este enfoque garantiza la reproducibilidad del análisis y la claridad en cada una de las fases que conforman el procedimiento metodológico.

En cuanto a su contribución educativa, el estudio se orienta a fortalecer la formación del ingeniero industrial mediante la integración de conceptos de análisis de supervivencia, estadística aplicada y confiabilidad, transformando un caso académico clásico en una guía metodológica que permita, en un futuro, analizar datos de falla reales bajo un enfoque riguroso y replicable. Si bien el estudio no evalúa equipos industriales específicos, los lineamientos generados pueden ser extrapolados posteriormente a contextos reales de mantenimiento, siempre que se cuente con datos adecuados.

## **Limitaciones**

La primera y más importante limitación del estudio es que no utiliza datos provenientes de una empresa ni de sistemas industriales reales, sino el conjunto de datos de Aarset (1987). Aunque este conjunto es ampliamente aceptado como referencia académica para validar modelos con riesgo no monótono, su utilización restringe el análisis a un contexto simulado y no permite evaluar variabilidad operacional, factores ambientales o condiciones de uso reales.

Otra limitación se relaciona con la ausencia de censura en los datos, lo cual facilita el análisis estadístico, pero reduce la complejidad respecto a escenarios industriales donde la censura es común y puede afectar de manera importante la estimación de parámetros y la estabilidad de los modelos (Varshney et al., 2025; Elsayhar et al., 2025). Esto implica que el procedimiento propuesto deberá ajustarse en estudios futuros que incorporen censura a la derecha o datos incompletos.

Asimismo, el estudio se limita al análisis estadístico de confiabilidad, sin incluir aspectos económicos, logísticos o administrativos de la gestión de mantenimiento, como costos de reemplazo, optimización de inventarios o políticas de mantenimiento basadas en disponibilidad. Aunque estos aspectos son relevantes, su incorporación se reserva para trabajos posteriores.

Finalmente, la generalización de los resultados queda restringida al ámbito metodológico y formativo. Si bien el procedimiento propuesto es válido y robusto, su aplicación en industrias reales dependerá de la calidad, cantidad y estructura de los datos disponibles, así como de la capacidad del analista para interpretar adecuadamente los modelos y adaptar la metodología a condiciones operativas específicas.

En conjunto, estas limitaciones no disminuyen la relevancia del estudio; por el contrario, destacan su naturaleza académica y su aporte como referencia metodológica sólida para futuras investigaciones y aplicaciones industriales.

### **1.8 Descripción general de la estructura de la tesis**

La presente tesis se organiza en cinco capítulos que desarrollan de manera progresiva los fundamentos conceptuales, la metodología propuesta, el análisis estadístico y la interpretación crítica de los resultados en el marco de la Ingeniería de Confiabilidad. Esta estructura responde al propósito central del estudio: construir un procedimiento metodológico riguroso para modelar tiempos de falla con comportamiento tipo curva de bañera mediante distribuciones de probabilidad avanzadas, y analizar su pertinencia estadística e ingenieril utilizando el conjunto de datos clásico de Aarset (1987).

- **El Capítulo I** introduce el contexto general del estudio y plantea las bases que orientan el desarrollo de la investigación. Se presentan la importancia

de la confiabilidad industrial en entornos productivos modernos, los antecedentes teóricos y aplicados, el planteamiento del problema y las brechas existentes entre la teoría estadística y su aplicación en el análisis de fallas con riesgo no monótono. Asimismo, se formulan los objetivos general y específicos, se establecen las hipótesis, y se delimitan los alcances y limitaciones del estudio, junto con la justificación académica, profesional y metodológica del proyecto.

- **El Capítulo II**, correspondiente al Marco Teórico, expone los fundamentos conceptuales que sustentan la tesis. Se desarrollan los principios de la Ingeniería de Confiabilidad, los modelos probabilísticos clásicos y extendidos utilizados para representar tiempos de falla, y la descripción matemática de la curva de bañera, así como de sus implicaciones en mantenimiento y gestión de activos. Se revisan las distintas familias de distribuciones utilizadas para modelar riesgo no monótono, los criterios de selección de modelos (AIC, BIC, K-S), los métodos de estimación de parámetros y las técnicas no paramétricas de supervivencia, proporcionando el sustento teórico necesario para la interpretación posterior de los resultados.
- **El Capítulo III** presenta la Metodología empleada en el estudio. Se describe el enfoque cuantitativo adoptado, las etapas del procedimiento metodológico desarrollado y los criterios de selección del conjunto de datos de Aarset (1987). Se detalla el proceso para la caracterización exploratoria del comportamiento de falla, el ajuste de distribuciones avanzadas, la estimación de parámetros mediante máxima verosimilitud, y la comparación de modelos con base en criterios estadísticos e ingenieriles. El capítulo incluye además el diseño del marco metodológico propuesto y los lineamientos que estructuran el análisis reproducible.
- **El Capítulo IV** presenta los Resultados obtenidos a lo largo de la aplicación del procedimiento metodológico. Se exponen los hallazgos del análisis exploratorio, la identificación preliminar de las fases de la curva de bañera y la descripción del comportamiento estadístico del conjunto de datos. Asimismo, se muestran los ajustes paramétricos obtenidos para cada modelo evaluado, las métricas de bondad de ajuste, las gráficas de comparación teórico-empírica y la estimación de indicadores clave como

MTTF, función de falla y confiabilidad residual. Este capítulo constituye la evidencia cuantitativa que permite evaluar la pertinencia de los modelos y fundamenta la discusión posterior.

- **El Capítulo V**, denominado Discusión, constituye la fase interpretativa de la tesis. En esta sección se analizan y contrastan los resultados con los fundamentos teóricos, antecedentes y objetivos establecidos. Se examinan las implicaciones técnicas de los hallazgos, se identifican coincidencias y discrepancias con estudios previos, y se evalúa el aporte del procedimiento metodológico al análisis de fallas con riesgo no monótono. Además, se discuten las limitaciones del estudio, las posibles aplicaciones prácticas del modelo en contextos industriales reales y las contribuciones del trabajo a la formación del ingeniero industrial. El capítulo cierra con la relación directa de los resultados con los objetivos planteados y con propuestas de líneas futuras de investigación.

En conjunto, estos cinco capítulos conforman una estructura lógica, coherente y metodológicamente rigurosa que permite abordar de manera integral el problema de modelación de fallas con comportamiento tipo curva de bañera, proporcionando una base científica y aplicada para fortalecer la práctica de la Ingeniería de Confiabilidad y la toma de decisiones basada en datos.

## Capítulo II. Marco teórico

### 2.1 Introducción

El presente capítulo tiene como finalidad establecer los fundamentos teóricos, conceptuales y científicos que sustentan este trabajo de tesis. Su propósito es proporcionar un marco sólido que permita comprender el fenómeno de estudio desde una perspectiva integral, articulando los avances de la estadística aplicada, la ingeniería de confiabilidad y la modelación probabilística con las necesidades identificadas en el planteamiento del problema.

En los entornos industriales actuales, caracterizados por altos niveles de automatización, integración digital y exigencias de continuidad operativa, la predicción precisa del comportamiento de falla constituye un elemento estratégico para la competitividad. En este contexto, el análisis estadístico de los tiempos hasta la falla y el empleo de distribuciones de probabilidad adecuadas adquieren un papel fundamental, particularmente cuando los sistemas presentan patrones no monótonos asociados a la curva de bañera (O'Connor & Kleyner, 2023; Elsayhar, Rahman & van Bussel, 2025). La comprensión rigurosa de estos comportamientos es indispensable para optimizar el mantenimiento, reducir la incertidumbre operativa y mejorar la disponibilidad de los activos.

Este capítulo cumple dos funciones esenciales. En primer lugar, sustenta teóricamente la hipótesis planteada en el capítulo anterior, demostrando por qué las distribuciones avanzadas ofrecen una representación más precisa del riesgo de falla que los modelos tradicionales. En segundo lugar, organiza los conceptos clave que guiarán el desarrollo metodológico de la tesis, proporcionando coherencia entre el marco conceptual, la selección de modelos y el análisis de resultados.

Para ello, el capítulo se estructura en tres apartados principales: a) el estado del arte, que revisa críticamente investigaciones relevantes sobre distribuciones avanzadas y modelación probabilística; b) los fundamentos teóricos, que explican las bases matemáticas y estadísticas del análisis de confiabilidad; y c) el marco conceptual, que define los términos operativos utilizados a lo largo del estudio. En conjunto, este capítulo establece la base académica que permitirá comprender y justificar el enfoque metodológico

aplicado en la modelación de la curva de bañera en sistemas industriales contemporáneos.

## 2.2 Estado del arte

El análisis de confiabilidad ha experimentado un desarrollo significativo a lo largo de las últimas décadas, impulsado por la necesidad de comprender y predecir el comportamiento de las fallas de los sistemas industriales en un contexto de creciente complejidad tecnológica. Tradicionalmente, la modelación del tiempo hasta la falla se ha basado en distribuciones de probabilidad que permiten representar matemáticamente el deterioro de los componentes y estimar métricas como la función de supervivencia, la función de riesgo y el tiempo medio hasta la falla. Sin embargo, numerosos estudios han demostrado que los modelos clásicos ofrecen una representación limitada cuando el riesgo de falla presenta comportamientos no monótonos, característica central de la curva de bañera (Lai, Xie & Murthy, 2001; O'Connor & Kleyner, 2023).

La distribución Weibull, introducida en la ingeniería de confiabilidad a mediados del siglo XX, se consolidó como el modelo paramétrico más utilizado debido a su flexibilidad para representar tasas de falla crecientes o decrecientes según el valor del parámetro de forma (Wang, Sha, Gu & Xu, 2014). Su versatilidad, interpretabilidad clara y facilidad de estimación mediante métodos como la máxima verosimilitud la han convertido en la referencia estándar para diagnósticos preliminares de fallas. No obstante, investigaciones recientes coinciden en que la Weibull clásica resulta insuficiente para capturar los tres regímenes característicos de la curva de bañera, ya que su función de falla presenta un único cambio de pendiente (Du & Sun, 2022; Wang, Yu & Li, 2015). Este hallazgo ha motivado el desarrollo de distribuciones extendidas que incorporan parámetros adicionales o transformaciones flexibles para representar con mayor precisión comportamientos complejos.

Entre estas alternativas destacan la Weibull modificada, la Modified Weibull, la Improved Modified Weibull y la Beta–Weibull, distribuciones que han mostrado un desempeño superior en el modelado de sistemas con patrones multimodales o no monótonos de riesgo (Jiang et al., 2023; Almetwally et al., 2025). La Beta–Weibull, en particular, se ha posicionado como una de las familias más robustas para capturar curvaturas complejas de la función de falla,

permitiendo ajustar con precisión fases de mortalidad temprana, meseta operativa y desgaste acelerado. Su capacidad para generar una amplia gama de formas mediante cuatro parámetros ha sido validada en sectores como manufactura, metalmecánica, aeroespacial y sistemas hidráulicos.

Asimismo, la literatura ha documentado el uso creciente de modelos de intervalo finito, distribuciones Power-X y familias compuestas como la Weibull-Pareto, especialmente en aplicaciones donde la vida útil del componente tiene un límite físico bien definido o donde la distribución presenta colas pesadas que afectan la estimación del riesgo (Almetwally et al., 2025; Ramaki et al., 2025). Estos modelos ofrecen una mayor flexibilidad estadística, aunque requieren métodos de estimación más complejos y conjuntos de datos suficientemente representativos para garantizar estabilidad en los parámetros.

En paralelo al desarrollo de modelos paramétricos, los enfoques no paramétricos y semiparamétricos han jugado un papel fundamental en el diagnóstico preliminar del comportamiento del riesgo. Métodos como Kaplan-Meier y Nelson-Aalen permiten estimar la función de supervivencia y la función de falla acumulada sin asumir una forma específica de la distribución, resultando especialmente útiles en presencia de datos censurados o tamaños muestrales reducidos (Torabi et al., 2022). Estos métodos son comúnmente utilizados para detectar visualmente si el comportamiento de la función de falla corresponde a un patrón de bañera antes de seleccionar un modelo paramétrico definitivo. En esta línea, el trabajo clásico de Aarset (1987) se ha convertido en un referente metodológico para ilustrar patrones de falla no monótonos y para validar la necesidad de modelos flexibles en el análisis de tiempos de falla.

Además de los modelos basados directamente en tiempos de falla, la literatura reciente ha enfatizado la importancia de los procesos estocásticos de degradación como alternativa para capturar el deterioro físico progresivo de los componentes. Modelos como el proceso de Wiener con drift dependiente del tiempo, el proceso Gamma no estacionario y los modelos de Browniano fraccional han demostrado una elevada capacidad para representar sistemas cuyo deterioro no puede observarse directamente a través de datos de falla, sino mediante señales continuas como vibración, temperatura, fricción o desgaste de herramientas (Giorgio, Piscopo & Pulcini, 2023; Lee & Park, 2021). Aunque estos enfoques ofrecen un marco más realista para algunos sistemas, su complejidad

matemática y los requerimientos de datos limitan su adopción en contextos educativos y en industrias con infraestructura limitada.

En el marco de la Industria 4.0, se ha observado un incremento notable en la integración de modelos probabilísticos con algoritmos de aprendizaje automático. Diversos autores han propuesto esquemas híbridos que combinan distribuciones paramétricas con redes neuronales, máquinas de soporte vectorial o modelos de boosting, especialmente en aplicaciones donde los datos son ruidosos, presentan alta dimensionalidad o exhiben múltiples factores influyentes (Wu et al., 2023; Kharazian et al., 2025). Aunque estas aproximaciones muestran mejoras en predicción, la literatura coincide en que su interpretabilidad es más limitada y que su adopción depende de una adecuada formación estadística del analista.

Los estudios revisados coinciden en identificar diversas problemáticas persistentes en el campo:

- La falta de metodologías comparativas estandarizadas para seleccionar el modelo probabilístico más adecuado en presencia de la función de falla no monótona.
- La desconexión entre criterios estadísticos de ajuste (AIC, BIC, K–S) y métricas operativas relevantes para la industria (MTTF, disponibilidad, costo del ciclo de vida).
- La escasa adopción de modelos avanzados en entornos industriales reales debido a la falta de guías prácticas.
- La insuficiente formación estadística aplicada en programas educativos de Ingeniería Industrial, lo que limita la capacidad de los egresados para modelar adecuadamente la curva de bañera y tomar decisiones basadas en evidencia (Martínez-Reyes, 2021; Ramírez-Cortés, 2020; Hernández-Santiago, 2019).

Por otra parte, la literatura reporta avances significativos en modelos capaces de representar tasas de falla con múltiples inflexiones, propios de sistemas con deterioro complejo, ciclos de carga variables o efectos combinados de desgaste y fallas tempranas. La Tabla 1 muestra algunas de las distribuciones más empleadas.

**Tabla 1.** Comparación de las distribuciones utilizadas para modelar curvas de bañera.

| Distribución                         | Nº de parámetros | Capacidad de modelar bañera | Ventajas                                       | Limitaciones                                |
|--------------------------------------|------------------|-----------------------------|------------------------------------------------|---------------------------------------------|
| Weibull clásica                      | 2                | Baja                        | Interpretación simple                          | Una sola fase dominante                     |
| Modified Weibull                     | 3                | Media                       | Permite curvaturas flexibles                   | Inestabilidad con pocos datos               |
| Improved Modified Weibull            | 4                | Alta                        | Ajuste superior en no-monotonía                | Estimación compleja                         |
| Beta–Weibull                         | 4                | Muy alta                    | Excelente ajuste a bañera (Jiang et al., 2023) | Sensible a censura                          |
| Power–X                              | 3                | Alta                        | Control independiente de forma y escala        | Requiere calibración rigurosa               |
| Weibull–Pareto                       | 4                | Alta                        | Colas pesadas útiles en desgaste tardío        | Sobremodelación en muestras pequeñas        |
| Modelos de variable intervalo finito |                  | Alta                        | Representan vida útil acotada                  | Necesidad de límites físicos bien definidos |

También, la literatura reciente confirma un avance significativo en la construcción de distribuciones avanzadas que superan las limitaciones de los modelos tradicionales al capturar con mayor precisión las características de la curva de bañera. Entre las más estudiadas se encuentran:

- Weibull modificada: mejora el ajuste en fases tempranas y en transiciones hacia el desgaste (Jiang et al., 2023).
- Power–Lindley y Power-X: permiten modelar hazard functions multimodales con alta variabilidad (Almetwally et al., 2025).
- Weibull–Pareto compuesta: captura colas pesadas y desgaste acelerado, útil en equipos sometidos a fricción extrema (Rahman et al., 2024).
- Beta–Weibull: representa de manera flexible comportamientos en fases iniciales y finales (Kumar & Singh, 2022).

- Q-Weibull: derivada de la entropía Tsallis, incorpora efectos de memoria y dependencia temporal (Silva et al., 2024).
- Log-Logistic generalizada: especialmente útil en procesos de fatiga y en sistemas con ciclos térmicos repetitivos (Zhang et al., 2022).

En síntesis, el estado del arte evidencia que, aunque existen modelos avanzados capaces de representar con precisión la curva de bañera y comportamientos complejos del riesgo, la aplicación práctica de estos modelos continúa enfrentando desafíos metodológicos, educativos y operativos. Esta situación confirma la pertinencia de desarrollar procedimientos didácticos, reproducibles y metodológicamente rigurosos que faciliten la selección, ajuste e interpretación de distribuciones avanzadas para el análisis de confiabilidad, tal como se propone en la presente tesis.

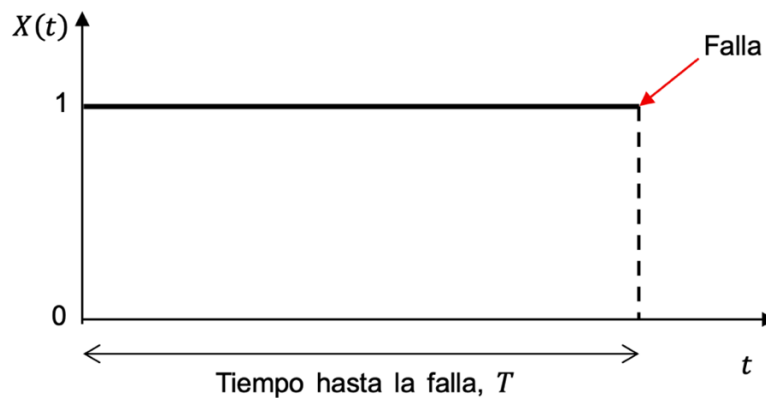
## 2.3 Fundamentos teóricos

### Tiempo hasta la falla (Time-to-Failure)

El tiempo hasta la falla o vida útil de un producto, componente, dispositivo, equipo o sistema es el periodo que transcurre desde que el producto se pone en operación hasta que falla por primera vez. Al menos en cierta medida, el tiempo hasta la falla está sujeto a variaciones aleatorias. Por lo tanto, es natural interpretar el tiempo hasta la falla como una variable aleatoria,  $T$ . Principalmente se utiliza el término tiempo hasta la falla, pero en algunos casos también se emplea el término vida útil. La relación entre la variable de estado  $X(t)$  y el tiempo hasta la falla  $T$  se muestra en la Figura 1. A menos que se indique lo contrario, siempre se asume que el producto es nuevo y se encuentra en estado de funcionamiento cuando se pone en marcha en el tiempo  $t = 0$ .

Obsérvese que el tiempo hasta la falla no siempre se mide en tiempo calendario. También puede medirse mediante conceptos de tiempo más indirectos, tales como:

- Número de veces que se acciona un interruptor.
- Número de kilómetros recorridos por un automóvil.
- Número de rotaciones de un rodamiento.
- Número de ciclos de un equipo que opera de manera periódica.



**Figura 1.** La variable de estado y el tiempo hasta la falla de un producto.

A partir de estos ejemplos, se observa que el tiempo hasta la falla puede ser, con frecuencia, una variable discreta. Sin embargo, una variable discreta puede aproximarse mediante una variable continua. Aquí, salvo que se indique lo contrario, se asume que el tiempo hasta la falla  $T$  está distribuido de manera continua.

### Conjunto de Datos

Considere un experimento en el que  $n$  productos idénticos e independientes se ponen en operación en el tiempo  $t = 0$ . Se dejan los  $n$  productos sin intervención y se observan los tiempos hasta la falla de cada producto. El resultado de este experimento es el conjunto de datos  $[t_1, t_2, \dots, t_n]$ . A este tipo de conjunto de datos también se le denomina conjunto de datos histórico, para resaltar que proviene de operaciones pasadas de los productos y que los tiempos han sido registrados y se conocen.

En teoría de probabilidad y de confiabilidad, se acepta que no se puede saber de antemano cuál será el resultado de un experimento futuro; por lo tanto, se trata de predecir la probabilidad de ocurrencia de cada posible resultado con base en datos históricos. Estas probabilidades son válidas cuando:

1. Los experimentos pasados y futuros pueden considerarse idénticos e independientes (realizados de la misma manera, bajo las mismas condiciones y sin dependencias entre los resultados).
2. Los experimentos pasados se han repetido un gran número de veces.

A menudo es posible identificar algunas tendencias comunes y ciertas variaciones en el conjunto de datos histórico, lo que nos permite hacer predicciones útiles con incertidumbres manejables para experimentos futuros. En términos generales, esto requiere tres pasos principales:

1. Análisis de datos, para extraer tendencias y variaciones relevantes.
2. Modelado, para estructurar la información en un formato que permita realizar cálculos probabilísticos para nuevos productos.
3. Cuantificación y cálculo de probabilidades.

A continuación, se proporciona una visión general que permite comprender la información que proviene de las fases de análisis de datos y modelado para un solo producto no reparable. Se presentan las cantidades de interés que pueden calcularse y la forma de hacerlo. Como ejemplo, consideremos el conjunto de datos de  $n = 60$  tiempos hasta la falla observados en la Tabla 2.

**Tabla 2.** Conjunto de datos históricos

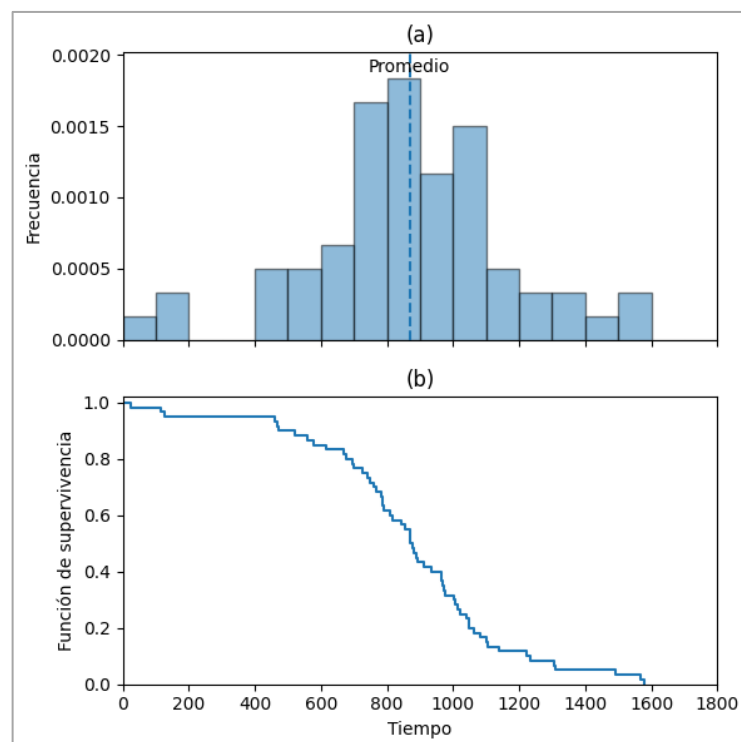
|      |      |      |      |      |      |
|------|------|------|------|------|------|
| 23   | 114  | 125  | 459  | 468  | 472  |
| 520  | 558  | 577  | 616  | 668  | 678  |
| 696  | 700  | 724  | 742  | 748  | 761  |
| 768  | 784  | 785  | 786  | 792  | 811  |
| 818  | 845  | 854  | 868  | 870  | 871  |
| 878  | 881  | 889  | 892  | 912  | 935  |
| 964  | 965  | 970  | 971  | 976  | 1001 |
| 1006 | 1013 | 1020 | 1041 | 1048 | 1049 |
| 1065 | 1084 | 1102 | 1103 | 1139 | 1224 |
| 1235 | 1304 | 1310 | 1491 | 1567 | 1577 |

La distribución de frecuencias de los tiempos de falla registrados se obtiene dividiendo el número de fallas en cada intervalo entre el número total de fallas; con ello se obtiene el número relativo de fallas que ocurren en dichos

intervalos. El histograma resultante se denomina distribución de frecuencia relativa. El histograma se construye de manera que el área de cada barra sea igual al porcentaje del total de fallas que ocurrieron en ese intervalo de tiempo; el área total bajo el histograma es 100% ( $= 1$ ).

La distribución de frecuencia relativa puede utilizarse para estimar e ilustrar cantidades de confiabilidad, tales como el promedio empírico, la desviación estándar empírica y la probabilidad de sobrevivir un tiempo dado. El promedio (empírico) se muestra en la Figura 2a.

En la práctica, el histograma puede presentar uno o más valores máximos y una desviación estándar razonable alrededor de dichos máximos. Diferentes modos de falla y/o causas de falla pueden dar lugar a más de un valor máximo. El histograma de la Figura 2 muestra que los tiempos de falla se distribuyen alrededor de la media y que existen algunas fallas tempranas que ocurrieron poco tiempo después del arranque.



**Figura 2.** Histograma y curva de supervivencia del conjunto de datos (Tabla 2).

### Distribución Empírica y Función de Supervivencia

Otra forma de presentar el conjunto de datos de la Tabla 2 es construir una función de supervivencia empírica. Esto se realiza ordenando los tiempos de

falla, comenzando por el menor y terminando con el mayor. Para cada tiempo de falla, se grafica la proporción (es decir, el porcentaje) de productos que sobrevivieron a dicho tiempo. La función resultante, naturalmente, decrece de 1 a 0. La proporción de productos que sobrevivieron, por ejemplo, hasta un tiempo  $t_i$ , puede utilizarse para estimar la probabilidad de que un producto sobreviva al tiempo  $t_i$  en un experimento futuro. La función de supervivencia empírica para el conjunto de datos de la Tabla 2 se muestra en la Figura 2b.

La función de supervivencia empírica puede utilizarse para estimar probabilidades de interés en experimentos futuros; sin embargo, es más común ajustar una función continua a la función empírica y emplear dicha función continua en los estudios de confiabilidad.

### **Características Generales de las Distribuciones de Tiempo hasta la Falla**

Supongamos que el tiempo hasta la falla  $T$  es una variable aleatoria continua, con Función de Densidad de Probabilidad (FDP)  $f(t)$  y Función de Distribución de Acumulada (FDA)  $F(t)$ .

$$F(t) = \Pr(T \leq t) = \int_0^t f(u)du \quad \text{para } t > 0. \quad (1)$$

El evento  $T \leq t$  ocurre cuando el producto falla antes del tiempo  $t$ , por lo que  $F(t)$  es la probabilidad de que el producto falle en el intervalo  $(0, t]$ . La función de densidad  $f(t)$  es, a partir de (1), la derivada de  $F(t)$ :

$$f(t) = \frac{d}{dt}F(t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta t)}{\Delta t}.$$

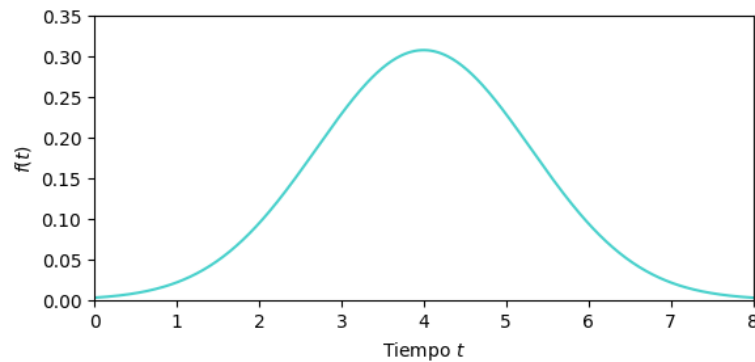
Esto implica que, cuando  $\Delta t$  es pequeño:

$$\Pr(t < T \leq t + \Delta t) \approx f(t) \Delta t. \quad (2)$$

Cuando, en el tiempo  $t = 0$ , se observa hacia el futuro, por lo tanto, la  $\Pr(t < T \leq t + \Delta t)$  indica la probabilidad de que el producto falle en el intervalo corto  $(t, t + \Delta t]$ . Cuando esta probabilidad es alta (baja), la densidad de falla  $f(t)$

también es alta (baja). Por esta razón, a  $f(t)$  se le llama función de densidad de falla.

Un ejemplo de una curva de la FDP se muestra en la Figura 3. La unidad de tiempo utilizada en la figura no se especifica; podría ser, por ejemplo, un año o 10 000 horas.



**Figura 3.** Función de Densidad de Probabilidad,  $f(t)$  para el tiempo hasta la falla  $T$ .

Para que  $f(t)$  sea una FDP válida, debe satisfacer dos condiciones:

1.  $f(t) \geq 0$  para todo  $t \geq 0$
2.  $\int_0^{\infty} f(t)dt = 1$

Cuando se especifica una función de densidad de probabilidad, normalmente sólo se indica la parte donde es distinta de cero, asumiéndose implícitamente que es igual a cero en cualquier región no especificada. Como el tiempo hasta la falla  $T$  no puede tomar valores negativos,  $f(t)$  solo se define para valores no negativos de  $t$ .

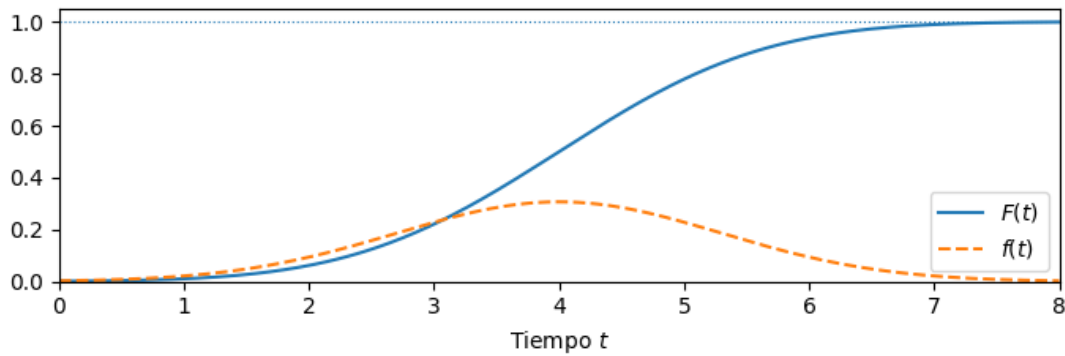
Para una variable aleatoria continua, la probabilidad de que  $T$  sea exactamente igual para  $t$  siempre es cero; esto es  $\Pr(T = t) = 0$  para todos los valores específicos de  $t$ . Esto significa que  $\Pr(T \leq t) = \Pr(T < t)$  y  $\Pr(T \geq t) = \Pr(T > t)$ .

Dado que  $f(t) \geq 0$  para todo  $t$ , la FDA debe satisfacer:

1.  $0 \leq F(t) \leq 1$ , porque  $F(t)$  representa una probabilidad.
2.  $\lim_{t \rightarrow 0} F(t) = 0$

3.  $\lim_{t \rightarrow \infty} F(t) = 1$
4.  $F(t_1) \geq F(t_2)$  cuando  $t_1 > t_2$ ; es decir,  $F(t)$  es una función no decreciente de  $t$ .

La FDA  $F(t)$  y la FDP  $f(t)$  para una misma distribución se muestran en la Figura 4. La densidad (línea discontinua) es la misma que en la Figura 3, pero la escala del eje vertical es distinta.

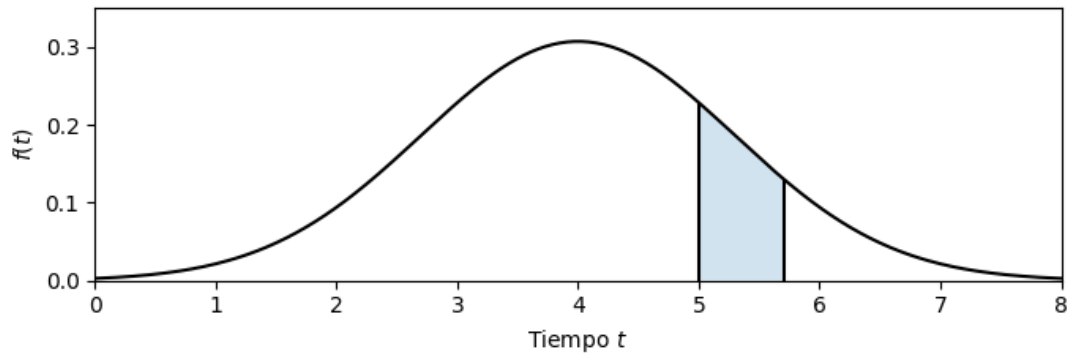


**Figura 4.** La FDA  $F(t)$  y FDP  $f(t)$ .

La probabilidad de que ocurra una falla en el intervalo  $(t_1, t_2]$  es:

$$\Pr(t_1 < T \leq t_2) = F(t_2) - F(t_1) = \int_{t_1}^{t_2} f(u) du. \quad (3)$$

Esta cantidad corresponde al área sombreada bajo la curva  $f(t)$  en la Figura 5 cuando  $t_1 = 5$  y  $t_2 = 5.7$  unidades de tiempo. Dependiendo de los valores de  $t_1$ ,  $t_2$  y de  $f(t)$  en  $(t_1, t_2]$ , el área azul cambiará, y también cambiará la probabilidad de falla en ese intervalo.



**Figura 5.** Cálculo integral de la probabilidad de fallar dentro del intervalo  $(t_1, t_2] = (5.0, 5.7]$ .

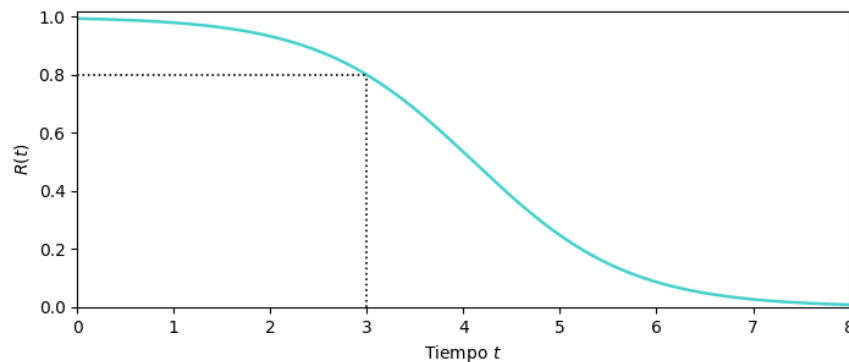
### Función de supervivencia

La Función de Supervivencia (FS) de un ítem se define como

$$R(t) = 1 - F(t) = \Pr(T > t) \quad (4)$$

o, de manera equivalente,

$$R(t) = 1 - \int_0^t f(u) du = \int_t^{\infty} f(u) du \quad (5)$$



**Figura 6.** Función de supervivencia  $R(t)$ .

Por lo tanto,  $R(t)$  es la probabilidad de que el ítem no falle en el intervalo de tiempo  $(0, t]$ ; en otras palabras, la probabilidad de que el ítem sobreviva al intervalo  $(0, t]$  y continúe funcionando en el tiempo  $t$ .

La función de supervivencia también se denomina Función de Probabilidad de Supervivencia (FPS). Algunos autores definen la confiabilidad mediante  $R(t)$  y, en consecuencia, la llaman función de confiabilidad. Esta es

también la razón para utilizar el símbolo  $R(t)$ . La función de supervivencia correspondiente a la función de densidad de probabilidad mostrada en la Figura 3 se presenta en la Figura 6.

En la Figura 6, la línea punteada indica que la probabilidad de que el ítem sobreviva 3 unidades de tiempo es aproximadamente 0.80 (=80%). También puede leerse de forma inversa, encontrando que el tiempo correspondiente a una supervivencia del 80% es aproximadamente 3 unidades de tiempo.

### **Función de tasa de falla o Función de falla**

La probabilidad de que un ítem falle en el intervalo de tiempo  $(t, t + \Delta t]$ , sabiendo que el ítem está funcionando en el tiempo  $t$ , es

$$\Pr(t < T \leq t + \Delta t \mid T > t) = \frac{\Pr(t < T \leq t + \Delta t)}{\Pr(T > t)} = \frac{F(t + \Delta t) - F(t)}{R(t)}.$$

Al dividir esta probabilidad entre la longitud del intervalo de tiempo  $\Delta t$  y hacer que  $\Delta t \rightarrow 0$ , se obtiene la Función de Tasa de Falla (FF)  $h(t)$  del ítem:

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta t \mid T > t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{R(t)} = \frac{f(t)}{R(t)} \end{aligned} \quad (6)$$

Esto implica que cuando  $\Delta t$  es pequeño,

$$\Pr(t < T \leq t + \Delta t \mid T > t) \approx h(t)\Delta t$$

Debido a que  $R(t)$  es una probabilidad y cumple  $R(t) \leq 1$  para todo  $t$ , la ecuación anterior implica que  $h(t) \geq f(t)$  para todo  $t \geq 0$ .

Las relaciones entre las funciones  $f(t)$ ,  $F(t)$ ,  $R(t)$  y  $h(t)$  se presentan en la Tabla 3.

**Tabla 3.** Relación entre las funciones  $F(t)$ ,  $f(t)$ ,  $R(t)$  y  $h(t)$ .

| Función  | $F(t)$                              | $f(t)$                               | $R(t)$                   | $h(t)$                      |
|----------|-------------------------------------|--------------------------------------|--------------------------|-----------------------------|
| $F(t) =$ | -                                   | $\int_0^t f(u) du$                   | $1 - R(t)$               | $1 - e^{-\int_0^t h(u) du}$ |
| $f(t) =$ | $\frac{d}{dt}F(t)$                  | -                                    | $-\frac{d}{dt}R(t)$      | $h(t)e^{-\int_0^t h(u) du}$ |
| $R(t) =$ | $1 - F(t)$                          | $\int_t^\infty f(u) du$              | -                        | $e^{-\int_0^t h(u) du}$     |
| $h(t) =$ | $\frac{\frac{dF(t)}{dt}}{1 - F(t)}$ | $\frac{f(t)}{\int_t^\infty f(u) du}$ | $-\frac{d}{dt}\log R(t)$ | -                           |

### La curva de la bañera

A partir de la ecuación  $R(t) = e^{-\int_0^t h(u) du}$  se observa que la función de supervivencia  $R(t)$  queda determinada de manera única por la FF  $h(t)$ . Para determinar la forma de  $h(t)$  para un tipo específico de ítems, puede llevarse a cabo el siguiente experimento:

Se colocan  $n$  ítems idénticos y no reparables en operación en el tiempo  $t = 0$  y se registra el instante en el que falla cada ítem. Supóngase que la última falla ocurre en el tiempo  $t_{\max}$ . Se divide el eje del tiempo en intervalos disjuntos de igual longitud  $\Delta t$ . Comenzando desde  $t = 0$ , los intervalos se numeran como  $j = 1, 2, \dots$ . Para cada intervalo se registra:

- El número de ítems,  $n(j)$  que fallan en el intervalo  $j$ .
- Los tiempos observados de funcionamiento de los ítems individuales  $(t_{1j}, t_{2j}, \dots, t_{nj})$  en el intervalo  $j$ . En consecuencia,  $t_{ij}$  representa el tiempo durante el cual el ítem  $i$  ha estado funcionando en el intervalo de tiempo  $j$ . El valor  $t_{ij}$  es, por lo tanto, igual a cero si el ítem  $i$  ha fallado antes del intervalo  $j$ , donde  $j = 1, 2, \dots, m$ .

De este modo,  $\sum_{i=1}^n t_{ij}$  corresponde al tiempo total de funcionamiento de los ítems en el intervalo  $j$ . Ahora

$$h(j) = \frac{n(j)}{\sum_{i=1}^n t_{ij}}.$$

Es decir, el número de fallas por unidad de tiempo de funcionamiento en el intervalo  $j$ . Esta constituye una estimación natural de la tasa de falla en el intervalo  $j$  para los ítems que se encuentran funcionando al inicio de dicho intervalo.

Sea  $v(j)$  el número de ítems que están funcionando al inicio del intervalo  $j$ . La tasa de falla en el intervalo  $j$  puede aproximarse como

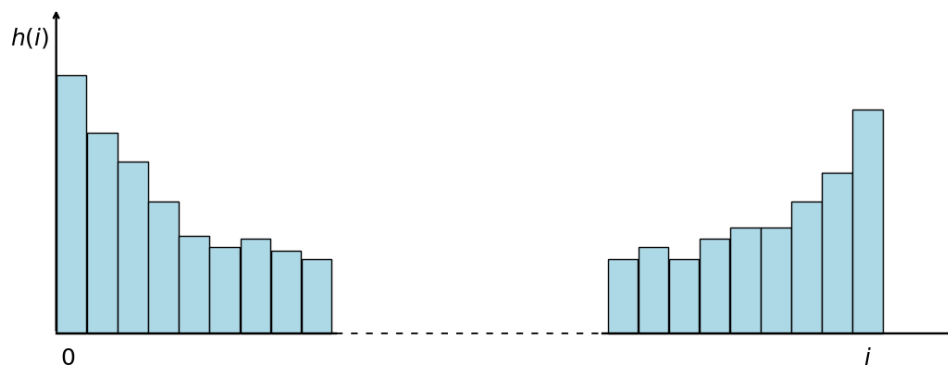
$$h(j) \approx \frac{n(j)}{v(j)\Delta t},$$

y, por lo tanto,

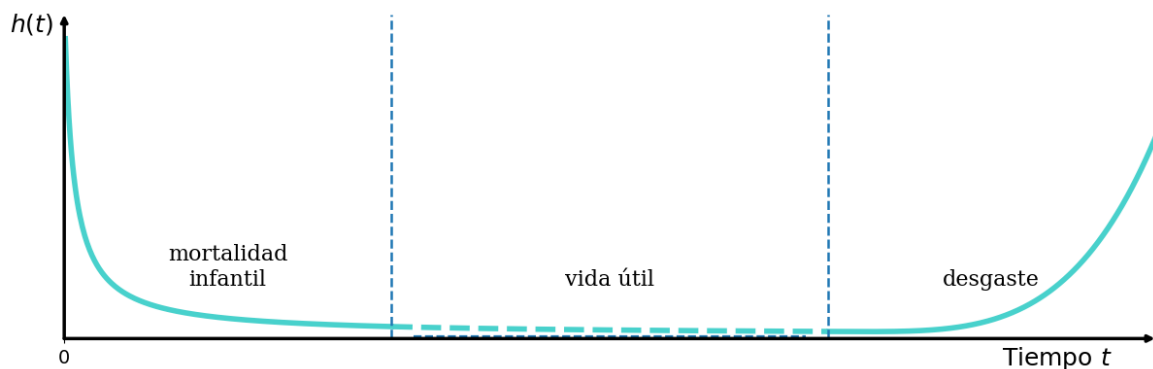
$$h(j)\Delta t \approx \frac{n(j)}{v(j)}.$$

Un histograma que represente  $h(i)$  como función de  $i$  suele presentar la forma mostrada en la Figura 7. Si el número de ítems  $n$  es grande, pueden utilizarse intervalos de tiempo pequeños. Al hacer que  $\Delta t \rightarrow 0$ , se espera que la función escalonada  $h(i)$  tienda hacia una curva “suave”, como se muestra en la Figura 8, la cual constituye una estimación de la función de tasa de falla  $h(t)$ .

Esta curva se denomina comúnmente curva de la bañera, debido a su forma característica. La tasa de falla suele ser elevada en la fase inicial. Esto puede explicarse por la presencia de defectos no detectados en los ítems, los cuales se manifiestan rápidamente cuando estos ítems se ponen en operación; las fallas asociadas se denominan fallas por mortalidad infantil. Una vez que el ítem ha superado el periodo de mortalidad infantil, la tasa de falla suele estabilizarse en un nivel aproximadamente constante durante un cierto intervalo de tiempo, hasta que comienza a incrementarse a medida que los ítems inician su proceso de desgaste.



**Figura 7.** Curva de bañera empírica.



**Figura 8.** Curva de bañera

A partir de la forma de la curva de la bañera, el tiempo hasta la falla de un ítem puede dividirse en tres intervalos típicos: el periodo de mortalidad infantil, el periodo de vida útil, y el periodo de desgaste.

El periodo de vida útil también se denomina periodo de fallas aleatorias. En ocasiones, los ítems son probados en fábrica antes de ser distribuidos a los usuarios, y de este modo gran parte de los problemas asociados a la mortalidad infantil se eliminan antes de que los ítems sean puestos en uso. Para la mayoría de los ítems mecánicos, la FF suele presentar una ligera tendencia creciente durante el periodo de vida útil.

### Tiempo medio hasta la falla

Para el conjunto de datos de la Tabla 2, el tiempo promedio hasta la falla es una métrica que describe la localización central de los tiempos de falla. Este puede calcularse empíricamente como la suma de los tiempos de falla observados dividida entre el número  $n$  de ítems que han fallado.

$$t = \frac{1}{n} \sum_{i=1}^n t_i. \quad (7)$$

En teoría de la probabilidad, la ley de los grandes números establece que, si  $n$  tiende a infinito, la media empírica  $\bar{t}$  se estabiliza alrededor de un valor constante que ya no depende de  $n$ . Este valor se denomina valor esperado o media de  $T$ , y se denota como  $E(T)$ . En teoría de la confiabilidad, este valor recibe el nombre de Tiempo Medio Hasta la Falla (MTTF, por sus siglas en inglés, Mean Time To Failure).

$$\text{MTTF} = E(T) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n t_i \quad (8)$$

### Ley de los grandes números

Sean  $X_1, X_2, \dots$  una secuencia de variables aleatorias independientes con una distribución común, y sea  $E(X_i) = \mu$ . Entonces, con probabilidad 1,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mu \text{ cuando } n \rightarrow \infty \quad (9)$$

Esta definición es equivalente a la presentada en la ecuación (10), la cual puede interpretarse como una versión continua del límite de la media empírica: cada posible tiempo de falla  $t$  se multiplica por su frecuencia de ocurrencia  $f(t) dt$ , y la suma se reemplaza por una integral.

$$\text{MTTF} = E(T) = \int_0^{\infty} t f(t) dt \quad (10)$$

Obsérvese que el MTTF únicamente proporciona información sobre la localización central de los tiempos de falla y no aporta información acerca de cómo estos se dispersan alrededor de la media. Por lo tanto, el MTTF ofrece mucha menos información que el histograma mostrado en la Figura 2; sin embargo, proporciona una referencia útil para una primera evaluación y es ampliamente utilizado en aplicaciones de ingeniería de confiabilidad.

El MTTF puede derivarse a partir de otras métricas de confiabilidad. Dado que  $f(t) = -R'(t)$ ,

$$\text{MTTF} = - \int_0^{\infty} t R'(t) dt$$

Mediante integración por partes, se obtiene

$$\text{MTTF} = -[tR(t)]_0^{\infty} + \int_0^{\infty} R(t) dt$$

Si  $\text{MTTF} < \infty$ , puede demostrarse que  $[tR(t)]_0^{\infty} = 0$ . En ese caso,

$$\text{MTTF} = \int_0^{\infty} R(t) dt \quad (11)$$

Con frecuencia resulta más sencillo determinar el MTTF mediante la ecuación (11) que mediante la ecuación (10).

### Vida media residual

Considere un ítem que se pone en operación en el tiempo  $t = 0$  y que continúa funcionando en el tiempo  $x$ . El ítem falla en el tiempo aleatorio  $T$ . La vida residual del ítem, cuando se sabe que aún está funcionando en el tiempo  $x$ , es  $T - x$ , como se muestra en la Figura 9.



**Figura 9.** Vida residual de un ítem que continúa funcionando en el tiempo  $x$ .

La Vida Media Residual (MRL, por sus siglas en inglés, Mean Residual Life),  $\text{MRL}(x)$ , del ítem a la edad  $x$  se define como

$$\text{MRL}(x) = E(T - x \mid T > x),$$

es decir, el valor medio de la variable aleatoria  $T - x$  cuando se sabe que  $T > x$ . Este valor medio puede determinarse a partir de la función de supervivencia condicional indicada en la ecuación (11), de la siguiente manera:

$$\text{MRL}(x) = \mu(x) = \int_x^\infty R(t \mid x) dt = \frac{1}{R(x)} \int_x^\infty R(t) dt \quad (12)$$

Obsérvese que  $\text{MRL}(x)$  representa el MTTF adicional, es decir, la vida media restante de un ítem que ha alcanzado la edad  $x$ . Esto significa que, cuando el ítem ha alcanzado la edad  $x$ , su edad media al fallar es  $x + \text{MRL}(x)$ .

Asimismo, debe observarse que  $\text{MRL}(x)$  se aplica a un ítem genérico que ha alcanzado la edad  $x$ . No se dispone de información adicional sobre el ítem particular ni sobre su historial en el intervalo  $(0, x)$ . El conocimiento sobre una posible degradación del ítem en la edad  $x$  es el mismo que se tenía cuando el ítem se puso en operación en el tiempo  $t = 0$ .

En el tiempo  $t = 0$ , el ítem es nuevo y se tiene  $\mu(0) = \mu = \text{MTTF}$ . En ocasiones resulta de interés estudiar la función

$$g(x) = \frac{\text{MRL}(x)}{\text{MTTF}} = \frac{\mu(x)}{\mu}. \quad (13)$$

Cuando un ítem ha sobrevivido hasta el tiempo  $x$ , la función  $g(x)$  proporciona el valor de  $\text{MRL}(x)$  como un porcentaje del MTTF inicial. Por ejemplo, si  $g(x) = 0.60$ , entonces la vida media residual  $\text{MRL}(x)$  en el tiempo  $x$  corresponde al 60% de la vida media residual en el tiempo  $t = 0$ .

### **Algunas Distribuciones de Tiempo hasta la Falla**

Algunas distribuciones paramétricas utilizadas para modelar el tiempo hasta la falla son:

1. La distribución exponencial
2. La distribución gamma

3. La distribución de Weibull
4. La distribución normal (gaussiana)
5. La distribución lognormal

### Distribución Exponencial

Considere un producto que se pone en operación en el tiempo  $t = 0$ . El tiempo hasta la falla  $T$  del producto tiene la siguiente función de densidad de probabilidad:

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t > 0, \lambda > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (14)$$

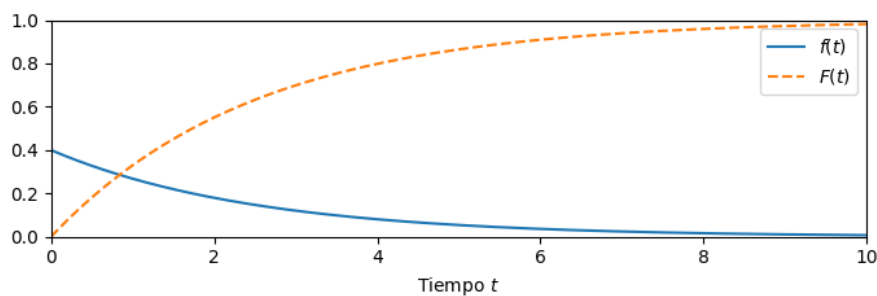
Esta distribución se denomina distribución exponencial con parámetro  $\lambda$ , y se denota como:  $T \sim \exp(\lambda)$ .

### Función de supervivencia para la exponencial

La función de supervivencia se define como:

$$R(t) = \Pr(T > t) = \int_t^{\infty} f(u) du = e^{-\lambda t} \quad \text{para } t > 0. \quad (15)$$

La FDP  $f(t)$  y la FS  $R(t)$  para la distribución exponencial se muestran en la Figura 10.



**Figura 10.** FDP  $f(t)$  (línea continua) y FDA  $F(t)$  (línea discontinua) para la distribución exponencial ( $\lambda = 0.4$ ).

### MTTF para la exponencial

El tiempo medio hasta la falla de la distribución exponencial es:

$$\text{MTTF} = \int_0^{\infty} R(t) dt = \frac{1}{\lambda} = \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}, \quad (16)$$

La varianza del tiempo  $T$  hasta la falla es:

$$\text{var}(T) = \frac{1}{\lambda^2}. \quad (17)$$

Obsérvese que cuando el MTTF aumenta (o disminuye), la varianza hace lo mismo. Esta es una limitación de la distribución exponencial, ya que hace imposible ajustar de manera independiente la media y la varianza para que se adapten a un conjunto de datos históricos.

La probabilidad de que un producto sobreviva su propio MTTF es:

$$R(\text{MTTF}) = R\left(\frac{1}{\lambda}\right) = e^{-1} \approx 0.3679 \text{ para todos los valores de } \lambda$$

Cualquier producto con una distribución exponencial del tiempo hasta la falla sobrevivirá su MTTF con una probabilidad de aproximadamente 36.8%.

### Función de Tasa de Falla para la exponencial

La tasa de falla es:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \quad (18)$$

Por lo tanto, un producto con una distribución exponencial del tiempo hasta la falla tiene una tasa de falla constante e independiente del tiempo. Dado que existe una correspondencia uno a uno entre la distribución y la función de tasa de falla, cualquier producto con tasa de falla constante tiene una distribución exponencial del tiempo hasta la falla.

La Figura 10 indica que la distribución exponencial puede ser una representación realista del tiempo hasta la falla de un producto durante su periodo de vida útil, al menos para ciertos tipos de productos.

Los resultados de la ecuación (16) y (18) concuerdan bien con el uso cotidiano de estos conceptos. Si un producto tiene en promedio  $\lambda = 4$  fallas/año, entonces su MTTF es de  $\frac{1}{4}$  de año.

La tasa de falla acumulada correspondiente es  $h(t) = \lambda t$  y puede representarse como una línea recta con pendiente  $\lambda$ .

### Mediana del tiempo hasta la falla

La mediana del tiempo hasta la falla de la distribución exponencial se obtiene de la ecuación  $R(t_m) = 0.50$  y es:

$$t_m = \frac{\log 2}{\lambda} \approx \frac{0.693}{\lambda} = 0.693 \text{ MTTF.} \quad (19)$$

Esto significa que, para un producto con tasa de falla constante, existe un 50% de probabilidad de que falle antes de alcanzar el 69.3% de su MTTF.

### Cambio de escala temporal

Consideremos un producto con tiempo hasta la falla  $T \sim \exp(\lambda)$ . Suponga que se cambia la unidad de tiempo para medir  $T$ ; por ejemplo, de horas a días. Este cambio de escala se expresa como  $T_1 = aT$  para alguna constante  $a$ .

La función de sobrevivencia del tiempo hasta la falla  $T_1$  en la nueva escala de tiempo es:

$$R_1(t) = \Pr(T_1 > t) = \Pr(aT > t) = \Pr\left(T > \frac{t}{a}\right) = e^{-\frac{\lambda t}{a}}$$

Esto significa que  $T_1 \sim \exp\left(\frac{\lambda}{a}\right)$  con MTTF es:

$$\text{MTTF}_1 = \frac{a}{\lambda} = a \text{ MTTF,}$$

lo cual es el resultado esperado. Esto muestra que la distribución exponencial es cerrada respecto a cambios de escala, es decir:

$$T \sim \exp(\lambda) \Rightarrow aT \sim \exp\left(\frac{\lambda}{a}\right) \text{ para todas estas constantes } a > 0. \quad (20)$$

### Probabilidad de Falla en un Intervalo Corto de Tiempo

La serie de Maclaurin de la función exponencial es:

$$e^{-\lambda t} = \sum_{x=0}^{\infty} \frac{(-\lambda t)^x}{x!} = 1 - \lambda t + \frac{(\lambda t)^2}{2} - \frac{(\lambda t)^3}{6} + \dots$$

Cuando  $\lambda t$  es pequeño,  $(\lambda t)^x$  para  $x = 2, 3, \dots$  son despreciables, y podemos usar la aproximación:

$$e^{-\lambda t} \approx 1 - \lambda t \text{ cuando } \lambda t \text{ es pequeño} \quad (21)$$

Consideremos un intervalo corto de tiempo  $(t, t + \Delta t]$ . La probabilidad de que un producto con tiempo hasta la falla  $T \sim \exp(\lambda)$  falle dentro de este intervalo es:

$$\begin{aligned} \Pr(t < T \leq t + \Delta t) &= \Pr(T \leq t + \Delta t) - \Pr(T \leq t) = \\ 1 - e^{-\lambda(t+\Delta t)} - (1 - e^{-\lambda t}) &= e^{-\lambda t} - e^{-\lambda(t+\Delta t)} \approx \lambda \Delta t. \end{aligned} \quad (22)$$

La probabilidad de que un producto con tasa de falla constante  $\lambda$  falle en un intervalo corto de duración  $\Delta t$  es aproximadamente:  $\lambda \Delta t$ . La aproximación es suficientemente precisa cuando  $\Delta t$  es muy pequeño.

### Estructura en Serie de Componentes Independientes

Consideremos una estructura en serie compuesta por  $n$  componentes independientes, cada uno con tasa de falla constante  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Una estructura en serie falla cuando ocurre la primera falla entre sus componentes, por lo que el tiempo hasta la falla  $T_s$  de la estructura es:

$$T_s = \min(T_1, T_2, \dots, T_n) = \min_{i=1,2,\dots,n} T_i.$$

La función de sobrevivencia de la estructura es:

$$\begin{aligned} R_s(t) &= \Pr\left(\min_{i=1,2,\dots,n} T_i > t\right) = \Pr\left(\bigcap_{i=1}^n T_i > t\right) \\ &= \Pr[(T_1 > t) \cap (T_2 > t) \cap \dots \cap (T_n > t)] = \prod_{i=1}^n \Pr(T_i > t) \\ &= \prod_{i=1}^n e^{-\lambda_i t} = e^{-(\sum_{i=1}^n \lambda_i)t} = e^{-\lambda_s t}, \end{aligned} \quad (23)$$

donde  $\lambda_s = \sum_{i=1}^n \lambda_i$ . Esto demuestra que el tiempo hasta la falla de la estructura,  $T_s$ , está exponencialmente distribuido con tasa de falla  $\lambda_s = \sum_{i=1}^n \lambda_i$ .

Para el caso especial en que los  $n$  componentes independientes son idénticos, de modo que  $\lambda_i = \lambda$  para  $i = 1, 2, \dots, n$ , el tiempo hasta la falla de la estructura en serie está distribuido exponencialmente con tasa de falla  $\lambda_s = n\lambda$ . El tiempo medio hasta la falla de la estructura en serie es:

$$\text{MTTF}_s = \frac{1}{\lambda_s} = \frac{1}{n} \frac{1}{\lambda},$$

es decir, el MTTF de la estructura en serie es igual al MTTF de un solo componente dividido por el número de componentes en la estructura.

### Función de Supervivencia Condicional y Vida Media Residual

La función de supervivencia condicional de un producto cuyo tiempo hasta la falla es  $T \sim \exp(\lambda)$  es:

$$\begin{aligned} R(t | xt) &= \Pr(T > t + x | T > xt) = \frac{\Pr(T > t + x)}{\Pr(T > xt)} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda xt}} = e^{-\lambda tx} = \Pr(T > t) = R(tx). \end{aligned} \quad (24)$$

Por lo tanto, la FS de un producto que ha estado funcionando durante  $x$  unidades de tiempo es igual a la de un producto nuevo. Un producto usado (que sigue funcionando) y un producto nuevo tienen exactamente la misma probabilidad de sobrevivir un intervalo adicional de duración  $t$ . La vida residual media para la distribución exponencial es:

$$\text{MRL}(xt) = \int_0^{\infty} R(t | xt) dt = \int_0^{\infty} R(tx) dtx = \text{MTTF}.$$

La vida media residual,  $\text{MRL}(t)$ , de un producto con una distribución exponencial del tiempo hasta la falla es, por lo tanto, igual a su MTTF, independientemente de la edad  $xt$  del producto. El producto es, en consecuencia, “como nuevo” mientras continúe funcionando, y a esto se debe que con frecuencia se diga que la distribución exponencial no tiene memoria.

Asumir que el tiempo hasta la falla sigue una distribución exponencial implica que:

- Un producto usado es estocásticamente “como nuevo”, por lo que no existe razón para reemplazar un producto que aún está funcionando.
- Para estimar la función de sobrevivencia, el MTTF, y otros parámetros, es suficiente recopilar datos sobre el número de horas observadas en operación y el número de fallas. La edad específica de los productos no es relevante en este contexto.

La distribución exponencial es la más utilizada en el análisis de confiabilidad aplicada para analizar tiempos hasta la falla. La razón de ello es su simplicidad matemática y el hecho de que produce modelos realistas de tiempo hasta la falla para ciertos tipos de productos.

### **La Diferencia entre una Variable Aleatoria y un Parámetro**

Un experimento estocástico generalmente se lleva a cabo con el fin de observar y medir una o más variables aleatorias, como el tiempo hasta la falla  $T$ . Al

observar  $T$ , obtenemos un número, por ejemplo, 5000 horas. Experimentos idénticos producirán números diferentes.

La variación o incertidumbre en estos valores puede describirse mediante una distribución estadística  $F(t)$ . Como base del experimento, la distribución normalmente no se especifica por completo, sino que depende de una o más variables conocidas como parámetros. Los parámetros suelen representarse en la distribución con letras griegas. Un ejemplo es el parámetro  $\lambda$  de la distribución exponencial.

En estadística, un parámetro es una variable que no puede medirse directamente en un experimento, sino que debe estimarse a partir de los valores observados (números) de la variable aleatoria. Después de un experimento, medimos la variable aleatoria, pero estimamos el parámetro. Diferentes experimentos usualmente producirán estimaciones ligeramente distintas del parámetro. La regla o fórmula utilizada para estimar un parámetro se denomina estimador del parámetro, y puede evaluarse mediante su valor medio y su desviación estándar.

### Ejemplo sobre una Bomba rotativa

Una bomba rotativa tiene una tasa de falla constante  $\lambda = 4.28 \times 10^{-4} \text{ h}^{-1}$ . La probabilidad de que la bomba sobreviva un mes ( $t = 730$  horas) en operación continua es:

$$R(t) = e^{-\lambda t} = e^{-4.28 \times 10^{-4} \cdot 730} \approx 0.732.$$

El MTTF es:

$$\text{MTTF} = \frac{1}{\lambda} = \frac{1}{4.28 \times 10^{-4}} \text{ h} \approx 2336 \text{ h} \approx 3.2 \text{ meses}.$$

Supongamos que la bomba ha estado funcionando sin fallas durante sus primeros dos meses ( $t_1 = 1460$  horas) de operación. La probabilidad de que falle durante el siguiente mes ( $t_2 = 730$  horas) es:

$$\Pr(T \leq t_1 + t_2 \mid T > t_1) = \Pr(T \leq t_2) = 1 - e^{-4.28 \times 10^{-4} \cdot 730} \approx 0.268$$

porque la bomba está “como nueva”, cuando sabemos que sigue funcionando en el tiempo  $t_1$ .

### Ejemplo sobre la probabilidad de que un producto falle antes que otro

Consideremos una estructura de dos componentes independientes con tasas de falla  $\lambda_1$  y  $\lambda_2$ , respectivamente. La probabilidad de que el componente 1 falle antes que el componente 2 es:

$$\begin{aligned}\Pr(T_2 > T_1) &= \int_0^{\infty} \Pr(T_2 > t \mid T_1 = t) f_{T_1}(t) dt \\ &= \int_0^{\infty} e^{-\lambda_2 t} \lambda_1 e^{-\lambda_1 t} dt \\ &= \lambda_1 \int_0^{\infty} e^{-(\lambda_1 + \lambda_2)t} dt = \frac{\lambda_1}{\lambda_1 + \lambda_2}.\end{aligned}$$

Este resultado puede generalizarse fácilmente a una estructura de  $n$  componentes independientes con tasas de falla  $\lambda_1, \dots, \lambda_n$ . La probabilidad de que el componente  $j$  sea el primero en fallar es:

$$\Pr(\text{Componente } j \text{ falla primero}) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i} \quad (25)$$

### Mezcla de Distribuciones Exponenciales

Supongamos que el mismo tipo de productos se produce en dos plantas diferentes. Se asume que los productos son independientes y que tienen tasas de falla constantes. El proceso de producción es ligeramente distinto en cada planta, por lo que los productos tendrán diferentes tasas de falla. Sea  $\lambda_i$  la tasa de falla de los productos provenientes de la planta  $i$ , para  $i = 1, 2$ . Los productos se mezclan antes de venderse. Una fracción  $p$  proviene de la planta 1, y el resto  $(1 - p)$  proviene de la planta 2. Si seleccionamos un producto al azar, la función de sobrevivencia de dicho producto es:

$$R(t) = p R_1(t) + (1 - p) R_2(t) = p e^{-\lambda_1 t} + (1 - p) e^{-\lambda_2 t}.$$

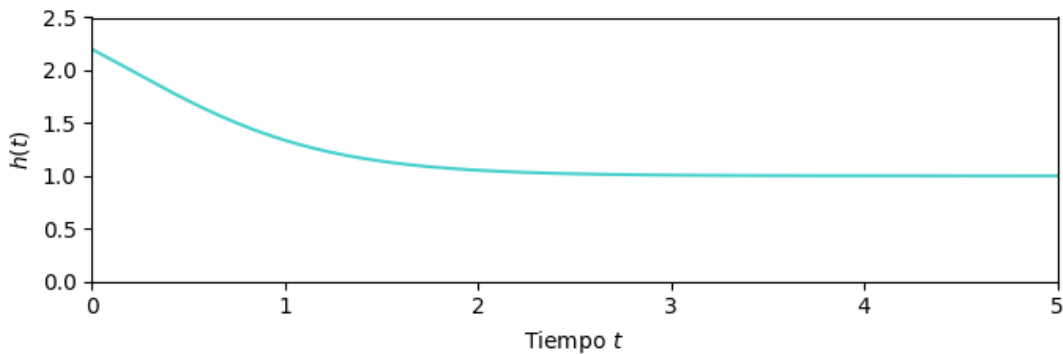
El MTTF se define como:

$$\text{MTTF} = \frac{p}{\lambda_1} + \frac{1-p}{\lambda_2},$$

y la función de tasa de falla es

$$h(t) = \frac{p\lambda_1 e^{-\lambda_1 t} + (1-p)\lambda_2 e^{-\lambda_2 t}}{p e^{-\lambda_1 t} + (1-p) e^{-\lambda_2 t}} \quad (26)$$

La función de tasa de falla, que se muestra en la Figura 11, se observa que es decreciente. Si asumimos que  $\lambda_1 > \lambda_2$ , las fallas tempranas deberían presentar una tasa de falla cercana a  $\lambda_1$ . Después de un tiempo, todos los componentes “débiles” habrán fallado y permanecerán únicamente componentes con una tasa de falla menor  $\lambda_2$ .



**Figura 11.** La función de tasa de falla de la mezcla de dos distribuciones exponenciales ( $\lambda_1 = 1, \lambda_2 = 3$ , y  $p = 0.4$ ).

### Tasa de Falla Constantemente Escalonada

Consideremos un producto que opera solo en intervalos específicos. Cuando no está en funcionamiento, permanece en un modo de espera (standby), que puede estar energizado o no energizado. Un ejemplo de este tipo de producto es una bomba de calor residencial. Cuando la temperatura de la habitación es baja, la bomba de calor se activa bajo demanda desde un termostato, y cuando la temperatura es alta, la bomba se detiene y entra en modo de espera. El producto puede fallar al arrancar (bajo demanda) con una

probabilidad  $p$ . Cuando está funcionando, tiene una tasa de falla constante  $\lambda_r$ , y en modo de espera tiene una tasa de falla constante  $\lambda_s$ .

Si podemos registrar el número  $n$  de demandas de arranque por unidad de tiempo (por ejemplo, por semana) y la fracción  $v$  del tiempo en que el producto está funcionando, podemos calcular una tasa de falla promedio  $\lambda_t$  del producto como:

$$\lambda_t = \lambda_d + v\lambda_r + (1 - v)\lambda_s, \quad (27)$$

donde  $\lambda_d = np$  es el número de fallas de arranque por unidad de tiempo.

### La Distribución de Weibull

La distribución de Weibull es una de las distribuciones de tiempo hasta la falla más utilizadas en el análisis de confiabilidad. La distribución lleva el nombre del profesor sueco Waloddi Weibull (1887–1979), quien desarrolló este modelo para describir la resistencia de materiales. La distribución de Weibull es muy flexible y, mediante una elección adecuada de parámetros, puede modelar muchos tipos de comportamientos de tasas de falla.

### Distribución de Weibull de Dos Parámetros

El tiempo hasta la falla  $T$  de un producto se dice que sigue una distribución Weibull con parámetros  $\alpha(> 0)$  y  $\theta(> 0)$  si la función de distribución está dada por:

$$F(t) = \Pr(T \leq t) = \begin{cases} 1 - e^{-(t/\theta)^\alpha}, & \text{para } t > 0, \\ 0 & \text{en otro caso.} \end{cases} \quad (28)$$

La distribución Weibull de dos parámetros se escribe comúnmente como  $T \sim \text{Weibull}(\alpha, \theta)$ . La correspondiente función de densidad de probabilidad es:

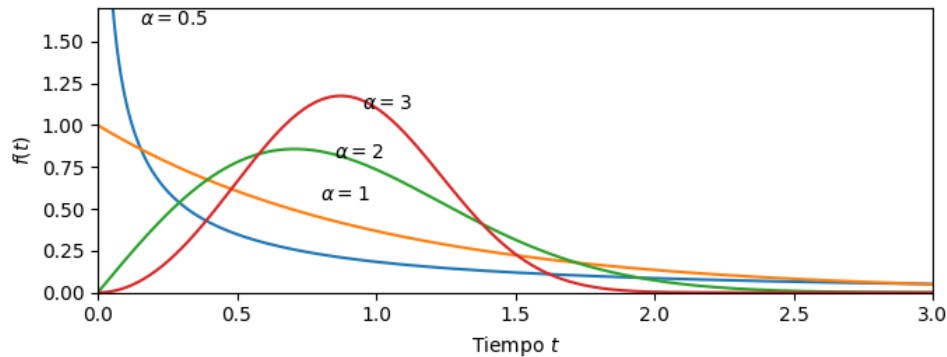
$$f(t) = \frac{d}{dt}F(t) = \frac{\alpha}{\theta} \left(\frac{t}{\theta}\right)^{\alpha-1} e^{-(t/\theta)^\alpha}, \text{ para } t > 0. \quad (29)$$

Donde  $\theta$  es un parámetro de escala medido en unidades de tiempo, y  $\alpha$  es una constante adimensional llamada parámetro de forma. Obsérvese que cuando  $\alpha = 1$ , la distribución de Weibull es igual a la distribución exponencial con  $\lambda = \frac{1}{\theta}$ .

### Elección de parámetros

Los parámetros en la ecuación (29) se eligen así porque esta es la parametrización predeterminada en R. Muchos autores prefieren en cambio los parámetros  $\alpha$  y  $\lambda = \frac{1}{\theta}$ , en cuyo caso la función de distribución se escribe  $F(t) = 1 - e^{-(\lambda t)^\alpha}$ . De este modo, el caso especial para  $\alpha = 1$  se convierte directamente en la distribución exponencial  $\exp(\lambda)$ . Ambas parametrizaciones dan los mismos resultados, y por lo tanto es cuestión de hábito y conveniencia elegir una u otra. Tanto  $\theta$  como  $\lambda$  se denominan parámetros de escala.

La FDP  $f(t)$  para la distribución Weibull se muestra en la Figura 12 para una selección de valores de  $\alpha$ .



**Figura 12.** Función de densidad de probabilidad de la distribución de Weibull para valores seleccionados del parámetro de forma  $\alpha(\theta = 1)$ .

### Función de Supervivencia para la Weibull

La función de supervivencia de  $T \sim \text{Weibull}(\alpha, 0)$  es:

$$R(t) = \Pr(T > 0) = e^{-(t/\theta)^\alpha} \text{ para } t > 0. \quad (30)$$

### Función de Tasa de Falla para la Weibull

La función de tasa de falla de  $T \sim \text{Weibull}(\alpha, 0)$  es:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\alpha}{\theta} \left(\frac{t}{\theta}\right)^{\alpha-1} \text{ para } t > 0. \quad (31)$$

Obsérvese que la tasa de falla también puede escribirse como:

$$h(t) = \alpha \theta^{-\alpha} t^{\alpha-1} \text{ para } t > 0.$$

### Interpretación de la forma de la tasa de falla

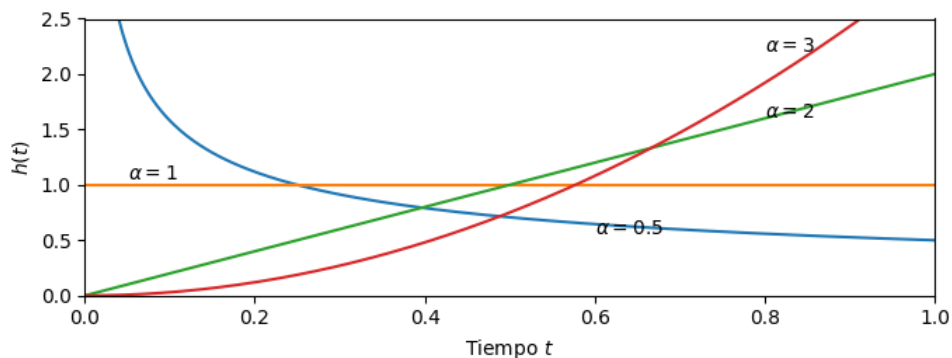
- Cuando  $\alpha = 1$ , la tasa de falla es constante.
- Cuando  $\alpha > 1$ , la tasa de falla es creciente.
- Cuando  $0 < \alpha < 1$ , la tasa de falla es decreciente.
- Cuando  $\alpha = 2$  la función de tasa de falla es linealmente creciente; este caso se conoce como la distribución de Rayleigh.

La Figura 13 muestra la tasa de falla  $h(t)$  de la distribución Weibull para valores seleccionados del parámetro de forma  $\alpha$ . Observe que:

$\alpha < 1 \Rightarrow h(t)$  decreciente en el tiempo

$\alpha = 1 \Rightarrow h(t)$  constante

$\alpha > 1 \Rightarrow h(t)$  creciente en el tiempo



**Figura 13.** Función de tasa de falla de la distribución de Weibull, con  $\theta = 1$  y cuatro valores diferentes del parámetro de forma ( $\alpha$ ).

Se observa que la función de tasa de falla es discontinua como función del parámetro de forma  $\alpha$  en  $\alpha = 1$ . Es importante estar consciente de esta discontinuidad en los cálculos numéricos, porque, por ejemplo,  $\alpha = 0.999$ ,  $\alpha = 1.000$  y  $\alpha = 1.001$  producen funciones de tasa de falla significativamente diferentes para valores pequeños de  $t$ .

Asuma que  $T \sim \text{Weibull}(\alpha, 0)$  y considere la variable  $T^\alpha$ . La función de supervivencia de  $T^\alpha$  es:

$$\Pr(T^\alpha > t) = \Pr\left(T > t^{\frac{1}{\alpha}}\right) = \exp\left(-\frac{t}{\theta^\alpha}\right),$$

lo cual significa que  $T^\alpha$  está distribuido exponencialmente con una tasa de falla constante  $\lambda = \frac{1}{\theta^\alpha}$ .

El parámetro  $\theta$  se denomina tiempo de vida característico de la distribución de Weibull. A partir de la ecuación (30), se obtiene:

$$R(\theta) = e^{-1} = \frac{1}{e} \approx 0.368, \text{ para todo } \alpha > 0.$$

Esto significa que, para cualquier valor del parámetro de forma  $\alpha$ , el producto sobrevivirá hasta tiempo  $\theta$  con una probabilidad del 36.8%.

### MTTF para la Weibull

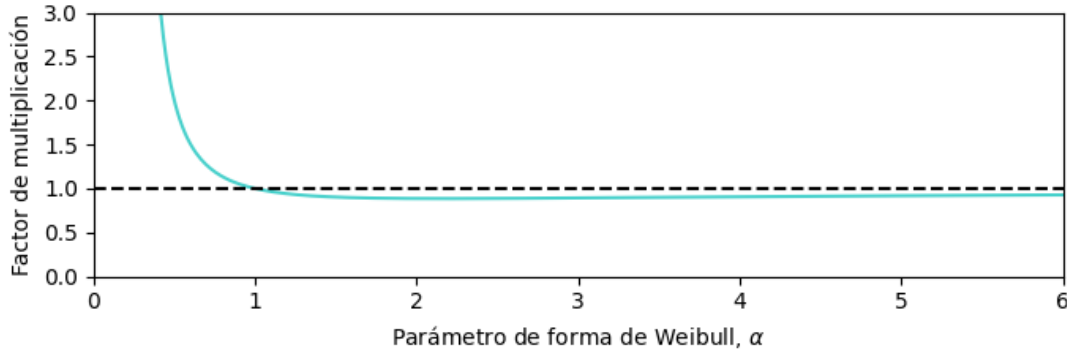
El MTTF de la distribución Weibull de dos parámetros es:

$$\text{MTTF} = \int_0^\infty R(t) dt = \theta \Gamma\left(1 + \frac{1}{\alpha}\right). \quad (32)$$

El MTTF es igual al tiempo de vida característico  $\theta$  multiplicado por un factor que depende del parámetro de forma  $\alpha$ . Este factor,  $\Gamma\left(1 + \frac{1}{\alpha}\right)$ , varía con  $\alpha$ , como se muestra en la Figura 14, lo que indica que el MTTF es ligeramente menor que  $\theta$  cuando  $\alpha \geq 1$ .

El tiempo hasta la falla mediano  $t_m$  de la distribución Weibull es:

$$R(t_m) = 0.50 \Rightarrow t_m = \theta(\log 2)^{\frac{1}{\alpha}} \quad (33)$$



**Figura 14.** El factor de proporcionalidad del MTTF como función de  $\alpha$ .

La varianza de  $T$

$$\text{var}(T) = \theta^2 \left[ \Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]. \quad (34)$$

Obsérvese que  $\frac{\text{MTTF}}{\sqrt{\text{var}(T)}}$  es independiente de  $\theta$ .

La distribución de Weibull también surge como una distribución límite para el mínimo de un gran número de variables aleatorias independientes, idénticamente distribuidas y no negativas. Por ello, la distribución de Weibull se conoce frecuentemente como la distribución del eslabón más débil.

La distribución de Weibull ha sido ampliamente utilizada en el análisis de confiabilidad de semiconductores, rodamientos, motores, soldaduras por puntos, organismos biológicos, etc. La distribución de Weibull es discutida en detalle por Murthy et al. (2003) y McCool (2012). Asimismo, es una distribución básica en el software R y está disponible en varios paquetes (se puede consultar el paquete Weibull-R).

### Ejemplo sobre una Válvula estranguladora

El tiempo hasta la falla  $T$  de una válvula estranguladora variable se asume con una distribución Weibull con parámetro de forma  $\alpha = 2.25$  y parámetro de escala

$\theta = 8695$  horas. La válvula sobrevivirá seis meses ( $t = 4380$  horas) en operación continua con probabilidad:

$$R(t) = \exp \left[ - \left( \frac{t}{\theta} \right)^\alpha \right] = \exp \left[ - \left( \frac{4380}{8695} \right)^{2.25} \right] \approx 0.808.$$

MTTF se define como

$$\text{MTTF} = \theta \Gamma \left( 1 + \frac{1}{\alpha} \right) = 8695 \Gamma(1.44) \text{ h} \approx 7701 \text{ h},$$

y el tiempo mediano hasta la falla es:

$$t_m = \theta (\log 2)^{\frac{1}{\alpha}} \approx 7387 \text{ h}.$$

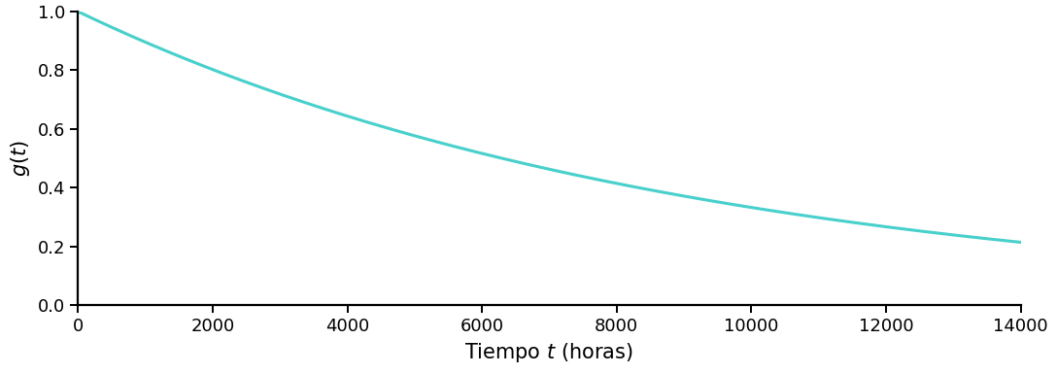
Una válvula que ha sobrevivido los primeros seis meses ( $t_1 = 4380$  h) sobrevivirá los siguientes seis meses ( $t_2 = 4380$  h) con probabilidad:

$$R(t_1 + t_2 | t_1) = \frac{R(t_1 + t_2)}{R(t_1)} = \frac{\exp \left[ - \left( \frac{t_1 + t_2}{\theta} \right)^\alpha \right]}{\exp \left[ - \left( \frac{t_1}{\theta} \right)^\alpha \right]} \approx 0.448,$$

lo cual es significativamente menor que la probabilidad de que una válvula nueva sobreviva seis meses.

Por lo tanto, la MRL cuando la válvula ha estado funcionando durante seis meses ( $x = 4380$  h), la vida residual media es:

$$\text{MRL}(x) = \frac{1}{R(x)} \int_0^\infty R(t + x) dt \approx 4448 \text{ h}.$$



**Figura 15.** La función de vida residual media escalada  $g(t) = \text{MRL} \frac{x}{\text{MTTF}}$  para la distribución de Weibull con parámetros  $\alpha = 2.25$  y  $\theta = 8760$  horas.

La  $\text{MRL}(x)$  no puede expresarse de manera cerrada en forma analítica en este caso, y por lo tanto fue calculada utilizando una computadora. La función  $g(x) = \frac{\text{MRL}(x)}{\text{MTTF}}$  se muestra en la Figura 15.

### Estructura en Serie de Componentes Independientes

Consideremos una estructura en serie de  $n$  componentes. Los tiempos hasta la falla  $T_1, T_2, \dots, T_n$  de los  $n$  componentes se asumen independientes y con distribución Weibull:

$$T_i \sim \text{Weibull}(\alpha, \theta_i) \text{ para } i = 1, 2, \dots, n.$$

Una estructura en serie falla tan pronto como el primer componente falle. El tiempo hasta la falla de la estructura,  $T_s$ , es entonces:

$$T_s = \min\{T_1, T_2, \dots, T_n\}.$$

Función de sobrevivencia de la estructura en serie

$$\begin{aligned} R_s(t) &= \Pr(T > t) = \Pr\left(\min_{1 \leq i \leq n} T_i > t\right) = \prod_{i=1}^n \Pr(T_i > t) = \prod_{i=1}^n \exp\left[-\left(\frac{t}{\theta_i}\right)^\alpha\right] = \\ &= \exp\left[-\sum_{i=1}^n \left(\frac{t}{\theta_i}\right)^\alpha\right] = \exp\left[-\sum_{i=1}^n \left(\frac{1}{\theta_i}\right)^\alpha t^\alpha\right]. \end{aligned}$$

Una estructura en serie de componentes independientes con tiempos hasta la falla Weibull y el mismo parámetro de forma  $\alpha$  tiene nuevamente un tiempo hasta la falla Weibull, con parámetro de escala  $\theta_s = \frac{1}{\left(\sum_{i=1}^n \left(\frac{1}{\theta_i}\right)^\alpha\right)^{\frac{1}{\alpha}}}$  y el

parámetro de forma permanece sin cambios.

### Componentes Idénticos

Cuando todos los componentes  $1, 2, \dots, n$  tienen la misma distribución, de modo que  $\theta_i = \theta$  para  $i = 1, 2, \dots, n$ , entonces la estructura en serie tiene un tiempo hasta la falla Weibull con un parámetro de escala:  $\theta_s = \frac{\theta}{n^{\frac{1}{\alpha}}}$  y un parámetro de forma  $\alpha$ .

### Ejemplo numérico

Consideremos una estructura en serie de  $n$  componentes independientes e idénticos, cada uno con tiempos hasta la falla distribuidos Weibull con los mismos parámetros que en el Ejemplo 3,  $\alpha = 2.25$  y  $\theta = 8695$  horas. El MTTF de la estructura en serie es:

$$\text{MTTF}_s = \theta_s \Gamma\left(1 + \frac{1}{\alpha}\right),$$

donde:

$$\theta_s = \frac{\theta}{n^{\frac{1}{\alpha}}}.$$

Para una estructura en serie de  $n = 5$  componentes, el tiempo medio hasta la falla es:

$$\text{MTTF}_s = \frac{8695}{5^{\frac{1}{2.25}}} \Gamma\left(1 + \frac{1}{2.25}\right) \text{ h} \approx 3766.3 \text{ h}.$$

En la Figura 16, se muestra  $MTTF_s$  como función de  $n$ , es decir, del número de componentes idénticos en la estructura en serie.

### Distribución Weibull de Tres Parámetros

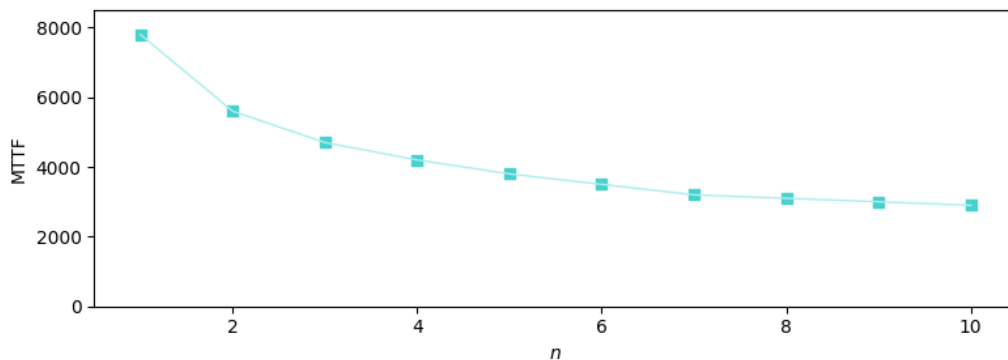
La distribución de Weibull que hemos discutido hasta ahora es una distribución de dos parámetros, con parámetro de forma  $\alpha > 0$  y parámetro de escala  $\theta > 0$ . Una extensión natural de esta distribución es la distribución Weibull de tres parámetros  $(\alpha, \theta, \xi)$ , cuya FDA es:

$$F(t) = \Pr(T \leq t) = \begin{cases} 1 - e^{-\left(\frac{t-\xi}{\theta}\right)^\alpha} & \text{para } t > \xi, \\ 0 & \text{en otro caso.} \end{cases} \quad (35)$$

La correspondiente FDP es:

$$f(t) = \frac{d}{dt}F(t) = \frac{\alpha}{\theta} \left(\frac{t-\xi}{\theta}\right)^{\alpha-1} e^{-\left(\frac{t-\xi}{\theta}\right)^\alpha} \text{ para } t > \xi.$$

El tercer parámetro  $\xi$  es a veces llamado parámetro de garantía (guarantee) o parámetro umbral (threshold), porque la probabilidad de que ocurra una falla antes del tiempo  $\xi$  es 0.



**Figura 16.**  $MTTF_s$  como función de  $n$ , el número de componentes independientes e idénticos en una estructura en serie (Ejemplo 4).

Dado que  $(T - \xi)$  tiene obviamente una distribución Weibull de dos parámetros  $(\alpha, \theta)$ , la media y la varianza de la distribución Weibull de tres parámetros  $(\alpha, \theta, \xi)$  se obtienen a partir de la ecuación (32) y (34).

$$\text{MTTF} = \xi + \theta \Gamma\left(1 + \frac{1}{\alpha}\right).$$

$$\text{var}(T) = \theta^2 \left[ \Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right].$$

En aplicaciones de confiabilidad, cuando se hace referencia a la distribución Weibull, por lo general se entiende la familia de dos parámetros, a menos que se indique lo contrario.

## 2.4 Marco conceptual

En esta sección se reúnen las definiciones clave necesarias para sustentar la modelación estadística de los tiempos de falla en sistemas industriales, particularmente aquellos cuyo comportamiento se ajusta a la curva de bañera. Las definiciones que se presentan a continuación integran conceptos formales de estadística, confiabilidad y mantenimiento industrial, y constituyen la base operativa para el análisis desarrollado en esta investigación.

### Tiempo hasta la falla (Time to Failure, TTF)

El tiempo hasta la falla, denotado como la variable aleatoria continua  $T$ , representa el periodo de operación que transcurre desde que un componente inicia su funcionamiento hasta que ocurre una falla. Esta variable es la base de todos los modelos de vida útil (Dodson & Nolan, 2021). Formalmente, se asume que:

$$T \geq 0, T \in \mathbb{R}^+ \quad (36)$$

En ingeniería industrial, el TTF permite describir estadísticamente la durabilidad de los componentes y establecer políticas de mantenimiento basadas en evidencia empírica.

### Función de Distribución Acumulada $F(t)$

La función de distribución acumulada, definida como:

$$F(t) = P(T \leq t), \quad (37)$$

indica la probabilidad de que el componente falle en o antes del tiempo  $t$ . Esta función permite cuantificar qué proporción de la población habrá fallado para un intervalo de operación dado (Meeker & Escobar, 2020). En sistemas metalmecánicos, valores altos de  $F(t)$  durante las primeras horas de operación suelen asociarse con fenómenos de mortalidad temprana.

### **Función de Densidad de Probabilidad $f(t)$**

La función de densidad de probabilidad, denotada como  $f(t)$ , describe la distribución puntual del tiempo hasta la falla:

$$f(t) = \frac{dF(t)}{dt}, f(t) \geq 0, \int_0^{\infty} f(t) dt = 1 \quad (38)$$

En términos ingenieriles,  $f(t)$  permite identificar los periodos donde la frecuencia relativa de fallas es mayor, facilitando la planeación de inspecciones o monitoreo intensivo (Dodson & Nolan, 2021).

### **Función de Supervivencia o Confiabilidad $R(t)$**

La función de confiabilidad, definida como:

$$R(t) = P(T > t) = 1 - F(t) \quad (39)$$

representa la probabilidad de que el componente continúe funcionando después del tiempo  $t$ . Esta función es fundamental para la planeación del mantenimiento, ya que permite evaluar la probabilidad de supervivencia de los equipos bajo diferentes horizontes temporales (Dhillon, 2017).

Cuando  $R(t)$  decrece lentamente, el componente posee alta confiabilidad; por el contrario, una disminución abrupta señala condiciones de deterioro acelerado.

### Función Falla o Función de Tasa Falla $h(t)$

La tasa de falla es uno de los conceptos centrales de la ingeniería de confiabilidad. Describe la probabilidad condicional de que ocurra una falla en el intervalo  $[t, t + \Delta t]$ , dado que el componente ha sobrevivido hasta el tiempo  $t$ :

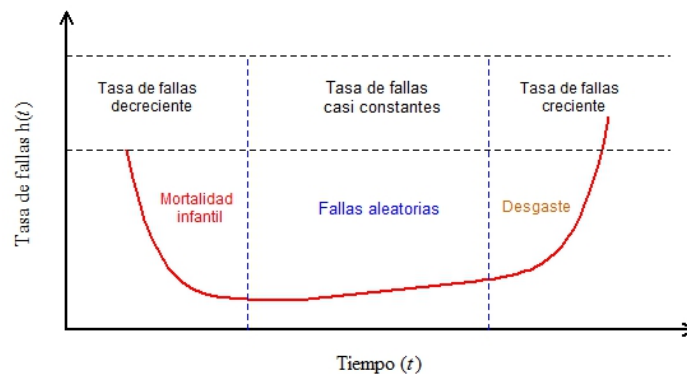
$$h(t) = \frac{f(t)}{R(t)} \quad (40)$$

La forma de  $h(t)$  determina el comportamiento del ciclo de vida del componente (Meeker & Escobar, 2020). La relación entre el riesgo y el tiempo permite clasificar los mecanismos dominantes de falla y seleccionar la distribución estadística adecuada para modelarlos.

### Curva de bañera (Bathtub Curve)

La curva de bañera es la representación clásica del comportamiento de la tasa de falla en sistemas industriales (Lai, Xie & Murthy, 2001):

- Fase de mortalidad temprana:  $h(t)$  decreciente
- Fase de vida útil:  $h(t)$  casi constante
- Fase de desgaste:  $h(t)$  creciente



**Figura 17.** Curva de bañera para tasa de fallas (González-Hernández et al., 2025).

Esta forma no monótona constituye la motivación central del proyecto, dada la necesidad de utilizar distribuciones avanzadas capaces de capturar estos regímenes heterogéneos de riesgo (Lai, Xie & Murthy, 2001).

## Medidas estadísticas de confiabilidad: MTTF, MTBF y disponibilidad

- Tiempo medio hasta la falla

$$MTTF = \int_0^{\infty} t f(t) dt \quad (41)$$

Representa la expectativa matemática del tiempo de vida de equipos no reparables (Dhillon, 2017).

- Tiempo medio entre fallas

$$MTBF = MTTF + MTTR \quad (42)$$

Incluye el tiempo de reparación MTTR y es útil para equipos reparables (Meeker & Escobar, 2020).

- Disponibilidad instantánea y estacionaria

$$A = \frac{MTBF}{MTBF + MTTR} \quad (43)$$

La disponibilidad expresa la proporción del tiempo en que un sistema está operativo y funcional. Esta medida está estrechamente vinculada con la planificación del mantenimiento (Dhillon, 2017).

## Censura en datos de confiabilidad

En análisis de vida es común que no todas las fallas sean observadas; esto se denomina censura. Existen diversos tipos:

- Censura derecha: el componente no ha fallado al finalizar la observación.
- Censura izquierda: se desconoce cuándo inició el deterioro.
- Censura por intervalos: la falla ocurrió entre dos inspecciones.

La presencia de censura exige métodos de estimación robustos, como máxima verosimilitud modificada o algoritmos EM (Meeker & Escobar, 2020).

### **Distribuciones de probabilidad empleadas en el análisis de confiabilidad**

- Distribución exponencial: Adecuada para equipos electrónicos y sistemas sin desgaste (Dodson & Nolan, 2021). Modelo de riesgo constante:

$$h(t) = \lambda. \quad (44)$$

- Distribución Weibull: Una de las más versátiles:

$$h(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1} \quad (45)$$

Permite modelar mortalidad temprana, vida útil y desgaste según el parámetro de forma  $\beta$  (Dhillon, 2017):

1.  $\beta < 1 \rightarrow$  mortalidad temprana
2.  $\beta = 1 \rightarrow$  tasa constante
3.  $\beta > 1 \rightarrow$  desgaste

- Distribuciones extendidas para modelar curvas de bañera:
  - Weibull modificada
  - Power-X
  - Weibull-Pareto compuesta
  - Q-Weibull
  - Procesos Wiener de degradación

Su capacidad para representar tasas de falla no monótonas fundamenta su relevancia en esta tesis (Garg, 2020).

### **Mantenimiento basado en condición (CBM)**

El Condition-Based Maintenance utiliza señales de vibración, temperatura, presión o ruido para estimar el riesgo de falla y programar intervenciones antes

de que ocurra un evento crítico. Su adopción depende de la capacidad del modelo estadístico para determinar de forma precisa el comportamiento de la función  $h(t)$  (Dhillon, 2017).

### **Confiabilidad como propiedad del sistema industrial**

La confiabilidad no sólo depende del diseño del componente, sino de factores operativos como:

- condiciones de carga,
- nivel de vibración,
- condiciones ambientales,
- calidad del mantenimiento.

Su medición y modelación se apoya en los conceptos antes definidos:  $T$ ,  $f(t)$ ,  $F(t)$ ,  $R(t)$ ,  $h(t)$ , MTTF y disponibilidad (Bracke, 2024).

### **Síntesis conceptual**

El marco conceptual establece las definiciones fundamentales que permiten modelar matemáticamente el comportamiento de los tiempos de falla y representar la curva de bañera mediante distribuciones avanzadas. Estos conceptos serán empleados en el capítulo metodológico para justificar la selección de modelos, realizar la estimación de parámetros y evaluar la capacidad predictiva de cada distribución en el análisis de confiabilidad industrial.

## **Capítulo III. Marco metodológico para el análisis de tiempos de falla**

### **3.1 Selección de la metodología de la tesis**

La selección de la metodología constituye un elemento central dentro del desarrollo de este proyecto de tesis, ya que permite garantizar que las actividades de análisis, modelado y validación estadística de los tiempos de falla se realicen bajo un marco riguroso, replicable y coherente con los principios de la ingeniería de confiabilidad. Como señalan Hernández-Sampieri y Mendoza (2018), toda investigación debe estructurarse a partir de decisiones metodológicas claras que definan el tipo de estudio, su alcance, enfoque y procedimientos analíticos, asegurando así que la solución propuesta responda adecuadamente al planteamiento del problema.

En este sentido, el presente trabajo adopta una metodología orientada al desarrollo y validación de un procedimiento aplicado para la selección y ajuste de distribuciones de probabilidad que modelen fallas con comportamiento tipo curva de bañera, con énfasis en herramientas estadísticas y en el uso del software R como plataforma de análisis.

#### **Tipo de proyecto**

De acuerdo con las clasificaciones metodológicas tradicionales, este estudio se enmarca como investigación aplicada y tecnológica. Según Vásquez et al. (2023), la investigación aplicada se orienta a resolver problemas prácticos mediante el uso de fundamentos científicos, mientras que la investigación tecnológica se enfoca en el diseño, implementación y evaluación de métodos, procesos o sistemas que mejoran la práctica profesional. En el contexto del planteamiento del problema, este proyecto:

- Es aplicada, porque su finalidad es resolver una necesidad real: la falta de una metodología práctica para que los ingenieros industriales puedan modelar fallas y estimar métricas de confiabilidad relevantes para la toma de decisiones.

- Es tecnológica, porque desarrolla un procedimiento sistematizado en seis fases que incorpora técnicas estadísticas avanzadas, herramientas de modelado probabilístico y funciones especializadas del software R para analizar datos de fallas no censurados.

El producto final es una guía metodológica operativa para uso académico y profesional.

### **Propósito o alcance del estudio**

La naturaleza del problema requiere un enfoque metodológico que permita describir, explicar y fundamentar técnicamente el comportamiento de los tiempos de falla. Conforme a Hernández-Sampieri (2018), los alcances pueden categorizarse como exploratorios, descriptivos, correlacionales o explicativos, dependiendo del nivel de conocimiento existente y del objetivo de la investigación. En este proyecto, el estudio posee dos alcances complementarios:

- Descriptivo: Porque busca caracterizar el comportamiento de un conjunto de tiempos de falla, describiendo:
  - la forma empírica del hazard  $h(t)$ ,
  - la supervivencia  $R(t)$ ,
  - las fases de la curva de bañera,
  - y las propiedades estadísticas del conjunto de datos.

El análisis descriptivo permite detallar cómo se comporta el fenómeno, sin necesidad de establecer relaciones causales.

- Explicativo: El estudio es también explicativo, en tanto pretende fundamentar por qué ciertas distribuciones probabilísticas representan mejor la dinámica real del riesgo de falla y cómo dicha representación permite inferir métricas como MTTF, vida remanente o puntos de inflexión de la curva de bañera.

Este doble alcance responde a la necesidad de transformar datos empíricos en información técnica útil para la ingeniería industrial.

## **Enfoque o método del proyecto**

Conforme a los criterios de Hernández-Sampieri (2018), un estudio utiliza enfoque cuantitativo cuando emplea mediciones numéricas, métodos estadísticos y análisis estructurado para responder preguntas de investigación. En esta tesis, el enfoque es claramente cuantitativo, debido a que:

- Trabaja con datos numéricos (tiempos de falla).
- Aplica técnicas estadísticas como Kaplan–Meier, estimación de hazard, máxima verosimilitud (MLE), AIC, BIC, KS y AD.
- Utiliza software especializado (R) para replicabilidad y procesamiento.
- Genera resultados matemáticos interpretables para decisiones de ingeniería.

Además, el diseño es no experimental de corte transversal, ya que los datos no son manipulados, sino analizados tal como fueron registrados en un periodo específico. Este enfoque es el más adecuado para el análisis de confiabilidad, dado que los modelos probabilísticos requieren estimaciones basadas en evidencia empírica directa.

## **Técnicas e instrumentos de recolección de datos**

La recolección de datos para este estudio se basa en registros históricos de fallas, metodología ampliamente reconocida como una fuente primaria válida para el análisis de confiabilidad en ingeniería (Jardine & Tsang, 2013). En este proyecto, los tiempos de falla provienen de un conjunto de datos clásico utilizado en la literatura científica para ilustrar comportamientos tipo curva de bañera. Dichos datos fueron tomados del artículo de Aarset (1987), uno de los trabajos más citados en el ámbito del análisis de función de falla no monótona y un referente fundamental para la identificación de tasas de falla con forma de bañera.

El vector empleado corresponde exclusivamente a datos de falla no censurados, lo que permite aplicar de manera directa modelos paramétricos tradicionales (Weibull, lognormal, gamma) y distribuciones no monótonas (como Beta–Weibull o Q–Weibull), facilitando la validación comparativa mediante

criterios de bondad de ajuste. Las principales técnicas e instrumentos de recolección y validación de datos son:

- Observación documental y análisis de registros técnicos: se revisaron los datos presentados por Aarset (1987) para extraer los tiempos de falla y comprender su contexto de uso como ejemplo estándar de hazard no monótono.
- Software especializado: se empleó principalmente R, dado que constituye una plataforma robusta para análisis estadístico avanzado. En etapas posteriores se utilizaron funciones como `fitdist()`, `muhaz()`, `optim()` y herramientas de diagnóstico visual para la selección y evaluación de modelos probabilísticos.

### **Procedimiento para el análisis de datos**

El procedimiento metodológico está estructurado en seis fases, diseñadas para asegurar un análisis exhaustivo y replicable, alineado con estándares de ingeniería de confiabilidad:

1. Preparación y comprensión de los datos.
2. Selección preliminar del modelo probabilístico.
3. Ajuste paramétrico mediante el método de Estimación de Máxima Verosimilitud (MLE por sus siglas en inglés).
4. Evaluación de calidad del modelo.
5. Interpretación de los resultados.
6. Documentación y comunicación técnica.

Este proceso asegura coherencia metodológica y permite transformar modelos estadísticos en decisiones prácticas para ingeniería industrial.

### **3.2 Propuesta del procedimiento metodológico para la modelación de tiempos de falla en forma de curva de bañera**

La modelación de tiempos de falla con comportamiento tipo curva de bañera exige un enfoque metodológico estructurado que integre los fundamentos estadísticos de la ingeniería de confiabilidad con las necesidades reales del

entorno industrial. Debido a que la curva de bañera presenta fases diferenciadas como mortalidad temprana, vida útil estable y desgaste, su análisis requiere técnicas capaces de captar patrones no monótonos en la tasa de riesgo, así como procedimientos rigurosos para seleccionar, ajustar y evaluar distribuciones de probabilidad adecuadas. En respuesta a la brecha formativa identificada en el planteamiento del problema, esta tesis propone un procedimiento metodológico de seis fases orientado a guiar al estudiante o profesional en cada etapa del análisis de fallas.

El enfoque integra análisis exploratorio, métodos no paramétricos, teoría de distribuciones, estimación por máxima verosimilitud y criterios de bondad de ajuste, complementados con una interpretación en ingeniería en el contexto del comportamiento de la confiabilidad. Además, se incorpora el uso del software R como herramienta central debido a su capacidad para procesar datos de falla, estimar hazard empíricos, ajustar modelos clásicos y no monótonos, y generar comparaciones visuales reproducibles.

El objetivo central es proporcionar una guía práctica que permita al ingeniero industrial estimar métricas como  $R(t)$ ,  $h(t)$ , MTTF y MRL, fortaleciendo así la vinculación entre formación académica y práctica profesional basada en datos. A continuación, se presentan las fases de la guía metodológica para analizar tiempos de falla, posteriormente se realiza una descripción de cada fase.

## **Fase 1. Preparación y comprensión de los datos**

### **1. Definir el contexto del sistema a analizar**

- Tipo de equipo o componente.
- Condiciones de operación.
- Dominio físico de fallas.

### **2. Construcción del vector de tiempos de falla**

- Fallas no censuradas/fallas + censura.
- Validación de calidad de datos.

### **3. Análisis descriptivo de los tiempos de falla**

- Estadísticos básicos: media, mediana, IQR, curtosis, asimetría.
- Histogramas, boxplots, diagramas de frecuencia.

### **4. Estimación no paramétrica de $R(t)$ y $h(t)$**

- Supervivencia empírica (Kaplan–Meier).
- Función de falla empírica (muhaz).
- Identificación preliminar de fases:
  - mortalidad temprana,
  - vida útil estable,
  - desgaste.

## **Fase 2. Selección preliminar del modelo probabilístico**

- 1. Diagnóstico de forma con base en la distribución de los datos**
  - Sesgo, colas, variabilidad.
  - Forma empírica de la función de falla.
- 2. Definición del catálogo de distribuciones candidatas**
  - Tradicionales: Weibull, Lognormal, Gamma, Exponencial.
  - Extensivas: Log-logistic, Generalizada Gamma.
  - No monótonas: Beta–Weibull, Power–Lindley, Q-Weibull, Weibull–Pareto.
- 3. Clasificación preliminar del tipo de la función de falla**
  - Monótona creciente
  - Monótona decreciente
  - Bañera
  - Unimodal

## **Fase 3. Ajuste paramétrico del modelo a través del método de Estimación de Máxima Verosimilitud (MLE)**

- 1. Ajuste de distribuciones clásicas con MLE automático**
  - fitdistrplus, flexsurv.
- 2. Ajuste de distribuciones no monótonas**
  - Construcción explícita de FDP, FDA,  $R(t)$  y  $h(t)$ .
  - Optimización numérica (optim).
- 3. Obtención de parámetros estimados**
  - Estimación puntual.
  - Error estándar.
  - Intervalos de confianza.
  - Interpretación física de parámetros.

#### **4. Cálculo de las métricas de desempeño:**

- $R(t)$  (función de supervivencia).
- $h(t)$  (función de falla).
- MTTF.
- Supervivencia condicional.
- Vida media residual (MRL).

### **Fase 4. Evaluación de la calidad del modelo**

#### **1. Comparación de criterios de información**

- AIC
- BIC
- Log-likelihood

#### **2. Pruebas de bondad de ajuste**

- Kolmogorov–Smirnov
- Anderson–Darling

#### **3. Contraste visual sistemático**

- $R(t)$  teórica vs  $R(t)$  empírica.
- $h(t)$  teórica vs  $h(t)$  empírica.
- FDP teórica vs histograma.
- QQ-plot, PP-plot.

#### **4. Selección del modelo óptimo**

- Mínimo AIC/BIC.
- Buen ajuste visual.
- Coherencia con el comportamiento físico (curva de bañera).

### **Fase 5. Interpretación de los resultados en el contexto de ingeniería de confiabilidad**

#### **1. Identificación de las fases de la curva de bañera**

- $t_1$  (fin de mortalidad temprana).
- $t_2$  (inicio de desgaste).

#### **2. Estimación de métricas de confiabilidad**

- MTTF.
- Tasa de falla acumulada.
- Probabilidad de falla antes de un tiempo  $t$  dado.

- Vida media residual (MRL).

### **3. Implicaciones para mantenimiento**

- Política de reemplazo óptima.
- Intervalos de inspección.
- Estrategias preventivas/predictivas.

## **Fase 6. Documentación y comunicación técnica**

1. Plantillas para reportes de ingeniería
2. Visualizaciones estandarizadas
3. Conclusiones técnicas
4. Recomendaciones basadas en datos

## **Fase 1. Preparación y comprensión de los datos**

La primera fase del análisis de confiabilidad constituye la base metodológica que permite garantizar que los modelos estadísticos aplicados a la curva de bañera representen fielmente el comportamiento real de los equipos analizados. En Ingeniería Industrial, el entendimiento profundo del proceso, del componente y del entorno operativo es esencial, ya que la confiabilidad no depende únicamente de una función estadística, sino también de las condiciones físicas, mecánicas y humanas en las que opera el sistema (Dhillon, 2023). Por ello, esta fase integra cuatro actividades fundamentales: la definición del contexto del sistema, la construcción del vector de tiempos de falla, el análisis descriptivo de los datos y la estimación no paramétrica de la confiabilidad y de la función de falla.

Dado que en esta tesis se empleará el lenguaje del software R como herramienta computacional para el análisis de confiabilidad, a lo largo de las presentes fases se incluyen sugerencias de funciones, paquetes y estructuras de código que permiten implementar cada actividad descrita.

### **F-1.1. Definir el contexto del sistema a analizar**

#### **Tipo de equipo o componente**

El análisis de confiabilidad debe partir de la caracterización del producto, componente o equipo industrial que se esté analizando: su función dentro del proceso productivo, su criticidad, el tipo de tecnología incorporada y el modo general de operación. Los equipos pueden pertenecer a categorías mecánicas

(bombas, compresores, motores), eléctricas (variadores, transformadores) o electromecánicas (sistemas CNC, robots industriales). Cada categoría posee patrones de falla distintivos que influyen la forma de la función de falla  $h(t)$  (Jardine & Tsang, 2013).

Aunque esta etapa es fundamentalmente conceptual, R puede emplearse para organizar la información del sistema en estructuras tipo dataframe, lo que permite posteriormente aplicar análisis por subgrupos. Funciones como `data.frame()` o las rutinas de manipulación del paquete `dplyr` resultan útiles para este propósito.

### **Condiciones de operación**

Factores como carga aplicada, ciclos de trabajo, temperatura, vibración, lubricación y condiciones ambientales influyen directamente en la degradación del sistema. La confiabilidad  $R(t)$  tiende a disminuir más rápidamente en ambientes térmicamente exigentes, mientras que los modos de fallo por fatiga mecánica dependen del número de ciclos operativos. Estos factores deben documentarse para interpretar correctamente la forma de  $h(t)$ .

Estas características pueden describirse y resumirse en R mediante funciones como `summary()`, o visualizarse con `boxplot()` y otras funciones gráficas básicas, lo que permite detectar condiciones extremas que podrían sesgar la interpretación de los tiempos de vida.

### **Dominio físico de fallas**

La identificación del dominio físico (desgaste, corrosión, fractura por fatiga, sobrecalentamiento, fallas eléctricas, degradación dieléctrica, etc.) permite establecer hipótesis preliminares sobre la naturaleza de la falla. Por ejemplo:

- Fallas infantiles suelen relacionarse con defectos de manufactura.
- Periodos de vida útil estable con fallas aleatorias.
- Fase de desgaste con mecanismos acumulativos no lineales.

Estos conocimientos son esenciales antes de seleccionar un modelo probabilístico.

### **F-1.2. Construcción del vector de tiempos de falla**

Una vez definido el sistema, se procede a construir el vector  $T$ , que contiene los tiempos de falla observados. Por lo tanto, este vector puede incluir:

**Fallas no censuradas (eventos completos):** Representan componentes que efectivamente fallaron durante el periodo de observación. Su tiempo corresponde al ciclo de vida real.

**Datos censurados:** Cuando el componente no falló antes del final de la observación, se clasifica como censura por la derecha. Este tipo de datos es típico en mantenimiento preventivo, paros programados o equipos retirados por decisión operativa.

En el caso de los datos utilizados en esta tesis, todos los tiempos registrados corresponden únicamente a fallas observadas, sin presencia de censura. En términos de implementación computacional, la representación estándar de este tipo de datos en R se realiza mediante la función `Surv()` del paquete `survival`, especificando que todos los eventos son fallas mediante un vector indicador igual a 1.

### **Validación de calidad de datos**

Se aplican filtros para eliminar registros inconsistentes, se pueden utilizar como las funciones `subset()` y `unique()`, así como en los procedimientos de depuración disponibles en el paquete `dplyr`. Estas validaciones aseguran que el vector  $T$  sea consistente y representativo, condición indispensable para garantizar la fiabilidad de los resultados que se obtendrán en las fases siguientes, ejemplos de datos que deben eliminarse son:

- tiempos negativos,
- duplicados mal registrados,
- censuras mal clasificadas,
- equipos nuevos con tiempos cero,
- errores por reinicios de ciclo.

El análisis de confiabilidad sólo es sólido si el vector  $T$  es consistente y representativo.

### F-1.3. Análisis descriptivo de los tiempos de falla

#### Estadísticos básicos

Esta etapa provee la primera aproximación a la forma del comportamiento de los datos (tiempos de falla). Algunos estadísticos que se calculan son los siguientes:

- Media:

$$\bar{t} = \frac{1}{n} \sum t_i \quad (46)$$

- Mediana (robusta ante valores extremos).
- Rango intercuartílico (IQR) para evaluar dispersión.
- Asimetría y curtosis para comprender la forma.
- Coeficiente de variación para detectar variabilidad relativa.

Para el cálculo de estadísticos básicos como media, mediana e IQR, pueden utilizarse funciones propias de R como `mean()`, `median()` y `IQR()`. Para evaluar la asimetría y la curtosis, R dispone de funciones como `skewness()` y `kurtosis()` del paquete `e1071`, que permiten identificar desviaciones significativas respecto a distribuciones simétricas.

#### Visualizaciones

En cuanto a las visualizaciones, R proporciona herramientas como `hist()` para histogramas, `boxplot()` para detección de valores atípicos y `ecdf()` para graficar la frecuencia acumulada. Estas representaciones gráficas permiten observar multimodalidad, colas pesadas o comportamientos característicos de desgaste, aspectos que orientan la selección de modelos probabilísticos en fases posteriores. Algunas de las gráficas útiles son:

- Histogramas permiten observar multimodalidad o sesgo.
- Boxplots detectan valores atípicos (outliers) y asimetría.

- Diagramas de frecuencia acumulada ayudan a intuir el comportamiento de supervivencia.

Este análisis preliminar orienta la selección de distribuciones candidatas, por ejemplo:

- curvas con asimetría positiva: Weibull, lognormal.
- distribuciones más simétricas: normal (raro en confiabilidad).
- colas pesadas: Weibull-Pareto, log-logística.

#### **F-1.4. Estimación no paramétrica de $R(t)$ y $h(t)$**

##### **Supervivencia empírica (Kaplan–Meier)**

La supervivencia empírica mediante Kaplan–Meier permite estimar la confiabilidad  $R(t)$  sin asumir ninguna distribución teórica y es especialmente útil cuando existen datos censurados, como equipos que no fallan durante el periodo de observación. Este estimador construye una curva de supervivencia basada únicamente en los datos reales, lo que facilita identificar patrones iniciales del comportamiento del sistema, incluyendo posibles fases de la curva de bañera. Además, la curva Kaplan–Meier sirve como referencia para comparar el ajuste de modelos paramétricos posteriores y permite obtener medidas como el MTTF empírico y la supervivencia condicional, fundamentales para el análisis de confiabilidad y la toma de decisiones de mantenimiento. La función de supervivencia o confiabilidad empírica se calcula mediante el estimador de Kaplan–Meier:

$$\hat{R}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (47)$$

donde:

- $d_i$  = número de fallas en  $t_i$ .
- $n_i$  = número de componentes en riesgo inmediatamente antes de  $t_i$ .

La estimación de la función de confiabilidad  $R(t)$  mediante el estimador de Kaplan–Meier se puede implementar utilizando la función `survfit()` del paquete `survival`. Esta herramienta permite calcular la supervivencia empírica incorporando datos censurados y generar la curva Kaplan–Meier, la cual es la base para identificar patrones como mortalidad temprana, estabilidad y desgaste. La función `summary()` permite obtener valores puntuales de  $R(t)$ , mientras que las opciones gráficas de `plot()` facilitan su interpretación visual.

### **Función de falla empírica $h(t)$**

La función de falla empírica  $h(t)$  permite identificar cómo cambia el riesgo instantáneo de falla a lo largo del tiempo, dado que el componente ha sobrevivido hasta ese instante. A partir de los datos observados, este indicador revela si el sistema presenta mortalidad temprana, un periodo de riesgo constante o una fase de desgaste, lo cual es fundamental para reconocer patrones propios de la curva de bañera. Su análisis orienta la selección del modelo probabilístico más adecuado y ayuda a validar si las distribuciones ajustadas reflejan correctamente el comportamiento real del riesgo, apoyando decisiones de mantenimiento y estimación de vida remanente. La función de falla empírica puede estimarse con métodos de suavizamiento, como *muha*, para visualizar si el patrón es:

- decreciente (mortalidad temprana),
- constante (vida útil estable),
- creciente (desgaste),
- o una combinación (curva de bañera completa).

Para estimar el comportamiento de la falla se puede emplear una aproximación no paramétrica basada en técnicas de suavizamiento tipo kernel. En particular, la función de falla empírica se puede calcular utilizando la función `muha`() del paquete `muha` del software R (Hess, 2024), la cual implementa el estimador propuesto por Müller y Wang (1994) para obtener una tasa de falla suavizada a partir de los tiempos de vida y los datos censurados. Esta herramienta permite identificar patrones de riesgo característicos como: mortalidad temprana, riesgo constante o desgaste. Además, proporciona una

representación detallada de la dinámica temporal de  $h(t)$  sin asumir una distribución teórica específica.

### **Identificación preliminar de fases**

A partir del comportamiento de  $R(t)$  y  $h(t)$  se identifican tres regiones críticas:

1. Mortalidad temprana:  $h(t)$  decreciente.
2. Vida útil estable:  $h(t)$  aproximadamente constante.
3. Desgaste:  $h(t)$  creciente.

Esta clasificación es esencial para seleccionar una distribución adecuada en fases posteriores. La interpretación de estas fases se apoya, en el caso de esta tesis, en las visualizaciones obtenidas mediante `plot(survfit())` y `plot(muhaz())`, que permiten contrastar empíricamente el comportamiento esperado según la física de fallas del sistema.

### **Fase 2. Selección preliminar del modelo probabilístico**

La selección del modelo probabilístico representa una etapa esencial dentro del análisis de confiabilidad, ya que permite identificar qué funciones de distribución podrían reproducir con mayor fidelidad el comportamiento empírico de las fallas observadas en el producto, componente, equipo o sistema. Esta etapa constituye un puente entre la caracterización inicial de los datos (Fase 1) y el ajuste paramétrico por máxima verosimilitud (Fase 3), y su objetivo principal es acotar el conjunto de modelos candidatos con base en la forma empírica de la función de falla  $h(t)$  y de la estructura estadística subyacente. Como señalan Meeker y Escobar (2022), la selección adecuada de distribuciones es crucial, ya que un modelo mal elegido produce estimaciones sesgadas de confiabilidad, MTTF y vida remanente.

#### **F-2.1. Diagnóstico de forma con base en la distribución de los datos**

El diagnóstico de la forma de los datos constituye el primer filtro para seleccionar las distribuciones probabilísticas. Este análisis se enfoca en el sesgo, la forma de las colas y la variabilidad, elementos que influyen directamente en la pertinencia de una u otra distribución (Lawless, 2003). En sistemas industriales,

los tiempos de falla suelen presentar asimetría positiva, típicamente asociada con procesos de degradación acumulativa o mecanismos multiplicativos de daño. Este sesgo puede cuantificarse mediante el coeficiente clásico de asimetría:

$$\text{Asimetria} = \frac{1}{n} \sum_{i=1}^n \left( \frac{t_i - \bar{t}}{s} \right)^3. \quad (48)$$

Esta asimetría puede cuantificarse mediante funciones como `skewness()` y `kurtosis()` del paquete `e1071` en R, que permiten determinar la forma empírica del conjunto de datos.

Asimismo, la forma empírica de la función de falla  $h(t)$ , estimada mediante técnicas de suavizamiento kernel, puede obtenerse en R utilizando la función `muhaaz()` del paquete `muhaaz`. Esta herramienta genera una curva suavizada del riesgo instantáneo que facilita la identificación de patrones como decrecimiento inicial, riesgo estable o desgaste acelerado. Según Jardine y Tsang (2013), el hazard empírico constituye uno de los criterios más confiables para determinar si un sistema experimenta mortalidad temprana, una fase aleatoria o desgaste acelerado.

En conjunto, estas funciones permiten caracterizar estadísticamente la distribución de los tiempos de falla y descartar modelos teóricos incompatibles con la evidencia empírica. El objetivo de este diagnóstico no es seleccionar el modelo definitivo, sino reducir el universo de posibilidades a un conjunto coherente con la física de fallas y con la estructura estadística observada.

### **F-2.2. Definición del catálogo de distribuciones candidatas**

Con base en la evidencia empírica, se establece una cartera de distribuciones candidatas agrupadas en tres categorías: tradicionales, extensivas y no monótonas. Esta clasificación está alineada con los lineamientos propuestos en la literatura moderna de confiabilidad (Meeker & Escobar, 2022; Murthy et al., 2004).

En R, la exploración preliminar de estas distribuciones puede realizarse mediante el uso del paquete `fitdistrplus`, particularmente con la función `fitdistr()`,

que permite ajustar varios modelos paramétricos y comparar su desempeño. Además, las funciones `denscomp()`, `qqcomp()` y `ppcomp()` permiten contrastar visualmente el ajuste de cada distribución candidata respecto a los datos empíricos.

### Distribuciones tradicionales

Las distribuciones clásicas ofrecen interpretabilidad y una estructura matemática conocida. En el entorno R, estas distribuciones se ajustan fácilmente mediante `fitdist()`, especificando el nombre del modelo (“weibull”, “lnorm”, “gamma”, “exp”). Entre las principales se encuentran:

- Weibull, cuya función de falla es:

$$h(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1}, \quad (49)$$

permite representar riesgos decrecientes, constantes o crecientes.

- Lognormal, recomendada cuando los mecanismos de daño presentan efectos multiplicativos.
- Gamma, apropiada para procesos de desgaste progresivo.
- Exponencial, útil únicamente cuando la función de falla es estrictamente constante.

La comparación preliminar de estas alternativas puede realizarse en R mediante `gofstat()` o parámetros como AIC y BIC.

### Distribuciones extensivas

Estas distribuciones amplían la flexibilidad de las tradicionales y permiten capturar comportamientos intermedios o colas más pesadas:

- Log-logística, adecuada para sistemas con alta variabilidad y colas pesadas.
- Generalized Gamma, una familia versátil que R puede ajustar mediante extensiones del paquete `flexsurv`, particularmente con la función `flexsurvreg()` (Meeker & Escobar, 2022).

Estas distribuciones son útiles cuando la función de falla es monótono pero los datos requieren mayor adaptación que la ofrecida por los modelos clásicos.

### **Distribuciones no monótonas**

Cuando la función de falla empírico presenta patrones complejos, como bañera o unimodal, las distribuciones tradicionales se vuelven insuficientes. En R, este tipo de modelos puede explorarse mediante:

- `flexsurvreg()` para distribuciones generalizadas.
- Funciones personalizadas dentro de `fitdist()` cuando la distribución no esté incluida por defecto, o ampliaciones disponibles en paquetes especializados.

Entre las alternativas no monótonas destacan:

Beta–Weibull

Power–Lindley

Q–Weibull

Weibull–Pareto

Estas distribuciones permiten modelar sistemas con múltiples etapas de riesgo y colas complejas, representando de manera más realista la física de fallas.

### **F-2.3. Clasificación preliminar del tipo de la función de falla**

Una vez estimado la función de falla empírica mediante `muhaz()`, y observada su forma general, se clasifica en una de las categorías propuestas por Murthy et al. (2004):

- Monótona creciente, típica de procesos de deterioro acumulativo.
- Monótona decreciente, relacionada con fallas infantiles o defectos iniciales.

- Bañera, donde coexisten mortalidad temprana, vida útil estable y desgaste acelerado.
- Unimodal, cuando existe un único pico de riesgo debido a modos de falla competitivos.

Esta clasificación orienta la selección del subconjunto final de distribuciones a ajustar en la Fase 3 y evita la aplicación de modelos cuya estructura matemática es incompatible con el comportamiento empírico del sistema.

### **Fase 3. Ajuste paramétrico del modelo mediante Estimación de Máxima Verosimilitud (MLE)**

La tercera fase del análisis de confiabilidad consiste en ajustar formalmente los modelos probabilísticos seleccionados mediante el método de Estimación de Máxima Verosimilitud (MLE, por sus siglas en inglés, Maximum Likelihood Estimation,). Este procedimiento permite obtener los valores de los parámetros que hacen más probable observar los datos reales registrados, proporcionando así una representación matemática coherente con el comportamiento empírico del sistema. El MLE es el estándar en análisis de vida debido a su consistencia, eficiencia y capacidad para proporcionar estimadores interpretables desde una perspectiva física (Meeker & Escobar, 2022).

Dado que la tesis emplea R como herramienta analítica principal, en esta fase se mencionan las funciones pertinentes para implementar cada procedimiento, el desarrollo computacional se incluirá en los anexos.

#### **F-3.1. Ajuste de distribuciones clásicas mediante MLE automático**

El ajuste de las distribuciones clásicas como la Weibull, Lognormal, Gamma y Exponencial, se realiza mediante métodos automáticos disponibles en R, principalmente con el paquete `fitdistrplus`, a través de la función `fitdist()`, que implementa el método de MLE para una amplia variedad de modelos. Esta función permite obtener los parámetros estimados, sus errores estándar y diversas métricas de bondad de ajuste.

Asimismo, el paquete `flexsurv`, mediante la función `flexsurvreg()`, ofrece una implementación más flexible del MLE, particularmente útil cuando se trabaja

con datos censurados o cuando se requiere ajustar distribuciones extendidas o variantes paramétricas complejas. Ambos paquetes generan salidas compatibles con gráficos comparativos como densidad ajustada, probabilidad acumulada y curvas de supervivencia teórica frente a las empíricas.

En esta etapa, los modelos tradicionales son ajustados primero porque sirven como línea base para comparar posteriormente el desempeño de modelos más flexibles o no monótonos.

### **F-3.2. Ajuste de distribuciones no monótonas**

Cuando los datos presentan una estructura de la función de falla no monótona (es decir, con forma de bañera, unimodal o con múltiples cambios de tendencia), los modelos tradicionales son insuficientes. En estos casos, el ajuste requiere pasos adicionales.

Para este tipo de modelos, R permite definir de manera explícita la FDP, la FDA, la FS  $R(t)$  y la FF  $h(t)$ . Estas definiciones se incorporan en una función de verosimilitud personalizada, la cual se maximiza utilizando herramientas de optimización numérica.

La función `optim()` es la herramienta central para este proceso, ya que permite maximizar la verosimilitud o minimizar la función negativa de la verosimilitud según los parámetros del modelo. Esta función admite distintas estrategias de búsqueda como Nelder–Mead o BFGS, lo que permite abordar modelos altamente no lineales.

Modelos modificados como la distribución Beta–Weibull, Power–Lindley, Q–Weibull o Weibull–Pareto se ajustan mediante esta estrategia, lo que proporciona estimaciones precisas de los parámetros incluso cuando la estructura matemática no está disponible directamente en paquetes predefinidos.

### **F-3.3. Obtención de parámetros estimados**

Una vez completado el ajuste por el método MLE, se obtienen los parámetros estimados de cada distribución evaluada. Tanto `fitdist()` como `flexsurvreg()` proporcionan directamente:

- Estimaciones puntuales de los parámetros, como escala, forma o parámetros adicionales propios de distribuciones extendidas.
- Errores estándar, necesarios para evaluar la precisión de los estimadores.
- Intervalos de confianza, que permiten cuantificar la incertidumbre asociada a cada parámetro.

En modelos ajustados manualmente mediante `optim()`, estos elementos se pueden obtener a partir de la matriz Hessiana calculada numéricamente, para lo cual R ofrece funciones como `optimHess()`, que permiten derivar las varianzas asintóticas de los estimadores. Posteriormente, se interpreta cada parámetro desde el punto de vista físico: por ejemplo, en la distribución Weibull, el parámetro  $\beta$  indica el tipo de hazard (mortalidad temprana, vida útil estable o desgaste), mientras que el parámetro  $\eta$  representa una medida de vida característica. En distribuciones extendidas, los parámetros adicionales suelen controlar la forma del hazard o la pesadez de las colas.

#### F-3.4. Cálculo de métricas de desempeño del modelo

Una vez estimados los parámetros, se procede al cálculo de las métricas fundamentales para el análisis de confiabilidad:

- **$R(t)$ : función de supervivencia**

Las funciones `psurvreg()`, `pweibull()`, `plnorm()`, entre otras, permiten calcular la supervivencia teórica según la distribución ajustada. En modelos personalizados,  $R(t)$  se obtiene directamente de la FDA definida por el usuario.

- **$h(t)$ : función de falla**

En distribuciones estándar,  $h(t)$  puede obtenerse usando las funciones de densidad (`dweibull()`, `dlnorm()`) y supervivencia mediante la relación:

$$h(t) = \frac{f(t)}{R(t)} \quad (50)$$

Para modelos no monótonos,  $h(t)$  se calcula con las expresiones matemáticas definidas junto con la función de verosimilitud.

- **MTTF: Tiempo Medio Hasta el Fallo (Mean Time To Failure)**

En R, el MTTF se calcula mediante las fórmulas teóricas específicas de cada distribución o, en modelos personalizados, mediante integración numérica usando funciones como `integrate()`.

- **Supervivencia condicional**

Para un tiempo  $s$  dado, se calcula mediante:

$$R(t | t > s) = \frac{R(t)}{R(s)} \quad (51)$$

Este valor se obtiene evaluando la función de supervivencia ajustada en los puntos  $s$  y  $t$ .

- **Vida media residual (MRL)**

La MRL puede obtenerse mediante integración:

$$MRL(t) = E[T - t | T > t] \quad (52)$$

En R se calcula integrando numéricamente la supervivencia o utilizando expresiones cerradas según la distribución.

#### **Fase 4. Evaluación de la calidad del modelo**

La cuarta fase del análisis de confiabilidad tiene como finalidad determinar qué modelo probabilístico describe con mayor precisión el comportamiento real de los datos. Esta etapa integra criterios estadísticos formales, pruebas de bondad de ajuste, comparaciones visuales y la verificación de coherencia física con el mecanismo de fallas. Según Meeker y Escobar (2022), la evaluación rigurosa del ajuste es crucial, ya que seleccionar un modelo inapropiado conduce a estimaciones erróneas de MTTF, vida remanente, hazard y otros indicadores clave para la toma de decisiones en Ingeniería Industrial.

#### F-4.1. Comparación de criterios de información

El primer paso para evaluar la calidad del ajuste de los datos (tiempos de falla) consiste en comparar el desempeño estadístico de los modelos mediante criterios de información. Los más utilizados son:

- **AIC (Akaike Information Criterion):** Mide la calidad relativa del modelo penalizando la complejidad. Un AIC menor indica mejor ajuste.
- **BIC (Bayesian Information Criterion):** Penaliza aún más los modelos con muchos parámetros. También se selecciona el valor más bajo.
- **Log-likelihood:** Representa cuán probable es que el modelo genere los datos observados. Valores más altos indican un mejor desempeño.

En R, tanto `fitdist()` como `flexsurvreg()` reportan automáticamente el AIC, BIC y log-likelihood para cada modelo ajustado, lo que permite establecer comparaciones objetivas entre alternativas tradicionales y extendidas.

#### F-4.2. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste permiten evaluar si los datos observados pueden considerarse como realizaciones plausibles del modelo teórico ajustado, las pruebas más comunes son:

- **Kolmogorov–Smirnov (K–S):** Compara la distancia máxima entre las funciones de distribución empírica y teórica. En R, esta prueba puede aplicarse mediante la función `ks.test()` cuando se define la FDA del modelo ajustado.
- **Anderson–Darling (A–D):** Otorga mayor peso a las discrepancias en las colas de la distribución, siendo especialmente útil en confiabilidad. En R puede emplearse la función `ad.test()` del paquete `kSamples` o `nortest`, según el modelo.

Estas pruebas aportan evidencia estadística complementaria para aceptar o rechazar un modelo, aunque no deben utilizarse como único criterio decisorio, sino en conjunto con otros elementos del análisis.

### F-4.3. Contraste visual sistemático

La evaluación visual es un componente esencial en la selección del modelo, ya que permite identificar discrepancias que los criterios numéricos podrían no capturar. Esta fase incluye varios tipos de gráficos comparativos:

- **$R(t)$  teórica vs.  $R(t)$  empírica:** Permite verificar si la curva paramétrica ajustada reproduce adecuadamente la supervivencia observada. En R, los objetos generados por `survfit()` pueden compararse con las curvas teóricas obtenidas mediante funciones como `plot()` y las evaluaciones de la distribución ajustada.
- **$h(t)$  teórica vs.  $h(t)$  empírica:** Es fundamental para detectar incoherencias en sistemas con curvas de bañera, riesgo creciente o riesgo decreciente. La estimación empírica mediante `muhatz()` puede contrastarse con el hazard teórico obtenido a partir de las funciones de densidad y supervivencia del modelo ajustado.
- **FDP teórica vs. histograma de datos:** Permite observar si la forma del modelo se ajusta a la distribución empírica. En R, el paquete `fitdistrplus` incluye funciones como `denscomp()` para este tipo de comparaciones.
- **QQ-plot y PP-plot:** Son herramientas estándar que permiten evaluar la alineación entre cuantiles o probabilidades empíricas y teóricas. El paquete `fitdistrplus` ofrece funciones como `qqcomp()` y `ppcomp()`, que generan comparaciones uniformes para múltiples modelos.

Estos contrastes visuales permiten evaluar no solo el desempeño global del modelo, sino también su ajuste en áreas críticas como las colas o los puntos de inflexión, que son relevantes para aplicaciones de mantenimiento y análisis de riesgo.

### F-4.4. Selección del modelo óptimo

Una vez realizados los análisis estadísticos, las pruebas de bondad de ajuste y los contrastes visuales, se procede a seleccionar el modelo óptimo. Esta selección se basa en tres criterios principales:

1. **Mínimo AIC/BIC:** El modelo elegido debe poseer los valores más bajos de AIC y BIC entre los candidatos, sin comprometer la interpretabilidad.
2. **Buen ajuste visual:** Las curvas teóricas deben coincidir razonablemente con las empíricas, sin divergencias significativas en los extremos o en zonas críticas.
3. **Coherencia con el comportamiento físico:** El modelo debe ser compatible con la física del fallo y con la forma empírica de la función de falla.

Por ejemplo:

- Si  $h(t)$  presenta forma de bañera, deben descartarse modelos monótonos.
- Si  $h(t)$  es claramente creciente, una distribución Weibull con  $\beta > 1$  es físicamente coherente.

En R, la combinación de métricas estadísticas, pruebas y gráficos permite consolidar de manera objetiva la selección del modelo final. Este modelo será utilizado posteriormente en la Fase 5 para interpretar los resultados en términos de confiabilidad, vida remanente y estrategias de mantenimiento.

## **Fase 5. Interpretación de los resultados en el contexto de ingeniería de confiabilidad**

La quinta fase del análisis constituye el puente entre los resultados estadísticos obtenidos en las fases previas y su aplicación en la ingeniería. En confiabilidad, un modelo ajustado carece de valor si no se traduce en información accionable para la toma de decisiones en mantenimiento, operación y gestión del riesgo. Por ello, en esta fase se interpretan los parámetros del modelo, se identifican las fases de la curva de bañera, se calculan métricas clave de confiabilidad y se derivan implicaciones prácticas.

### **F-5.1. Identificación de las fases de la curva de bañera**

Una vez seleccionado el modelo óptimo en la Fase 4, la siguiente etapa consiste en analizar la forma de la función de falla  $h(t)$  y determinar los puntos críticos de la curva de bañera:

- $t_1$ : fin de la mortalidad temprana; corresponde al tiempo en el que la tasa de fallas deja de decrecer y se aproxima a un valor estable. Este punto indica la finalización de los fallos imputables a defectos de fabricación, instalación o puesta en marcha.
- $t_2$ : inicio de la fase de desgaste; representa el momento en que  $h(t)$  cambia de estabilidad a crecimiento, señalando el inicio del deterioro acelerado debido a fatiga, degradación de materiales o envejecimiento funcional.

En R, estos puntos pueden identificarse observando la curva de la función de falla teórica generada a partir del modelo paramétrico. Para ello se emplean funciones como `summary()` y métodos de predicción en objetos ajustados con `flexsurv`, así como comparaciones visuales con la función de falla empírica producido por `muhaaz()`. Por lo tanto, el análisis conjunto de  $h(t)$  empírico y teórico permite garantizar que la interpretación sea consistente con la física del sistema.

### F-5.2. Estimación de métricas de confiabilidad

Con un modelo ajustado, es posible cuantificar indicadores clave que permiten describir el desempeño operativo del componente o equipo. Entre las métricas más importantes se encuentran:

- **MTTF:** Representa el tiempo promedio de operación antes de que ocurra una falla. En R, puede calcularse directamente a partir del modelo ajustado mediante funciones específicas del paquete `flexsurv` o mediante integración numérica utilizando `integrate()` cuando no existe una expresión cerrada.
- **Tasa de falla acumulada:** Indica el número esperado de fallas hasta un tiempo  $t$  dado. En R puede obtenerse evaluando la función de falla acumulada del modelo, disponible para muchos modelos en `flexsurv` o calculable manualmente cuando se trata de modelos extendidos.
- **Probabilidad de falla antes de un tiempo  $t$ :** Se calcula como:

$$F(t) = 1 - R(t) \quad (53)$$

En R, esta métrica puede obtenerse mediante las funciones de distribución acumulada del modelo ajustado, como `pweibull()`, `plnorm()`, `pgamma()` o sus equivalentes en modelos más complejos mediante `predict()` en `flexsurv`.

- **Vida media residual (MRL):** La vida media residual indica cuánto tiempo adicional puede esperarse que funcione el sistema dado que ya ha sobrevivido hasta un tiempo  $t$ . En R, esta métrica puede obtenerse utilizando las rutinas de supervivencia residual disponibles en `flexsurv`, o mediante integración directa de la función de supervivencia teórica.

Cada una de estas métricas permite evaluar el comportamiento temporal del equipo y apoyar decisiones técnicas basadas en probabilidad, especialmente en contextos industriales donde la disponibilidad operativa es crítica.

### F-5.3. Implicaciones para mantenimiento

Las conclusiones de la Fase 5 no son meramente estadísticas: deben traducirse en acciones concretas para mejorar la gestión de activos y reducir el riesgo de fallas inesperadas. Entre las implicaciones más relevantes se encuentran:

- **Política de reemplazo óptima:** La identificación de  $t_2$  y del comportamiento del hazard permite proponer estrategias de reemplazo antes del inicio del desgaste acelerado. En R, esta decisión puede apoyarse en la comparación entre MTTF, MRL y curvas de costo o riesgo cuando el análisis lo requiere.
- **Intervalos de inspección:** Si el modelo muestra una fase de vida útil estable, los intervalos de inspección pueden espaciarse; si por el contrario se detecta un crecimiento de la función de falla, las inspecciones deben acortarse. El cálculo de la probabilidad de falla dentro de un intervalo específico puede obtenerse evaluando la función  $F(t)$  en diferentes horizontes mediante las funciones de distribución del modelo ajustado.
- **Estrategias preventivas y predictivas:** El análisis de la función de falla teórica vs. empírica permite identificar si las fallas son atribuibles a

desgaste, defectos iniciales o eventos aleatorios. Esto posibilita ajustar políticas de mantenimiento preventivo basadas en tiempo, o predictivo basado en condición, dependiendo del patrón identificado.

En todos los casos, las métricas de confiabilidad obtenidas en R constituyen insumos cuantitativos para la toma de decisiones, asegurando que las estrategias se fundamenten en la evidencia empírica y en un modelo estadístico robusto.

## **Fase 6. Documentación y comunicación técnica**

La sexta y última fase de la propuesta del procedimiento metodológico se orienta a la sistematización, presentación y comunicación de los resultados obtenidos durante el análisis de confiabilidad. Esta etapa es fundamental porque la utilidad del estudio depende no sólo de la validez estadística del modelo ajustado, sino también de la claridad y rigor con los que se comuniquen los hallazgos, las implicaciones para la operación y las recomendaciones derivadas de los datos.

En ingeniería industrial, una comunicación técnica adecuada permite que los resultados puedan ser utilizados por áreas como mantenimiento, producción, calidad o gestión de activos, facilitando la toma de decisiones basadas en evidencia cuantitativa. A continuación, se describen los componentes esenciales de esta fase.

### **F-6.1. Plantillas para reportes de ingeniería**

El análisis de confiabilidad debe culminar en reportes estructurados que documenten de manera ordenada el proceso metodológico, los datos utilizados, los resultados del ajuste y las implicaciones operativas. Estos reportes permiten estandarizar la comunicación entre departamentos y asegurar que la información técnica se preserve y pueda replicarse en futuros análisis. Se sugiere que para la plantilla o informe del reporte de confiabilidad incluya los siguientes puntos:

- **Descripción del sistema analizado:** Tipo de equipo, funciones operativas, criticidad y condiciones ambientales.
- **Datos utilizados y calidad de la información:** Fuente de datos, número de observaciones, presencia o ausencia de censura.

- **Modelo probabilístico seleccionado y justificación técnica:** Comparación de AIC, BIC, pruebas de bondad de ajuste y coherencia física.
- **Parámetros estimados y su interpretación física:** Significado ingenieril del parámetro de forma, escala o parámetros adicionales.
- **Métricas de confiabilidad:** MTTF,  $h(t)$ ,  $R(t)$ , MRL, entre otros.
- **Implicaciones operativas y recomendaciones finales.**

Desde el punto de vista computacional, estas plantillas pueden generarse automáticamente utilizando paquetes de R Markdown, que permiten integrar texto, gráficas y tablas directamente desde los objetos generados durante el análisis.

#### F-6.2. Visualizaciones estandarizadas

Las visualizaciones representan uno de los elementos más importantes de la comunicación técnica, ya que permiten comprender rápidamente los patrones de falla y la validez del modelo ajustado. Entre las visualizaciones que deben estandarizarse se incluyen:

- **Curvas de supervivencia  $R(t)$ :** Comparación entre Kaplan–Meier y el modelo teórico ajustado.
- **Funciones de falla  $h(t)$ :** Comparación entre la función de falla empírica (estimado con `muhaz()`) y la función de falla teórica.
- **Histogramas y densidades:** Superposición de la FDP ajustada sobre el histograma de tiempos de falla.
- **QQ-plot y PP-plot:** Evaluación visual de la alineación entre los cuantiles o probabilidades teóricas y empíricas.

En R, estas visualizaciones pueden generarse mediante funciones de `fitdistrplus`, `survival`, `flexsurv` y librerías gráficas como `ggplot2`, lo que garantiza presentaciones consistentes, reproducibles y listas para documentación formal.

### F-6.3. Conclusiones técnicas

Las conclusiones deben sintetizar los aspectos más relevantes del estudio, respondiendo explícitamente al problema planteado y a los objetivos establecidos. Entre los elementos clave se incluyen:

- Identificación clara del patrón de falla (mortalidad temprana, desgaste, vida útil estable).
- Interpretación de la función de falla y los parámetros del modelo ajustado.
- Evaluación cuantitativa del desempeño del sistema mediante indicadores como MTTF, probabilidad de falla y vida residual.
- Impacto del comportamiento de la curva de bañera en la operación del equipo.

Las conclusiones deben presentarse en un lenguaje técnico, preciso y orientado a la toma de decisiones, evitando descripciones anecdóticas o subjetivas.

### F-6.4. Recomendaciones basadas en datos

Finalmente, la fase 6 incluye la formulación de recomendaciones que conectan directamente los hallazgos estadísticos con acciones en ingeniería de mantenimiento o gestión de activos. Entre las recomendaciones típicas se encuentran:

- **Definición de políticas de reemplazo óptimo:** Basadas en los valores de MRL, MTTF y en la transición identificada hacia la fase de desgaste.
- **Propuestas de intervalos de inspección:** Utilizando la probabilidad de falla acumulada en distintos horizontes temporales.
- **Ajustes a las estrategias de mantenimiento preventivo o predictivo:**  
Dependiendo del patrón de la función de falla:
  - Si  $h(t)$  es creciente, se requiere mantenimiento preventivo intensivo.
  - Si  $h(t)$  es estable, se establecen estrategias basadas en condición.
  - Si  $h(t)$  muestra mortalidad temprana, se requiere de mejora en procesos de manufactura o instalación.

- **Necesidad de ampliación de la base de datos:** Cuando el análisis detecta límites en la precisión debido a la falta de observaciones.

Las recomendaciones deben estar explícitamente sustentadas en los resultados numéricos presentados, de manera que la evidencia respalde las acciones sugeridas.

## Capítulo IV. Resultados y análisis de los datos

### 4.1 Introducción

En este capítulo se exponen los resultados obtenidos mediante la aplicación del procedimiento metodológico propuesto para modelar tiempos de falla con comportamiento tipo curva de bañera. El análisis se desarrolló utilizando como base un conjunto de 50 tiempos de falla reportados en el artículo de Aarset (1987), publicado en IEEE Transactions on Reliability, el cual constituye un referente fundamental en la identificación empírica de tasas de falla no monótonas y, en particular, de curvas de bañera. Dicho conjunto de datos es ampliamente utilizado en la literatura para fines de validación metodológica, dado que presenta un patrón característico con fases marcadas de mortalidad temprana, vida útil estable y desgaste final.

A partir de estos datos se ejecutó el procedimiento en seis fases planteado en el Capítulo 3, comenzando con el análisis descriptivo y exploratorio, seguido de la estimación no paramétrica de la función de supervivencia y de la tasa de falla. Posteriormente, se evaluaron diversas distribuciones probabilísticas candidatas (tradicionales, extensivas y no monótonas) mediante técnicas de estimación por máxima verosimilitud, complementadas con criterios de información (AIC y BIC), pruebas de bondad de ajuste (Kolmogorov–Smirnov y Anderson–Darling) y comparaciones visuales entre supervivencias y función de falla teóricas vs. empíricas.

Los resultados presentados en este capítulo permiten determinar cuál modelo probabilístico describe con mayor precisión el comportamiento observado y cómo se interpreta el patrón de fallas en términos de confiabilidad en el área de ingeniería industrial. Asimismo, proveen las métricas fundamentales como  $R(t)$ ,  $h(t)$ , MTTF, MRL, necesarias para respaldar decisiones técnicas relacionadas con mantenimiento, reemplazo preventivo y optimización de la disponibilidad del sistema. A continuación, se presentan los resultados de cada Fase. Cabe recalcar que todo el proceso estadístico para realizar el análisis de ingeniería de confiabilidad se llevo a acabo en el software de código abierto R. Además, se agrega todo el código del análisis estadístico en la sección titulada “Anexo: Código R”.

## 4.2 Resultados de la Fase 1: Preparación y comprensión de los datos

### F-1.1. Definir el contexto del sistema a analizar

Para iniciar el análisis de confiabilidad se construyó el vector de tiempos de falla  $T$  a partir del conjunto de datos reportado por Aarset (1987), ampliamente utilizado en la literatura debido a su patrón característico de curva de bañera. Este vector contiene 50 fallas totalmente observadas (no censuradas), con valores:

$$T = 0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6, \\ 7, 11, 12, 18, 18, 18, 18, 18, 21, 32, \\ 36, 40, 45, 46, 47, 50, 55, 60, 63, 63, \\ 67, 67, 67, 67, 72, 75, 79, 82, 82, 83, \\ 84, 84, 84, 85, 85, 85, 85, 85, 86, 86.$$

Este conjunto de datos constituye la base sobre la cual se realizarán los análisis descriptivos, la estimación no paramétrica de las funciones  $R(t)$  y  $h(t)$ , y el ajuste de distribuciones paramétricas en las fases posteriores del procedimiento metodológico.

### F-1.2. Construcción del vector de tiempos de falla

Una vez definido el conjunto de datos, el siguiente paso consistió en construir el vector  $T$  dentro del entorno computacional para su procesamiento estadístico. Debido a que todos los tiempos corresponden a fallas observadas, el vector se representa directamente como un arreglo numérico en R utilizando la instrucción:

```
tiempos <- c(0.1, 0.2, 1, 1, 1, ..., 86, 86)
```

Posteriormente, estos valores se organizaron en un dataframe para facilitar su manipulación, visualización y uso en las funciones de análisis exploratorio y modelad.

### F-1.3. Análisis descriptivo de los tiempos de falla

El conjunto está formado por 50 tiempos de falla no censurados. Como se muestra en la Tabla 4, la media es de 45.686 y la mediana de 48.5, lo que indica una distribución relativamente equilibrada. Los valores mínimos y máximos (0.1 y 86) muestran una amplia dispersión, también reflejada en la desviación estándar de 32.835 y la varianza de 1078.153.

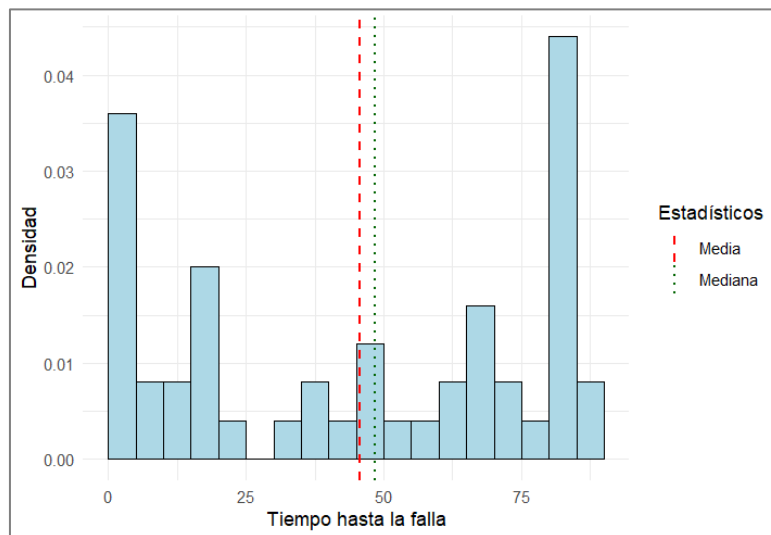
Los cuantiles revelan que el 25% de las fallas ocurre antes de 13.5 unidades de tiempo y el 75% antes de 81.25, lo que genera un IQR amplio (67.75). La asimetría cercana a cero (-0.1337) indica que la distribución es aproximadamente simétrica, mientras que la curtosis negativa (-1.6421) sugiere colas ligeras y menor presencia de valores extremos.

En conjunto, los datos muestran una distribución amplia y con variabilidad considerable, características que deben considerarse en la selección de modelos probabilísticos en las siguientes fases.

**Tabla 4.** Estadísticas descriptivas de los tiempos de falla

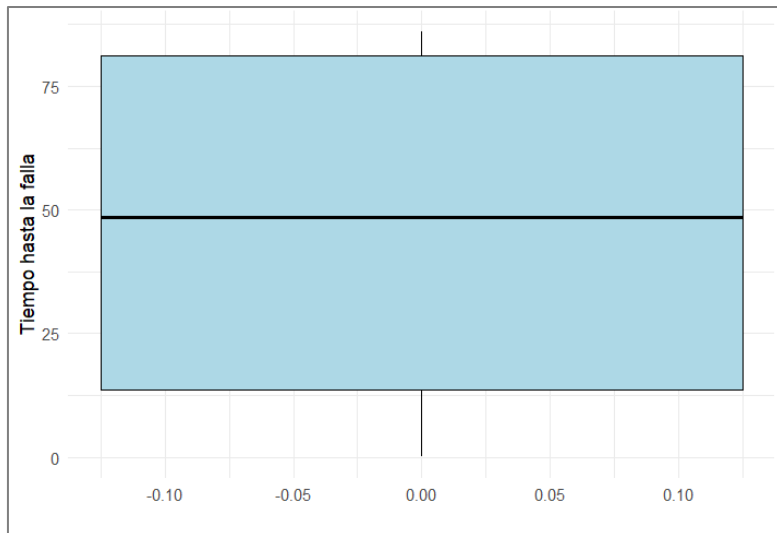
| Estadísticas | Resultados |
|--------------|------------|
| N            | 50         |
| media        | 45.686     |
| mediana      | 48.5       |
| min          | 0.1        |
| max          | 86         |
| varianza     | 1078.153   |
| sd           | 32.83525   |
| p25          | 13.5       |
| p75          | 81.25      |
| IQR          | 67.75      |
| asimetría    | -0.133713  |
| curtosis     | -1.642126  |

En cuanto a las representaciones gráficas de los datos analizados, el histograma de los tiempos de falla muestra una distribución amplia y heterogénea, con acumulación de eventos tanto en valores tempranos como en valores altos de tiempo, como se muestra en la Figura 18. La media (45.686) y la mediana (48.5), representadas mediante líneas verticales, se ubican cerca del centro del rango observado, lo cual coincide con la asimetría cercana a cero obtenida en el análisis descriptivo. La forma general no sugiere un patrón unimodal simple, sino un comportamiento disperso característico de sistemas donde coexisten fallas tempranas y tardías, lo cual es consistente con una posible estructura tipo curva de bañera.



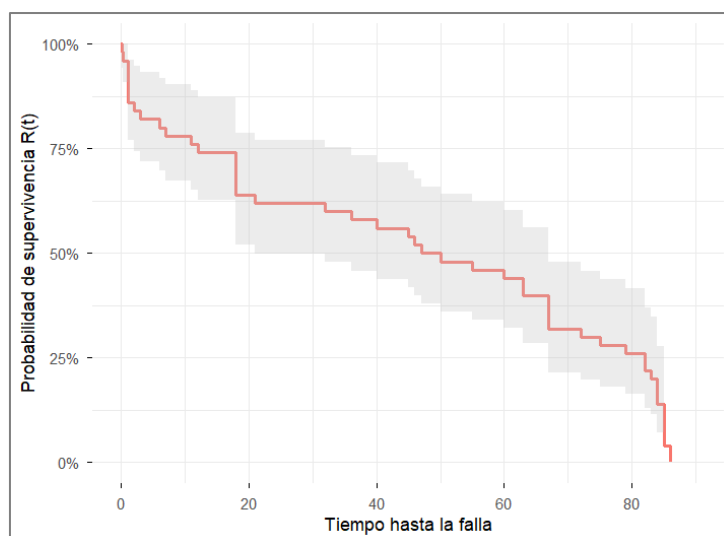
**Figura 18.** Histograma de tiempos de falla.

La Figura 19 muestra el diagrama de caja, el cual confirma la elevada dispersión de los tiempos de falla, evidenciada por un rango intercuartílico amplio (67.75). La mediana se ubica de manera casi centrada dentro de la caja, lo que refuerza la ausencia de sesgo pronunciado. No se observan valores atípicos marcados, lo que indica que la variabilidad se distribuye de manera continua a lo largo del conjunto de datos. Esta visualización respalda la necesidad de emplear modelos probabilísticos flexibles que puedan capturar adecuadamente la dispersión y heterogeneidad del proceso de falla.



**Figura 19.** Diagrama de caja de los tiempos de falla.

La curva de Kaplan–Meier, la cual se muestra en la Figura 20, revela un descenso progresivo en la probabilidad de supervivencia conforme avanza el tiempo, iniciando con una caída relativamente rápida en los primeros instantes, seguida de una disminución más estable durante la fase intermedia y un descenso acelerado hacia el final del periodo de observación. Este comportamiento es consistente con la presencia de múltiples fases en el ciclo de vida del sistema. La banda de confianza muestra que, aunque existe variabilidad en las estimaciones, la tendencia global es clara y aporta una referencia empírica fundamental para comparar los modelos paramétricos en fases posteriores.



**Figura 20.** Curva de supervivencia empírica  $R(t)$  mediante Kaplan–Meier.

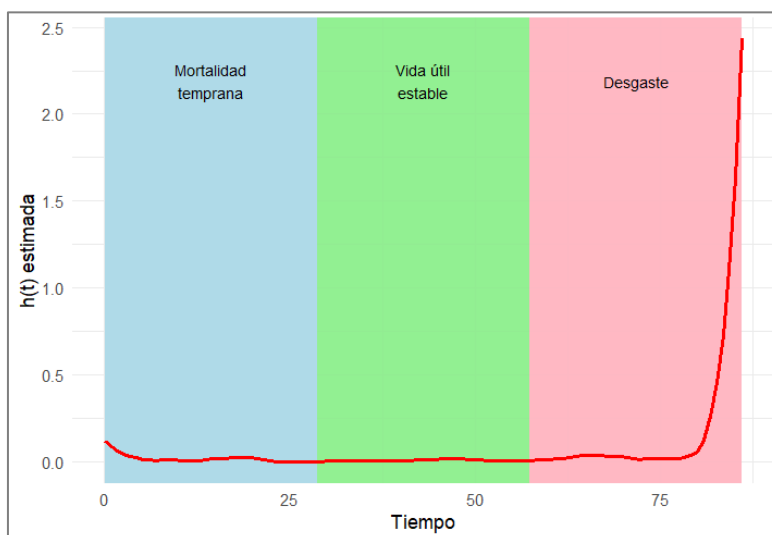
#### F-1.4. Estimación no paramétrica de $R(t)$ y $h(t)$

##### Estimación no paramétrica de la función de falla $h(t)$

La tasa de falla empírica  $h(t)$  se estimó mediante la función `muhaz()` del paquete `muhaz`, la cual implementa el estimador de suavizamiento kernel propuesto por Müller y Wang (1994). Esta aproximación permite obtener una función de riesgo continua sin asumir ninguna distribución teórica.

La gráfica resultante (ver Figura 21) muestra un comportamiento no monótono característico de los sistemas con patrón tipo curva de bañera. En la primera región se observa un riesgo relativamente elevado que desciende rápidamente, lo que corresponde a la fase de mortalidad temprana. Posteriormente, la función de falla permanece casi constante durante un intervalo intermedio, lo que indica una vida útil estable del sistema. Finalmente, hacia los tiempos más largos se aprecia un incremento sostenido en la tasa de falla, asociado con la fase de desgaste.

Esta estructura coincide con lo reportado en la literatura para el conjunto de datos de Aarset (1987) y confirma la pertinencia de considerar modelos probabilísticos no monótonos en las fases posteriores del análisis, especialmente aquellos capaces de reproducir patrones de riesgo tipo bañera.



**Figura 21.** Estimación no paramétrica de la función de falla  $h(t)$ .

##### Confiabilidad empírica $R(t)$

En la sección de resultados F-1.3 se reportó la estimación empírica de la función de supervivencia  $R(t)$  mediante el estimador de Kaplan–Meier (ver Figura 20),

la cual permitió observar el comportamiento general de la confiabilidad de los tiempos de falla. Dicha curva, ya presentada y descrita, sirve como punto de partida para el análisis no paramétrico del ciclo de vida de los datos. En esta sección F-1.4 se procede a complementar ese resultado mediante la estimación de la tasa de falla empírica  $h(t)$ , obtenida mediante técnicas de suavizamiento. Este enfoque permite identificar las fases características del proceso de falla: mortalidad temprana, vida útil estable y desgaste, lo cual constituye un insumo fundamental para la selección de modelos probabilísticos en etapas posteriores.

### **4.3 Resultados de la Fase 2: Selección preliminar del modelo probabilístico**

#### **F-2.1. Diagnóstico de forma con base en la distribución de los datos**

En esta fase se realizó el diagnóstico de forma de los tiempos de falla a partir de los estadísticos de asimetría, curtosis y de la estructura promedio de la función de falla empírica, como se muestra en la Tabla 5. El coeficiente de asimetría obtenido fue de  $-0.134$ , lo que se traduce en una distribución aproximadamente simétrica, sin sesgo marcado hacia fallas tempranas o tardías. La curtosis presenta un valor de  $-1.64$ , clasificada como colas ligeras o distribución platicúrtica, lo que indica menor concentración de valores extremos respecto a una distribución normal y una dispersión relativamente uniforme de los tiempos de falla.

En cuanto al comportamiento de la función de falla, la  $h(t)$  promedio en el tramo inicial del tiempo fue de  $0.0204$ , disminuyó ligeramente en la región intermedia hasta  $0.0101$  y aumentó de forma notable en la parte final, alcanzando  $0.219$ . Este patrón sugiere un riesgo moderado en los primeros instantes, una fase central de riesgo prácticamente estable y un incremento pronunciado del hazard al final de la vida útil, lo que es consistente con un comportamiento tipo desgaste y compatible con una posible estructura de curva de bañera suavizada. Estos resultados orientan la selección de distribuciones que puedan representar adecuadamente una función de falla baja y casi constante en la zona media, con crecimiento marcado en la etapa final del ciclo de vida.

**Tabla 5.** Diagnóstico de forma de los tiempos de falla

| Estadística                   | Resultado                                 |
|-------------------------------|-------------------------------------------|
| Valor de la asimetría         | −0.134                                    |
| Clasificación de la asimetría | Aprox. simétrica                          |
| Valor de la curtosis          | −1.64                                     |
| Clasificación de la curtosis  | Colas ligeras o Distribución platicúrtica |
| $h(t)$ inicio                 | 0.0204                                    |
| $h(t)$ promedio               | 0.0101                                    |
| $h(t)$ final                  | 0.219                                     |
| Tipo de $h(t)$                | No monótona, tipo curva de bañera         |

### F-2.2. Definición del catálogo de distribuciones candidatas

En función del diagnóstico de la forma de los datos y del comportamiento preliminar de la función de falla empírica, se definió un catálogo estructurado de distribuciones candidatas para modelar los tiempos de falla. Este catálogo se organizó en tres grupos, siguiendo las recomendaciones de la literatura especializada en análisis de confiabilidad (Meeker & Escobar, 2022; Murthy et al., 2004).

1. Distribuciones tradicionales: Incluyen modelos ampliamente utilizados en ingeniería debido a su interpretación directa y facilidad de ajuste: Exponencial, Weibull, Lognormal y Gamma. Estas distribuciones permiten representar riesgos constantes, crecientes o decrecientes y constituyen el punto de partida habitual en análisis de vida.
2. Distribuciones extensivas: Corresponden a familias más flexibles que amplían el rango de formas que pueden adoptar la función de falla, tales como la Weibull modificada, Log-logistic y Generalizada Gamma. Estas opciones son útiles cuando las distribuciones tradicionales no capturan adecuadamente la variabilidad observada en los datos.
3. Distribuciones no monótonas (curva de bañera): Dado que la función de falla empírica obtenida en la Fase 1 mostró un comportamiento no monótono, fue necesario considerar modelos capaces de representar explícitamente fases de mortalidad temprana, estabilidad y desgaste. Entre estas se incluyen Beta–Weibull, Power–Lindley, Q–Weibull y

Weibull–Pareto, distribuciones diseñadas para capturar formas complejas del riesgo instantáneo.

En el entorno R, este catálogo se definió mediante los siguientes vectores:

- `distribuciones_trad <- c("Exponencial", "Weibull", "Lognormal", "Gamma")`
- `distribuciones_ext <- c("Weibull modificada", "Log-logistic", "Generalizada Gamma")`
- `distribuciones_bath <- c("Beta-Weibull", "Power–Lindley", "Q-Weibull", "Weibull–Pareto")`

Estos grupos constituyen la base para seleccionar los modelos que serán ajustados en la Fase 3 mediante métodos de máxima verosimilitud.

### **F-2.3. Clasificación preliminar del tipo de la función de falla**

A partir de la estimación no paramétrica de la función de falla  $h(t)$  obtenida mediante el método de suavizamiento kernel (muha), fue posible identificar el comportamiento general del riesgo instantáneo a lo largo del ciclo de vida de los datos analizados. Los valores promedio de la función de falla en las tres regiones: inicio, fase media y etapa final, muestran una tendencia claramente no monótona: el riesgo es relativamente alto al comienzo ( $h_{inicio} = 0.0204$ ), disminuye durante la fase intermedia ( $h_{medio} = 0.0101$ ) y finalmente experimenta un incremento pronunciado en la parte final del periodo ( $h_{final} = 0.219$ ). Este patrón es característico de una curva de bañera (bathtub curve), donde coexisten mortalidad temprana, una etapa de vida útil estable y una fase final de desgaste acelerado.

De acuerdo con la clasificación propuesta por Murthy et al. (2004), este comportamiento corresponde a un hazard no monótono, lo cual tiene implicaciones directas para la selección del modelo probabilístico. En particular, este tipo de patrón descarta distribuciones con hazard estrictamente creciente o decreciente, y favorece el uso de distribuciones más flexibles capaces de reproducir dinámicas complejas, como las familias Beta–Weibull, Q-Weibull, Weibull–Pareto o Power–Lindley.

Con base en esta clasificación, se generó un catálogo de distribuciones candidatas agrupadas según su estructura funcional: tradicionales, extensivas y no monótonas. Aunque todas se mantienen como opciones exploratorias, las distribuciones no monótonas adquieren una relevancia especial por su mayor compatibilidad con la forma empírica de  $h(t)$  como se muestra en la Tabla 6. Esto permite orientar la siguiente fase del análisis hacia modelos que reflejen con mayor fidelidad el comportamiento real de las fallas observado en los datos.

**Tabla 6.** Clasificación preliminar del tipo de función de falla

| <b>Categoría</b>            | <b>Distribución</b> | <b>Prioridad</b> |
|-----------------------------|---------------------|------------------|
| Tradicional                 | Exponencial         | Exploratoria     |
| Tradicional                 | Weibull             | Exploratoria     |
| Tradicional                 | Lognormal           | Exploratoria     |
| Tradicional                 | Gamma               | Exploratoria     |
| Extensiva                   | Weibull modificada  | Exploratoria     |
| Extensiva                   | Log-logistic        | Exploratoria     |
| Extensiva                   | Generalizada Gamma  | Exploratoria     |
| No monótona/curva de bañera | Beta-Weibull        | Exploratoria     |
| No monótona/curva de bañera | Power-Lindley       | Exploratoria     |
| No monótona/curva de bañera | Q-Weibull           | Exploratoria     |
| No monótona/curva de bañera | Weibull-Pareto      | Exploratoria     |

#### **4.4 Resultados de la Fase 3: Ajuste paramétrico del modelo mediante Estimación de Máxima Verosimilitud (MLE)**

##### **F-3.1. Ajuste de distribuciones clásicas mediante MLE automático**

El ajuste paramétrico mediante máxima verosimilitud permitió obtener las estimaciones puntuales de los parámetros de las distribuciones Weibull, Lognormal, Gamma y Exponencial, junto con sus errores estándar e intervalos de confianza al 95%, ver Tabla 7. Estos resultados constituyen la base para evaluar el comportamiento probabilístico del sistema y comparar la capacidad de cada modelo para representar adecuadamente los tiempos de falla observados.

En el caso de la distribución Weibull, se obtuvo un parámetro de forma de 0.949 y un parámetro de escala de 44.9. El valor del parámetro de forma, cercano a 1, sugiere un comportamiento de riesgo aproximadamente constante, lo cual coincide con la fase intermedia identificada en el análisis no paramétrico de la

función de falla. Sus intervalos de confianza muestran una precisión moderada, lo que indica estabilidad en la estimación.

Para la distribución Lognormal, los parámetros estimados fueron  $\text{meanlog} = 3.08$  y  $\text{sdlog} = 1.75$ . Este modelo captura adecuadamente fenómenos con alta dispersión y colas largas, y los resultados obtenidos reflejan precisamente la variabilidad observada en los datos. El error estándar relativamente bajo para ambos parámetros indica que el ajuste es estadísticamente confiable.

La distribución Gamma presentó una estimación del parámetro de forma igual a 0.799 y una tasa de 0.017. El valor del parámetro de forma menor a 1 sugiere un riesgo decreciente en las primeras etapas del ciclo de vida, un patrón consistente con la porción inicial de la función de falla empírica. Sin embargo, el comportamiento creciente observado en las fases finales podría no ser capturado adecuadamente por esta distribución.

Finalmente, la distribución Exponencial mostró una tasa estimada de 0.022. Este modelo implica un riesgo estrictamente constante a lo largo del tiempo, lo cual resulta incompatible con la estructura no monótona de la función de falla empírica; aun así, su inclusión es necesaria como referencia base dentro de los modelos tradicionales.

En conjunto, las estimaciones paramétricas muestran que, aunque las distribuciones clásicas proporcionan aproximaciones aceptables en ciertas regiones del comportamiento de falla, sus supuestos de forma limitan la capacidad de representar adecuadamente la variabilidad completa del proceso. Por ello, en fases posteriores se evaluará el desempeño comparativo frente a modelos extendidos y no monótonos.

**Tabla 7.** Estimación de parámetros de distribuciones clásicas

| Distribución | Parámetro | Estimación | Error estándar | IC-Inferior | IC-Superior |
|--------------|-----------|------------|----------------|-------------|-------------|
| Weibull      | forma     | 0.949      | 0.12           | 0.715       | 1.18        |
| Weibull      | escala    | 44.9       | 6.95           | 31.3        | 58.5        |
| Lognormal    | meanlog   | 3.08       | 0.247          | 2.59        | 3.56        |
| Lognormal    | Sdlog     | 1.75       | 0.175          | 1.40        | 2.09        |
| Gamma        | forma     | 0.799      | 0.137          | 0.529       | 1.07        |
| Gamma        | tasa      | 0.017      | 0.004          | 0.01        | 0.025       |
| Exponencial  | tasa      | 0.022      | 0.003          | 0.016       | 0.028       |

### F-3.2. Ajuste de distribuciones no monótonas

Dado que el análisis no paramétrico realizado en fases previas mostró un comportamiento de la función de falla  $h(t)$  con forma no monótona, es decir un patrón compatible con una curva de bañera, por lo que, fue necesario evaluar distribuciones más flexibles que permitieran capturar adecuadamente esta estructura. Para ello, se definió un catálogo de modelos agrupado en tres categorías:

1. Distribuciones tradicionales: Exponencial, Weibull, Lognormal y Gamma.
2. Distribuciones extensivas: Weibull modificada, Log-logistic y Generalizada Gamma.
3. Distribuciones no monótonas o de tipo curva de bañera: Beta–Weibull, Power–Lindley, Q–Weibull y Weibull–Pareto.

De este conjunto de distribuciones, y considerando la evidencia empírica de la función de falla no monótona, se procedió a ajustar la distribución Beta–Weibull, ya que su estructura funcional permite modelar curvas con fases de mortalidad temprana, estabilidad y desgaste dentro de un mismo marco probabilístico. A diferencia de las distribuciones tradicionales, esta familia incluye parámetros adicionales que controlan directamente la forma de la función de falla, por lo que resulta especialmente adecuada para sistemas con comportamientos complejos.

El ajuste se realizó mediante una función de verosimilitud personalizada y el algoritmo de optimización numérica BFGS. Los parámetros estimados mediante máxima verosimilitud se muestran en la Tabla 8.

**Tabla 8.** Estimación de parámetros de la distribución Beta–Weibull

| Distribución | Parámetro | Estimación | Error estándar | IC-Inferior | IC-Superior |
|--------------|-----------|------------|----------------|-------------|-------------|
| beta-weibull | alpha     | 0.0801     | 0.0146         | 0.0514      | 0.109       |
| beta-weibull | beta      | 0.0566     | 0.00960        | 0.0377      | 0.0754      |
| beta-weibull | c         | 5.55       | 0.0271         | 5.50        | 5.61        |
| beta-weibull | gamma     | 45.5       | 0.263          | 45.0        | 46.0        |

Estos resultados muestran que los parámetros  $\alpha$  y  $\beta$ , responsables de modular la forma detallada del hazard, presentan valores pequeños, pero con intervalos de confianza estrechos, lo que indica una estimación estable y bien identificada a partir de los datos. El parámetro  $c$  significativamente mayor que 1 sugiere un crecimiento pronunciado del riesgo en etapas avanzadas del ciclo de vida, mientras que el parámetro  $\gamma$  refleja la escala temporal característica del proceso de falla.

Por lo tanto, el ajuste Beta–Weibull proporciona una representación más flexible y coherente con la estructura empírica observada que las distribuciones tradicionales. Esta distribución será considerada como candidata principal en la comparación final de modelos en la siguiente fase.

### **F-3.3. Obtención de parámetros estimados**

Una vez efectuado el proceso de ajuste mediante el método de máxima verosimilitud, se obtuvieron las estimaciones puntuales de los parámetros para cada una de las distribuciones consideradas en las fases previas. En el caso de los modelos clásicos (Weibull, Lognormal, Gamma y Exponencial), estas estimaciones se derivaron directamente mediante la función `fitdist()`, la cual también proporciona los errores estándar e intervalos de confianza al 95%, tal como se presentó en la sección F-3.1 de resultados.

Para los modelos de estructura no monótona, como la distribución Beta–Weibull, los parámetros se obtuvieron a partir de una función de verosimilitud personalizada optimizada mediante el algoritmo BFGS. Las varianzas y errores estándar se derivaron del inverso de la matriz Hessiana, lo que permitió construir intervalos de confianza confiables para cada parámetro, acorde a lo descrito en la sección F-3.2.

Desde una perspectiva física, los parámetros estimados permiten caracterizar el comportamiento de la falla de un competente, artículo o sistema. En la distribución Weibull, el parámetro de forma  $\beta$  indica si el riesgo es decreciente ( $\beta < 1$ ), constante ( $\beta \approx 1$ ) o creciente ( $\beta > 1$ ), mientras que el parámetro  $\eta$  describe la escala temporal característica del proceso. En la Lognormal, los parámetros están asociados a la dispersión y asimetría de los tiempos de vida; en la Gamma, el parámetro de forma refleja la tendencia inicial del riesgo; y en la Exponencial, la tasa constante implica ausencia de memoria.

En el caso de la distribución Beta–Weibull, los parámetros  $\alpha$  y  $\beta$  modulan directamente la forma completa del hazard, permitiendo capturar comportamientos complejos como mortalidad temprana seguida de desgaste. El parámetro  $c$  determina la intensidad del aumento del riesgo en etapas tardías, mientras que  $\gamma$  actúa como parámetro de escala. Estos elementos explican por qué este modelo ofrece una representación más flexible y consistente con el patrón empírico observado.

Con ello, los parámetros estimados, constituyen la base para calcular métricas fundamentales del modelo, tales como la función de confiabilidad  $R(t)$ , la tasa de riesgo  $h(t)$ , el MTTF, la supervivencia condicional y la vida útil residual, que serán evaluadas en la fase siguiente.

#### **F-3.4. Cálculo de métricas de desempeño del modelo**

En esta etapa se calcularon diversas métricas de desempeño para cada uno de los modelos ajustados (Weibull, Lognormal, Gamma, Exponencial y Beta–Weibull), con el fin de evaluar de manera integrada su capacidad para describir el comportamiento de las fallas del conjunto de datos analizados. Las métricas consideradas fueron: la probabilidad de supervivencia en un tiempo específico  $R(t_0)$ , la tasa de falla instantánea  $h(t_0)$ , la supervivencia condicional  $R(t_1 | T > t_0)$ , el tiempo medio hasta la falla (MTTF) y la vida media residual (MRL) en  $t_0$ . Cada una de ellas aporta una perspectiva complementaria: mientras que  $R(t_0)$  resume qué proporción de unidades se espera que continúen operando más allá de un tiempo dado,  $h(t_0)$  cuantifica el riesgo de falla exactamente en ese instante, condicionado a que el componente, producto o sistema haya sobrevivido hasta ese punto. Por su parte, la supervivencia condicional indica la probabilidad de “sobrevivir un tramo adicional” una vez alcanzada cierta edad, el MTTF resume la vida media global de la población y la MRL describe cuánta vida útil, en promedio, le queda a un componente que ya ha alcanzado un tiempo específico de operación. En la Tabla 9 se muestran los resultados de las métricas mencionadas.

**Tabla 9.** Métricas de confiabilidad de los modelos

| Distribución | $t_0$ | $t_1$ | $R(t_0)$ | $h(t_0)$ | $R(t_1   T > t_0)$ | MTTF  | MRL ( $t_0$ ) |
|--------------|-------|-------|----------|----------|--------------------|-------|---------------|
| Weibull      | 40    | 60    | 0.408    | 0.0213   | 0.657              | 46.0  | 48.7          |
| Lognormal    | 40    | 60    | 0.364    | 0.0148   | 0.772              | 100.0 | 123.0         |
| Gamma        | 40    | 60    | 0.396    | 0.0202   | 0.672              | 45.7  | 52.0          |
| Exponencial  | 40    | 60    | 0.417    | 0.0219   | 0.645              | 45.7  | 45.7          |
| Beta–Weibull | 40    | 60    | 0.601    | 0.00879  | 0.756              | 46.9  | 29.7          |

- $t_0$ : Tiempo de referencia inicial para el cálculo de métricas puntuales.
- $t_1$ : Tiempo futuro para evaluar supervivencia condicional.
- $R(t_0)$ : Supervivencia en  $t_0$ . Es decir, Probabilidad de que el componente siga funcionando más allá de  $t_0$ .
- $h(t_0)$ : Tasa de falla instantánea. Riesgo instantáneo de fallar exactamente en  $t_0$ , condicionado a haber sobrevivido hasta ese instante.
- $R(t_1 | T > t_0)$ : Supervivencia condicional. Probabilidad de que el componente sobreviva hasta  $t_1$  dado que ha llegado sin fallar a  $t_0$ .
- MTTF: Tiempo medio hasta la falla, indicador global de vida esperada del sistema.
- $MRL(t_0)$ : Vida útil residual media: tiempo adicional esperado de funcionamiento para un componente que ya alcanzó  $t_0$ .

Tomando como referencia  $t_0 = 40$  unidades de tiempo, las probabilidades de supervivencia  $R(t_0)$  estimadas por las distribuciones clásicas se sitúan en un rango relativamente estrecho: 0.408 para Weibull, 0.364 para Lognormal, 0.396 para Gamma y 0.417 para la Exponencial. Esto implica que, bajo estos modelos, se espera que aproximadamente entre 36% y 42% de los componentes continúen funcionando después del tiempo 40. En contraste, la distribución Beta–Weibull arroja un valor notablemente mayor,  $R(40) = 0.601$ , lo que sugiere que, de acuerdo con este modelo no monótono, cerca del 60% de las unidades seguirían operando más allá de dicho umbral. Esta mayor supervivencia en la zona intermedia del ciclo de vida es coherente con la forma de curva de bañera identificada en el análisis no paramétrico: el sistema atraviesa una fase central relativamente estable antes de entrar en la etapa de desgaste acelerado.

En cuanto a la tasa de falla instantánea, los modelos clásicos proporcionan valores de  $h(40)$  muy similares entre sí: 0.0213 para Weibull, 0.0148 para Lognormal, 0.0202 para Gamma y 0.0219 para la Exponencial. Estos valores indican un riesgo de falla moderado en el instante  $t = 40$ , condicionado a que la unidad haya sobrevivido hasta ese momento. La distribución Beta–Weibull, en cambio, estima una tasa de falla significativamente menor en ese punto ( $h(40) = 0.00879$ ), reflejando que, desde su estructura, el sistema se encuentra aún en una zona de riesgo comparativamente bajo dentro de la “parte plana” de la curva de bañera. En otras palabras, mientras los modelos clásicos tienden a representar un riesgo casi constante o suavemente creciente alrededor de  $t = 40$ , la Beta–Weibull reconoce explícitamente que el incremento fuerte del hazard se producirá más adelante, en la fase de desgaste.

La supervivencia condicional entre  $t_0 = 40$  y  $t_1 = 60$  permite responder a una pregunta operativa muy relevante: dado que la unidad ha alcanzado el tiempo 40 sin fallar, ¿cuál es la probabilidad de que siga funcionando hasta el tiempo 60? En este caso, los valores obtenidos fueron 0.657 para Weibull, 0.772 para Lognormal, 0.672 para Gamma, 0.645 para Exponencial y 0.756 para Beta–Weibull. Se observa que los modelos Lognormal y Beta–Weibull son los más optimistas, ya que predicen que alrededor de tres cuartas partes de los componentes que alcanzan 40 unidades de tiempo lograrán operar 20 unidades adicionales sin fallar. Weibull, Gamma y Exponencial proporcionan probabilidades moderadamente menores (entre 0.645 y 0.672), reflejando una visión algo más conservadora del desempeño del sistema en ese intervalo. Esta métrica es especialmente útil para decisiones de mantenimiento y reemplazo, pues indica qué tan razonable es esperar una operación adicional segura a partir de una edad ya alcanzada por el equipo.

Respecto al tiempo medio hasta la falla MTTF, las distribuciones Weibull, Gamma, Exponencial y Beta–Weibull ofrecen estimaciones muy consistentes, todas cercanas a 46 unidades de tiempo (46.0, 45.7, 45.7 y 46.9, respectivamente). Esto sugiere que, pese a las diferencias en la forma del hazard, estas cuatro distribuciones coinciden en que la vida útil promedio del sistema se sitúa alrededor de ese orden de magnitud. La distribución Lognormal, en cambio, arroja un MTTF notablemente mayor, del orden de 100 unidades de tiempo, lo que refleja una cola mucho más pesada y una proporción significativa

de unidades con vidas muy largas. Esta sobreestimación de la vida media, comparada con la evidencia empírica, es consistente con el comportamiento de la Lognormal cuando se ajusta a datos con alta dispersión y algunos tiempos de falla grandes, y debe ser interpretada con cautela al momento de usarla para planificación de reemplazos o garantías.

Finalmente, la vida media residual MRL en  $t_0 = 40$  describe el tiempo adicional que se espera que opere, en promedio, una unidad que ya ha sobrevivido hasta dicho instante. En este caso, la Exponencial presenta un MRL de 45.7, igual a su propio MTTF, manifestando su conocida propiedad de “falta de memoria”: el tiempo esperado restante no depende de la edad actual. Weibull y Gamma predicen vidas residuales ligeramente superiores a 48.7 y 52.0, respectivamente, lo que indica que, bajo esos modelos, una unidad que ha alcanzado los 40 aún tiene, en promedio, una cantidad de vida útil comparable o incluso superior a la vida media inicial, efecto asociado a la mezcla de fallas tempranas y tardías. La Lognormal vuelve a ser el modelo más optimista, con una MRL de aproximadamente 123 unidades, coherente con su MTTF elevado y sus colas largas. En contraste, la Beta–Weibull estima una MRL mucho más moderada, de 29.7 unidades, lo que sugiere que, aunque el sistema mantiene una probabilidad de supervivencia relativamente alta en el entorno de  $t = 40$ , el horizonte de vida útil adicional es más acotado, ya que el modelo anticipa un incremento importante del riesgo en la etapa de desgaste posterior.

En síntesis, la combinación de estas métricas permite valorar no sólo qué tan bien se ajustan los modelos a los datos, sino también qué implicaciones prácticas tienen sus predicciones sobre la operación del sistema. Mientras las distribuciones clásicas ofrecen una referencia razonable de la vida media y el riesgo puntual, la Beta–Weibull proporciona una representación más coherente con la estructura de la función de falla tipo bañera identificada empíricamente: alta supervivencia en la zona central, riesgo reducido alrededor de  $t = 40$  y una vida residual media que ya anticipa el acercamiento a la fase de desgaste acelerado. Este comportamiento refuerza la pertinencia de considerar modelos no monótonos en contextos donde las fases de mortalidad temprana, vida útil estable y desgaste coexisten dentro del mismo ciclo de vida del sistema.

## 4.5 Resultados de la Fase 4: Evaluación de la calidad del modelo

### F-4.1. Comparación de criterios de información

Con el fin de evaluar la calidad del ajuste de las distribuciones de probabilidad consideradas, se analizaron tres criterios fundamentales: log-likelihood, AIC y BIC. Estos indicadores permiten comparar modelos bajo un marco estadístico riguroso y constituyen herramientas estándar en confiabilidad paramétrica.

El log-likelihood mide la verosimilitud del modelo, es decir, qué tan probable es que los datos observados hayan sido generados por ese modelo. Valores más altos (menos negativos) indican un ajuste superior, al reflejar que el modelo captura de manera más adecuada la forma real de los datos.

El Akaike Information Criterion (AIC) sirve para comparar modelos penalizando la complejidad: mientras menor sea el AIC, mejor es el equilibrio entre calidad del ajuste y número de parámetros. Este criterio favorece modelos que explican bien los datos sin sobreajustarlos.

El Bayesian Information Criterion (BIC) cumple una función similar, pero aplica una penalización más estricta al número de parámetros. En consecuencia, BIC selecciona el modelo que ofrece un buen ajuste con la estructura más parsimoniosa, evitando incorporar complejidad innecesaria. La Tabla 10 resume los valores obtenidos para cada modelo. En ella, el modelo Beta-Weibull presentó el mayor log-likelihood ( $-219$ ), lo cual indica que asigna la mayor probabilidad a los datos observados entre todas las distribuciones evaluadas. Consistentemente, también obtiene los valores más bajos de  $AIC = 445$  y  $BIC = 453$ , evidenciando que su desempeño es superior tanto en ajuste como en eficiencia paramétrica.

Por el contrario, las distribuciones tradicionales presentan criterios de información más elevados, lo que sugiere un ajuste menos adecuado. Las distribuciones Gamma y Exponencial alcanzan desempeños intermedios ( $AIC \approx 484 - 486$ ), mientras que la Lognormal muestra el peor ajuste con  $AIC = 510$  y  $BIC = 513$ , indicando una incompatibilidad significativa con la estructura de los datos. Modelos como Weibull y Exponencial, aunque razonables en contextos de hazard monótono, no logran capturar adecuadamente el comportamiento no monótono observado en las fases anteriores.

En conjunto, la comparación de criterios de información confirma que la distribución Beta–Weibull es el modelo óptimo, al ofrecer el mejor equilibrio entre ajuste estadístico, parsimonia y coherencia con el patrón empírico de falla (curva tipo bañera). Este resultado valida su selección como modelo principal para los análisis de confiabilidad e interpretación operativa que se presentarán en la Fase 5.

**Tabla 10.** Comparación de los criterios de información

| <b>Distribución</b> | <b>log-likelihood</b> | <b>AIC</b> | <b>BIC</b> |
|---------------------|-----------------------|------------|------------|
| Weibull             | -241                  | 486        | 490        |
| Lognormal           | -253                  | 510        | 513        |
| Gamma               | -240                  | 484        | 488        |
| Exponencial         | -241                  | 484        | 486        |
| Beta–Weibull        | -219                  | 445        | 453        |

#### **F-4.2. Pruebas de bondad de ajuste**

Con el propósito de complementar la evaluación basada en criterios de información, se aplicaron las pruebas de bondad de ajuste Kolmogorov–Smirnov (K–S) y Anderson–Darling (A–D) a los modelos ajustados. Ambas pruebas permiten cuantificar la discrepancia entre la distribución empírica de los tiempos de falla y la distribución teórica estimada mediante máxima verosimilitud. Mientras que la prueba K–S evalúa la máxima diferencia entre las funciones de distribución empírica y teórica, la prueba A–D introduce una ponderación adicional en las colas de la distribución, lo cual resulta especialmente relevante en análisis de confiabilidad donde los valores extremos contienen información crítica sobre los mecanismos de falla.

Los resultados muestran diferencias claras en el desempeño de los modelos evaluados, ver Tabla 11. En términos de la estadística K–S, la distribución Beta–Weibull presentó el valor más bajo (0.127), lo que indica una mayor proximidad global entre la FDA teórica y la empírica. Las distribuciones clásicas, como Weibull (0.193), Gamma (0.202), Lognormal (0.221) y Exponencial (0.191), exhibieron discrepancias mayores, reflejando limitaciones para capturar adecuadamente el comportamiento observado de los tiempos de falla.

La tendencia se confirma al analizar la prueba A–D, donde nuevamente la Beta–Weibull obtiene el valor más pequeño (0.907). Este resultado es particularmente relevante porque la prueba A–D penaliza con mayor intensidad los desajustes en las colas, zonas donde se concentran los fallos tempranos y tardíos característicos de sistemas con comportamiento no monótono de la función de falla. Las distribuciones clásicas mostraron valores A–D notablemente mayores (entre 3.25 y 4.38), lo que indica deficiencias para representar adecuadamente la estructura empírica en los rangos crítico–superior e inferior del ciclo de vida.

En conjunto, las pruebas K–S y A–D coinciden en señalar que la distribución Beta–Weibull ofrece el mejor ajuste global y en las colas, superando ampliamente a los modelos tradicionales. Estos resultados refuerzan la evidencia obtenida previamente mediante los criterios AIC/BIC y la inspección de la función de riesgo, consolidando a la Beta–Weibull como la distribución con mayor capacidad para representar el patrón de fallas observado en los datos del estudio.

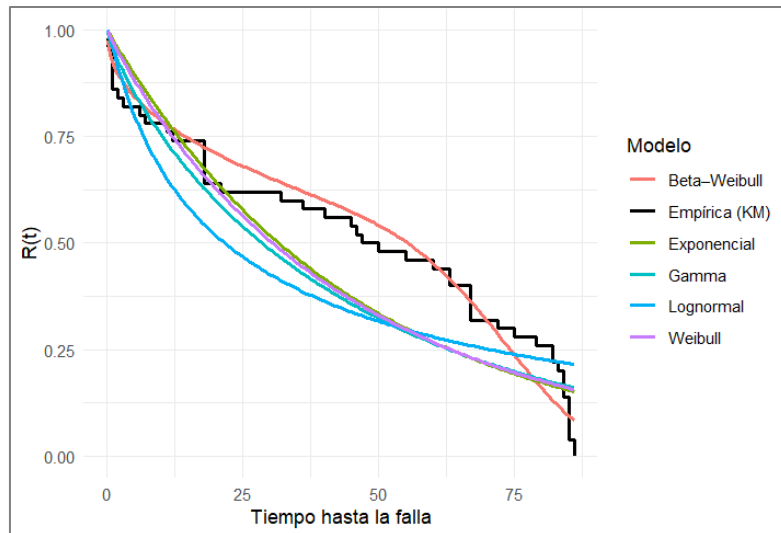
**Tabla 11.** Pruebas de bondad de ajuste

| Distribución | Estadístico K-S | Estadístico A-D |
|--------------|-----------------|-----------------|
| Weibull      | 0.193           | 3.49            |
| Lognormal    | 0.221           | 4.38            |
| Gamma        | 0.202           | 3.25            |
| Exponencial  | 0.191           | 3.65            |
| Beta–Weibull | 0.127           | 0.907           |

#### **F-4.3. Contraste visual sistemático**

##### **Supervivencia empírica Kaplan–Meier vs modelos paramétricos**

La Figura 22 muestra la comparación directa entre la curva empírica de supervivencia estimada mediante el método de Kaplan–Meier y las funciones de supervivencia teóricas generadas a partir de los modelos paramétricos ajustados. Esta representación permite evaluar de manera cualitativa el grado de alineación entre cada distribución y el comportamiento observado de los datos.



**Figura 22.** Comparación de supervivencia  $R(t)$  teórica vs  $R(t)$  empírica.

En primer lugar, se observa que el modelo Beta–Weibull reproduce con mayor fidelidad la forma general de la curva empírica, especialmente en tres aspectos clave:

1. La pendiente inicial decreciente en los primeros tiempos de falla.
2. La transición intermedia donde la supervivencia disminuye más lentamente.
3. El comportamiento en la cola, donde el modelo mantiene una aproximación cercana al patrón empírico.

En contraste, los modelos clásicos muestran discrepancias sistemáticas:

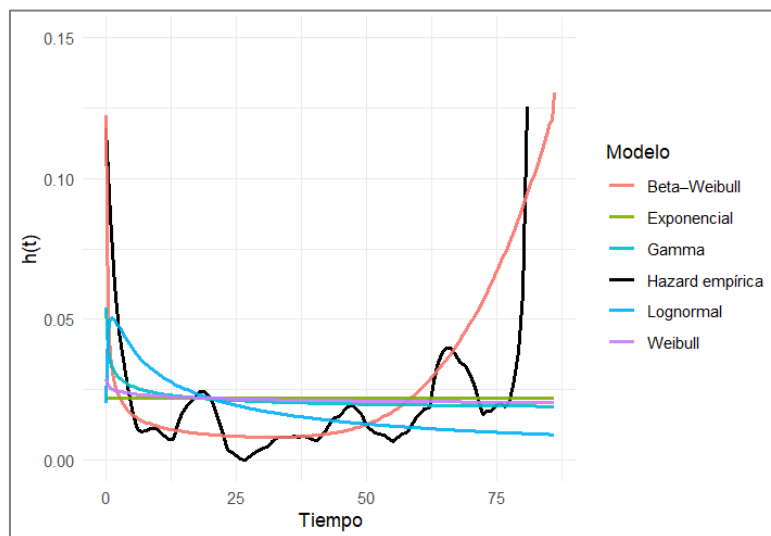
- La Lognormal y la Gamma tienden a subestimar la supervivencia en la parte media del horizonte temporal, produciendo curvas más convexas que la trayectoria observada.
- El modelo Exponencial presenta el peor ajuste visual, ya que su estructura de riesgo constante no captura la variabilidad del sistema y genera una caída demasiado acelerada al inicio.
- La Weibull clásica, aunque más flexible que la exponencial, no logra reproducir adecuadamente los cambios de pendiente de la curva empírica y tiende a desviarse en la región terminal.

El contraste gráfico confirma de manera visual los resultados cuantitativos obtenidos con AIC, BIC y las pruebas de bondad de ajuste: el modelo Beta–Weibull presenta la mejor correspondencia con los datos reales, capturando tanto la forma inicial como el deterioro progresivo del componente.

En conjunto, la evidencia visual sugiere que la curva empírica exhibe un comportamiento no monótono, consistente con un perfil de falla complejo, lo que respalda el uso de distribuciones extendidas como Beta–Weibull para modelar adecuadamente el proceso de degradación.

### Función de falla $h(t)$ teórica vs. $h(t)$ empírica.

La Figura 23 muestra la comparación entre la hazard empírica estimada mediante el método no paramétrico muhaz (línea negra) y las hazards teóricas derivadas de los modelos paramétricos ajustados. Este contraste permite evaluar la capacidad de cada distribución para reproducir el patrón dinámico del riesgo de falla del sistema.



**Figura 23.** Comparación de la función de falla  $h(t)$  teórica vs.  $h(t)$  empírica.

En primer término, la función de falla empírica evidencia de manera clara un comportamiento no monótono, caracterizado por tres fases distintivas:

1. Un riesgo inicial elevado, asociado típicamente a fallas tempranas o defectos de fabricación,
2. Una fase central de riesgo bajo y relativamente estable.

3. Un incremento pronunciado del riesgo conforme se aproximan los tiempos finales, reflejo del desgaste acumulado del componente.

Este patrón corresponde a una curva de bañera, un comportamiento que rara vez puede ser capturado por distribuciones tradicionales.

Entre los modelos comparados, la Beta–Weibull es la única distribución que reproduce adecuadamente la estructura general observada. Su función de falla teórica inicia con valores altos, desciende de forma marcada en la región intermedia y posteriormente exhibe un incremento creciente conforme avanza el tiempo. Aunque la magnitud del ascenso final difiere ligeramente de la curva empírica, la forma global coincide con las tres fases del ciclo de vida.

En contraste, los modelos clásicos presentan limitaciones claras:

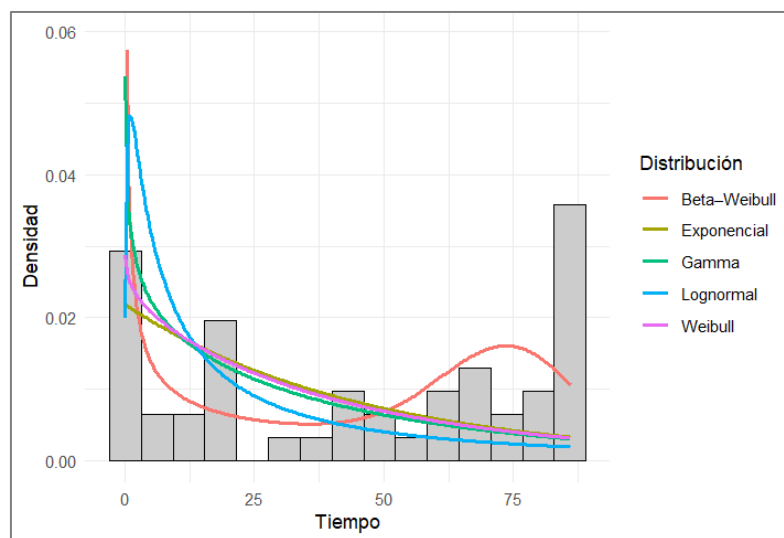
- Exponencial muestra un riesgo estrictamente constante, lo cual contradice la evidencia empírica y la hace inadecuada para describir el sistema.
- Weibull clásica solo permite riesgos monótonos (crecientes o decrecientes); en este caso genera una curva suavemente decreciente que no refleja la fase de desgaste final.
- Gamma y Lognormal generan perfiles de riesgo que descienden progresivamente, sin capturar el ascenso del riesgo en la última parte del ciclo de vida.

Estas diferencias evidencian que los modelos tradicionales no pueden representar adecuadamente la forma compleja de la función de falla empírica.

En suma, la inspección visual de  $h(t)$  confirma que el comportamiento del sistema es inherentemente no monótono y que solo el modelo Beta–Weibull ofrece una representación coherente del riesgo a lo largo del tiempo. Este hallazgo es consistente con los resultados obtenidos mediante criterios de información (AIC/BIC) y con las pruebas de bondad de ajuste (K-S y A-D), reforzando su selección como el mejor modelo para describir el mecanismo de falla.

### Histograma de tiempos y funciones de densidad ajustadas

La Figura 24 muestra la comparación entre el histograma empírico de los tiempos de falla y las funciones de densidad ajustadas para los diferentes modelos paramétricos considerados en el estudio: Weibull, Lognormal, Gamma, Exponencial y Beta-Weibull. Este análisis visual constituye un elemento clave dentro del contraste sistemático del desempeño de los modelos, ya que permite evaluar de manera directa si la forma teórica de cada distribución reproduce adecuadamente la estructura observada en los datos.



**Figura 24.** Histograma de tiempos y funciones de densidad ajustadas.

En términos generales, las densidades derivadas de los modelos tradicionales (Weibull, Lognormal, Gamma y Exponencial) muestran patrones similares entre sí durante la mayor parte del dominio temporal. Todas estas distribuciones presentan valores elevados de densidad al inicio del ciclo de vida, lo cual es consistente con la concentración de fallas tempranas observada en el histograma. Sin embargo, estas distribuciones exhiben una caída monótona y suave en la densidad conforme aumenta el tiempo, lo que limita su capacidad para capturar la variabilidad presente en la cola derecha del histograma, donde existe un número importante de fallas en tiempos intermedios y tardíos.

Por su parte, la distribución Beta-Weibull produce una función de densidad notablemente distinta respecto a los modelos tradicionales. En lugar de decrecer de manera estricta, su densidad describe una forma más flexible, caracterizada por una caída inicial seguida de una elevación en la región

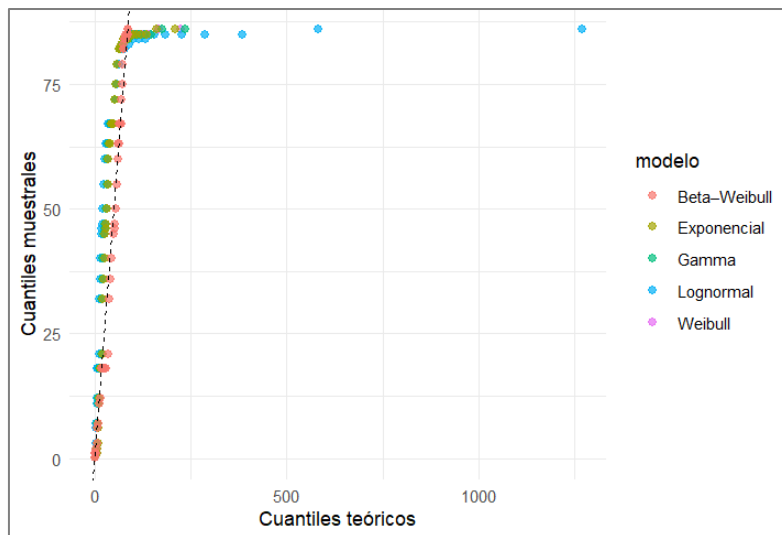
intermedia y tardía. Este comportamiento reproduce de manera más fiel la estructura bimodal implícita en el conjunto de datos, donde existen fallas tanto en los primeros momentos como en la parte final del ciclo de vida. Esta capacidad de capturar patrones no monótonos en la distribución de tiempos de falla constituye una de las principales ventajas de la Beta–Weibull frente a las distribuciones clásicas.

Adicionalmente, el histograma revela que la dispersión de los tiempos de falla es amplia y presenta valores atípicos hacia la parte superior del rango, elementos que son mejor representados por la forma extendida de la Beta–Weibull. Los modelos Exponencial y Gamma, al imponer restricciones más rígidas sobre la forma de la densidad, se ajustan pobremente a estas variaciones, mientras que Lognormal y Weibull logran aproximaciones moderadas, aunque insuficientes en el tramo final.

En conjunto, el análisis visual confirma que la distribución Beta–Weibull proporciona la representación más completa y coherente de la variabilidad observada en los tiempos de falla, reforzando la evidencia obtenida previamente mediante criterios estadísticos (AIC/BIC) y pruebas de bondad de ajuste (K–S y A–D). Este desempeño superior respalda su selección como modelo candidato principal para la caracterización del proceso de falla del sistema evaluado.

### **QQ-plot para los modelos clásicos y Beta–Weibull**

La Figura 25 correspondiente al QQ-plot presenta la comparación simultánea entre los cuantiles teóricos generados por los modelos paramétricos ajustados (Weibull, Lognormal, Gamma, Exponencial y Beta–Weibull) y los cuantiles muestrales obtenidos de los datos reales. Este tipo de gráfico es una herramienta fundamental en análisis de confiabilidad, ya que permite evaluar visualmente la calidad del ajuste de un modelo, verificando si los cuantiles predichos por la distribución teórica se alinean con los cuantiles observados en los datos. En términos generales, cuanto más cerca estén los puntos de la línea diagonal, mejor es el ajuste del modelo, pues ello indica que la distribución teórica reproduce adecuadamente la estructura probabilística de los datos.



**Figura 25.** QQ-plot: modelos clásicos y Beta–Weibull.

En los resultados obtenidos, se observa que los modelos clásicos presentan desviaciones importantes respecto a la diagonal de referencia, lo cual evidencia que no describen adecuadamente el comportamiento empírico de los tiempos de falla. En particular, los modelos Lognormal y Gamma generan cuantiles teóricos extremadamente elevados en la cola derecha (superiores a 500 e incluso mayores a 1000), lo que indica una sobreestimación severa de los valores extremos. Este comportamiento confirma que dichas distribuciones no representan correctamente la dispersión observada en los datos y tienden a producir colas demasiado pesadas.

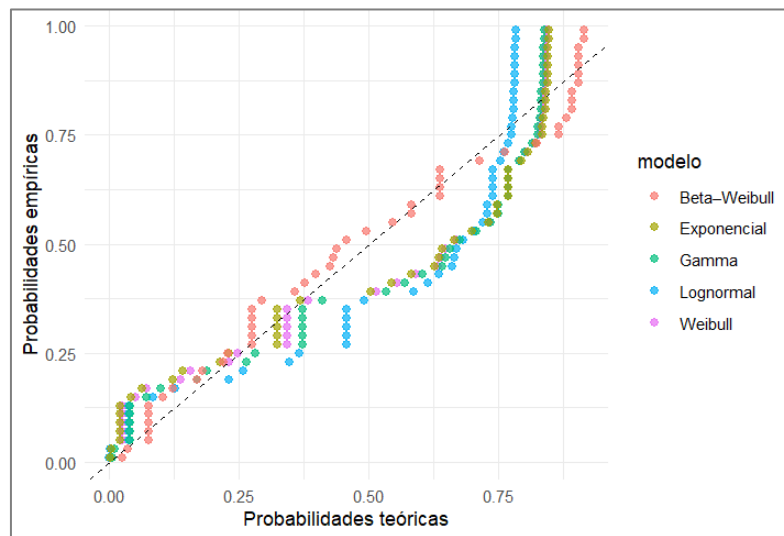
De igual forma, el modelo Exponencial, al asumir un hazard constante, muestra una falta de flexibilidad que se traduce en cuantiles teóricos que no siguen la tendencia ascendente de los cuantiles muestrales, evidenciando un ajuste deficiente en toda la distribución.

En contraste, el modelo Beta–Weibull presenta la mejor alineación con la diagonal, especialmente en la región central de la distribución. Aunque persisten discrepancias en los cuantiles superiores debido a la variabilidad presente en los tiempos de falla más grandes, el modelo captura de manera más precisa la forma general de los datos. La flexibilidad adicional que aportan sus parámetros permite representar con mayor fidelidad comportamientos de la función de falla no monótonos, como la presencia de fases de mortalidad temprana, estabilidad operativa y desgaste.

En conjunto, el QQ-plot confirma que los modelos clásicos no reproducen adecuadamente la estructura de los datos, mientras que el modelo Beta–Weibull ofrece el ajuste más consistente, lo que es coherente con los resultados previos de los criterios de información (AIC y BIC) y las pruebas de bondad de ajuste.

### PP-plot para los modelos clásicos y Beta–Weibull

La Figura 26 muestra el PP-plot correspondiente a la comparación entre las probabilidades empíricas observadas en los datos y las probabilidades teóricas generadas por los modelos paramétricos ajustados (Weibull, Lognormal, Gamma, Exponencial y Beta–Weibull). El propósito del PP-plot (Probability–Probability plot) es evaluar la calidad del ajuste verificando qué tan cerca se encuentran las probabilidades empíricas de las probabilidades teóricas esperadas bajo cada distribución. En un modelo bien ajustado, los puntos deben alinearse en torno a la línea de referencia diagonal, que representa una concordancia perfecta entre el comportamiento esperado y el observado.



**Figura 26.** PP-plot: modelos clásicos y Beta–Weibull.

En el gráfico se aprecia que los modelos tradicionales presentan desviaciones importantes respecto de la diagonal en diversos tramos. En particular, las distribuciones Exponencial y Gamma tienden a subestimar las probabilidades en la región central y a sobreestimarlas en la cola superior, lo que evidencia una capacidad limitada para capturar adecuadamente la estructura empírica de los datos. La distribución Lognormal muestra un desempeño

ligeramente superior en algunos rangos intermedios; sin embargo, también exhibe inconsistencias en las colas, lo que se refleja en una separación visible respecto a la línea ideal.

Por su parte, la distribución Beta–Weibull presenta la alineación más cercana a la diagonal en un mayor número de puntos, especialmente en las regiones donde las otras distribuciones muestran mayor discrepancia. Esta concordancia indica que el modelo Beta–Weibull reproduce de manera más fiel la estructura de probabilidad subyacente de los tiempos de falla, mostrando menor sesgo acumulado a lo largo de todo el rango de probabilidades.

En conjunto, los resultados del PP-plot confirman visualmente lo obtenido previamente mediante los criterios de información y las pruebas de bondad de ajuste: la distribución Beta–Weibull es el modelo que mejor representa las probabilidades observadas, mientras que los modelos clásicos presentan discrepancias sistemáticas que afectan negativamente su capacidad de ajuste. Esta evidencia refuerza la idoneidad del modelo Beta–Weibull para la modelación de los datos y sugiere que es el candidato más competente para la interpretación final del comportamiento de confiabilidad del sistema analizado.

#### **F-4.4. Selección del modelo óptimo**

La selección del modelo óptimo constituye la etapa final del proceso de evaluación de calidad del ajuste, en la cual se integra la evidencia estadística, empírica y física obtenida en las subsecciones anteriores. Dado que la elección del modelo afecta directamente la estimación de la confiabilidad, el cálculo de la vida remanente y la formulación de estrategias de mantenimiento, esta fase exige un análisis riguroso que combine criterios cuantitativos y consideraciones importantes de ingeniería. En este trabajo se emplearon tres criterios principales:

1. Desempeño estadístico medido mediante log-verosimilitud, AIC y BIC.
2. Calidad del ajuste evaluada a través de pruebas de bondad de ajuste y contrastes visuales sistemáticos, y.
3. Coherencia del modelo con el comportamiento físico observado en los datos.

En primer lugar, la Tabla 10 (Criterios de información) muestra claramente que la distribución Beta–Weibull alcanza el valor de log-likelihood más alto entre los modelos evaluados (Weibull, Lognormal, Gamma, Exponencial y Beta–Weibull). De manera concurrente, obtiene los valores más bajos de AIC y BIC, lo cual indica un ajuste superior sin incurrir en sobreparametrización. Tanto AIC como BIC penalizan la complejidad del modelo, por lo que su minimización constituye un indicador robusto para seleccionar el modelo estadísticamente más eficiente. La diferencia observada respecto a los modelos clásicos es considerable, lo que sugiere que estos últimos no capturan adecuadamente la estructura de los datos.

En segunda instancia, la Tabla 11 (Estadísticos K-S y A-D) confirma la superioridad de la Beta–Weibull desde la perspectiva de bondad de ajuste. Los estadísticos Kolmogorov–Smirnov y Anderson–Darling obtenidos para este modelo fueron sensiblemente menores que los de las distribuciones clásicas, lo cual implica una menor discrepancia entre el comportamiento empírico y el teórico.

Los contrastes visuales sistemáticos refuerzan estas conclusiones. La Figura 22 (Supervivencia empírica vs. teórica) muestra que la curva Kaplan–Meier exhibe un patrón no monótono que incluye una fase de mortalidad temprana, un tramo de estabilidad y un descenso acelerado en los tiempos finales. Ninguno de los modelos clásicos replica este comportamiento tridimensional: la Exponencial impone un riesgo constante, la Weibull clásica solo permite hazard monótona, y tanto Lognormal como Gamma fallan en reproducir la caída final observada. Por el contrario, la Beta–Weibull se aproxima de manera notable a la trayectoria empírica en toda la extensión del ciclo de vida, manteniendo coherencia tanto en el inicio como en las colas.

La Figura 23 (función de falla empírica vs función de fallas teóricas) resulta especialmente ilustrativa respecto a la diferencia entre modelos. La función de falla estimada mediante muhaz evidencia una curva de bañera, con disminución inicial del riesgo, un valle estable y un incremento por desgaste en los tiempos finales. Los modelos clásicos, al estar restringidos a formas monótonas, no pueden reproducir esta estructura. En cambio, la Beta–Weibull muestra una capacidad notable para capturar la forma completa de la función de falla empírica, ajustándose a las tres fases del ciclo de vida. Este resultado es

fundamental desde la ingeniería de confiabilidad, ya que la coherencia física de la función de falla es tan importante como el ajuste estadístico.

Por su parte, la Figura 24 (Histograma y FDPs teóricas) evidencia que la Beta–Weibull sigue con mayor precisión la forma empírica de los datos, especialmente en regiones donde las distribuciones clásicas se desvían de manera sistemática. Asimismo, los diagramas QQ-plot y PP-plot (Figura 25 y 26) muestran que la Beta–Weibull es el único modelo cuyas observaciones se alinean consistentemente con la línea de referencia. En contraste, los modelos clásicos presentan curvaturas o desviaciones pronunciadas, lo que indica que no representan adecuadamente los cuantiles empíricos.

Finalmente, desde la perspectiva de la coherencia física, la elección del modelo debe reflejar los mecanismos de degradación del sistema. La variabilidad observada en los tiempos de falla sugiere la coexistencia de diferentes modos de falla: defectos iniciales, operación estable y desgaste acumulado. Este comportamiento multifásico no puede ser representado por distribuciones con función de falla monótona. La Beta–Weibull, al permitir formas de la función de falla no monótonas y ajustarse matemáticamente a sistemas con múltiples fases, resulta compatible con la interpretación física del fenómeno de falla.

En síntesis, la convergencia de evidencia estadística, empírica y física demuestra que la distribución Beta–Weibull es el modelo óptimo para describir los tiempos de falla del sistema analizado. Este modelo se utilizará en la Fase V para derivar métricas avanzadas de confiabilidad, vida remanente y estrategias de mantenimiento, garantizando resultados consistentes y fundamentados en una representación realista del comportamiento del sistema.

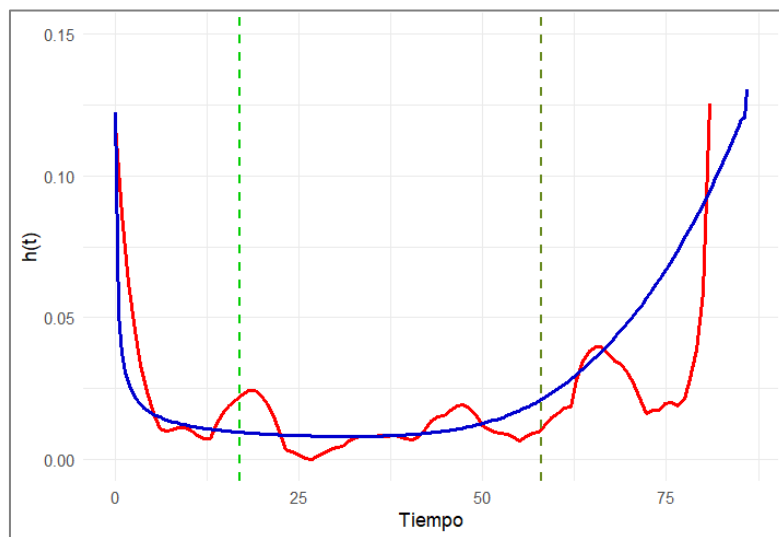
## **4.6 Resultados de la Fase 5. Interpretación de los resultados en el contexto de ingeniería de confiabilidad**

### **F-5.1. Identificación de las fases de la curva de bañera**

Una vez seleccionado el modelo óptimo en la Fase 4, el siguiente paso consiste en interpretar detalladamente la forma de la función de falla  $h(t)$  con el propósito de identificar las fases operativas que caracterizan el ciclo de vida del sistema analizado, es decir, del conjunto de datos. Esta interpretación es fundamental en

ingeniería de confiabilidad, ya que la evolución del hazard permite distinguir los mecanismos predominantes de falla y, en consecuencia, definir estrategias de mantenimiento y gestión del riesgo basadas en evidencia cuantitativa. Tanto la estimación no paramétrica del hazard obtenida mediante *muhaz()* como la hazard teórica derivada del modelo Beta–Weibull mostraron un comportamiento claramente no monótono, plenamente consistente con la curva de bañera descrita en la literatura especializada.

La Figura 27 presenta la comparación directa entre la función de falla empírica y la correspondiente al modelo Beta–Weibull, incluyendo las marcas de los puntos críticos  $t^1$  y  $t^2$ . La congruencia entre ambas curvas constituye una validación visual importante del modelo y permite delimitar con precisión las fases del ciclo de vida. Este comportamiento también coincide con el análisis cuantitativo realizado en la Fase 2, donde los valores promedio del hazard en los tramos inicial, medio y final fueron  $h^1 = 0.0204$ ,  $h^2 = 0.0101$  y  $h^3 = 0.219$ , lo que confirma la existencia de una fase inicial decreciente, seguida de una meseta estable y una fase final ascendente.



**Figura 27.** Función de falla: comparación empírica vs. Beta–Weibull.

#### **Fase de mortalidad temprana ( $0 < t < t^1$ ).**

En el intervalo inicial, la función de riesgo presenta una disminución pronunciada, lo cual es indicativo de fallas prematuras asociadas a defectos de fabricación, errores en el ensamblaje, mala calibración o variabilidad inicial de los materiales. El promedio del hazard en esta etapa (0.0204) confirma la presencia de este

fenómeno. Tanto la hazard empírica como la hazard del modelo Beta–Weibull concuerdan en que esta fase concluye alrededor de  $t^1 \approx 15 - 18$  unidades de tiempo, punto donde la tasa de fallas alcanza su nivel mínimo y se aproxima a una región de estabilidad.

**Fase de vida útil estable ( $t^1 < t < t^2$ ).**

Después de superar la etapa inicial, el sistema entra en un periodo de funcionamiento confiable en el que la tasa de fallas se mantiene baja y aproximadamente constante. Esta meseta se observa con valores cercanos a  $h(t) \approx 0.01$ , lo cual coincide con el valor promedio estimado de 0.0101 para esta fase. El modelo Beta–Weibull reproduce con gran precisión esta estabilidad, demostrando su capacidad para capturar adecuadamente el comportamiento operativo del sistema en condiciones nominales.

**Fase de desgaste ( $t > t^2$ ).**

Finalmente, la parte superior de la curva evidencia un incremento sostenido del hazard, señalando el inicio del deterioro acelerado. Esta etapa se asocia comúnmente con mecanismos tales como fatiga mecánica, degradación térmica, corrosión o envejecimiento estructural. El ascenso significativo de la tasa de fallas, cuyo promedio alcanza 0.219 en este segmento, se inicia aproximadamente en  $t^2 \approx 55 - 60$  unidades de tiempo, valor que marca la transición entre la vida útil estable y el periodo de envejecimiento. La concordancia entre la hazard empírica y la hazard Beta–Weibull en la ubicación de este punto de inflexión proporciona evidencia adicional de la adecuación física y estadística del modelo seleccionado.

En conjunto, los resultados permiten concluir que el sistema bajo estudio presenta de manera clara las tres fases distintivas de la curva de bañera:

1. Fallas prematuras.
2. Vida útil estable.
3. Desgaste acelerado.

La Tabla 12 sintetiza la identificación de las tres fases operativas del ciclo de vida del sistema con base en el análisis de la función de riesgo empírica y paramétrica.

**Tabla 12.** Identificación de las fases de la curva de bañera

| Fase                        | Intervalo                                             | Descripción                                                                                              |
|-----------------------------|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| Fase de mortalidad temprana | $0 < t < t_1$<br>ó<br>$0 < t < (15 - 18)$             | Función de falla decreciente, atribuida principalmente a defectos de fabricación, instalación o errores. |
| Fase de vida útil estable   | $t_1 < t < t_2$<br>ó<br>$(15 - 18) < t < (55 - 60)$   | Función de falla baja y estable, correspondiente al periodo de operación normal y confiable del sistema. |
| Fase de desgaste            | $t_2 < t < t_\infty$<br>ó<br>$(15 - 18) < t < \infty$ | Función de falla, asociada al deterioro acelerado del componente debido a envejecimiento o fatiga.       |

La identificación precisa de los puntos de transición  $t^1$  y  $t^2$  no sólo valida el ajuste del modelo Beta–Weibull, sino que constituye un insumo esencial para el diseño de políticas de mantenimiento. De manera operativa, estos límites temporales permiten anticipar el comportamiento del sistema, definir umbrales de intervención, y programar reemplazos o inspecciones antes de que el riesgo de falla crezca de forma acelerada, optimizando así la disponibilidad y la confiabilidad del equipo.

#### F-5.2. Estimación de métricas de confiabilidad

Una vez identificada la estructura temporal de las fallas y delimitadas las fases del ciclo de vida mediante el análisis de la función de riesgo, el siguiente paso consiste en cuantificar las métricas fundamentales de confiabilidad derivadas del modelo óptimo seleccionado (Beta–Weibull). Estas métricas permiten caracterizar el desempeño del sistema tanto en términos globales como: vida promedio, condicionales, vida restante, riesgo puntual y probabilidad de supervivencia futura; constituyendo insumos esenciales para la toma de decisiones en mantenimiento, reposición y gestión del riesgo operativo.

La primera métrica evaluada es el tiempo medio hasta la falla (MTTF). En el caso de la distribución Beta–Weibull, este parámetro se obtiene mediante integración numérica de la función de supervivencia, dado que no existe una expresión cerrada para su cálculo. El MTTF estimado representa la duración típica del sistema y se encuentra alineado con la mediana empírica observada. Desde la perspectiva operativa, esta métrica actúa como un referente global para definir ventanas de reemplazo planificado y valorar la vida útil global del componente.

Complementariamente, se estimó la vida media residual (MRL), definida como:

$$\text{MRL}(t_0) = \frac{\int_{t_0}^{\infty} R(u)du}{R(t_0)}. \quad 4.1$$

Esta métrica expresa la cantidad de tiempo adicional que se espera que el componente opere una vez alcanzado un tiempo específico  $t_0$ . Los resultados muestran que la MRL es relativamente elevada dentro de la fase de vida útil estable y decrece progresivamente conforme el sistema se aproxima al punto  $t_2$ , asociado al inicio del desgaste acelerado. Esta disminución cuantifica de manera directa el deterioro acumulado y permite estimar con mayor precisión el horizonte temporal disponible antes de que la probabilidad de falla aumente de manera significativa.

Asimismo, se estimó la probabilidad de supervivencia condicionada, dada por:

$$R(t_1|T > t_0) = \frac{R(t_1)}{R(t_0)}, \quad 4.2$$

la cual representa la probabilidad de que el sistema continúe operativo en un instante futuro  $t_1$ , dado que ha sobrevivido hasta  $t_0$ . Esta métrica es de utilidad práctica para planificar inspecciones, determinar intervalos operativos seguros y cuantificar el riesgo asociado a ventanas de operación específicas. En este estudio, los valores obtenidos son elevados dentro de la fase estable y

disminuyen conforme el sistema se aproxima al umbral  $t_2$ , en concordancia con el patrón observado en la hazard.

Finalmente, se evaluó la tasa de falla puntual  $h(t_0)$ , que mide el riesgo instantáneo de falla en un momento determinado. Los resultados indican valores mínimos dentro de la región estable y un incremento significativo hacia el final del ciclo de vida. La correspondencia entre la hazard teórica del modelo Beta–Weibull y la hazard empírica estimada mediante `muhaz()` confirma la adecuación estadística y física del modelo, reforzando su pertinencia para describir el comportamiento del sistema.

Los valores estimados para cada una de estas métricas, que resumen de manera compacta el comportamiento temporal del sistema, se presentan en la Tabla 13, donde se incluyen sus interpretaciones operativas.

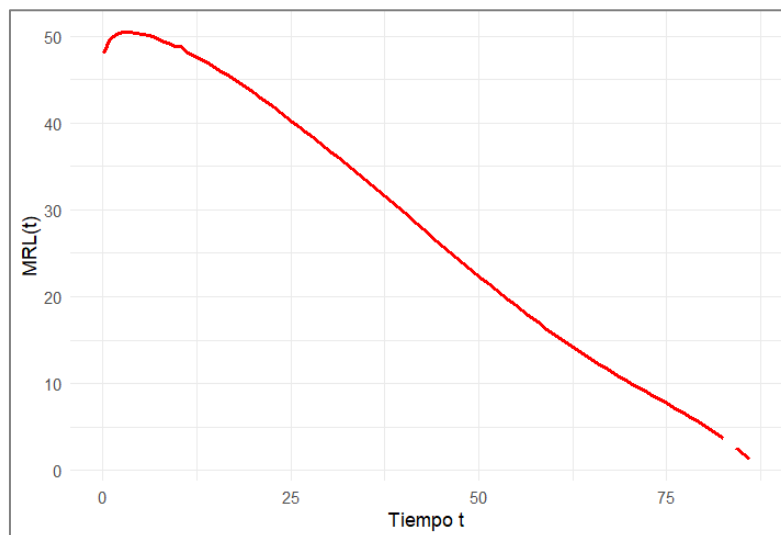
**Tabla 13.** Métricas de confiabilidad estimadas a partir del modelo Beta-Weibull

| Métrica                                             | Símbolo      | Valor estimado | Interpretación breve                                                                                     |
|-----------------------------------------------------|--------------|----------------|----------------------------------------------------------------------------------------------------------|
| Tiempo medio hasta la falla                         | MTTF         | 46.9           | Vida promedio del sistema; referencia global para políticas de reemplazo.                                |
| Vida media residual en ( $t_0 = 40$ )               | MRL(40)      | 29.7           | Tiempo adicional esperado de operación dado que el sistema ha llegado a 40.                              |
| Riesgo puntual en ( $t_0 = 40$ )                    | $h(40)$      | 0.00879        | Nivel instantáneo de riesgo en ( $t = 40$ ); indica que el sistema aún está en zona de baja falla.       |
| Probabilidad de supervivencia condicional (40 a 60) | $R(60   40)$ | 0.756          | Probabilidad de que el sistema opere de 40 a 60 sin fallar; útil para planear inspecciones y reemplazos. |

En conjunto, las métricas obtenidas brindan una caracterización cuantitativa integral que refuerza la comprensión operativa del proceso de deterioro. La coherencia entre los valores estimados y los patrones derivados del análisis de la función de falla demuestra que el modelo Beta–Weibull captura de manera precisa la dinámica temporal de falla y constituye una base sólida para la formulación de políticas óptimas de mantenimiento. Mientras el MTTF establece un horizonte global de desempeño, la MRL proporciona una medida dinámica de vida útil remanente; la supervivencia condicionada facilita la

evaluación del riesgo en periodos definidos y la tasa de falla puntual permite detectar el inicio del desgaste acelerado. La integración de estas métricas permite anticipar fallas, optimizar recursos y maximizar la disponibilidad del sistema.

Por otra parte, la Figura 28 presenta la evolución de la función  $MRL(t)$  estimada mediante el modelo Beta–Weibull. El comportamiento estrictamente decreciente es coherente con sistemas sujetos a envejecimiento o deterioro acumulativo, alcanzando un valor máximo cercano a 50 unidades de tiempo en los primeros instantes y descendiendo hasta valores inferiores a 10 unidades alrededor de  $t \approx 75$ . Este patrón confirma que, aunque la hazard se mantiene baja durante la fase estable, la vida remanente disminuye continuamente, lo cual justifica la programación de intervenciones preventivas antes de que el sistema cruce el umbral  $t_2$ .



**Figura 28.**  $MRL(t)$  bajo el modelo Beta–Weibull.

### F-5.3. Implicaciones para mantenimiento

La interpretación de las métricas de confiabilidad obtenidas a partir del modelo Beta–Weibull no constituye un ejercicio meramente estadístico, sino que representa el fundamento técnico para definir estrategias de mantenimiento orientadas a reducir la probabilidad de fallas inesperadas, optimizar el uso del activo y mejorar la disponibilidad operativa del sistema. La existencia de tres fases claramente diferenciadas (mortalidad temprana, vida útil estable y

desgaste acelerado) implica que las decisiones de mantenimiento deben adaptarse dinámicamente al comportamiento temporal del riesgo.

Un primer aspecto crítico consiste en la política de reemplazo óptima. La identificación precisa del punto  $t_2$  ( $\approx 55 - 60$  unidades de tiempo), asociado al inicio del deterioro acelerado, permite establecer un umbral preventivo antes del cual el reemplazo o la intervención mayor resulta más costo-efectiva. Para justificar estas decisiones, se contrastan métricas como el MTTF y la vida media residual (MRL), además de curvas de costo-riesgo cuando el análisis lo requiere. En este marco, el modelo Beta–Weibull proporciona una base cuantitativa sólida para recomendar reemplazos programados antes de que el riesgo crezca de forma significativa.

Asimismo, la determinación de intervalos óptimos de inspección depende directamente del comportamiento de la función de falla. Durante la fase de vida útil estable, donde la tasa de falla es baja y relativamente constante, los intervalos de inspección pueden espaciarse sin comprometer la seguridad operativa. En contraste, conforme la función de falla incrementa al aproximarse  $t_2$ , las inspecciones deben acortarse y orientarse a la detección temprana de mecanismos de desgaste. Para ello resulta útil el cálculo de la probabilidad de falla dentro de un intervalo específico, obtenida al evaluar la función acumulada  $F(t)$  del modelo ajustado en diferentes horizontes temporales.

Otro conjunto de decisiones deriva de la distinción entre mantenimiento preventivo basado en tiempo y mantenimiento predictivo basado en condición. El contraste entre la función de falla empírica y la función de falla teórica de la Beta–Weibull permite identificar si las fallas predominantes en el sistema provienen de defectos iniciales, desgaste progresivo o variabilidad aleatoria. En sistemas donde el riesgo es creciente y el deterioro se acelera con el tiempo, las estrategias basadas en condición (apoyadas en monitoreo continuo, umbrales de MRL o estimaciones de la función de falla instantánea) permiten intervenir de forma más eficiente que los esquemas puramente calendarizados.

En conjunto, estas métricas conforman un marco de decisión integral que permite anticipar fallas, optimizar recursos y maximizar la disponibilidad operativa del sistema. La Tabla 14 sintetiza las implicaciones operativas derivadas del análisis de confiabilidad y muestra cómo cada métrica se vincula con decisiones estratégicas de mantenimiento.

**Tabla 14.** Implicaciones operativas del modelo Beta–Weibull para el mantenimiento

| Métrica               | Interpretación operativa             | Implicación para mantenimiento                                                            |
|-----------------------|--------------------------------------|-------------------------------------------------------------------------------------------|
| $t_1 \approx 15 - 18$ | Fin de la mortalidad temprana        | Se pueden reducir inspecciones tempranas y estabilizar condiciones de operación.          |
| $t_1 \approx 55 - 60$ | Inicio del desgaste acelerado        | Punto crítico para planear reemplazo preventivo o intensificar monitoreo.                 |
| MTTF $\approx 46.9$   | Vida promedio del sistema            | Referencia global para definir ventanas de reemplazo programado.                          |
| MRL( $t$ )            | Vida útil remanente condicional      | Permite decidir si un componente continúa en operación o debe reemplazarse.               |
| $h(t)$                | Riesgo instantáneo de falla          | Determina la urgencia y frecuencia de las inspecciones.                                   |
| $R(t_1 t_0)$          | Probabilidad de supervivencia futura | Útil para planear intervalos operativos seguros y evaluar riesgo en ventanas específicas. |

#### 4.7 Resultados de la Fase 6. Documentación y comunicación técnica

La sexta fase del procedimiento metodológico se orienta a consolidar y comunicar de manera ordenada, rigurosa y replicable los resultados obtenidos en el análisis de confiabilidad. Aunque las fases previas se centraron en la construcción estadística del modelo, su validación y su interpretación ingenieril, la pertinencia del estudio depende en última instancia de la forma en que dicha información se documenta y se comunica a los actores responsables de la toma de decisiones. En ingeniería industrial, la transferencia efectiva de resultados hacia áreas como mantenimiento, operación, producción, calidad y gestión de activos constituye un componente crítico para garantizar la continuidad operativa de los equipos y la implementación de estrategias sustentadas en evidencia cuantitativa.

Esta fase, por tanto, no es meramente descriptiva; constituye la formalización del conocimiento generado a partir del análisis y su transformación en herramientas de gestión. Para ello, se desarrollan cuatro componentes fundamentales: plantillas de reporte, visualizaciones estandarizadas, conclusiones técnicas y recomendaciones basadas en datos.

### F-6.1. Plantillas para reportes de ingeniería

La documentación estructurada es esencial para asegurar la trazabilidad del análisis, la reproducibilidad de los resultados y la estandarización de la comunicación técnica entre departamentos. Con este fin, se propone una plantilla formal que puede adoptarse como estándar institucional en estudios de confiabilidad. Esta plantilla integra los elementos técnicos derivados de las Fases 1 a 5, asegurando que los reportes mantengan consistencia, claridad y rigor metodológico.

La Tabla 15 presenta la estructura recomendada para reportes de confiabilidad derivados del presente procedimiento. Esta plantilla permite que cualquier ingeniero encargado de mantenimiento o análisis estadístico pueda comprender el sistema estudiado, replicar el modelo, verificar decisiones y utilizar los resultados en la planeación operativa.

**Tabla 15.** Estructura sugerida para un reporte técnico de confiabilidad

| Sección del reporte           | Contenido mínimo                          | Relación con el procedimiento | Responsable típico          |
|-------------------------------|-------------------------------------------|-------------------------------|-----------------------------|
| 1.Descripción del sistema     | Función, criticidad, entorno operativo    | F-1.1                         | Ingeniería / Producción     |
| 2.Datos utilizados            | Fuente, cantidad, censura, depuración     | F-1.2–1.3                     | Confiabilidad               |
| 3.Análisis exploratorio       | Histogramas, curva KM, estadísticos       | Fase 1                        | Confiabilidad               |
| 4.Selección de distribuciones | Justificación AIC/BIC y coherencia física | Fase 2                        | Confiabilidad               |
| 5.Ajuste paramétrico          | Parámetros MLE, IC95%, error estándar     | Fase 3                        | Estadística / Confiabilidad |
| 6.Evaluación del modelo       | QQ-plot, PP-plot, hazard, pruebas GOF     | Fase 4                        | Confiabilidad               |
| 7.Métricas de confiabilidad   | MTTF, $h(t^0)$ , $R(t)$ , MRL( $t$ )      | F-5.2                         | Confiabilidad               |
| 8.Interpretación ingenieril   | Fases del ciclo de vida, curva de bañera  | F-5.1–5.3                     | Ingeniería                  |
| 9.Recomendaciones operativas  | Reemplazo, inspecciones, mantenimiento    | F-5.3 y F-6.4                 | Mantenimiento               |
| 10.Anexos                     | Código R, tablas, gráficas, datos         | Todas                         | Confiabilidad               |

### F-6.2. Visualizaciones estandarizadas

Las visualizaciones constituyen una de las herramientas más valiosas para comunicar patrones de falla, validar los supuestos del modelo, comparar distribuciones y detectar inconsistencias. La estandarización de estas visualizaciones asegura que todos los análisis futuros mantengan uniformidad metodológica y facilita la comprensión de los resultados por parte de personal no especializado en estadística.

En la Tabla 16 se resumen las figuras que deben incorporarse en cualquier estudio de confiabilidad basado en la metodología desarrollada en este trabajo.

**Tabla 16.** Visualizaciones recomendadas y su propósito técnico

| Visualización                       | Propósito                                     | Relación con el análisis | Figura en esta tesis |
|-------------------------------------|-----------------------------------------------|--------------------------|----------------------|
| $R(t)$ : Kaplan–Meier vs modelo     | Validar supervivencia teórica                 | F4.3.1                   | Fig. 22              |
| $h(t)$ : hazard empírico vs teórico | Confirmar forma del hazard y curva de bañera  | F4.3.2                   | Fig. 23              |
| Histograma + densidades             | Comparación con FDP ajustada                  | F4.3.3                   | Fig. 24              |
| QQ-plot conjunto                    | Evaluar discrepancia en cuantiles             | F4.3.4                   | Fig. 25              |
| PP-plot conjunto                    | Verificar ajuste en probabilidades acumuladas | F4.3.5                   | Fig. 26              |
| Curva de vida residual $MRL(t)$     | Evaluar deterioro progresivo                  | F-5.2                    | Fig. 28              |

Estas visualizaciones proporcionan un soporte gráfico imprescindible para la comunicación institucional del análisis. Además, permiten identificar patrones de confiabilidad que pueden no ser evidentes únicamente a partir de métricas numéricas.

### F-6.3. Conclusiones técnicas

Las conclusiones técnicas integran de manera cualitativa y cuantitativa los hallazgos más relevantes del estudio, relacionando explícitamente los resultados estadísticos con su interpretación ingenieril. Esta sección debe responder de

forma concreta al problema de investigación planteado y demostrar que el modelo Beta–Weibull seleccionado ofrece una representación válida del comportamiento real del sistema. Las conclusiones deben basarse en los siguientes elementos:

1. Identificación del patrón de falla, confirmando la presencia de una curva de bañera caracterizada por mortalidad temprana, vida útil estable y desgaste acelerado.
2. Interpretación de parámetros, donde el parámetro de forma y los parámetros adicionales de la distribución Beta–Weibull reflejan fenómenos físicos como defectos iniciales y envejecimiento progresivo.
3. Métricas de desempeño, como MTTF,  $MRL(t)$ , riesgo puntual  $h(t^0)$  y probabilidad condicionada  $R(t^1|t^0)$ , que cuantifican la vida operativa y el riesgo futuro.
4. Impacto operativo, centrado en cómo la transición hacia el desgaste ( $t^2$ ) influye en la planificación de mantenimiento y la disponibilidad del equipo.

La curva  $MRL(t)$  presentada en la Figura 28 se convierte en un argumento central para justificar decisiones basadas en condición y reemplazo anticipado.

#### **F-6.4. Recomendaciones basadas en datos**

La última sección conecta directamente los hallazgos cuantitativos con decisiones operativas. Las recomendaciones no deben formularse de manera genérica, sino respaldarse con las métricas obtenidas y con las características observadas del sistema bajo estudio.

Para facilitar la relación entre métricas y decisiones, se incluye la Tabla 17, que resume las acciones sugeridas en mantenimiento según los valores del modelo Beta–Weibull.

**Tabla 17.** Relación entre métricas de confiabilidad y decisiones operativas

| Métrica              | Interpretación de estado                   | Umbral típico   | Acción recomendada                                        |
|----------------------|--------------------------------------------|-----------------|-----------------------------------------------------------|
| MTTF                 | Vida promedio del sistema                  | $t \approx 47$  | Definir reemplazo planeado antes de la fase de desgaste.  |
| MRL( $t_0$ )<br>alta | Componente aún con vida útil significativa | $> 25$          | Continuar operación con inspecciones rutinarias.          |
| MRL( $t_0$ )<br>baja | Cercano a desgaste acelerado               | $< 15$          | Programar inspección mayor o reemplazo.                   |
| $R(t^1 t^0)$<br>alta | Bajo riesgo futuro                         | $\geq 0.8$      | Mantener estrategia de inspección y operación actual.     |
| $R(t^1 t^0)$<br>baja | Riesgo creciente                           | $\leq 0.5$      | Reducir intervalos de inspección y preparar intervención. |
| $h(t)$<br>creciente  | Inicio de desgaste                         | $t \approx t_2$ | Aumentar frecuencia de inspecciones y preparar reemplazo. |
| $h(t)$ estable       | Vida útil nominal                          | $t_1 < t < t_2$ | Mantener mantenimiento basado en condición.               |

## Capítulo V. Discusión

### 5.1 Interpretación de los principales hallazgos

Los resultados obtenidos evidencian que los datos de tiempos de falla analizados presentan un comportamiento claramente no monótono de la tasa de riesgo, consistente con la estructura clásica de la curva de bañera. La identificación explícita de las fases de mortalidad temprana, vida útil estable y desgaste confirmó que el uso exclusivo de modelos tradicionales, particularmente la distribución Weibull estándar, resulta insuficiente para describir adecuadamente la dinámica temporal del riesgo en sistemas con múltiples mecanismos de falla. Este hallazgo resulta especialmente relevante desde una perspectiva académica, ya que valida empíricamente uno de los supuestos centrales de la tesis, la cual es la necesidad de utilizar modelos probabilísticos más flexibles cuando el comportamiento de la función de falla presenta cambios de tendencia a lo largo del ciclo de vida.

El contraste entre la función de falla empírica estimada mediante métodos no paramétricos y las funciones de falla teóricas derivadas de los modelos paramétricos permitió observar diferencias sustantivas en la capacidad de ajuste de los modelos evaluados. La distribución Beta–Weibull mostró una mayor capacidad para reproducir los cambios de concavidad y las transiciones entre fases características de la curva de bañera, mientras que los modelos clásicos tendieron a suavizar o distorsionar estos comportamientos. Este resultado refuerza la idea de que la selección del modelo probabilístico no debe basarse únicamente en criterios de simplicidad o tradición, sino en su coherencia estadística y su congruencia con la interpretación ingenieril del fenómeno de falla.

Asimismo, las métricas de confiabilidad derivadas del modelo seleccionado como el Tiempo Medio Hasta la Falla MTTF, la Vida media residual MRL y la probabilidad de supervivencia  $R(t)$ , mostraron coherencia interna y correspondencia con el comportamiento observado de la tasa de falla. En particular, la evolución decreciente de la vida residual media aportó evidencia cuantitativa del deterioro progresivo del sistema, confirmando su utilidad como

indicador para anticipar la transición hacia la fase de desgaste. Este resultado destaca la importancia de complementar los indicadores tradicionales de confiabilidad con métricas dinámicas que permitan una interpretación más profunda del estado del sistema a lo largo del tiempo.

Los hallazgos obtenidos guardaron una relación directa con los planteamientos teóricos clásicos y contemporáneos revisados en el marco teórico. La presencia de un patrón tipo curva de bañera coincidió con los trabajos fundacionales de Lai, Xie y Murthy, quienes señalaron que numerosos sistemas industriales presentan tasas de riesgo no monótonas asociadas a distintos mecanismos físicos de falla. De igual manera, los resultados reforzaron lo documentado en estudios recientes que advierten sobre las limitaciones de la distribución Weibull estándar para modelar datos con dos cambios de tendencia en la función de falla, tal como lo han señalado Jiang et al. y Almetwally et al. al proponer distribuciones modificadas y familias más flexibles.

El uso de estimadores no paramétricos como herramienta complementaria permitió validar empíricamente la forma de la tasa de riesgo sin imponer supuestos estructurales previos, en concordancia con enfoques metodológicos recomendados en la literatura para reducir sesgos derivados de una selección prematura del modelo. Esta estrategia fortaleció la validez del procedimiento metodológico propuesto y permitió establecer una conexión clara entre la exploración inicial de los datos y la selección final del modelo probabilístico. A diferencia de estudios que se limitan a comparar indicadores numéricos de ajuste, la presente investigación integró explícitamente la interpretación física de las fases del ciclo de vida con la evaluación estadística, lo que constituye una aportación metodológica relevante en el ámbito académico.

Desde la perspectiva de la Ingeniería Industrial, los resultados obtenidos tienen implicaciones directas para la formación profesional y la práctica analítica del ingeniero. El procedimiento metodológico desarrollado demostró que es posible estructurar un análisis de confiabilidad riguroso, reproducible y conceptualmente claro, integrando herramientas estadísticas avanzadas sin perder interpretabilidad. Este enfoque contribuye a cerrar la brecha identificada en el planteamiento del problema, donde se señaló la desconexión entre los

contenidos estadísticos impartidos en la formación universitaria y su aplicación real en el análisis de fallas industriales.

El énfasis en la selección fundamentada de modelos, la interpretación de la función de falla y el uso de métricas dinámicas de confiabilidad fortalece la capacidad del ingeniero para traducir resultados estadísticos en información útil para la toma de decisiones, incluso cuando el estudio se desarrolla con fines metodológicos y no como una implementación directa en una empresa específica. En este sentido, la investigación aporta valor académico al proponer una guía estructurada que puede ser utilizada como referencia didáctica en cursos de confiabilidad, estadística aplicada o ingeniería del mantenimiento.

No obstante, el estudio presentó limitaciones que deben ser reconocidas para contextualizar adecuadamente sus alcances. El análisis se basó en un conjunto de datos de referencia ampliamente utilizado en la literatura académica, lo cual permitió validar la metodología propuesta, pero limita la generalización directa de los resultados a contextos industriales específicos sin una adaptación previa. Asimismo, el enfoque se centró exclusivamente en el modelado estadístico de la confiabilidad, sin incorporar variables económicas, organizacionales o humanas que también influyen en la gestión del mantenimiento. De igual forma, se asumió independencia entre eventos de falla, lo cual podría no reflejar completamente la complejidad de sistemas con interdependencias estructurales o funcionales.

A pesar de estas limitaciones, los resultados permitieron dar cumplimiento al objetivo general de la investigación, al desarrollarse un procedimiento metodológico claro y reproducible para modelar tiempos de falla con comportamiento tipo curva de bañera mediante distribuciones de probabilidad avanzadas. Los objetivos específicos también fueron alcanzados, al caracterizar el comportamiento estadístico de los tiempos de falla, comparar modelos probabilísticos con base en criterios técnicos y evaluar métricas de confiabilidad que facilitaron la interpretación del desempeño del sistema a lo largo de su ciclo de vida.

Finalmente, los hallazgos obtenidos abren diversas líneas para futuras investigaciones, entre ellas la aplicación del procedimiento metodológico propuesto a bases de datos industriales reales, la integración con modelos de degradación o técnicas de aprendizaje automático, y la incorporación de análisis

económicos del ciclo de vida. En síntesis, la discusión permitió concluir que el uso de distribuciones de probabilidad avanzadas y de un enfoque metodológico estructurado constituye una alternativa sólida para el estudio de sistemas industriales con comportamiento tipo curva de bañera, fortaleciendo el vínculo entre la estadística aplicada y la ingeniería de confiabilidad, y contribuyendo a la formación de ingenieros industriales con una visión analítica, crítica y orientada a la toma de decisiones basada en datos.

## Conclusiones

El desarrollo de la presente tesis permitió atender de manera sistemática el problema académico identificado en el capítulo inicial, consistente en la ausencia de una metodología aplicada, clara y reproducible que facilite a estudiantes y egresados de Ingeniería Industrial la modelación de tiempos de falla con comportamiento en forma de curva de bañera mediante distribuciones de probabilidad adecuadas. A lo largo del estudio fue posible integrar fundamentos teóricos, herramientas estadísticas y criterios de interpretación ingenieril en un procedimiento metodológico coherente, orientado al análisis riguroso de la confiabilidad en sistemas industriales.

En relación con el objetivo general de la tesis, se concluyó que este fue cumplido de manera satisfactoria. Se diseñó y validó un procedimiento metodológico estructurado para la modelación de tiempos de falla utilizando distribuciones de probabilidad capaces de representar comportamientos no monótonos del riesgo o falla. Dicho procedimiento permitió identificar las fases características de la curva de bañera, seleccionar modelos probabilísticos acordes con la forma empírica de la función de falla, estimar parámetros de manera consistente y evaluar métricas de confiabilidad relevantes para la interpretación del desempeño del sistema. Con ello, se demostró que es posible superar las limitaciones asociadas al uso exclusivo de la distribución Weibull estándar mediante un enfoque metodológico más flexible y técnicamente fundamentado.

Respecto a los objetivos específicos, los resultados mostraron que el comportamiento estadístico de los tiempos de falla analizados pudo caracterizarse adecuadamente mediante análisis exploratorio y estimadores no paramétricos, lo que permitió diagnosticar la presencia de una tasa de falla no monótona y delimitar las fases del ciclo de vida del sistema. Asimismo, la comparación entre diferentes distribuciones de probabilidad evidenció que uno de los modelos avanzados presenta una mayor capacidad para reproducir la forma real de la función de falla y describir de manera coherente los mecanismos de deterioro subyacentes. Además, la estimación e interpretación de métricas como el tiempo medio hasta la falla, la vida media residual y la probabilidad de

supervivencia condicionada permitió traducir los resultados estadísticos en información relevante desde una perspectiva de confiabilidad y mantenimiento.

Entre los principales hallazgos del estudio se confirmó que la selección del modelo probabilístico influye de manera directa en la interpretación del comportamiento de una falla en un producto, dispositivo, equipo o sistema. El análisis realizado evidenció que los sistemas con patrón tipo curva de bañera no pueden describirse adecuadamente mediante modelos de riesgo monótono, y que el uso de distribuciones flexibles permite capturar con mayor fidelidad las transiciones entre las fases de mortalidad temprana, vida útil estable y desgaste acelerado. Este resultado reforzó la necesidad de complementar los criterios puramente estadísticos de ajuste con una interpretación física e ingenieril del fenómeno de falla.

En cuanto a los aportes del estudio, la principal contribución se ubica en el ámbito académico de la Ingeniería Industrial. La tesis propuso un procedimiento metodológico didáctico, estructurado y reproducible que integra estadística aplicada, análisis de supervivencia e ingeniería de confiabilidad, constituyendo una herramienta útil para la formación de estudiantes y para el desarrollo de investigaciones posteriores en el área. A diferencia de trabajos exclusivamente aplicados, este estudio aportó valor al sistematizar un proceso metodológico que puede ser replicado, adaptado y ampliado en distintos contextos educativos y, potencialmente, industriales. De manera indirecta, el procedimiento desarrollado también ofrece una base técnica susceptible de transferencia a organizaciones interesadas en fortalecer sus análisis de mantenimiento y gestión de activos.

No obstante, se reconoce que el estudio se desarrolló a partir de un conjunto de datos de referencia ampliamente utilizado en la literatura académica, lo cual limita la generalización inmediata de los resultados a sistemas industriales específicos. Asimismo, el análisis se concentró en el componente estadístico de la confiabilidad, sin incorporar de manera explícita variables económicas, organizacionales o humanas que también influyen en la gestión del mantenimiento.

A partir de los resultados obtenidos, se identificaron implicaciones relevantes para trabajos futuros. El procedimiento metodológico desarrollado puede servir como base para investigaciones posteriores que incorporen datos

industriales reales, análisis de costos del ciclo de vida, modelos de degradación o enfoques híbridos que integren técnicas de aprendizaje automático. De igual manera, los hallazgos abren la posibilidad de fortalecer la enseñanza de la ingeniería de confiabilidad en los programas de Ingeniería Industrial, promoviendo una formación más aplicada, crítica y alineada con las demandas actuales del sector productivo.

En síntesis, la presente tesis demostró que el análisis de confiabilidad de productos o sistemas con comportamiento en forma de curva de bañera requiere metodologías más robustas que las tradicionalmente empleadas. Al proponer y validar un procedimiento metodológico basado en distribuciones de probabilidad avanzadas, este trabajo realizó una contribución significativa al ámbito académico de la Ingeniería Industrial, fortaleciendo el vínculo entre la estadística aplicada y la toma de decisiones en confiabilidad. Los resultados confirmaron que una modelación adecuada del riesgo o falla no sólo incrementa la precisión analítica, sino que también enriquece la comprensión ingenieril del fenómeno de falla, sentando bases sólidas para futuras aplicaciones y desarrollos en la disciplina.

Finalmente, como parte de los aportes metodológicos de esta investigación, el procedimiento estadístico desarrollado fue documentado de manera reproducible mediante scripts en el lenguaje R, los cuales se incluyen en los anexos de la tesis. Esta documentación permite la verificación de los resultados obtenidos, así como la replicación y adaptación del análisis en futuros estudios de confiabilidad industrial con comportamiento tipo curva de bañera.

## Referencias bibliográficas

- Aarset, M. V. (1987). *How to identify a bathtub hazard rate*. IEEE Transactions on Reliability, 36(1), 106–108.
- Abdulshahed, A. M., Longstaff, A. P., Fletcher, S., & Myers, A. (2022). Integration of probabilistic reliability models into predictive maintenance systems for metalworking machinery. *Procedia Manufacturing*, 58, 123–132.
- Ahn, J., & Kang, H. (2024). Hybrid Weibull–Pareto model for electronic component reliability in automotive manufacturing. *Reliability Engineering & System Safety*, 240, 109125.
- Ahn, J., & Kang, H. (2024). Hybrid Weibull–Pareto models for reliability of industrial components. *Reliability Engineering & System Safety*, 243, 109789.
- Alazemi, S., & Elsayed, E. A. (2023). Reliability estimation of repairable systems under varying operational conditions. *Reliability Engineering & System Safety*, 230, 109156.
- Almetwally, E. M., Ahmed, A., & Ramadan, S. (2025). A new flexible probability model for engineering reliability data. *Reliability Engineering & System Safety*, 239, 109122.
- Almetwally, E. M., El-Sagheer, A., & Ramadan, M. (2025). Flexible probability distributions for modeling bathtub-shaped failure rates. *Reliability Engineering & System Safety*, 259, 110402.
- Almetwally, E. M., Mahmoud, M., & Abu-Zinadah, H. (2025). Power-X family of distributions: Properties and applications in reliability. *Mathematics*, 13(1), 155.
- Almetwally, E. M., Mohamed, I., & Rashad, A. (2025). A new Power-X family for modeling heavy-tailed reliability data. *Reliability Engineering & System Safety*, 254, 110334.
- Almetwally, E. M., Qureshi, A. M., & Shah, S. H. (2025). Modified Power-X family for modeling bathtub-shaped failure rates. *Reliability Engineering & System Safety*, 247, 110929.

- Bouslah, B., Gharbi, A., & Jemai, Z. (2023). Integration of maintenance optimization and reliability modeling in industrial systems. *Reliability Engineering & System Safety*, 234, 109202.
- Bracke, S. (2024). *Reliability engineering data analytics: Modeling and risk prediction*. Springer.
- Crespo, J., & Sánchez, L. (2022). Probabilistic reliability assessment for manufacturing systems. *Journal of Industrial Statistics*, 45(3), 212–229.
- Cui, X., Liu, D., & Xu, Z. (2024). Probabilistic assessment of degradation and lifetime prediction in complex systems. *Journal of Statistical Computation and Simulation*, 94(5), 978–995.
- Dey, S., Singh, R. S., & Upadhyay, S. K. (2019). Alpha-power transformed inverse Lindley distribution: A distribution with an upside-down bathtub-shaped hazard function. *Journal of Computational and Applied Mathematics*, 348, 130–145.
- Dhillon, B. S. (2017). *Engineering systems reliability, safety, and maintenance: An integrated approach*. CRC Press.
- Dhillon, B. S. (2017). *Engineering reliability: New techniques and applications*. Springer.
- Dhillon, B. S. (2023). *Engineering Reliability: New Techniques and Applications*. Springer.
- Ding, Y., Lin, Z., & Xue, F. (2024). Bayesian reliability updating for degradation-based maintenance optimization. *Reliability Engineering & System Safety*, 256, 110401.
- Dodson, B., & Nolan, D. (2021). *Reliability engineering handbook*. CRC Press.
- Du, W., & Sun, J. (2022). Statistical modeling for bathtub-shaped failure rates in complex systems. *Journal of Statistical Planning and Inference*, 221, 137–152.
- Ebeling, C. (2021). *An introduction to reliability and maintainability engineering* (3rd ed.). Waveland Press.
- El-Sayed, R., & Hassan, K. (2024). Hybrid Weibull–Pareto distribution for industrial failure modeling. *Mechanical Systems and Signal Processing*, 194, 110764.

- Elsahhar, A. M., Rahman, M., & van Bussel, G. J. W. (2025). Analysis of Failure and Maintenance Records in Aging Wind Farms to Inform End-of-Life Asset Management. *Wind Energy*, 28(6), e70024.
- Elsahhar, A., Ghoneim, M., & Omar, A. (2025). Reliability assessment of mechanical systems under varying operational stress. *Engineering Failure Analysis*, 161, 112420.
- El-Sagheer, A. A., & Ramadan, M. Z. (2025). Reliability engineering education and advanced probabilistic modeling for industrial applications. *Education for Chemical Engineers*, 40, 1–12.
- Elsahhar, H., Ahmed, M., & Kassem, R. (2025). Reliability-based maintenance optimization in manufacturing systems. *Journal of Manufacturing Systems*, 76, 32–45.
- Elsayed, E. A. (2023). *Reliability engineering* (3rd ed.). Wiley.
- García-Treviño, M., & Ríos, L. (2024). Hybrid predictive maintenance framework using machine learning and reliability analysis. *Ingeniería y Ciencia*, 20(3), 45–62.
- Garg, R. (2020). *Probability distributions in reliability analysis*. In Reliability Engineering and Sciences (pp. 45–72). Springer.
- Giorgio, M., Piscopo, M., & Pulcini, G. (2023). Stochastic degradation models for bathtub hazard functions. *European Journal of Operational Research*, 310(2), 591–604.
- Giorgio, M., Piscopo, V., & Pulcini, G. (2023). Stochastic degradation models in reliability engineering. *Applied Stochastic Models in Business and Industry*, 39(3), 451–466.
- Giorgio, M., Piscopo, V., & Pulcini, G. (2023). Stochastic models for degradation and maintenance optimization in industrial systems. *Reliability Engineering & System Safety*, 238, 108992.
- Goh, T. N., Chong, C. Y., & Kuo, W. (2023). Smart reliability engineering for Industry 4.0 systems. *International Journal of Production Research*, 61(12), 3981–3995.
- Gómez, J., Hernández, M., & Ruiz, P. (2023). Application of Weibull models in metal-mechanic systems under fatigue stress. *Revista Ingeniería Mecánica*, 41(2), 87–104.

- González-Hernández, I. J., Méndez-González, L. C., & Soto-Campos, C. A. (2025). Ingeniería de confiabilidad y analítica de datos. En *El arte de la ciencia de datos: Aplicaciones prácticas* (pp. 185–214). Universidad Autónoma del Estado de Hidalgo.
- González, J., & Ramírez, F. (2023). Análisis de confiabilidad y mantenimiento basado en la distribución Weibull en líneas de producción metalmecánicas. *Revista Mexicana de Ingeniería Industrial*, 15(2), 66–81.
- Hernández-Sampieri, R., Fernández, C., & Baptista, M. (2022). *Metodología de la investigación* (7ª ed.). McGraw-Hill.
- Hernández-Sampieri, R., & Mendoza, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Hernández-Santiago, J. (2019). *Modelado de la confiabilidad de transformadores eléctricos mediante distribuciones de Weibull* [Tesis de Maestría, UNAM].
- Hess, K. R. (2024). muhaz: Hazard Function Estimation in R (Versión 1.2.6) [Paquete de software]. CRAN. <https://CRAN.R-project.org/package=muhaz>
- Hu, B., Wang, X., & Chen, L. (2023). IoT-enabled adaptive reliability analysis for predictive maintenance. *Sensors and Actuators A: Physical*, 358, 114245.
- ISO 14224:2016. *Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment*. ISO.
- ISO 55000:2014. *Gestión de activos — Panorama general, principios y terminología*. ISO.
- ISO 9000:2015. *Sistemas de gestión de la calidad — Fundamentos y vocabulario*. ISO.
- Jardine, A. K. S., & Tsang, A. H. C. (2013). *Maintenance, replacement, and reliability: Theory and applications* (2nd ed.). CRC Press.
- Jiang, J., Hu, X., & Li, C. (2023). Modified Weibull and Power-X families for advanced reliability prediction of industrial systems. *Reliability Engineering & System Safety*, 239, 109043.

- Jiang, W., Yu, Y., & Zhang, L. (2023). Modified Weibull models for bathtub-shaped failure rate systems. *Reliability Engineering & System Safety*, 232, 108935.
- Kaur, R., & Singh, J. (2024). Dynamic failure rate analysis for predictive maintenance in industrial environments. *Procedia Manufacturing*, 87, 551–563.
- Kerlinger, F. N., & Lee, H. B. (2020). *Foundations of behavioral research* (5th ed.). Wadsworth.
- Khakzad, N., & van Gelder, P. H. A. J. M. (2021). Risk-based maintenance strategies for complex industrial systems: A review. *Reliability Engineering & System Safety*, 212, 107593.
- Kharazian, Z., Lindgren, T., Magnússon, S., Steinert, O., & Reyna, O. A. (2025). SCANIA Component X dataset: A real-world multivariate time series dataset for predictive maintenance. *Scientific Data*, 12, 493.
- Khurodze, T., & Kakubava, M. (2025). Reliability modeling of industrial equipment under limited life-cycle data and censored observations. *International Journal of Industrial Engineering*, 32(2), 145–158.
- Khurodze, T., & Kakubava, N. (2025). Semi-Markov reliability models for repairable industrial systems with dependent failures. *European Journal of Operational Research*, 317(2), 481–494.
- Lai, C. D., Xie, M., & Murthy, D. N. P. (2001). Bathtub-shaped failure rate life distributions. *Quality and Reliability Engineering International*, 17(6), 481–493.
- Lai, C. D., Xie, M., & Murthy, D. N. P. (2001). Bathtub-shaped failure rate life distributions. In Balakrishnan, N., & Rao, C. R. (Eds.), *Handbook of Statistics* (Vol. 20, pp. 69–123). Elsevier.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). Wiley.
- Lee, J., Bagheri, B., & Kao, H. A. (2024). Industrial AI: Applications in smart manufacturing and reliability optimization. *Journal of Manufacturing Systems*, 70, 240–256.

- Li, J., Yu, Z., & Song, W. (2023). Deep learning-enhanced reliability modeling for CNC machine degradation. *Journal of Intelligent Manufacturing*, 34(5), 1823–1841.
- Li, M., Chen, R., & Zhou, P. (2024). Machine learning-enhanced reliability assessment in automotive manufacturing systems. *Engineering Applications of Artificial Intelligence*, 133, 108147.
- Liu, Y., Zhang, T., Wang, H., & Li, X. (2024). Advanced survival analysis for industrial reliability assessment under non-monotonic hazard rates. *IEEE Transactions on Reliability*, 73(1), 214–227.
- Martínez-Reyes, P. (2021). *Análisis de confiabilidad aplicada a maquinaria industrial mediante distribución de Weibull* [Tesis de Licenciatura, UAEH].
- Meeker, W. Q., & Escobar, L. A. (2022). *Statistical Methods for Reliability Data*. Wiley.
- Mohanty, S., & Laha, A. (2024). Advanced probabilistic models for time-to-failure prediction under limited data scenarios. *Reliability Engineering & System Safety*, 257, 110365.
- Montgomery, D. C. (2021). *Design and analysis of experiments* (10th ed.). Wiley.
- Moubray, J. (1997). *Reliability-centered maintenance* (2nd ed.). Butterworth-Heinemann.
- Müller, H. G., & Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50(1), 61–76.
- Murthy, D. N. P., Xie, M., & Jiang, R. (2004). Weibull models. Wiley.
- Niu, R., & Yang, Z. (2024). Data-driven reliability modeling for industrial maintenance optimization. *IEEE Transactions on Reliability*, 73(1), 231–243.
- Nor, A. K. M., Pedapati, S. R., & Muhammad, M. (2021). Reliability engineering applications in electronic, software, nuclear and aerospace industries: A 20-year review (2000–2020). *Ain Shams Engineering Journal*, 12(3), 3009–3019.
- Nor, M. N., Pedapati, S. R., & Muhammad, N. M. N. (2021). Reliability analysis and maintenance decision-making for industrial systems: A

- review of recent advances. *Journal of Quality in Maintenance Engineering*, 27(4), 611–630.
- O'Connor, P. D. T., & Kleyner, A. (2023). *Practical reliability engineering* (6th ed.). Wiley.
  - Oladokun, S. O., Akinola, A., & Bello, M. (2023). Power-X modeling of hydraulic press reliability in manufacturing industries. *International Journal of Industrial Engineering Computations*, 14(5), 812–826.
  - Pal, S., & Aselisewine, J. (2024). Hybrid machine learning and probabilistic approaches for predictive maintenance in manufacturing systems. *Computers & Industrial Engineering*, 187, 109817.
  - Patil, R., & Rao, M. (2022). *Reliability-centered industrial systems: Theory and implementation*. Springer Industrial Series.
  - Pham, H. (2023). *Reliability and statistical models: An engineering approach*. Springer.
  - *Probability distributions in reliability analysis*. (2020). Springer.
  - Rahman, M. M., Khan, F., Ahmed, S., & Imtiaz, S. (2024). Data-driven reliability and maintenance modeling in Industry 4.0: Challenges and opportunities. *Reliability Engineering & System Safety*, 240, 109516.
  - Ramaki, A., Tadj, L., & Belkacem, S. (2025). Hybrid reliability modeling using modified Weibull–Pareto distributions for industrial applications. *Journal of Quality and Reliability Engineering*, 1–15.
  - Ramaki, A., Bagheri, M., Kharazian, Z., & Dehghan, R. (2025). Statistical reliability modeling of complex industrial systems with non-monotonic failure behavior. *Quality and Reliability Engineering International*, 41(1), 89–105.
  - Ramaki, A. A., Almetwally, E., & Ramadan, A. (2025). Weibull–Pareto composite model for life data analysis. *Reliability Engineering & System Safety*, 245, 110901.
  - Ramírez, C., & Zhao, L. (2023). Reliability-centered and risk-based maintenance in metalworking industries. *Procedia Manufacturing*, 65, 521–529.

- Ramírez-Cortés, J. (2020). *Modelos de predicción de fallas industriales basados en aprendizaje automático y confiabilidad* [Tesis de Maestría, IPN].
- Rao, P., & Kumar, D. (2021). Reliability analysis of machining systems using probabilistic life models. *Journal of Manufacturing Science and Engineering*, 143(11), 111005.
- Relia, A., Han, J., & Zio, E. (2022). Stochastic degradation modeling of industrial components under variable loading. *Mechanical Systems and Signal Processing*, 178, 109266.
- Rezaei, M., Esmaeili, M., & Moradi, S. (2022). Probabilistic modeling of time-dependent failure behavior in industrial machines. *Journal of Statistical Computation and Simulation*, 92(10), 2065–2084.
- Serrano, M. A. (2022). *Simulación de confiabilidad y mantenimiento centrado en equipos hidráulicos industriales* [Tesis de Maestría, UNAM].
- Shah, A., Khan, R., Ahmad, Z., & Gul, M. (2022). Optimization of maintenance strategies using reliability-based models: An industrial perspective. *Journal of Manufacturing Systems*, 63, 75–88.
- Sharma, R., Gupta, A., & Reddy, K. (2023). Operational reliability assessment of complex industrial systems using probabilistic methods. *Mechanical Systems and Signal Processing*, 182, 109643.
- *Statistical reliability engineering*. (2022). Wiley.
- Sun, C., Li, J., & Xu, F. (2024). *Reliability engineering: Theory and practice*. Elsevier.
- Varshney, R., Kumar, D., & Singh, R. (2025). Reliability assessment and predictive maintenance optimization in industrial systems. *Mechanical Systems and Signal Processing*, 208, 110930.
- Varshney, S., Bajaj, M., & Choudhary, K. K. (2025). Performance analysis and reliability prediction of multi-state service systems. *Engineering Reports*, 7(5), e70268.
- Vásquez, A. A., Guanuchi, L. M., Cahuana, R., Vera, R., & Holgado, J. (2023). *Métodos de investigación científica*. Instituto Inudi.

- Wang, J., Sha, L., Gu, F., & Xu, Y. (2023). Reliability modeling and life prediction of industrial components using hybrid Weibull models. *Mechanical Systems and Signal Processing*, 206, 110887.
- Wang, X., Yu, C., & Li, Y. (2015). A finite interval lifetime distribution model for fitting bathtub-shaped failure rate curve. *Mathematical Problems in Engineering*, 2015, 954327.
- Wang, Y., Yu, X., & Li, J. (2015). Finite range distribution and its applications in reliability analysis. *Journal of Statistical Computation and Simulation*, 85(13), 2615–2633.
- Wu, Q., Li, Y., & Xu, G. (2023). Gamma process modeling for remaining useful life prediction in manufacturing equipment. *Reliability Engineering & System Safety*, 239, 109275.
- Zhang, W., Liu, C., & Wang, H. (2022). Failure mechanisms in metalworking systems under cyclic stress conditions. *Journal of Manufacturing Processes*, 84, 267–278.
- Zhang, Y., Wang, C., & Hu, T. (2023). Intelligent reliability-based predictive maintenance in smart factories. *Procedia CIRP*, 114, 321–330.
- Zhao, L., Li, M., & Wang, D. (2022). Integrating RCM and RBM for intelligent maintenance in smart manufacturing systems. *Reliability Engineering & System Safety*, 226, 108739.
- Zheng, L., Chen, Y., Li, S., & Zhao, W. (2024). Reliability-centered maintenance optimization using advanced lifetime distributions. *Applied Mathematical Modelling*, 126, 469–486.
- Zhou, X., Li, P., & Zhao, X. (2023). Hybrid stochastic degradation modeling in intelligent manufacturing. *Reliability Engineering & System Safety*, 241, 109326.
- Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., & da Trindade, E. (2020). Predictive maintenance in Industry 4.0: A systematic review. *Computers & Industrial Engineering*, 150, 106889.

## Anexo: Código en R

```
#=====
# ----- Fase 1: Preparación y comprensión de los datos -----
---
#=====

# ----- F-1.1. Definir el contexto del sistema a analizar -----

# --- Instalar librerías (packages) necesarias -----

# install.packages(c("tidyverse", "survival", "survminer", "e1071", "muhaaz"))

# --- Ejecutar solo una vez si no se tiene instalado los paquetes: -----

library(tidyverse) # Manejo de datos y gráficos (ggplot2)
library(survival)  # Análisis de supervivencia (Kaplan–Meier)
library(survminer) # Gráficos bonitos para curvas de supervivencia
library(e1071)     # Asimetría y curtosis
library(muhaaz)    # Estimación no paramétrica de la tasa de falla (hazard)

# ----- F-1.2. Construcción del vector de tiempos de falla -----

tiempos <- c(
  0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6, 7, 11, 12, 18, 18, 18, 18, 18, 21, 32, 36, 40, 45,
  46, 47, 50, 55, 60, 63, 63, 67, 67, 67, 67, 72, 75, 79, 82, 82, 83, 84, 84, 84, 85,
  85, 85, 85, 85, 86, 86
)

# ----- F-1.3. Análisis descriptivo de los tiempos de falla -----

# Estadísticos descriptivos básicos
resumen_basico <- data.frame(
```

```

n      = length(tiempos),
media  = mean(tiempos),
mediana = median(tiempos),
min     = min(tiempos),
max     = max(tiempos),
varianza = var(tiempos),
sd      = sd(tiempos),
p25     = quantile(tiempos, 0.25),
p75     = quantile(tiempos, 0.75),
IQR     = IQR(tiempos),
asimetria = skewness(tiempos),
curtosis = kurtosis(tiempos)
)

```

```
rownames(resumen_basico) <- "Estadísticas"
```

```
resumen_basico
```

```
# --- Gráficas exploratorias para visualizar el comportamiento de la variable
tiempo hasta la falla (T) ---
```

```
# --- Crear data frame a partir del vector de tiempos
```

```
df_tiempos <- data.frame(
  tiempo_falla = tiempos
)
```

```
df_tiempos
```

```
# --- Histograma de los tiempos de falla
```

```
media_t  <- mean(tiempos)
mediana_t <- median(tiempos)
```

```

ggplot(df_tiempos, aes(x = tiempo_falla)) +
  # Histograma en densidad y con color personalizado
  geom_histogram(
    aes(y = ..density..),
    binwidth = 5,
    boundary = 0,
    closed = "right",
    fill = "lightblue", # color de relleno del histograma
    color = "black"     # borde de las barras
  ) +
  # Línea de la media
  geom_vline(
    aes(xintercept = media_t, color = "Media", linetype = "Media"),
    linewidth = 0.8
  ) +
  # Línea de la mediana
  geom_vline(
    aes(xintercept = mediana_t, color = "Mediana", linetype = "Mediana"),
    linewidth = 0.8
  ) +
  # Escalas para que aparezca la leyenda de líneas
  scale_color_manual(
    name = "Estadísticos",
    values = c("Media" = "red", "Mediana" = "darkgreen")
  ) +
  scale_linetype_manual(
    name = "Estadísticos",
    values = c("Media" = "dashed", "Mediana" = "dotted")
  ) +
  labs(
    title = "Histograma de tiempos de falla con media y mediana",
    x = "Tiempo hasta la falla",
    y = "Densidad"
  ) +

```

```

theme_minimal()

# --- Boxplot de los tiempos de falla

ggplot(df_tiempos, aes(y = tiempo_falla)) +
  geom_boxplot(
    fill = "lightblue", # color de la caja
    color = "black",    # borde
    width = 0.25,
    outlier.color = "red", # color de los atípicos
    outlier.size = 2
  ) +
  labs(
    title = "Diagrama de caja de los tiempos de falla",
    x = "",
    y = "Tiempo hasta la falla"
  ) +
  theme_minimal()

# --- Graficar la curva de supervivencia con survminer

# Como los datos son NO censurados, creamos una variable evento = 1 para
# todos
df_surv <- df_tiempos |>
  mutate(evento = 1)

# Ajustar el modelo de supervivencia Kaplan–Meier
fit_km <- survfit(Surv(time = tiempo_falla, event = evento) ~ 1,
  data = df_surv)

# Graficar la curva de supervivencia con survminer
ggsurvplot(
  fit_km,
  data = df_surv,

```

```

conf.int = TRUE,          # banda de confianza
risk.table = FALSE,       # tabla de riesgo desactivada por ahora
xlab     = "Tiempo hasta la falla",
ylab     = "Probabilidad de supervivencia R(t)",
ggtheme  = theme_minimal(),
surv.scale = "percent",   # muestra R(t) en porcentaje
legend = "none",
xlim = c(0, 90)
)

# ----- F-1.4. Estimación no paramétrica de R(t) y h(t) -----

# --- Estimación no paramétrica para h(t)

# Calcular hazard usando muhaz
haz_est <- muhaz(
  times = tiempos,
  delta = rep(1, length(tiempos)), # 1 = evento observado (no censurados)
  bw.method = "global",           # método de suavizado
  min.time = min(tiempos),
  max.time = max(tiempos)
)
haz_est

# Convertir a data frame para ggplot
df_haz <- data.frame(
  tiempo = haz_est$est.grid,
  hazard = haz_est$haz.est
)
df_haz

# Definir cortes de las fases (por terciles del tiempo de hazard)
t_min <- min(df_haz$tiempo)
t_max <- max(df_haz$tiempo)

```

```

b1  <- t_min + (t_max - t_min) / 3
b2  <- t_min + 2 * (t_max - t_min) / 3

# Gráfico de hazard con franjas de fases
ggplot(df_haz, aes(x = tiempo, y = hazard)) +
  geom_rect(aes(xmin = t_min, xmax = b1, ymin = -Inf, ymax = Inf),
    fill = "lightblue", alpha = 0.2, inherit.aes = FALSE) +
  geom_rect(aes(xmin = b1, xmax = b2, ymin = -Inf, ymax = Inf),
    fill = "lightgreen", alpha = 0.2, inherit.aes = FALSE) +
  geom_rect(aes(xmin = b2, xmax = t_max, ymin = -Inf, ymax = Inf),
    fill = "lightpink", alpha = 0.2, inherit.aes = FALSE) +
  geom_line(color = "red", linewidth = 1) +
  annotate("text", x = (t_min + b1)/2, y = max(df_haz$hazard)*0.9,
    label = "Mortalidad\ntemprana", size = 3) +
  annotate("text", x = (b1 + b2)/2, y = max(df_haz$hazard)*0.9,
    label = "Vida útil\nestable", size = 3) +
  annotate("text", x = (b2 + t_max)/2, y = max(df_haz$hazard)*0.9,
    label = "Desgaste", size = 3) +
  labs(
    title = "Estimación no paramétrica de la tasa de falla h(t) con fases del ciclo de
vida",
    x    = "Tiempo",
    y    = "h(t) estimada"
  ) +
  theme_minimal()

# --- Estimación no paramétrica para R(t)

library(survival)
library(survminer)

# Datos
df_tiempos <- data.frame(tiempo_falla = tiempos)

```

```

# Todos son fallas observadas
df_surv <- df_tiempos |>
  dplyr::mutate(evento = 1)

# Estimador de Kaplan–Meier
fit_km <- survfit(Surv(time = tiempo_falla, event = evento) ~ 1,
  data = df_surv)

# Gráfica de R(t)
ggsurvplot(
  fit_km,
  data = df_surv,
  conf.int = TRUE,
  xlab = "Tiempo hasta la falla",
  ylab = "Probabilidad de supervivencia R(t)",
  ggtheme = theme_minimal(),
  surv.scale = "percent",
  legend = "none"
)

# (Opcional) Tabla con valores puntuales de R(t)
summary(fit_km)

#=====

# ----- Fase 2 — Selección y evaluación preliminar de
distribuciones de confiabilidad para modelar los tiempos de falla ---

#=====

# ----- F-2.1. Diagnóstico de forma con base en la distribución de
los datos -----

# 2.1.1 Resumen de forma con base en asimetría y curtosis

```

```
sk <- as.numeric(resumen_basico$asimetria)
ku <- as.numeric(resumen_basico$curtosis)
```

```
forma_asimetria <- dplyr::case_when(
  sk > 0.5 ~ "sesgo a la derecha",
  sk < -0.5 ~ "sesgo a la izquierda",
  TRUE    ~ "aprox. simétrica"
)
```

```
comportamiento_colas <- dplyr::case_when(
  ku > 0 ~ "colas pesadas / leptocúrtica",
  ku < 0 ~ "colas ligeras / platicúrtica",
  TRUE   ~ "mesocúrtica"
)
```

# 2.1.2 Resumen de la forma del hazard (inicio, medio, final)

```
grid <- haz_est$est.grid
h     <- haz_est$haz.est
```

```
q1 <- quantile(grid, 0.33)
q2 <- quantile(grid, 0.66)
```

```
haz_inicio <- mean(h[grid <= q1])
haz_medio  <- mean(h[grid > q1 & grid <= q2])
haz_final  <- mean(h[grid > q2])
```

```
haz_clasificacion <- dplyr::case_when(
  haz_inicio > haz_medio & haz_final > haz_medio ~
  "no_monotona_tipo_bañera",
  haz_inicio < haz_medio & haz_medio < haz_final ~ "creciente",
  haz_inicio > haz_medio & haz_medio > haz_final ~ "decreciente",
  TRUE                                           ~ "patron_complejo/indefinido"
)
```

### # 2.1.3 Tabla-resumen de diagnóstico de forma

```
resumen_forma <- tibble::tibble(  
  asimetria_valor      = sk,  
  asimetria_clasificacion = forma_asimetria,  
  curtosis_valor       = ku,  
  colas_clasificacion  = comportamiento_colas,  
  haz_inicio           = haz_inicio,  
  haz_medio            = haz_medio,  
  haz_final            = haz_final,  
  hazard_clasificacion = haz_clasificacion  
)
```

resumen\_forma

### # ----- F-2.2. Definición del catálogo de distribuciones candidatas -

```
distribuciones_trad <- c("Exponencial", "Weibull", "Lognormal", "Gamma")  
distribuciones_ext  <- c("Weibull modificada", "Log-logistic", "Generalizada  
Gamma")  
distribuciones_bath <- c("Beta-Weibull", "Power-Lindley", "Q-Weibull", "Weibull-  
Pareto")
```

### # ----- F-2.3. Clasificación preliminar del tipo de hazard -----

```
candidatos <- tibble::tibble(  
  categoria = c(  
    rep("Tradicional", length(distribuciones_trad)),  
    rep("Extensiva", length(distribuciones_ext)),  
    rep("No monótona / curva de bañera", length(distribuciones_bath))  
  ),  
  distribucion = c(distribuciones_trad, distribuciones_ext, distribuciones_bath)  
) |>  
  dplyr::mutate(  
    prioridad = dplyr::case_when(  
      #
```

```

# Si el hazard luce tipo bañera, se priorizan las no monótonas
categoria == "No monótona / curva de bañera" &
  haz_clasificacion == "no_monotona_tipo_bañera" ~ "Alta",
# Si es creciente/decreciente, las tradicionales/extensivas siguen siendo
relevantes
categoria == "Tradicional" &
  haz_clasificacion %in% c("creciente", "decreciente") ~ "Media",
categoria == "Extensiva" &
  haz_clasificacion %in% c("creciente", "decreciente",
"no_monotona_tipo_bañera") ~ "Media",
  TRUE ~ "Exploratoria"
)
)

```

candidatos

```

#=====

```

```

# ----- Fase 3: Ajuste paramétrico del modelo mediante
Estimación de Máxima Verosimilitud (MLE) -----

```

```

#=====

```

```

# ----- F-3.1. Ajuste de distribuciones clásicas mediante MLE
automático -----

```

```

library(fitdistrplus) # Ajuste de distribuciones clásicas
library(flexsurv)     # Modelos de supervivencia paramétricos
library(EnvStats)     # Funciones adicionales (opcional)
library(numDeriv)     # Hessiano numérico
library(tidyverse)

```

```

# --- Ajuste mle de distribuciones clásicas con fitdistrplus
# --- (Weibull, Lognormal, Gamma, Exponencial)

dists_clasicas <- c("weibull", "lnorm", "gamma", "exp")

ajustes_clasicos <- lapply(dists_clasicas, function(d) {
  fitdistrplus::fitdist(tiempos, distr = d)
})
names(ajustes_clasicos) <- dists_clasicas

# --- Tabla de parámetros, error estándar e IC95% para clásicas ---

extraer_info_mle <- function(nombre_dist, objeto_fit, nivel_conf = 0.95) {
  est <- objeto_fit$estimate
  se <- objeto_fit$sd
  z <- qnorm((1 + nivel_conf) / 2) # 1.96 para 95%

  ci_inf <- est - z * se
  ci_sup <- est + z * se

  tibble(
    distribucion = nombre_dist,
    parametro = names(est),
    estimacion = as.numeric(est),
    error_est = as.numeric(se),
    ci_inf = as.numeric(ci_inf),
    ci_sup = as.numeric(ci_sup)
  )
}

tabla_param_clasicas_rep <- tabla_param_clasicas |>
  dplyr::mutate(
    distribucion = dplyr::recode(
      distribucion,

```

```

    "weibull" = "Weibull",
    "lnorm"   = "Lognormal",
    "gamma"   = "Gamma",
    "exp"     = "Exponencial"
  ),
  estimacion = round(estimacion, 3),
  error_est  = round(error_est, 3),
  ci_inf     = round(ci_inf, 3),
  ci_sup     = round(ci_sup, 3)
)

```

tabla\_param\_clasicas\_rep

# ----- F-3.2. Ajuste de distribuciones no monótonas -----

```

# --- Ajuste mle de la distribución beta–weibull
#   (para curvas tipo bañera u otras formas flexibles)

```

```

# --- Funciones base Weibull

```

```

Fw_weibull <- function(x, c, gamma) {
  1 - exp(-(x / gamma)^c)
}

```

```

fw_weibull <- function(x, c, gamma) {
  (c / gamma) * (x / gamma)^(c - 1) * exp(-(x / gamma)^c)
}

```

```

# --- PDF, CDF, supervivencia y hazard Beta–Weibull
# el segundo parámetro se llama 'b' para no chocar con la función beta()

```

```

dbetaweibull <- function(x, alpha, b, c, gamma) {
  F0 <- Fw_weibull(x, c, gamma)
  f0 <- fw_weibull(x, c, gamma)
}

```

```
(1 / beta(alpha, b)) * f0 * F0^(alpha - 1) * (1 - F0)^(b - 1)
}
```

```
pbetaweibull <- function(x, alpha, b, c, gamma) {
  F0 <- Fw_weibull(x, c, gamma)
  pbeta(F0, alpha, b)
}
```

```
sbetaweibull <- function(x, alpha, b, c, gamma) {
  1 - pbetaweibull(x, alpha, b, c, gamma)
}
```

```
hbetaweibull <- function(x, alpha, b, c, gamma) {
  g <- dbetaweibull(x, alpha, b, c, gamma)
  Sx <- sbetaweibull(x, alpha, b, c, gamma)
  g / Sx
}
```

```
# --- Log-verosimilitud (datos no censurados)
```

```
loglik_betaweibull <- function(par_log, x) {
  alpha <- exp(par_log[1])
  b <- exp(par_log[2])
  c <- exp(par_log[3])
  gamma <- exp(par_log[4])

  g_vals <- dbetaweibull(x, alpha, b, c, gamma)
  g_vals[g_vals < 1e-12] <- 1e-12 # evitar log(0)

  -sum(log(g_vals)) # negativo de la log-verosimilitud
}
```

```

# --- Ajuste por MLE con optim

start_par <- log(c(alpha = 1, b = 1, c = 1, gamma = mean(tiempos)))

fit_bw <- optim(
  par    = start_par,
  fn     = loglik_betaweibull,
  x      = tiempos,
  method = "BFGS",
  hessian = TRUE
)

fit_bw$convergence # 0 = convergió bien

# Parámetros en escala original
est_par_bw <- exp(fit_bw$par)
names(est_par_bw) <- c("alpha", "beta", "c", "gamma") # nombre externo "beta"
para tabla

# --- Errores estándar e IC95% para Beta–Weibull

vcov_log_bw <- solve(fit_bw$hessian) # var-cov en escala log
se_bw      <- sqrt(diag(vcov_log_bw)) * est_par_bw

z <- qnorm(0.975)
ciL_bw <- est_par_bw - z * se_bw
ciU_bw <- est_par_bw + z * se_bw

tabla_param_bw <- tibble(
  distribucion = "beta-weibull",
  parametro    = names(est_par_bw),
  estimacion   = as.numeric(est_par_bw),
  error_est    = as.numeric(se_bw),

```

```

ci_inf    = as.numeric(ciL_bw),
ci_sup    = as.numeric(ciU_bw)
)

tabla_param_bw

# ----- F-3.3. Obtención de parámetros estimados -----

# --- Tabla única de parámetros para todas las distribuciones

tabla_parametros_todas <- bind_rows(
  tabla_param_clasicas,
  tabla_param_bw
)

tabla_parametros_todas

# ----- F-3.4. Cálculo de métricas de desempeño del modelo -----

# Tiempos de referencia para las métricas
t0 <- 40 # tiempo de referencia (puedes cambiarlo)
t1 <- 60 # tiempo futuro para  $R(t_1 | T > t_0)$ 
upper_int <- max(tiempos) * 5 # límite superior para integraciones numéricas

# --- 1) Supervivencia y hazard teórica para modelos clásicos -----

R_clasica <- function(dist_name, fit_obj, t) {
  est <- fit_obj$estimate

  if (dist_name == "weibull") {
    1 - pweibull(t, shape = est["shape"], scale = est["scale"])
  } else if (dist_name == "lnorm") {
    1 - plnorm(t, meanlog = est["meanlog"], sdlog = est["sdlog"])
  } else if (dist_name == "gamma") {

```

```

    1 - pgamma(t, shape = est["shape"], rate = est["rate"])
  } else if (dist_name == "exp") {
    1 - pexp(t, rate = est["rate"])
  } else {
    stop("Distribución no implementada en R_clasica()")
  }
}

f_clasica <- function(dist_name, fit_obj, t) {
  est <- fit_obj$estimate

  if (dist_name == "weibull") {
    dweibull(t, shape = est["shape"], scale = est["scale"])
  } else if (dist_name == "lnorm") {
    dlnorm(t, meanlog = est["meanlog"], sdlog = est["sdlog"])
  } else if (dist_name == "gamma") {
    dgamma(t, shape = est["shape"], rate = est["rate"])
  } else if (dist_name == "exp") {
    dexp(t, rate = est["rate"])
  } else {
    stop("Distribución no implementada en f_clasica()")
  }
}

# MTTF teórico para modelos clásicos (fórmulas cerradas)
mttf_clasica <- function(dist_name, fit_obj) {
  est <- fit_obj$estimate

  if (dist_name == "weibull") {
    scale <- est["scale"]; shape <- est["shape"]
    scale * gamma(1 + 1 / shape)
  } else if (dist_name == "lnorm") {
    meanlog <- est["meanlog"]; sdlog <- est["sdlog"]
    exp(meanlog + 0.5 * sdlog^2)
  }
}

```

```

} else if (dist_name == "gamma") {
  est["shape"] / est["rate"]
} else if (dist_name == "exp") {
  1 / est["rate"]
} else {
  NA_real_
}
}

# MRL(t0) =  $\int_{t0}^{\infty} R(u) du / R(t0)$ 
mrl_clasica <- function(dist_name, fit_obj, t0, upper = upper_int) {
  R_fun <- function(u) R_clasica(dist_name, fit_obj, u)
  R_t0 <- R_clasica(dist_name, fit_obj, t0)
  if (R_t0 <= 0) return(NA_real_)

  integ <- integrate(R_fun, lower = t0, upper = upper)
  integ$value / R_t0
}

metricas_clasica <- function(dist_name, fit_obj, t0, t1) {
  R_t0 <- R_clasica(dist_name, fit_obj, t0)
  R_t1 <- R_clasica(dist_name, fit_obj, t1)
  f_t0 <- f_clasica(dist_name, fit_obj, t0)
  h_t0 <- f_t0 / max(R_t0, 1e-12)
  R_cond <- if (R_t0 > 0) R_t1 / R_t0 else NA_real_

  tibble::tibble(
    distribucion = dist_name,
    t0           = t0,
    t1           = t1,
    R_t0         = R_t0,
    h_t0         = h_t0,
    R_cond_t1_t0 = R_cond,
    MTTF         = mttf_clasica(dist_name, fit_obj),

```

```

    MRL_t0    = mrl_clasica(dist_name, fit_obj, t0)
  )
}

```

```

metricas_clasicos <- purrr::map2_dfr(
  .x = names(ajustes_clasicos),
  .y = ajustes_clasicos,
  .f = ~ metricas_clasica(.x, .y, t0 = t0, t1 = t1)
) |>
dplyr::mutate(
  distribucion = dplyr::recode(
    distribucion,
    "weibull" = "Weibull",
    "lnorm"   = "Lognormal",
    "gamma"   = "Gamma",
    "exp"     = "Exponencial"
  )
)

```

# --- 2) Métricas para la distribución Beta–Weibull -----

```

alpha_hat <- est_par_bw["alpha"]
beta_hat  <- est_par_bw["beta"]
c_hat     <- est_par_bw["c"]
gamma_hat <- est_par_bw["gamma"]

```

```

R_bw <- function(t) {
  sbetaweibull(t,
    alpha = alpha_hat,
    b     = beta_hat,
    c     = c_hat,
    gamma = gamma_hat)
}

```

```

h_bw <- function(t) {
  hbetaweibull(t,
    alpha = alpha_hat,
    b      = beta_hat,
    c      = c_hat,
    gamma = gamma_hat)
}

# MTTF =  $\int_0^\infty R(t) dt$ 
mttf_bw <- function(upper = upper_int) {
  integ <- integrate(function(u) R_bw(u), lower = 0, upper = upper)
  integ$value
}

# MRL( $t_0$ ) =  $\int_{t_0}^\infty R(u) du / R(t_0)$ 
mrl_bw <- function(t0, upper = upper_int) {
  R_t0 <- R_bw(t0)
  if (R_t0 <= 0) return(NA_real_)
  integ <- integrate(function(u) R_bw(u), lower = t0, upper = upper)
  integ$value / R_t0
}

R_t0_bw <- R_bw(t0)
R_t1_bw <- R_bw(t1)
h_t0_bw <- h_bw(t0)
R_cond_bw <- if (R_t0_bw > 0) R_t1_bw / R_t0_bw else NA_real_

metricas_bw <- tibble::tibble(
  distribucion = "Beta-Weibull",
  t0           = t0,
  t1           = t1,
  R_t0         = R_t0_bw,
  h_t0         = h_t0_bw,
  R_cond_t1_t0 = R_cond_bw,

```

```

MTTF      = mttf_bw(),
MRL_t0    = mrl_bw(t0)
)

```

# --- 3) Tabla conjunta de métricas -----

```

tabla_metricas_modelos <- dplyr::bind_rows(
  metricas_clasicos,
  metricas_bw
)

```

tabla\_metricas\_modelos

# --- 4) Calculo de la función de riesgo teórica  $h(t)$  de cada modelo en un grid común de tiempos:

# --- se evaluarán las funciones  $f(t)$ ,  $F(t)$ ,  $R(t)$  y, finalmente,  $h(t)$  de cada distribución.

# Grid común de tiempos

```

t_grid <- seq(
  from = min(tiempos),
  to   = max(tiempos),
  length.out = 200
)

```

# Hazard teórico para distribuciones clásicas

```

hazard_clasica <- function(dist_name, fit_obj, tiempos_eval) {
  est <- fit_obj$estimate

  if (dist_name == "weibull") {
    shape <- est["shape"]
    scale <- est["scale"]
    f_t   <- dweibull(tiempos_eval, shape = shape, scale = scale)
    F_t   <- pweibull(tiempos_eval, shape = shape, scale = scale)
  }
}

```

```

} else if (dist_name == "lnorm") {
  meanlog <- est["meanlog"]
  sdlog <- est["sdlog"]
  f_t <- dlnorm(tiempos_eval, meanlog = meanlog, sdlog = sdlog)
  F_t <- plnorm(tiempos_eval, meanlog = meanlog, sdlog = sdlog)

} else if (dist_name == "gamma") {
  shape <- est["shape"]
  rate <- est["rate"]
  f_t <- dgamma(tiempos_eval, shape = shape, rate = rate)
  F_t <- pgamma(tiempos_eval, shape = shape, rate = rate)

} else if (dist_name == "exp") {
  rate <- est["rate"]
  f_t <- dexp(tiempos_eval, rate = rate)
  F_t <- pexp(tiempos_eval, rate = rate)

} else {
  stop("Distribución no implementada en hazard_clasica()")
}

R_t <- 1 - F_t
h_t <- f_t / pmax(R_t, 1e-12) # evitar división por 0

tibble(
  tiempo = tiempos_eval,
  hazard = h_t,
  distribucion = dist_name
)
}

hazards_clasicas <- purrr::map2_dfr(
  .x = names(ajustes_clasicos),

```

```

.y = ajustes_clasicos,
.f = ~ hazard_clasica(.x, .y, t_grid)
)

# Hazard teórica Beta–Weibull usando parámetros MLE
alpha_hat <- est_par_bw["alpha"]
beta_hat <- est_par_bw["beta"]
c_hat <- est_par_bw["c"]
gamma_hat <- est_par_bw["gamma"]

haz_bw_df <- tibble(
  tiempo = t_grid,
  hazard = hbetaweibull(t_grid,
                        alpha = alpha_hat,
                        b = beta_hat, # segundo parámetro
                        c = c_hat,
                        gamma = gamma_hat),
  distribucion = "beta-weibull"
)

# Unir TODAS las hazards en un solo data frame
hazards_todas <- bind_rows(
  hazards_clasicas,
  haz_bw_df
)

# Renombrar para que la leyenda salga "bonita"
hazards_todas <- hazards_todas |>
mutate(distribucion = recode(distribucion,
                             "weibull" = "Weibull",
                             "lnorm" = "Lognormal",
                             "gamma" = "Gamma",
                             "exp" = "Exponencial",
                             "beta-weibull" = "Beta–Weibull"))

```

```

#=====

# ----- Fase 4: Evaluación de la calidad del modelo -----

#=====

# ----- F-4.1. Comparación de criterios de información -----

# install.packages("ADGofTest")

library(ADGofTest) # para ad.test (Anderson–Darling)

# --- Tabla AIC, BIC y log-likelihood para todos los modelos

# Modelos clásicos (fitdistrplus guarda loglik, aic, bic)
info_clasicas <- tibble::tibble(
  distribucion_bruta = names(ajustes_clasicos),
  loglik = sapply(ajustes_clasicos, function(f) f$loglik),
  AIC = sapply(ajustes_clasicos, function(f) f$aic),
  BIC = sapply(ajustes_clasicos, function(f) f$bic)
)

# Beta–Weibull: usamos la log-verosimilitud de fit_bw
n_obs <- length(tiempos)
k_bw <- length(fit_bw$par) # nº de parámetros
loglik_bw <- -fit_bw$value # optim minimiza -logLik
AIC_bw <- -2 * loglik_bw + 2 * k_bw
BIC_bw <- -2 * loglik_bw + k_bw * log(n_obs)

info_bw <- tibble::tibble(
  distribucion_bruta = "beta-weibull",
  loglik = loglik_bw,
  AIC = AIC_bw,
  BIC = BIC_bw
)

```

```
)
```

```
# Unir todo y renombrar distribuciones (usamos dplyr::select explícito)
```

```
info_modelos_ABL <- bind_rows(info_clasicas, info_bw) |>
```

```
  dplyr::mutate(
```

```
    distribucion = dplyr::recode(
```

```
      distribucion_bruta,
```

```
      "weibull"    = "Weibull",
```

```
      "lnorm"      = "Lognormal",
```

```
      "gamma"      = "Gamma",
```

```
      "exp"        = "Exponencial",
```

```
      "beta-weibull" = "Beta-Weibull"
```

```
    )
```

```
  ) |>
```

```
  dplyr::select(distribucion, loglik, AIC, BIC)
```

```
info_modelos_ABL # <- tabla de criterios de información
```

```
# ----- F-4.2. Pruebas de bondad de ajuste (K-S y A-D) -----
```

```
gof_clasicos <- fitdistrplus::gofstat(ajustes_clasicos)
```

```
ks_clas <- gof_clasicos$ks
```

```
ad_clas <- gof_clasicos$ad
```

```
gof_tabla_clas <- tibble::tibble(
```

```
  distribucion = dplyr::recode(
```

```
    names(ks_clas),
```

```
    "weibull" = "Weibull",
```

```
    "lnorm"   = "Lognormal",
```

```
    "gamma"   = "Gamma",
```

```
    "exp"     = "Exponencial"
```

```
  ),
```

```

KS_estadistico = as.numeric(ks_clas),
AD_estadistico = as.numeric(ad_clas)
)

# --- KS y AD para Beta-Weibull -----

# 1) F_hat(t) con parámetros MLE (probabilidades teóricas)
alpha_hat <- est_par_bw["alpha"]
beta_hat  <- est_par_bw["beta"]
c_hat    <- est_par_bw["c"]
gamma_hat <- est_par_bw["gamma"]

F_bw <- pbetaweibull(
  tiempos,
  alpha = alpha_hat,
  b      = beta_hat,
  c      = c_hat,
  gamma = gamma_hat
)

# 2) Pruebas sobre F_bw ~ U(0,1)
ks_bw <- stats::ks.test(F_bw, "punif")      # Kolmogorov-Smirnov
ad_bw <- ADGofTest::ad.test(F_bw, null = "punif") # Anderson-Darling

gof_tabla_bw <- tibble::tibble(
  distribucion = "Beta-Weibull",
  KS_estadistico = as.numeric(ks_bw$statistic),
  AD_estadistico = as.numeric(ad_bw$statistic)
)

# --- Tabla de bondad de ajuste para todos los modelos ---

gof_modelos <- bind_rows(
  gof_tabla_clas,

```

```
gof_tabla_bw  
)
```

```
gof_modelos
```

```
# ----- F-4.2. (Version 2, solo se agrego el p-value) Pruebas de  
bondad de ajuste (K-S y A-D) -----  
# ----- No se utilizo
```

```
# 1) Función CDF teórica para los modelos clásicos
```

```
# CDF teórica para cada modelo clásico
```

```
F_clasica <- function(dist_name, fit_obj, x) {  
  est <- fit_obj$estimate  
  
  if (dist_name == "weibull") {  
    pweibull(x, shape = est["shape"], scale = est["scale"])  
  } else if (dist_name == "lnorm") {  
    plnorm(x, meanlog = est["meanlog"], sdlog = est["sdlog"])  
  } else if (dist_name == "gamma") {  
    pgamma(x, shape = est["shape"], rate = est["rate"])  
  } else if (dist_name == "exp") {  
    pexp(x, rate = est["rate"])  
  } else {  
    stop("Distribución no implementada en F_clasica()")  
  }  
}
```

```
# 2) K-S y A-D con estadístico y p-value para modelos clásicos
```

```
library(ADGofTest)
```

```
gof_clasicos_p <- purrr::map2_dfr(  
  .x = names(ajustes_clasicos),
```

```

.y = ajustes_clasicos,
.f = function(nombre, fit_obj) {
  # Probabilidades teóricas bajo el modelo ajustado
  U <- F_clasica(nombre, fit_obj, tiempos)

  ks_res <- stats::ks.test(U, "punif")
  ad_res <- ADGofTest::ad.test(U, null = "punif")

  tibble::tibble(
    distribucion = nombre,
    KS_estadistico = as.numeric(ks_res$statistic),
    KS_pvalue      = ks_res$p.value,
    AD_estadistico = as.numeric(ad_res$statistic),
    AD_pvalue      = ad_res$p.value
  )
}
)|>
dplyr::mutate(
  distribucion = dplyr::recode(
    distribucion,
    "weibull" = "Weibull",
    "lnorm"   = "Lognormal",
    "gamma"   = "Gamma",
    "exp"     = "Exponencial"
  )
)

```

# 3) K–S y A–D con p-values para Beta–Weibull

# Ya definidos F\_bw, ks\_bw, ad\_bw

```

gof_bw_p <- tibble::tibble(
  distribucion = "Beta–Weibull",
  KS_estadistico = as.numeric(ks_bw$statistic),

```

```

KS_pvalue      = ks_bw$p.value,
AD_estadistico = as.numeric(ad_bw$statistic),
AD_pvalue      = ad_bw$p.value
)

```

# 4) Tabla final de bondad de ajuste con p-values

```

gof_modelos_p <- dplyr::bind_rows(
  gof_clasicos_p,
  gof_bw_p
)

```

```
gof_modelos_p
```

```

#=====
# ----- F-4.3. Contraste visual sistemático -----

```

```

# F-4.3.1 Supervivencia empírica vs. teórica
# F-4.3.2 Hazard empírica vs. hazards teóricas
# F-4.3.3 PDF teórica vs. histograma de datos
# F-4.3.4 QQ-plot y PP-plot de los tiempos de falla
#=====

```

```
# ----- F-4.3.1 Supervivencia empírica (KM) vs teórica -----
```

```
# Supervivencia empírica Kaplan–Meier (ya tienes fit_km de Fase 1)
```

```

km_df <- data.frame(
  tiempo = fit_km$time,
  surv   = fit_km$surv
)

```

```
# Supervivencia teórica para modelos clásicos
```

```

superv_clasica <- function(dist_name, fit_obj, tiempos_eval) {
  est <- fit_obj$estimate

```

```

if (dist_name == "weibull") {
  shape <- est["shape"]; scale <- est["scale"]
  R_t <- 1 - pweibull(tiempos_eval, shape = shape, scale = scale)

} else if (dist_name == "lnorm") {
  meanlog <- est["meanlog"]; sdlog <- est["sdlog"]
  R_t <- 1 - plnorm(tiempos_eval, meanlog = meanlog, sdlog = sdlog)

} else if (dist_name == "gamma") {
  shape <- est["shape"]; rate <- est["rate"]
  R_t <- 1 - pgamma(tiempos_eval, shape = shape, rate = rate)

} else if (dist_name == "exp") {
  rate <- est["rate"]
  R_t <- 1 - pexp(tiempos_eval, rate = rate)

} else {
  stop("Distribución no implementada en superv_clasica()")
}

tibble::tibble(
  tiempo = tiempos_eval,
  supervivencia = R_t,
  distribucion = dist_name
)
}

t_grid_sup <- seq(min(tiempos), max(tiempos), length.out = 200)

sup_clasicas <- purrr::map2_dfr(
  .x = names(ajustes_clasicos),
  .y = ajustes_clasicos,
  .f = ~ superv_clasica(.x, .y, t_grid_sup)
)

```

```
)
```

```
sup_bw <- tibble::tibble(  
  tiempo      = t_grid_sup,  
  supervivencia = sbetaweibull(t_grid_sup,  
                                alpha = alpha_hat,  
                                b      = beta_hat,  
                                c      = c_hat,  
                                gamma = gamma_hat),  
  distribucion = "beta-weibull"  
)
```

```
sup_todas <- dplyr::bind_rows(sup_clasicas, sup_bw) |>  
  dplyr::mutate(  
    distribucion = dplyr::recode(  
      distribucion,  
      "weibull"    = "Weibull",  
      "lnorm"     = "Lognormal",  
      "gamma"     = "Gamma",  
      "exp"       = "Exponencial",  
      "beta-weibull" = "Beta-Weibull"  
    )  
  )
```

```
# Gráfico: Kaplan–Meier vs curvas teóricas
```

```
km_df_plot <- km_df |>  
  dplyr::mutate(distribucion = "Empírica (KM)")  
  
ggplot() +  
  geom_step(data = km_df_plot,  
            aes(x = tiempo, y = surv, color = distribucion),  
            linewidth = 1) +  
  geom_line(data = sup_todas,  
            aes(x = tiempo, y = supervivencia, color = distribucion),
```

```

        linewidth = 0.9) +
scale_color_manual(
  name = "Modelo",
  values = c(
    "Empírica (KM)" = "black",
    "Beta-Weibull" = "#F8766D",
    "Exponencial" = "#7CAE00",
    "Gamma" = "#00BFC4",
    "Lognormal" = "#00B0F6",
    "Weibull" = "#C77CFF"
  )
) +
labs(
  title = "Comparación de supervivencia: empírica (KM) vs modelos
paramétricos",
  x = "Tiempo hasta la falla",
  y = "R(t)"
) +
theme_minimal()

# ----- F-4.3.2 Hazard empírica vs hazards teóricas -----

# df_haz: hazard empírica (mu_haz) calculada en la Fase 1
# hazards_todas: hazards teóricas calculadas en F-3.4 (ya definido antes)

df_haz_plot <- df_haz |>
  dplyr::mutate(distribucion = "Hazard empírica")

ggplot() +
  geom_line(data = df_haz_plot,
    aes(x = tiempo, y = hazard, color = distribucion),
    linewidth = 1.1) +
  geom_line(data = hazards_todas,
    aes(x = tiempo, y = hazard, color = distribucion),

```

```

        linewidth = 0.9, alpha = 0.9) +
scale_color_manual(
  name = "Modelo",
  values = c(
    "Hazard empírica" = "black",
    "Beta-Weibull" = "#F8766D",
    "Exponencial" = "#7CAE00",
    "Gamma" = "#00BFC4",
    "Lognormal" = "#00B0F6",
    "Weibull" = "#C77CFF"
  )
) +
labs(
  title = "Hazard empírica vs hazards teóricas de los modelos ajustados",
  x = "Tiempo",
  y = "h(t)"
) +
theme_minimal() +
scale_y_continuous(limits = c(0, 0.15))

```

# ----- F-4.3.3 PDF teórica vs histograma de datos -----

# Grid de tiempos para evaluar las densidades

```
t_grid_dens <- seq(min(tiempos), max(tiempos), length.out = 300)
```

# Densidades de modelos clásicos

```
dens_clasica <- function(dist_name, fit_obj, tiempos_eval) {
  est <- fit_obj$estimate
```

```

  if (dist_name == "weibull") {
    f_t <- dweibull(tiempos_eval,
      shape = est["shape"],
      scale = est["scale"])
  } else if (dist_name == "lnorm") {

```

```

f_t <- dlnorm(tiempos_eval,
              meanlog = est["meanlog"],
              sdlog   = est["sdlog"])
} else if (dist_name == "gamma") {
  f_t <- dgamma(tiempos_eval,
                shape = est["shape"],
                rate   = est["rate"])
} else if (dist_name == "exp") {
  f_t <- dexp(tiempos_eval,
              rate = est["rate"])
} else {
  stop("Distribución no implementada en dens_clasica()")
}

tibble(
  tiempo      = tiempos_eval,
  densidad    = f_t,
  distribucion = dist_name
)
}

dens_clasicas <- purrr::map2_dfr(
  .x = names(ajustes_clasicos),
  .y = ajustes_clasicos,
  .f = ~ dens_clasica(.x, .y, t_grid_dens)
)

# Densidad Beta-Weibull
dens_bw <- tibble(
  tiempo      = t_grid_dens,
  densidad    = dbetaweibull(t_grid_dens,
                              alpha = alpha_hat,
                              b      = beta_hat,
                              c      = c_hat,

```

```

        gamma = gamma_hat),
  distribucion = "beta-weibull"
)

dens_todas <- bind_rows(dens_clasicas, dens_bw) |>
  mutate(distribucion = dplyr::recode(
    distribucion,
    "weibull"    = "Weibull",
    "lnorm"      = "Lognormal",
    "gamma"      = "Gamma",
    "exp"        = "Exponencial",
    "beta-weibull" = "Beta-Weibull"
  ))

ggplot() +
  geom_histogram(
    data = tibble(tiempos = tiempos),
    aes(x = tiempos, y = after_stat(density)),
    bins = 15,
    fill = "grey80",
    color = "black"
  ) +
  geom_line(
    data = dens_todas,
    aes(x = tiempo, y = densidad, color = distribucion),
    linewidth = 1
  ) +
  labs(
    title = "Histograma de tiempos y funciones de densidad ajustadas",
    x     = "Tiempo",
    y     = "Densidad",
    color = "Distribución"
  ) +
  theme_minimal() +

```

```

scale_y_continuous(limits = c(0, 0.060))

# =====
# F-4.3.4 QQ-plot y PP-plot conjuntos
# Modelos: Weibull, Lognormal, Gamma, Exponencial, Beta-Weibull
# =====

# 1) Probabilidades empíricas y tiempos ordenados
tiempos_ord <- sort(tiempos)
n          <- length(tiempos_ord)
p_emp      <- (1:n - 0.5) / n # prob. empíricas tipo Gringorten

# 2) Parámetros MLE ya calculados
# - modelos clásicos en 'ajustes_clasicos'
# - Beta-Weibull en 'est_par_bw'
alpha_hat <- est_par_bw["alpha"]
beta_hat  <- est_par_bw["beta"]
c_hat     <- est_par_bw["c"]
gamma_hat <- est_par_bw["gamma"]

# 3) Función cuantíl Beta-Weibull (no existe en R base, la construimos)
qbetaweibull <- function(p, alpha, b, c, gamma,
                        lower = 1e-8,
                        upper = max(tiempos) * 5) {
  sapply(p, function(pp) {
    uniroot(
      f = function(x) pbetaweibull(x, alpha, b, c, gamma) - pp,
      lower = lower,
      upper = upper
    )$root
  })
}

# 4) Cuantiles teóricos por modelo (para QQ-plot)

```

```

q_teorico_modelo <- function(nombre_bruto, p_vec) {
  if (nombre_bruto == "weibull") {
    est <- ajustes_clasicos[["weibull"]]$estimate
    qweibull(p_vec, shape = est["shape"], scale = est["scale"])

  } else if (nombre_bruto == "lnorm") {
    est <- ajustes_clasicos[["lnorm"]]$estimate
    qlnorm(p_vec, meanlog = est["meanlog"], sdlog = est["sdlog"])

  } else if (nombre_bruto == "gamma") {
    est <- ajustes_clasicos[["gamma"]]$estimate
    qgamma(p_vec, shape = est["shape"], rate = est["rate"])

  } else if (nombre_bruto == "exp") {
    est <- ajustes_clasicos[["exp"]]$estimate
    qexp(p_vec, rate = est["rate"])

  } else if (nombre_bruto == "beta-weibull") {
    qbetaweibull(p_vec,
      alpha = alpha_hat,
      b     = beta_hat,
      c     = c_hat,
      gamma = gamma_hat)

  } else {
    stop("Modelo no implementado en q_teorico_modelo()")
  }
}

```

# 5) CDF teórica por modelo (para PP-plot)

```

F_teorica_modelo <- function(nombre_bruto, t_vec) {
  if (nombre_bruto == "weibull") {
    est <- ajustes_clasicos[["weibull"]]$estimate
    pweibull(t_vec, shape = est["shape"], scale = est["scale"])
  }
}

```

```

} else if (nombre_bruto == "lnorm") {
  est <- ajustes_clasicos[["lnorm"]]$estimate
  plnorm(t_vec, meanlog = est["meanlog"], sdlog = est["sdlog"])

} else if (nombre_bruto == "gamma") {
  est <- ajustes_clasicos[["gamma"]]$estimate
  pgamma(t_vec, shape = est["shape"], rate = est["rate"])

} else if (nombre_bruto == "exp") {
  est <- ajustes_clasicos[["exp"]]$estimate
  pexp(t_vec, rate = est["rate"])

} else if (nombre_bruto == "beta-weibull") {
  pbetaweibull(t_vec,
    alpha = alpha_hat,
    b      = beta_hat,
    c      = c_hat,
    gamma = gamma_hat)

} else {
  stop("Modelo no implementado en F_teorica_modelo()")
}
}

# 6) Nombres internos y bonitos de los modelos
modelos_brutos <- c("weibull", "lnorm", "gamma", "exp", "beta-weibull")
modelos_etiquet <- c("Weibull", "Lognormal", "Gamma", "Exponencial", "Beta-
Weibull")

# ----- QQ-plot conjunto -----
qq_df <- purrr::map2_dfr(
  .x = modelos_brutos,
  .y = modelos_etiquet,

```

```

.f = function(m_bruto, m_label) {
  q_teo <- q_teorico_modelo(m_bruto, p_emp)
  tibble::tibble(
    modelo = m_label,
    q_teorico = q_teo,
    q_empirico = tiempos_ord
  )
}
)

ggplot(qq_df, aes(x = q_teorico, y = q_empirico, color = modelo)) +
  geom_point(alpha = 0.7, size = 1.8) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  labs(
    title = "QQ-plot conjunto: modelos clásicos y Beta–Weibull",
    x    = "Cuantiles teóricos",
    y    = "Cuantiles muestrales"
  ) +
  theme_minimal()

# ----- PP-plot conjunto -----
pp_df <- purrr::map2_dfr(
  .x = modelos_brutos,
  .y = modelos_etiquet,
  .f = function(m_bruto, m_label) {
    F_teo <- F_teorica_modelo(m_bruto, tiempos_ord)
    tibble::tibble(
      modelo  = m_label,
      F_teorica = F_teo,
      F_empirica = p_emp
    )
  }
)
)

```

```
ggplot(pp_df, aes(x = F_teorica, y = F_empirica, color = modelo)) +
  geom_point(alpha = 0.7, size = 1.8) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  labs(
    title = "PP-plot conjunto: modelos clásicos y Beta–Weibull",
    x    = "Probabilidades teóricas",
    y    = "Probabilidades empíricas"
  ) +
  theme_minimal()
```

# ----- FIGURA: Hazard tipo bañera con t1 y t2 -----

```
library(ggplot2)
```

```
# Valores estimados previamente
```

```
t1 <- 17    # fin de mortalidad temprana (aprox. 15–18)
```

```
t2 <- 58    # inicio de desgaste (aprox. 55–60)
```

```
df_haz_plot <- df_haz |>
```

```
  dplyr::mutate(distribucion = "Hazard empírica")
```

```
haz_bw_curve <- tibble::tibble(
```

```
  tiempo = t_grid,
```

```
  hazard = hbetaweibull(
```

```
    t_grid,
```

```
    alpha = alpha_hat,
```

```
    b     = beta_hat,
```

```
    c     = c_hat,
```

```
    gamma = gamma_hat
```

```
  ),
```

```
  distribucion = "Hazard Beta–Weibull"
```

```
)
```

```
ggplot() +
```

```

geom_line(data = df_haz_plot,
          aes(x = tiempo, y = hazard),
          color = "red", linewidth = 1.1) +

geom_line(data = haz_bw_curve,
          aes(x = tiempo, y = hazard),
          color = "#0000CD", linewidth = 1) +

# Líneas verticales t1 y t2
geom_vline(xintercept = t1, linetype = "dashed", color = "#00CD00", size = 0.8)
+
geom_vline(xintercept = t2, linetype = "dashed", color = "#698B22", size = 0.8)
+

# Etiquetas
annotate("text", x = t1, y = max(df_haz$hazard)*0.9,
         label = "t1: Fin de mortalidad temprana",
         color = "#00CD00", angle = 90, vjust = -0.5, size = 4) +

annotate("text", x = t2, y = max(df_haz$hazard)*0.9,
         label = "t2: Inicio de desgaste",
         color = "#698B22", angle = 90, vjust = -0.5, size = 4) +

labs(
  title = "Bathtub hazard function: comparación empírica vs. Beta–Weibull",
  x = "Tiempo",
  y = "h(t)"
) +
theme_minimal() +
scale_y_continuous(limits = c(0, 0.15))

```

```

#=====
# Gráfica de la vida residual media MRL(t) – modelo Beta–Weibull
# (correr después de haber definido: tiempos, est_par_bw)
#=====

library(tidyverse)

# 1) Parámetros MLE (ya los tienes, sólo los retomamos)
alpha_hat <- est_par_bw["alpha"]
beta_hat <- est_par_bw["beta"]
c_hat <- est_par_bw["c"]
gamma_hat <- est_par_bw["gamma"]

# 2) Supervivencia Beta–Weibull (si ya la tienes definida, omite esta parte)
Fw_weibull <- function(x, c, gamma) {
  1 - exp(-(x / gamma)^c)
}

pbetaweibull <- function(x, alpha, b, c, gamma) {
  F0 <- Fw_weibull(x, c, gamma)
  pbeta(F0, alpha, b)
}

sbetaweibull <- function(x, alpha, b, c, gamma) {
  1 - pbetaweibull(x, alpha, b, c, gamma)
}

R_bw <- function(t) {
  sbetaweibull(t,
    alpha = alpha_hat,
    b = beta_hat,
    c = c_hat,
    gamma = gamma_hat)
}

```

# 3) Definimos MRL(t) con protección de errores en integrate()

```
upper_int <- max(tiempos) * 5 # por ejemplo 5 veces el máximo observado
```

```
mrl_bw_puntual <- function(t0, upper = upper_int) {  
  R_t0 <- R_bw(t0)  
  
  # Si la supervivencia es ~0 o no es finita, devolvemos NA  
  if (!is.finite(R_t0) || R_t0 <= 0) return(NA_real_)  
  
  res <- tryCatch(  
    integrate(function(u) R_bw(u), lower = t0, upper = upper),  
    error = function(e) NA_real_  
  )  
  
  if (is.list(res)) {  
    return(res$value / R_t0)  
  } else {  
    return(NA_real_)  
  }  
}
```

# Versión vectorizada

```
mrl_bw_vec <- function(t_vec, upper = upper_int) {  
  sapply(t_vec, mrl_bw_puntual, upper = upper)  
}
```

# 4) Grid de tiempos para la curva MRL(t)

```
t_grid_mrl <- seq(  
  from = min(tiempos),  
  to   = max(tiempos),  
  length.out = 100 # 100 puntos es suficiente y más estable
```

```
)
```

```
mrl_vals <- mrl_bw_vec(t_grid_mrl)
```

```
df_mrl <- tibble(  
  tiempo = t_grid_mrl,  
  MRL    = mrl_vals  
)
```

```
# 5) Gráfica de MRL(t)
```

```
ggplot(df_mrl, aes(x = tiempo, y = MRL)) +  
  geom_line(linewidth = 1, color = "red") +  
  labs(  
    title = "Vida residual media MRL(t) bajo el modelo Beta–Weibull",  
    x     = "Tiempo t",  
    y     = "MRL(t)"  
  ) +  
  theme_minimal()
```