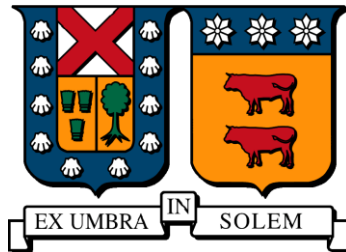


Universidad técnica Federico santa maría
Departamento ingeniería mecánica
Santiago-Chile



**“Metodología de Data Science para el análisis de
historiales de fallas y detenciones, para determinar
modelos de confiabilidad y disponibilidad.”**

Nicolás Andrés Pantoja Rojas

Memoria para optar al título de ingeniero civil mecánico

Profesor guía: Dr. Ing. David Godoy Ramos

Profesor correferente: Ing. Luis Guzmán Bonet

Junio 2020

Agradecimientos

Luego de tantos años estudiando es complejo terminar esta etapa, tanta gente que ha aportado su grano de arena para que todas las vivencias tuvieran el desenlace que estoy viendo a estas alturas, tantos profesores, profesoras, amigos y amigas que fueron importantes en menos o mas momentos, pero que formaron parte de mi vida como estudiante universitario, tanta gente a la que agradecerle... pero indiscutiblemente esta instancia se la debo a mi madre, tu hiciste lo imposible para que tus dos hijos pudieran estudiar, y conseguiste sacar a delante a una familia completa a costa de esfuerzo y sacrificio... muchas gracias por ser mi madre, te mereces este y mil reconocimientos más.

Abstract

La gestión del mantenimiento ha evolucionado lo largo del tiempo. Uno de los campos que permitió grandes avances corresponde al análisis estadístico, el cual fue ampliando sus horizontes gracias al constante aumento de la cantidad de datos registrados y/o disponibles. Hoy en día ya se ha vuelto notablemente más complejo el análisis debido a la enorme cantidad de datos con los cuales se trabaja, volviéndose necesario utilizar herramientas que faciliten su transformación en información. Por esto se debe trabajar acorde a la evolución del mantenimiento, agregándole más valor a la información obtenida de los datos y facilitando dicho proceso teniendo en cuenta el manejo de grandes volúmenes de datos. Para lo cual se debe implementar una metodología de análisis de historiales de fallas y detenciones mediante técnicas de Data Science, con la cual se pueda generar un modelo de confiabilidad, disponibilidad y de costos que permitan obtener el tiempo óptimo de intervención visto desde diferentes criterios. Pudiendo observarse que al aplicar esta metodología a un historial de 1.040 datos, que solo presenta una clasificación correspondiente al equipo, se logran clasificar 29 componentes y obtener parámetros para los modelos de confiabilidad y de disponibilidad, junto con los tiempos óptimos de intervención. Gracias a este trabajo se pueden gestionar políticas de mantenimiento con un respaldo de información y conocimiento más amplio, además de permitir el cuestionamiento de factores que antes podían estar ocultos (calidad de la gestión de los datos, calidad de los instrumentos de medición).

Tal como se menciona anteriormente, este análisis trabaja con un modelo para analizar datos de eventos, propio del RCM. El paso más lógico a continuar, tal como fue la evolución del mantenimiento, es complejizar el modelo analizado en esta metodología para incluir un modelo de CBM que permita el análisis por condición de los componentes.

Glosario

- **Activo fijo:** Es un elemento físico que posee una empresa, el cual generalmente hace referencia a un equipo o maquinaria requerida para la producción pero que formalmente se refiere a todos los elementos físicos que pueden ser valorados mediante una unidad monetaria.
- **TOI:** Corresponde al tiempo óptimo de intervención, el cual es el insumo principal para una política de mantenimiento a edad constante, es decir, aquella que busca identificar con que periodicidad constante realizar los mantenimientos preventivos.
- **Intervención:** Corresponde al momento en el cual se saca al activo físico de la producción para realizar un mantenimiento en este, ya sea por una razón preventiva o producto de una falla.
- **MTBF:** Corresponde al tiempo medio entre fallas, lo cual toma en cuenta solo las intervenciones correctivas.
- **MTTR:** Corresponde al tiempo medio para reparar.
- **MTBI:** Corresponde al tiempo medio entre las intervenciones, por lo cual toma en cuenta las intervenciones correctivas y preventivas.
- **Residuo:** Hace referencia a la diferencia entre la estimación y los valores registrados.

Tabla contenido

Agradecimientos.....	II
Abstract	III
Glosario	IV
Tabla contenido	V
Listado de figuras.....	VIII
Listado de tablas.....	X
Introducción.	1
Objetivos	3
1.1 Objetivo general.....	3
1.2 Objetivos específicos	3
1. Marco Teórico.....	1
1.1 Mantenimiento.....	4
1.1.1 Evolución del mantenimiento.....	4
1.1.2 Mantenimiento basado en confiabilidad (RCM).	5
1.1.3 Confiabilidad:	7
1.1.4 Mantenibilidad:.....	7
1.1.5 Disponibilidad:	8
1.1.6 Tipos de sistemas o activos físicos según su tipo de reparación.	10
1.1.7 Estimación de confiabilidad	14
1.1.8 Distribución de Weibull.....	16
1.1.9 Intervalos de confianza.....	17

1.1.10 Bondad de ajuste.....	18
1.1.11 Información para políticas de mantenimiento	19
1.2 Ciencia de datos (Data Science).	22
1.2.1 Recolección de datos.	25
1.2.2 Procesamiento de los datos.....	26
1.2.3 Limpieza de datos.....	26
1.2.4 Modelo estadístico y Resultados.	28
2. Metodología	30
2.1 Etapas de la metodología	30
2.2 Metodología unificada	32
3. Caso de estudio.....	34
3.1 Recolección de datos: Identificar formato	34
3.2 Recolección de datos: Ingresar datos	35
3.3 Procesamiento de datos: Filtrar datos faltantes.....	36
3.4 Procesamiento de datos: Clasificar registros	37
3.5 Limpieza de datos: Calcular TBF y TTR.....	39
3.6 Limpieza de datos: Identificar datos erróneos	40
3.7 Limpieza de datos: Identificar datos atípicos.....	40
3.8 Modelo estadístico: Test de Mann, Laplace, LR y MH	42
3.9 Modelo estadístico: Ajuste Weibull.....	43
3.10 Modelo estadístico: Valores influyentes	46
3.11 Modelo estadístico: Bondad de ajuste.....	48

3.12 Resultados	49
4. Conclusiones	51
5. Referencias	52
6. Anexo	54
6.1 Registros clasificados	54
6.2 Conteo ingresos validos e inválidos.....	56
6.3 Conteo valores atípicos.....	57
6.4 Calculo indicadores test para determinar modelo optimo a ser aplicado (RP o NHPP)	58
6.5 Parámetros estimación Weibull	59
6.6 Parámetros Weibull de dos y tres parámetros, junto con pruebas de bondad de ajuste	60
6.7 Resumen de resultados.....	61

Listado de figuras

<i>Figura 0.1: Diagrama de flujo para los capítulos y sus contenidos. Fuente: elaboración propia.</i>	1
<i>Figura 1.1: Ilustración de la evolución del mantenimiento, obtenido de la referencia [2].</i>	4
<i>Figura 1.2: esquema de los tiempos de operación obtenido de la referencia [1] y modificado.</i>	6
<i>Figura 1.3: Posibles tendencias del tiempo entre fallas extraído de la referencia [8]. Cada flecha es una falla.</i>	11
<i>Figura 1.4: Metodología para la selección del modelo de tiempo para fallar (time to failure model), extraído de la referencia [8].</i>	12
<i>Figura 1.5: Esquema simplificado del proceso de estimación de confiabilidad.</i>	14
<i>Figura 1.6: extracto de grafico obtenido de la referencia [1], representación de los diferentes modelos de confiabilidad utilizados para los diferentes ajustes de curvas más típicos.</i>	16
<i>Figura 1.7: Diagrama de Jack-Knife. Fuente: referencia [9].</i>	20
<i>Figura 1.8: Diagrama de Pareto extraído de la referencia [2].</i>	21
<i>Figura 1.9: Esquema del proceso de Data Science, extraído de la referencia [6], pagina 42.</i>	23
<i>Figura 1.10: “The 4 steps process to build models for component replacements” fuente: referencia [7].</i>	24
<i>Figura 1.11: Esquema de la metodología a aplicar para el análisis de historiales de falla.</i>	25
<i>Figura 1.12: diagrama de cajas demostrativo. Fuente: extraído de la referencia [10].</i>	27
<i>Figura 2.2: desglose sección recolección de datos. Fuente: elaboración propia.</i>	30
<i>Figura 2.3: desglose sección procesamiento de datos. Fuente: elaboración propia.</i>	30
<i>Figura 2.4: desglose sección limpieza de datos. Fuente: elaboración propia.</i>	31
<i>Figura 2.5: desglose sección modelo estadístico. Fuente: elaboración propia.</i>	31
<i>Figura 2.6: desglose sección resultado. Fuente: elaboración propia.</i>	32

<i>Figura 2.7: Diagrama de la metodología unificada para el análisis de historiales de falla. Fuente: elaboración propia.....</i>	<i>33</i>
<i>Figura 3.1: Diagrama de Pareto de los 20 equipos con más fallas del listado de equipos completo. Fuente: elaboración propia.....</i>	<i>36</i>
<i>Figura 3.2: Diagrama de Jack-knife del listado completo de equipos. Fuente: elaboración propia.</i>	<i>37</i>
<i>Figura 3.3: Diagrama de cajas para los 30 componentes Fuente: elaboración propia.....</i>	<i>41</i>
<i>Figura 3.4: Curvas de confiabilidad y probabilidad de fallas para el componente clasificado como cabezal. Fuente: elaboración propia.</i>	<i>45</i>
<i>Figura 3.5: Histograma de residuos sin eliminar valores atípicos. Fuente: elaboración propia.</i>	<i>46</i>
<i>Figura 3.6: Histograma de residuos luego de eliminar valores atípicos. Fuente: elaboración propia.</i>	<i>47</i>
<i>Figura 3.7: Histograma de residuos eliminando el máximo de valores sin considerar criterio de DFBETA. Fuente: elaboración propia.</i>	<i>48</i>

Listado de tablas

<i>Tabla 1.1: formato ideal de ingreso de información de mantenimiento.</i>	25
<i>Tabla 3.1: Extracto de los datos presentes en el historial de fallas a analizar. Fuente: base de datos de CODELCO.</i>	34
<i>Tabla 3.2: Extracto de los registros clasificados. Fuente: elaboración propia.</i>	38
<i>Tabla 3.3: Resumen de numero de fallas y detenciones de los componentes clasificados. Extracto de la Tabla 6.1 (anexo 6.1).</i>	39
<i>Tabla 3.4: Extracto de la Tabla 6.2: Número de registros validos luego del cálculo de TTR, incluyendo su respectivo porcentaje en relación al total de datos.</i>	40
<i>Tabla 3.5: Extracto de la Tabla 6.3: Numero de datos atípicos y el impacto de la eliminación de un dato con respecto al total de datos clasificados por componente.</i>	41
<i>Tabla 3.6: Extracto de la Tabla 6.4: Resultados de los test para determinar si aplicar un modelo para un proceso renovable o un proceso no homogéneo.</i>	42
<i>Tabla 3.7: extracto de la Tabla 6.5: Resultados de la estimación de un modelo de Weibull de dos parámetros.</i>	44
<i>Tabla 3.8: Resumen de datos eliminados para cada componente.</i>	46
<i>Tabla 3.9: extracto de la Tabla 6.6: resultados de las pruebas de bondad de ajuste tanto para el modelo de dos parámetros como para el de tres.</i>	49
<i>Tabla 3.10: Resumen resultados finales TOI, cálculos de disponibilidad y MTBF por componente. Extracto de la Tabla 6.7</i>	49
<i>Tabla 6.1: cantidad de registros de fallas y detenciones por componente clasificado.</i>	54
<i>Tabla 6.2: Número de registros validos luego del cálculo de TTR, incluyendo su respectivo porcentaje en relación al total de datos.</i>	56
<i>Tabla 6.3: Numero de datos atípicos y el impacto de la eliminación de un dato con respecto al total de datos clasificados por componente.</i>	57

<i>Tabla 6.4: Resultados de los test para determinar si aplicar un modelo para un proceso renovable o un proceso no homogéneo.....</i>	<i>58</i>
<i>Tabla 6.5: Resultados de la estimación de un modelo de Weibull de dos parámetros.</i>	<i>59</i>
<i>Tabla 6.6: resultados de las pruebas de bondad de ajuste tanto para el modelo de dos parámetros como para el de tres.</i>	<i>60</i>
<i>Tabla 6.7: Resumen resultados finales TOI, cálculos de disponibilidad y MTBF por componente.</i>	<i>61</i>

1. Introducción.

Con el avance de la tecnología, de los procesos de producción y de la civilización humana en sí, cada día se requiere de un mejor manejo de los recursos de los que se dispone, esta es una de las tantas razones por las cuales la gestión de los activos fue evolucionando para ser cada día más exigente, ya que un mantenimiento prematuro implica pérdidas de recursos, mal uso de estos, descarte prematuro de componentes funcionales, muchos sinónimos que pueden traducirse como un mal uso de los recursos, pero a su vez un mantenimiento tardío puede significar incluso la pérdida de vidas.

El aumento de dicha exigencia en la gestión del mantenimiento va de la mano con la evolución tecnológica, el incremento exponencial de información disponible y medios para trabajarla. Actualmente el foco de mayor nivel está puesto en la pronóstico de las fallas de los equipos, saber que va a fallar antes de que lo haga, lo cual requiere un manejo de grandes volúmenes de datos de forma eficiente y precisa.

Pero ¿Qué se necesita antes de dar ese salto? Pues así como el mantenimiento fue volviéndose más exigente y evolucionando, también se han desarrollado diversas herramientas de análisis de datos que permiten ampliar los horizontes de los análisis de la gestión de activos y antes de saltar a la inclusión de análisis de datos en tiempo real nace la necesidad de establecer una estructura de trabajo para tratar con grandes volúmenes de información imperfecta y sentar las bases lo más sólidas posibles para continuar con el flujo natural de la evolución del mantenimiento y pasar de la obtención de información reactiva a información preventiva que permita pronosticar eventos antes de que ocurran.

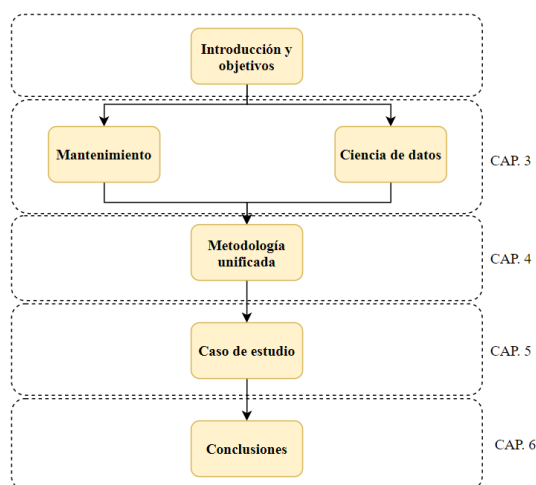


Figura 1.1: Diagrama de flujo para los capítulos y sus contenidos. Fuente: elaboración propia.

Para abordar esta temática se realizaron los pasos mostrados en el flujo de trabajo de la Figura 1.1, en la cual se puede apreciar que luego de la introducción se presentan los objetivos que delimitan el enfoque de este trabajo, posterior a esto se observa el capítulo 3, correspondiente al marco teórico en el cual se aborda el área de mantenimiento y de ciencia de datos. Ambas áreas abordadas en el marco teórico son unificadas en el capítulo 4 generando la metodología. Para poner a prueba la metodología se presenta un caso de estudio en el capítulo 5, con las posteriores conclusiones en el capítulo 6.

2. Objetivos

2.1 Objetivo general

Implementar una metodología para el análisis de historiales de fallas y detenciones, mediante técnicas de Data Science, definiendo las consideraciones necesarias para pasar de un grupo de eventos registrados a un listado de datos válidos para la determinación de un modelo de confiabilidad y disponibilidad, Utilizando esta metodología en un caso de estudio de Codelco, pero siendo replicable a otros análisis.

2.2 Objetivos específicos

- a) Definir la estructura de ingreso de datos que poseen la robustez necesaria para el posterior procesamiento de dichos datos aplicando técnicas de Data Science.
- b) Determinar las diversas consideraciones que permitan obtener modelos estadísticos dependiendo del nivel de detalle de la información ingresada, explicitando un mínimo de información requerida para este fin y un ideal para el mayor aprovechamiento de ésta.
- c) Modelar la confiabilidad y disponibilidad, considerando diversos modelos estadísticos (Weibull de dos parámetros, Weibull de tres parámetros).
- d) Estimar los tiempos óptimos de intervención, dando como resultado de los datos analizados un insumo de información útil para las políticas de mantenimiento.

3. Marco Teórico.

Los fundamentos teóricos de este trabajo se abordan desde dos grandes áreas, la primera corresponde a la gestión de activos fijos (mantenimiento), mientras que la segunda corresponde a la ciencia de datos (Data Science), a continuación, se tratarán los conceptos fundamentales de ambas áreas que son relevantes para éste trabajo.

3.1 Mantenimiento.

“La principal función del mantenimiento es sostener la funcionalidad de los equipos y el buen estado de las máquinas a través del tiempo.”[1] El mantenimiento posee diversas herramientas provenientes de múltiples áreas de conocimiento, las cuales son aplicadas con la finalidad descrita en el extracto mencionado al inicio. Pasando por herramientas estadísticas, logísticas, técnicas, entre otras. Dichas herramientas han ido evolucionando con el tiempo y, por ende, el mantenimiento en si ha evolucionado.

3.1.1 Evolución del mantenimiento



Figura 3.1: Ilustración de la evolución del mantenimiento, obtenido de la referencia [2].

Como se muestra en la **Figura 3.1**, en un comienzo el mantenimiento se basaba en las habilidades técnicas únicamente, reparando las fallas una vez ocurrían, lo que hoy en día se conoce como una política de mantenimiento correctivo o reactivo.

En la segunda generación, se buscaba mejorar los tiempos disponibles de las maquinas, para ello se intentó anticiparse a las fallas implementando la estrategia del mantenimiento preventivo al realizar inspecciones y reparaciones rutinarias. El problema que esto trajo fue un aumento en los costos, haciéndose cada vez más relevantes en comparación a los otros costos operacionales, siendo así el origen de la motivación de la gente para buscar métodos que permitan maximizar la vida útil de las máquinas y disminuir los costos.

Bajo la premisa anterior se da origen a la tercera generación, siendo esta generación aquella que presenta una mayor cantidad de objetivos, tal y como se aprecia en la **Figura 3.1**. Aquí se dio origen al mantenimiento por condición, el mantenimiento predictivo, algunas técnicas como el RCM (mantenimiento centrado en la confiabilidad) y el TPM (mantenimiento productivo total).

Luego de la tercera generación se definen algunas otras, pero es relevante hacer un alto aquí para este trabajo. En la tercera generación se dio un fuerte avance en la aplicación de la estadística como herramienta de la gestión del mantenimiento, permitiendo mejorar los insumos de información en la planificación a mediano y largo plazo. Si solamente tomamos en cuenta el aspecto de la evolución del insumo de información, en esta generación se comienzan a distinguir dos áreas principalmente, el mantenimiento basado en condición (CBM) y el mantenimiento centrado en la confiabilidad (RCM).

Este trabajo se centra en el RCM, ya que se considera que para la correcta aplicación de una política por condición con las herramientas que hoy en día se están aplicando (Data Mining, Artificial intelligence, Machine Learning, etc), es necesario generar una metodología base que permita mejorar el aprovechamiento de la información histórica de fallas, para así posteriormente ampliar dicha metodología al incluir la información de sensores y pasar de un análisis de RCM a uno CBM.

3.1.2 Mantenimiento basado en confiabilidad (RCM).

“El RCM es un enfoque sistemático del mantenimiento, que analiza los componentes dentro de un sistema, para clasificarlos según sea su criticidad, creando una lista válida de tareas de mantenimiento preventivo más riguroso en las máquinas cruciales, y dando tanta atención a los otros componentes del sistema como sea necesario para proveer operaciones con un costo y eficacia uniformes.” [2].

Tal como lo indica el párrafo anterior, la filosofía del RCM se fundamenta en los siguientes puntos (según referencia [1]):

1. Evaluar los componentes de los equipos, su estado y su función.
2. Identificar componentes críticos.
3. Selección de las políticas de mantenimiento adecuadas acorde al punto 2.

El primer punto se fundamenta en análisis particular de cada equipo, en el cual se plantean cuáles son los sistemas que los conforman, en qué estado se encuentran y en que otro estado se consideraría que están en falla. Para trabajar con el punto 2 y 3 se requieren un análisis de las variables de CMD (confiabilidad, mantenibilidad, disponibilidad). “La confiabilidad, la mantenibilidad y la disponibilidad son prácticamente las únicas medidas técnicas y científicas fundamentadas en cálculos matemáticos, estadísticos y probabilísticos que tiene el mantenimiento para su análisis” [1] La estimación de estas variables de fundamenta en el análisis de los tiempos en los que interactúan los eventos de fallas y reparaciones de cada elemento, haciendo mucho énfasis en que dichos tiempos deben ser los más representativos para el componente/equipo estudiado (horas de funcionamiento, caudal, cantidad de material o producto procesado, etc). Los tiempos relevantes se muestran en la **Figura 3.2**.

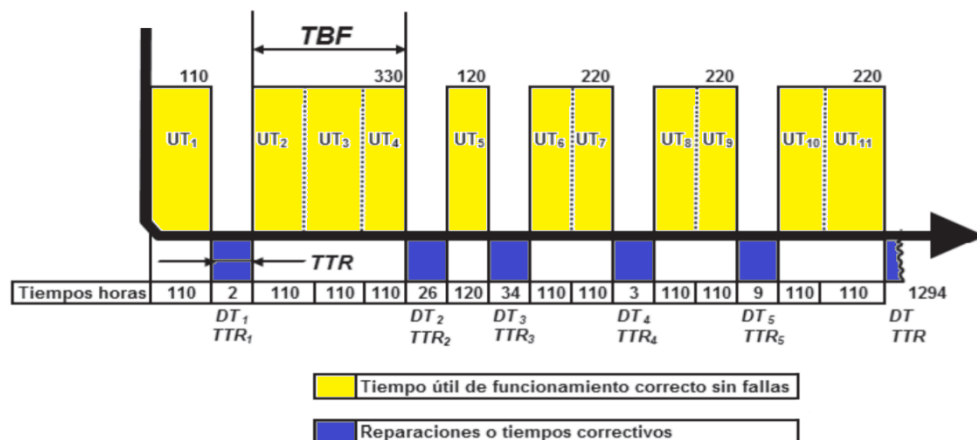


Figura 3.2: esquema de los tiempos de operación obtenido de la referencia [1] y modificado.

En la **Figura 3.2** se pueden observar los eventos de operación y mantenimiento graficados, los UT son los tiempos útiles de operación (Up time). Los TBF (Time between failures) son los tiempos entre las fallas, es decir, el tiempo que transcurre entre que el equipo fue devuelto a su condición de funcionamiento luego de una reparación y el momento en el que ocurre la siguiente falla. Los DT son los tiempos no útiles, uno de los tiempos asociados a esto es el TTR (time to repair) que es el tiempo para reparar.

Además de los tiempos descritos anteriormente, existen otros tiempos asociados a la logística, como viene siendo la llegada del componente necesario para efectuar la reparación, pero dichos tiempos no se tomaran en cuenta para este trabajo de forma directa.

Otro punto importante a tener en cuenta antes de comenzar a detallar las variables de CMD es la evolución de los métodos de predicción de estas, los cuales son los siguientes:

- Puntual (promedio simple).
- Distribuciones estadísticas.
- Modelos reparables y no reparables.
- Series temporales.

Estos métodos de predicción se detallarán más adelante según corresponda, pero es importante tenerlos en mente para saber el nivel de información requerida para poder trabajar con cada estimación, lo cual dependerá también de cual variable de CMD se esté calculando. Teniendo todo esto en mente y una comprensión de los tiempos y sus nomenclaturas, a continuación se definen las variables de CMD.

3.1.3 Confiabilidad:

La confiabilidad se puede definir como la probabilidad de supervivencia de un activo a un determinado tiempo, es decir, la probabilidad de que dicho equipo no falle en el instante temporal que se está analizando. Es importante recalcar que dicho tiempo a evaluar no tiene que ser estrictamente una unidad temporal (horas, minutos, etc.), debe ser una unidad que represente fielmente el funcionamiento del activo, si se trata de una turbina o una bomba puede ser que sea más representativo determinar su confiabilidad respecto al flujo y no al tiempo de operación. Tal como existe la confiabilidad $R(t)$ está su contraparte, la probabilidad de falla $F(t)$, estando relacionados ambos elementos de la siguiente forma:

$$R(t) = 1 - F(t) \quad (3.1)$$

3.1.4 Mantenibilidad:

La mantenibilidad se puede definir como la probabilidad de que el activo físico pueda alternar del estado de falla a un estado disponible para operar luego de realizarse el mantenimiento respectivo. Este indicador puede permitir dar una primera aproximación a los efectos del desgaste en el activo físico, o la falta de un mantenimiento más importante como el reemplazo del componente por uno nuevo.

3.1.5 Disponibilidad:

La disponibilidad es la probabilidad de que el equipo se encuentre apto para funcionar cuando sea requerido, es un indicador muy relevante ya que se asocia directamente con los niveles de producción u operación de los equipos, esto se debe a que a menor disponibilidad significa que los equipos operarán un menor tiempo, lo cual se traduce en costos por no producir (costos de ineficiencia). Más importante aún, si se trata de un equipo de seguridad o aquel que de su funcionamiento depende la vida de una persona, la probabilidad de que dicho equipo no esté disponible puede ocasionar la muerte de una persona.

Existen diferentes tipos de disponibilidad, en general se estima mediante una relación entre el tiempo disponible de la maquina o componente versus el tiempo total, pero dichos tiempos a considerar cambian según el tipo de disponibilidad que se desee obtener. Las ecuaciones para cada una de las disponibilidades se obtienen de la referencia [1].

La **disponibilidad genérica** es el cálculo más simple de disponibilidad con el que se trabaja, es útil para activos o empresas donde no se trabaja con cálculos o predicciones de CMD, consiste en una relación de cálculo puntual entre el tiempo disponible del activo y el tiempo total. A continuación se describe su ecuación.

$$D. Generica: A_G = \frac{MUT}{MUT + MDT} \quad (3.2)$$

En la ecuación (3.2), correspondiente a la disponibilidad generalizada, *MUT* es el tiempo medio disponible de la maquina o componente, extraído directamente de los registros como calculo puntual. *MDT* es el tiempo medio que la maquina o componente no está disponible.

La **disponibilidad inherente** es una primera aproximación que utiliza las estimaciones de CMD para controlar los mantenimientos no planificados, esta primera aproximación no considera los tiempos logísticos. Es un indicador muy utilizado como KPI en torno a las políticas de mantenimiento. A continuación se describe su ecuación.

$$D. inherente: A_I = \frac{MTBF}{MTBF + MTTR} \quad (3.3)$$

En la ecuación (3.3), correspondiente a la disponibilidad inherente, *MTBF* corresponde al tiempo medio entre fallas, este valor puede provenir de una estimación puntual o cualquiera de los otros métodos de estimación

descritos anteriormente (distribuciones estadísticas, series temporales) al igual que el $MTTR$, pero éste último corresponde al tiempo medio de las reparaciones.

La **disponibilidad alcanzada** es similar a la disponibilidad inherente, pero se incluye información de las tareas planeadas, siendo un indicador más global que los anteriores, al considerar los tiempos y las frecuencias de los mantenimientos preventivos. A continuación se describe su ecuación.

$$D. alcanzada: A_A = \frac{MTBM}{MTBM + \bar{M}} \quad (3.4)$$

En la ecuación (3.4), correspondiente a la disponibilidad alcanzada, $MTBM$ corresponde al tiempo entre mantenimientos y su cálculo se muestra en la ecuación (3.5), $\bar{M}$ corresponde al “tiempo medio de mantenimiento activo que se requiere para realizar una tarea de mantenimiento”[1] y la ecuación para su cálculo es la (3.6).

$$MTBM = \frac{1}{\frac{1}{MTBM_c} + \frac{1}{MTBM_p}} \quad (3.5)$$

Donde $MTBM_c$ es el tiempo medio entre los mantenimientos correctivos y $MTBM_p$ es el tiempo medio entre los mantenimientos preventivos.

$$\bar{M} = \frac{\frac{MTTR}{MTBM_c} + \frac{M_p}{MTBM_p}}{\frac{1}{MTBM_c} + \frac{1}{MTBM_p}} \quad (3.6)$$

Donde M_p es el tiempo medio para ejecutar las tareas proactivas de mantenimientos planeados.

La **disponibilidad operacional** es un indicador que requiere una base de datos mucho más robusta en torno al registro de los tiempos asociados a los mantenimientos, ya que en este caso se consideran los retrasos logísticos, para así poder controlarlos, ya que ninguno de los indicadores anteriores los toma en cuenta. A continuación se describe su ecuación.

$$D. operacional: A_O = \frac{MTBM}{MTBM + \bar{M}'} \quad (3.7)$$

Donde $\overline{M'}$ es el mismo de la ecuación (3.6), pero en esta oportunidad se incluyen los tiempos correspondientes a los retrasos logísticos donde corresponda (TTR, por ejemplo).

La **disponibilidad operacional generalizada** es como la disponibilidad operacional, pero está enfocada en equipos que no siempre se encuentran operando, por lo cual se toman en cuenta los tiempos en los cuales el activo físico está listo para funcionar, pero no lo está haciendo en el indicador de $MTBM$, transformándolo en $MTBM'$. A continuación, se describe su ecuación.

$$D. generalizada: A_{OG} = \frac{MTBM'}{MTBM' + \overline{M'}} \quad (3.8)$$

Teniendo ya una noción de CDM ahora es necesario introducir los tipos de sistemas que están clasificados para diferenciar los métodos más representativos al estimar su confiabilidad.

3.1.6 Tipos de sistemas o activos físicos según su tipo de reparación.

Para adentrarse en estas clasificaciones se deben considerar 2 grandes grupos, los sistemas (componentes, activos físicos, piezas, etc.) reparables y los no reparables. Un sistema reparable es aquel al que se le puede realizar un mantenimiento para modificar su confiabilidad sin requerir necesariamente el reemplazo del sistema, mientras que los sistemas no reparables son aquellos en los que ante la falla se reemplaza el sistema por uno nuevo. Los análisis clásicos de CDM asumen directamente que se está trabajando con sistemas no reparables. Dentro de los sistemas reparables existen 5 modelos principales (referencia [8]), los cuales se explican de forma general a continuación.

Procesos renovables (RP): Este proceso asume que luego de cada reparación, el sistema regresa a una condición igual a la que tenía cuando era nuevo, lo cual convierte este análisis en uno equivalente al sistema no reparable.

Procesos homogéneos de Poisson (HPP): Este proceso es un caso particular del RP, aquí se asume que los tiempos entre fallas son independientes y están idénticamente distribuidos de forma exponencial, lo cual es perfectamente representable con un modelo exponencial de tasa de fallas constante.

Proceso de ramificación de Poisson (BPP): En este caso se asume que los tiempos entre fallas están distribuidos idénticamente pero no son independientes. Este modelo no tiene muchas aplicaciones prácticas que se hayan mencionado en la literatura según la referencia [8].

Proceso renovable superpuesto (SRP): En este caso se trabaja en base a un conjunto de modelos RP superpuestos que dan como resultado un modelo global que no es RP, lo cual es entendible de forma lógica al analizar un sistema complejo de forma global, por ejemplo, se puede considerar un motor como un sistema SRP, el cual está compuesto de varios sistemas RP como bujías, poleas, pistones, filtros, correas, etc. Considerando que al momento de presentarse una falla en cualquiera de estos subsistemas se reemplazarán por uno nuevo.

Proceso no homogéneo de Poisson (NHPP): a diferencia de todos los modelos anteriores, en este caso no es posible representar una reparación equivalente a la representada en un sistema no reparable, sino que aquí se interpreta una reparación mínima.

La principal diferencia que es determinante al momento de decidir si aplicar o no una metodología de ajuste del modelo como Weibull es verificar la existencia de una tendencia en los tiempos entre fallas como se muestran en la **Figura 3.3** línea B o línea C, dicha tendencia determinaría la necesidad de aplicar una metodología no estacionaria (NHPP por ejemplo).

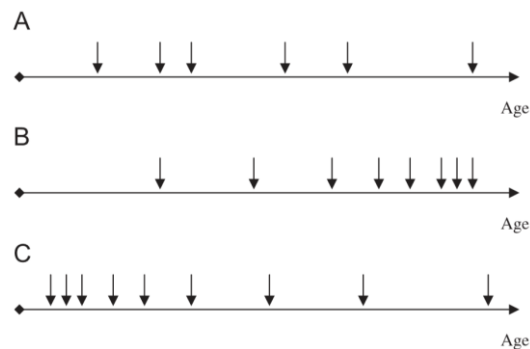


Figura 3.3: Posibles tendencias del tiempo entre fallas extraído de la referencia [8]. Cada flecha es una falla.

Bajo esta consideración, en la referencia [8], se plantea la metodología mostrada en la **Figura 3.4**. Esta metodología tiene una primera aproximación determinando si los datos son adecuados (tipo de datos y cantidad) para el objetivo del estudio, posterior a esto se comienza el análisis necesario para determinar qué modelo aplicar. Para determinar si utilizar un modelo renovable o un modelo NHPP se fundamenta en 4 test de hipótesis y un método gráfico, con lo cual puede determinar si existe una tendencia en los tiempos entre fallas y que tipo de tendencia es (creciente, decreciente, no monótona).

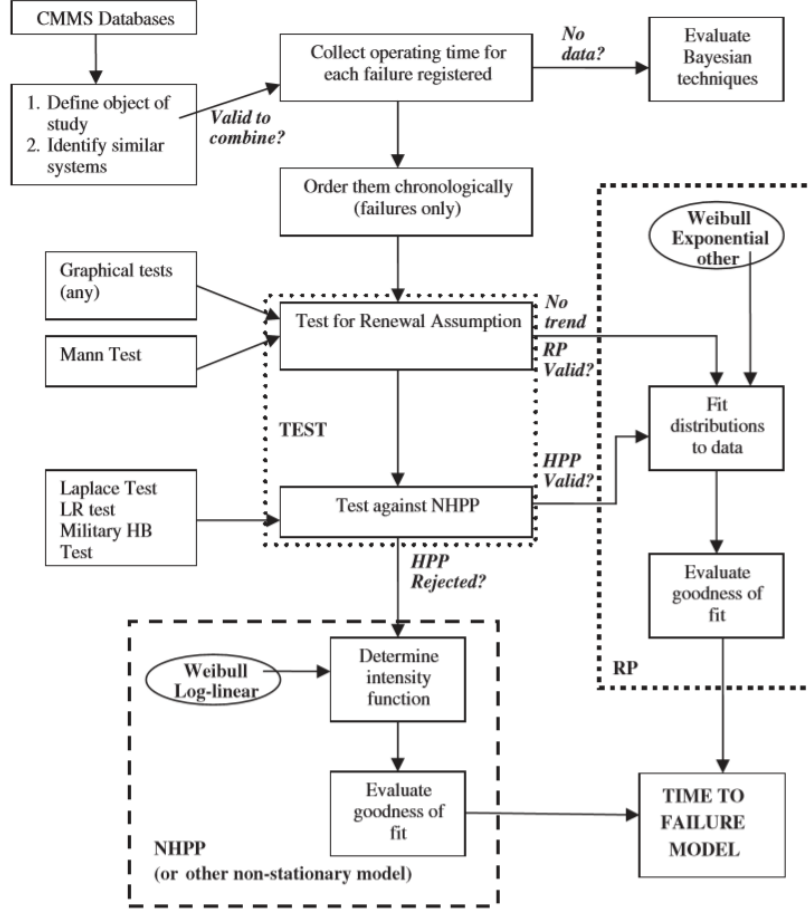


Figura 3.4: Metodología para la selección del modelo de tiempo para fallar (time to failure model), extraído de la referencia [8].

Test de Mann: la hipótesis nula de este test es que los datos pueden modelarse bajo un RP. Mientras que la hipótesis alternativa es que los datos presentan una tendencia monótona. En caso de que el estadístico (M) mostrado en la ecuación (3.9) sea muy distinto a $\frac{n(n-1)}{4}$ se rechaza la hipótesis nula (n corresponde al número de datos), en la referencia [8] no se cuantifica el “muy distinto a” por lo cual se considerará prudente desde una diferencia del 20%.

$$M = \sum_{i=1}^{n-1} \sum_{j=i+1}^n I(T_i < T_j) \quad (3.9)$$

En la ecuación (3.9) n corresponde al número de datos que se están analizando, T_j y T_i son tiempos entre fallas y $I(x)$ es un operador que toma el valor de 1 cuando se le asigna x como verdadero y toma el valor de 0 cuando se le asigna x como falso.

Test de Laplace: Este test tiene la hipótesis nula de que los datos se modelan como un proceso HPP con lo cual se puede asumir que los tiempos entre fallas están idénticamente distribuidos de forma exponencial, mientras que la hipótesis alternativa corresponde a que dichos datos corresponden a un proceso NHPP con una intensidad monótona. El estadístico de este test se muestra en la ecuación (3.10), la hipótesis nula se rechaza si el estadístico es distinto de cero, si es mayor que cero es un indicativo de una tendencia creciente (deterioro) mientras que si es menor que cero es indicativo de una tendencia decreciente.

$$L = \frac{\sum_{j=1}^{\hat{n}} T_j - \frac{1}{2} \hat{n}(b + a)}{\sqrt{\frac{1}{12} \hat{n}(b - a)^2}} \quad (3.10)$$

En la ecuación (3.10) aparecen algunos nuevos conceptos, $\hat{n}$ es igual a n si el proceso analizado es truncado en el tiempo, mientras que corresponde a $(n - 1)$ si el proceso es truncado a la falla. a y b son los límites del intervalo de tiempo analizado, siendo el límite inferior y el superior respectivamente.

Test de Lewis-Robinson (LR): este es un test derivado del anterior, por lo cual utiliza el estadístico mostrado en la ecuación (3.10) pero adaptado, tal como se muestra en la ecuación (3.11). La hipótesis nula corresponde a un proceso HPP, mientras que la hipótesis alternativa indica la presencia de una tendencia en los datos. Si el estadístico es distinto de cero, entonces se rechaza la hipótesis nula.

$$LR = \frac{L * \bar{X}}{\sigma(X)} \quad (3.11)$$

En la ecuación (3.11) $\bar{X}$ corresponde a la media de los tiempos entre falla, como calculo puntual, $\sigma(X)$ corresponde a la desviación estándar de los tiempos entre fallas.

Military handbook test (MH): La hipótesis nula de este test corresponde a un proceso HPP, mientras que la hipótesis alternativa corresponde a un proceso NHPP con intensidad monótona.

$$MH = 2 \sum_{j=1}^N \ln \left(\frac{T_{total}}{T_j} \right) \quad (3.12)$$

En la ecuación (3.12), correspondiente al estadístico de esta prueba de hipótesis, se utilizan los mismos parámetros que en todas las anteriores, pero este estadístico se contrasta con la prueba de Chi-cuadrado de

$2(N - 1)$ grados de libertad, para este test en específico se encuentra aplicado en la referencia [19] y descrito más a fondo en la referencia [23].

3.1.7 Estimación de confiabilidad



Figura 3.5: Esquema simplificado del proceso de estimación de confiabilidad.

En la **Figura 3.5** se muestra de forma simplificada el proceso necesario para estimar y modelar la confiabilidad, el primer paso es la obtención de los tiempos entre fallas (TBF), luego de eso se debe estimar de forma puntual la probabilidad de falla $F(t_i)$ y por último se realiza el proceso de ajuste de un determinado modelo de confiabilidad en el tiempo $R(t)$, el cual determinará los pasos requeridos para estimar sus parámetros correspondientes.

Para la estimación de la probabilidad de fallas existen diversos estimadores, a continuación se presentarán algunos de los más relevantes obtenidos de la referencia [1].

Estimador i – *kaésimo*, mostrado en la ecuación (3.13).

$$\hat{F}(t_i) = \frac{i}{N + 1} \quad (3.13)$$

Donde N corresponde al número total de datos, los cuales deben estar ordenados de forma ascendente, lo cual también aplica para los estimadores que se mostrarán a continuación.

Estimador de Bernard, mostrado en la ecuación (3.14).

$$\hat{F}(t_i) = \frac{i - 0.3}{N + 0.4} \quad (3.14)$$

Estimador de Kaplan-Meier, mostrado en la ecuación (3.15).

$$\hat{F}(t_i) = 1 - \prod_{i=1}^N \frac{N - i}{N - i + 1} \quad (3.15)$$

Esta estimación presenta el problema de que el último valor es igual a cero, por lo cual nace una modificación de este estimador, el cual se presenta a continuación.

Estimador de Kaplan-Meier modificado, mostrado en la ecuación (3.16).

$$\hat{F}(t_i) = 1 - \frac{N + 0.7}{N + 0.4} \prod_{i=1}^N \frac{N - i + 0.7}{N - i + 1.7} \quad (3.16)$$

De todos los estimadores, se tiene que el estimador de Bernard es el más conservador, entregando generalmente los valores más bajos de MTBF, mientras que el estimador de Kaplan-Meier generalmente otorga los valores más altos de MTBF al modelar la confiabilidad con dichos estimadores.

Además de los estimadores descritos recientemente, existe otro estimador importante a mencionar, el cual posee la característica diferenciadora de considerar las detenciones en su estimación de confiabilidad, no solo los eventos de falla. Este otro estimador corresponde al **estimador de Lewis** ([3] y [4]), el cual además realiza una estimación directa de la confiabilidad, mediante la ecuación (3.17).

$$\hat{R}(t_i) = \left(\frac{N + 1 - i}{N + 2 - i} \right)^{1 - \delta_i} \hat{R}(t_{i-1}) \quad (3.17)$$

Donde δ_i toma el valor de 1 cuando el registro corresponde a una censura (detención), y de 0 cuando corresponde a una falla. Se puede observar que es una ecuación recursiva, tomando en cuenta el valor anterior, por lo cual es importante destacar que este método considera que $R(0) = 1$.

Ahora es un buen momento para introducir los modelos de confiabilidad, pero para los efectos de esta memoria solamente se trabajará con Weibull de dos y tres parámetros gracias a su versatilidad, la cual se ve reflejada en la **Figura 3.6** (y más detalladamente en la referencia [20]), en la cual se pueden apreciar distintos comportamientos de la tasa de fallas en el tiempo, para cada uno de esos comportamientos existe una distribución adecuada, pero así mismo se observa como al variar uno de los parámetros de la distribución de Weibull (beta), esta se ajusta a cualquiera de los comportamientos mostrados en la **Figura 3.6**.

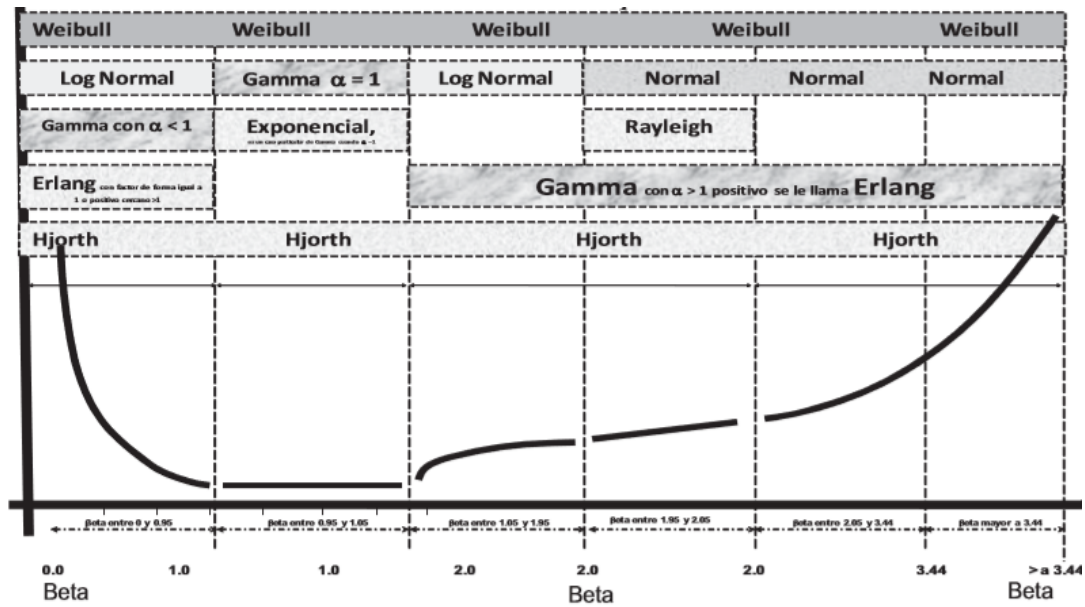


Figura 3.6: extracto de gráfico obtenido de la referencia [1], representación de los diferentes modelos de confiabilidad utilizados para los diferentes ajustes de curvas más típicos.

3.1.8 Distribución de Weibull.

La confiabilidad según la distribución de Weibull se define de la siguiente forma.

$$R(t) = e^{-\left(\frac{t-t_0}{\alpha}\right)^\beta} \quad (3.18)$$

Con lo cual se tiene que los parámetros de Weibull son α que corresponde al parámetro de escala, β que corresponde al parámetro de forma y t_0 que corresponde al parámetro de posición.

El parámetro de posición corresponde al parámetro diferenciador entre Weibull de dos parámetros y Weibull de tres parámetros, ya que en la primera distribución se asume como cero, con lo cual se ven sumamente simplificados los cálculos de los otros parámetros. Dos de los principales métodos de estimación de los parámetros para Weibull son:

- Método de regresión lineal con mínimos cuadrados
- Método de máxima verosimilitud

Ambos métodos se encuentran explicados en la referencia [1], y el método de regresión lineal se encuentra también detallado y ejemplificado en la referencia [20].

3.1.9 Intervalos de confianza

Para obtener los intervalos de confianza de los parámetros de Weibull se utilizan las mismas ecuaciones que ocupa el método de estimación de parámetros por máxima verosimilitud, tanto esta información como las ecuaciones específicas de los intervalos de confianza provienen de las referencias [21] y [22]. La ecuación (3.19) es el eje central de la metodología MLE.

$$L(t_i, \vec{\theta}) = \prod_{i=1}^N f(t_i, \vec{\theta}) \quad (3.19)$$

Donde $\vec{\theta}$ corresponde al vector de los parámetros de la distribución que se esté utilizando, f corresponde a la función de densidad de probabilidad de la distribución correspondiente. Para trabajar con esta ecuación de utiliza el logaritmo natural de ésta, reemplazando la función f por la correspondiente de la distribución de Weibull se tiene la ecuación (3.20).

$$l = \ln(L(T_i, \vec{\theta})) = \sum_{i=1}^N \left[\ln(\beta) + (\beta - 1) \ln(t_i - t_0) - \beta \ln(\alpha) - \left(\frac{t_i - t_0}{\alpha} \right)^\beta \right] \quad (3.20)$$

Para la estimación de los parámetros se deriva esta ecuación con respecto a cada uno de los parámetros y se iguala a cero, obteniendo un set de 3 ecuaciones para poder resolver el sistema de ecuaciones y obtener la estimación de los parámetros, mientras que para estimar los intervalos de confianza se requiere obtener la varianza de los parámetros, para ello se utiliza la ecuación (3.21), la cual posee la nomenclatura genérica de θ como estimador, para las siguientes se tiene θ_1 como α y θ_2 como β . Las resoluciones de las derivadas se muestran en las ecuaciones (3.22), (3.23), (3.24) y (3.25).

$$\begin{bmatrix} Var(\widehat{\theta_1}) & Cov(\widehat{\theta_1}, \widehat{\theta_2}) \\ Cov(\widehat{\theta_1}, \widehat{\theta_2}) & Var(\widehat{\theta_2}) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 l}{\partial \theta_1^2} & -\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} \\ -\frac{\partial^2 l}{\partial \theta_2 \partial \theta_1} & -\frac{\partial^2 l}{\partial \theta_2^2} \end{bmatrix}^{-1} \quad (3.21)$$

$$-\frac{\partial^2 l}{\partial \alpha^2} = -\frac{N\beta}{\alpha^2} + \frac{\beta^2 + \beta}{\alpha^{\beta+2}} \sum_{i=1}^N (t_i - t_0)^\beta \quad (3.22)$$

$$-\frac{\partial^2 l}{\partial \beta^2} = \frac{N}{\beta^2} + \sum_{i=1}^N \left(\frac{t_i - t_0}{\alpha} \right)^\beta \ln \left(\frac{t_i - t_0}{\alpha} \right)^2 \quad (3.23)$$

$$-\frac{\partial^2 l}{\partial \alpha \partial \beta} = \frac{N}{\alpha} - \left(\frac{1}{\alpha^{\beta+1}} \right) \sum_{i=1}^N (t_i - t_0)^\beta (1 + \beta \ln(t_i - t_0) - \beta \ln(\alpha)) \quad (3.24)$$

$$-\frac{\partial^2 l}{\partial \beta \partial \alpha} = \frac{N}{\alpha} - \left(\frac{1}{\alpha^{\beta+1}} \right) \sum_{i=1}^N (t_i - t_0)^\beta (1 + \beta \ln(t_i - t_0) - \beta \ln(\alpha)) \quad (3.25)$$

Con la resolución de las ecuaciones descritas previamente ya es posible obtener los intervalos de confianza de los estimadores de Weibull, se plantean dos ecuaciones para hacerlo, la ecuación (3.26) es la forma genérica de obtener un intervalo de confianza para los estimadores que se están analizando, pero en caso de que dicho estimador deba ser positivo se propone la ecuación (3.27) en la referencia [22].

$$\hat{\theta} - Z_{\frac{\alpha_s}{2}} \sqrt{\text{Var}(\hat{\theta})} < \theta < \hat{\theta} + Z_{\frac{\alpha_s}{2}} \sqrt{\text{Var}(\hat{\theta})} \quad (3.26)$$

$$\frac{e^{\frac{Z_{\frac{\alpha_s}{2}} \sqrt{\text{Var}(\hat{\theta})}}{\hat{\theta}}}}{\hat{\theta}} < \theta < \hat{\theta} e^{\frac{Z_{\frac{\alpha_s}{2}} \sqrt{\text{Var}(\hat{\theta})}}{\hat{\theta}}} \quad (3.27)$$

3.1.10 Bondad de ajuste

Luego de realizar la estimación de los parámetros del modelo que se desea aplicar es importante realizar una comprobación que permita identificar si dicho modelo y sus parámetros estimados realmente se ajustan adecuadamente al comportamiento de los datos, para esta comprobación se realizan diversos test de hipótesis que buscan comprobar la bondad del ajuste. Para este trabajo se tomarán en cuenta dos de las más conocidas pruebas de bondad de ajuste, la prueba de Kolmogorov-Smirnov y la prueba de Anderson-Darling.

Kolmogorov-Smirnov (K-S): Para esta prueba de hipótesis, la hipótesis nula es que los datos siguen una distribución especificada, mientras que la hipótesis alternativa es que dichos datos no siguen la distribución especificada. El estadístico para esta prueba se muestra en la ecuación (3.28), en caso de que dicho estadístico sea mayor que el valor crítico, entonces se rechaza la hipótesis nula. Los valores críticos para esta prueba se encuentran tabulados, tal como se muestra en la referencia [1].

$$D = \max_{1 \leq i \leq N} |F(Y_i) - E_i| \quad (3.28)$$

Donde $F(Y_i)$ es la función de distribución acumulada por probar, E_i corresponde a la función empírica K-S y N corresponde al número total de datos, lo que se traduce en puntos en ambas distribuciones.

Anderson-Darling (A-D): Esta prueba de hipótesis presenta más mismas hipótesis de la prueba de Kolmogorov-Smirnov, siendo la hipótesis nula que los datos siguen la distribución indicada y la hipótesis alternativa que los datos no siguen la distribución especificada. El estadístico de esta prueba se muestra en la ecuación (3.29).

$$A^2 = -N - \sum_{i=1}^N \left(\frac{2i-1}{N} \right) [\ln(F(Y_i)) + \ln(1 - F(Y_{N+1-i}))] \quad (3.29)$$

En caso de que el valor obtenido del estadístico sea mayor que el valor crítico tabulado, entonces se rechaza la hipótesis nula. Es importante considerar que para muestras pequeñas el estadístico recibe una modificación que depende de la distribución a estudiar, para el caso de la distribución de Weibull la modificación corresponde a la mostrada en la ecuación (3.30).

$$A_m^2 = A^2 \left(1 + \frac{0.2}{\sqrt{N}} \right) \quad (3.30)$$

3.1.11 Información para políticas de mantenimiento

El foco de este ítem es describir las herramientas que permiten entregar la información pertinente para trabajar con las políticas de mantenimiento luego de haber transformado los datos necesarios. Existen diversas herramientas pero para este trabajo el foco está en Jack Knife, Pareto y el tiempo óptimo de intervención (TOI). las primeras dos corresponden a herramientas claves para la priorización de equipos y componentes, permitiendo focalizar los esfuerzos en aquellos componentes y equipos que más lo requieren, y la última herramienta es más bien un indicador resultante que determina, bajo distintos criterios, cual es el tiempo óptimo para intervenir el componente que se está analizando.

Jack-Knife: Esta herramienta es útil para la priorización de activos físicos y/o sus modos de falla, ya que permite seccionar las fallas en cuatro cuadrantes con dos tipos de clasificaciones, si son fallas agudas leves, crónicas o críticas. Generalmente es una herramienta aplicada directamente en base a cálculos puntuales a diversos

equipos para una aproximación de clasificación de equipos y poder priorizar recursos, pero también puede aplicarse a los componentes de un equipo/sistema con la misma meta, priorizar componentes del equipo.

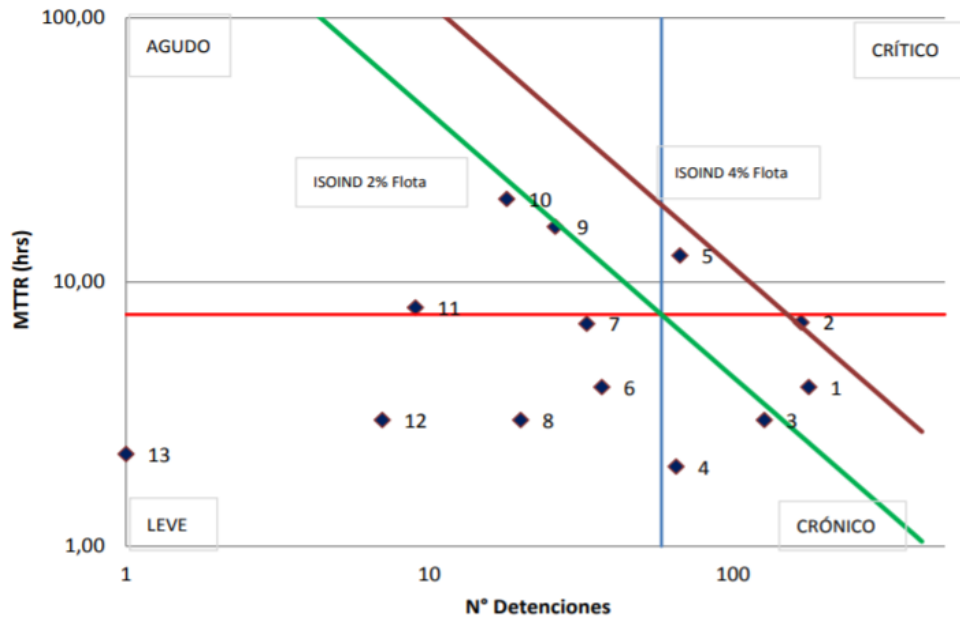


Figura 3.7: Diagrama de Jack-Knife. Fuente: referencia [9].

Esta herramienta tiene una representación gráfica como la mostrada en la **Figura 3.7**, para confeccionar esta grafica se tiene que; el eje x (frecuencia de falla) obedece a la ecuación **(3.31)** la cual corresponde al logaritmo del número de fallas del elemento a analizar (componente, equipo, modo de falla, etc.).

$$X = \log(Nro\ detenciones) \quad (3.31)$$

Mientras que el eje Y obedece a la ecuación **(3.32)**, la cual corresponde al logaritmo del tiempo medio para reparar del elemento a analizar.

$$Y = \log(MTTR_i) \quad (3.32)$$

Para dividir el grafico en los cuatro cuadrantes mostrados se definen dos límites, el primer límite corresponde a la división del eje Y en dos cuadrantes, lo cual se realiza con el cálculo puntual del MTTR de todos los elementos analizados y luego aplicado el logaritmo, tal como se muestra en la ecuación **(3.33)**.

$$Limite Y = \log \left(\frac{\text{Suma total del tiempo de detencion}}{\text{suma total del numero de detenciones}} \right) \quad (3.33)$$

El límite que divide el eje X es la proporción del número total de fallas con el número total de razones que se están analizando (total de equipos, componentes, etc.), al igual que en todas las ecuaciones anteriores, con el logaritmo aplicado producto del cambio de escala lineal a logarítmica, tal como se muestra en la ecuación (3.34).

$$Limite X = \log \left(\frac{\text{Suma total del numero de detenciones}}{\text{Nro total de elementos}} \right) \quad (3.34)$$

Pareto: Esta es otra herramienta de priorización, la cual permite identificar en que elementos que se están analizando se encuentra el mayor porcentaje de impacto, por ejemplo, se puede analizar el porcentaje de fallas de cada componente con respecto al total de fallas del equipo, pudiendo priorizar que componentes son más críticos. La representación gráfica de esta herramienta se muestra en la **Figura 3.8**.

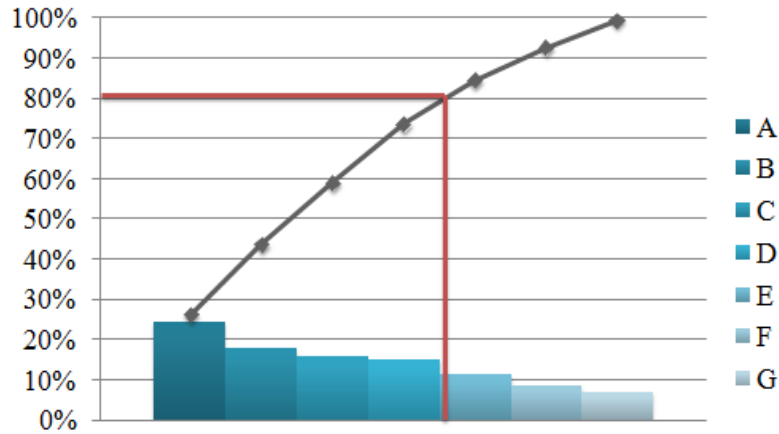


Figura 3.8: Diagrama de Pareto extraído de la referencia [2].

TOI: Para analizar este KPI existen diversos enfoques, pudiendo determinarse mediante costos, como un límite de confiabilidad del cual no se desea descender o con los indicadores de MTBF, MTBM, MTBI. Para el caso de costos se requiere la ecuación (3.35).

$$EC(tp) = \frac{C_p R(tp) + C_c F(tp)}{MTBM_{tp} + MTTR_{tp}} \quad (3.35)$$

Donde C_p corresponde al costo de los mantenimientos preventivos, C_c corresponde al costo de los mantenimientos correctivos, $MTBM_{tp}$ corresponde al tiempo medio entre mantenimientos evaluado al tiempo tp

y $MTTR_{tp}$ corresponde al tiempo medio para reparar evaluado al tiempo tp , estos últimos dos muestran su equivalencia de cálculo en la ecuación (3.36), donde aparecen dos nuevos términos, siendo $MTTR_p$ el tiempo medio para las reparaciones preventivas y $MTTR_e$ el tiempo medio para las reparaciones de emergencia.

$$EC(tp) = \frac{C_p R(tp) + C_e F(tp)}{(R(tp)tp + F(tp)MTBF(tp)) + (R(tp)MTTR_p + F(tp)MTTR_e)} \quad (3.36)$$

Por último, para el cálculo del $MTBF(tp)$ se requiere la ecuación (3.37), la cual proviene de la referencia [2].

$$MTBF(tp) = \frac{\sum_{i=0}^{tp} (F(t_i) - F(t_{i-1})) \left(\frac{t_i + t_{i-1}}{1} \right)}{F(tp)} \quad (3.37)$$

3.2 Ciencia de datos (Data Science).

La ciencia de los datos engloba múltiples disciplinas y busca precisamente transformar los datos que se encuentran disponibles en conocimiento. Si bien el trabajo específico a realizar dependerá exclusivamente del problema particular que se desee resolver, existen metodologías generales a analizar, como la que se muestra a continuación.

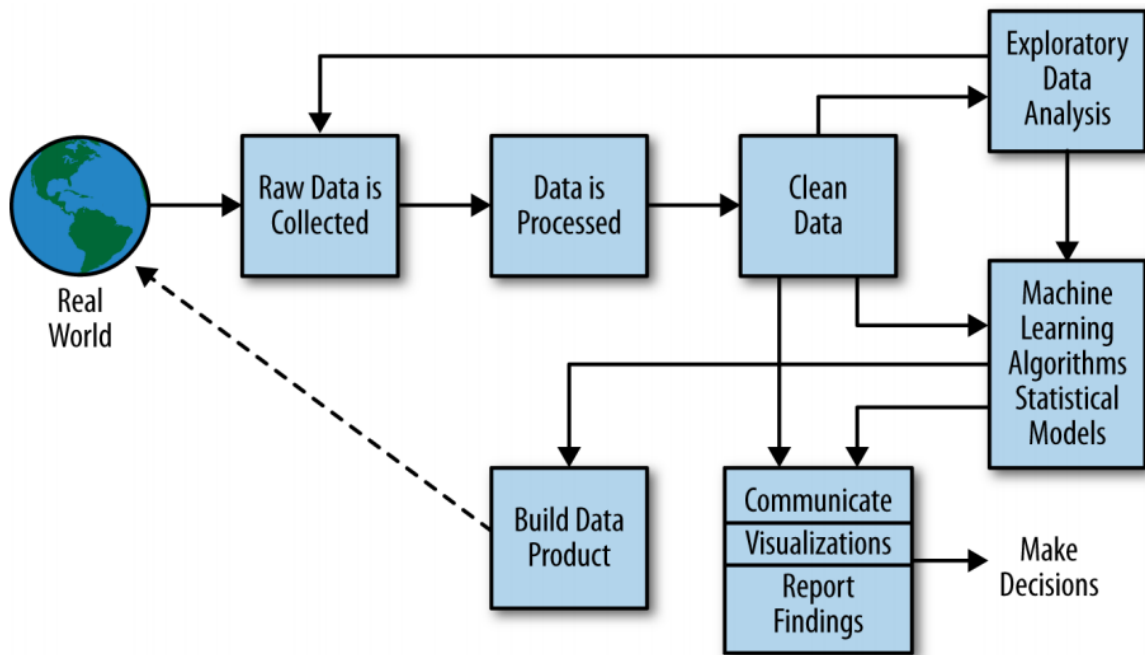


Figura 3.9: Esquema del proceso de Data Science, extraído de la referencia [6], pagina 42.

En la figura recién presentada se presenta una metodología o proceso compuesto por los siguientes pasos.

- Primero se tiene una recolección de datos en bruto que es extraída del mundo real, con las imperfecciones que esto implique y desde las múltiples fuentes que se puedan imaginar (por ejemplo Excel, emails, plantillas, informes, internet).
- Segundo, se procesa la información ingresada para poder limpiarla, en este punto existen diversas metodologías y recomendaciones, pero la finalidad es clara, se busca transformar los datos en algo trabajable según el modelo que se desee aplicar posteriormente, además de buscar un formato manejable por el programa con el cual se deseen realizar los posteriores pasos y transformar los datos en información.
- En tercer lugar, teniendo ya procesados los datos recibidos, se realiza un segundo procesado, o limpieza de dichos datos. Esta vez se analiza cómo actuar frente a datos incompletos, duplicados, posibles datos erróneos. También existe un nexo en este punto con la recolección de datos, ya que puede darse la necesidad de obtener más registros.
- Los pasos a continuación corresponden a la creación del modelo estadístico, en el cual se transforman los datos en información. Por último, se transforma dicha información en conocimiento y con esto se permite una toma de decisiones, siendo la finalidad de esta metodología.

Dependiendo de la labor específica que se deba realizar, existen diferentes metodologías específicas, por ejemplo, el área de Machine Learning, de Big Data y de Data Mining son 3 subáreas de la ciencia de datos que presentan sus propias adaptaciones metodológicas. Un ejemplo de una metodología de Data Mining es explicada en la referencia [7], y se aprecia en el siguiente esquema.

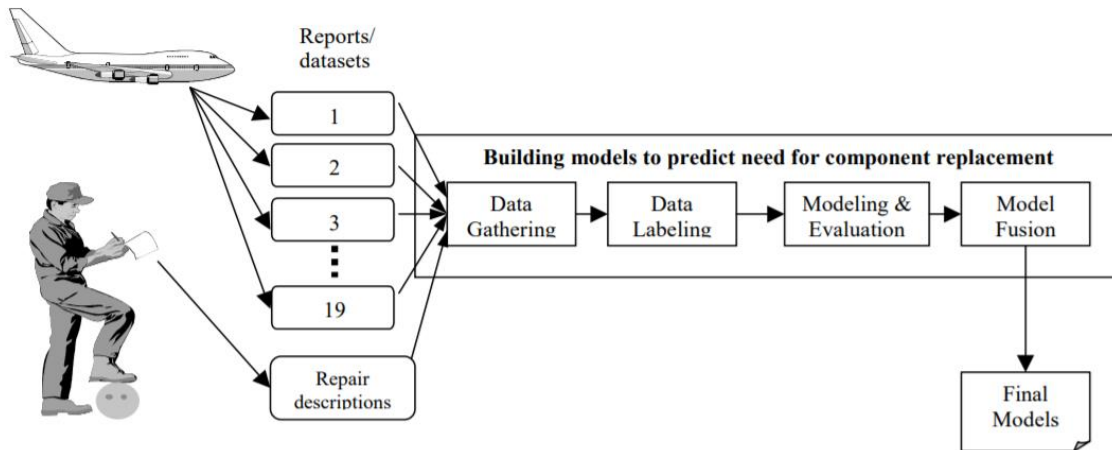


Figura 3.10: “The 4 steps process to build models for component replacements” fuente: referencia [7].

Este es un buen ejemplo a tomar en cuenta, ya que es desarrollado en sistemas en los cuales una falla puede significar la pérdida de vidas, por lo cual presentan un mayor registro de desarrollos en torno a los análisis de mantenimiento. Además, abarca el análisis de los dos tipos de datos, los datos de mantenimiento (historiales de falla, registros de mantenimiento) y los datos de sensores, con la finalidad de generar un modelo que permita determinar cuándo reemplazar cada componente.

Los pasos de esta metodología son básicamente 4, primero viene la recolección de los datos (Data Gathering), luego viene la clasificación de estos (Data Labeling) para poder alimentar los modelos de inteligencia artificial, con los cuales se generan los modelos de cada componente y se evalúa su precisión, por último se juntan todos los modelos en un modelo global (Model Fusion). Otro ejemplo interesante de esta metodología se presenta en la referencia [15], con un estudio mucho más exhaustivo de las tecnologías y etapas del proceso.

Si bien en este trabajo solo se desarrollarán modelos con datos de mantenimiento, es relevante comprender cuál sería el siguiente paso, lo cual se ve de forma tangible en el ejemplo anterior, en el cual, mediante una información de mantenimiento que presenta una calidad de ingreso tal que puede interactuar adecuadamente con la información de los sensores, se pueden generar modelos predictivos (con análisis por condición) para saber cuándo reemplazar los componentes.

Teniendo ambos ejemplos en mente se comienza a desarrollar la metodología a utilizar para el análisis de los historiales de falla. Se parte de una base muy similar a la mostrada en la **Figura 3.9**, teniéndose el siguiente esquema.



Figura 3.11: Esquema de la metodología a aplicar para el análisis de historiales de falla.

Para esta metodología, como se menciona anteriormente, el enfoque es el análisis de los historiales de falla, con la finalidad de extraer información relevante para la gestión del mantenimiento, a continuación se describen los pasos uno por uno.

3.2.1 Recolección de datos.

Esta primera etapa corresponde a la recolección de los datos, tal como lo indica su nombre, aquí se deben definir las fuentes de datos, los formatos de ingreso de los datos. Idealmente, casi obligatoriamente, es necesario definir también el formato ideal en el cual se requieren los datos para los siguientes pasos, para poder determinar si hay posibilidades de obtener dichos formatos de ingreso directamente de las fuentes de datos y simplificar los siguientes pasos.

Teniendo en cuenta lo descrito en la sección 3.1, se tiene en mente que el formato ideal del historial de fallas para los análisis que se quieren realizar es el que se muestra a continuación. Se hace énfasis en que se define como “ideal” dado que contienen todos los parámetros que hasta el momento se requieren para realizar todos los análisis propuestos.

Tabla 3.1: formato ideal de ingreso de información de mantenimiento.

ID equipo	ID componente	Clase de orden	Texto breve	Costo	Fecha inicio	Hora inicio	Fecha termino	Hora termino

Como se puede apreciar en la **Tabla 3.1**, para el formato ideal de los datos de mantenimiento se consideran los siguientes ítems:

- **Primero:** un código de identificación del equipo.
- **Segundo:** un código de identificación del componente.

- **Tercero:** la clase de orden (mantenimiento correctivo o preventivo).
- **Cuarto:** un texto breve que sirva a modo de descripción del evento.
- **Quinto:** el costo del evento.
- **Sexto:** la fecha de inicio del evento.
- **Séptimo:** la hora de inicio del evento.
- **Octavo:** la fecha de término del evento.
- **Noveno:** la hora de término del evento.

3.2.2 Procesamiento de los datos.

En esta etapa se busca transformar los formatos de los datos en los formatos que se consideren ideales para la posterior obtención de información. Las labores específicas y los métodos de trabajos dependerán exclusivamente de cada caso, ya que varían según el formato que posean los datos de ingreso en comparación con los formatos ideales para dichos datos. Todo esto se verá más a fondo al trabajar un caso específico. En esta etapa en particular se enfocan varias de las referencias utilizadas, dos muy destacables de ellas son las referencias [14] y [18].

3.2.3 Limpieza de datos.

Esta etapa corresponde a la limpieza de los datos. Luego de realizar el procesamiento de estos es necesario verificar de alguna forma la calidad de estos, intentando identificar posibles malos ingresos, datos incoherentes o simplemente aquellos que puedan ser un conflicto para la posterior generación del modelo estadístico.

Si se posee acceso directo a los datos y a las personas involucradas en el proceso que dio origen a dichos datos, es relevante conocer cuales datos presentan anomalías para así poder verificarlos y corregirlos en caso de encontrar la causa del error.

Para esta labor se tomarán en cuenta dos análisis genéricos, el análisis de datos atípicos (outliers) y el de datos influyentes en la regresión o modelo estadístico. Para el análisis de valores atípicos se consideran 3 métodos que permiten detectarlos, el rango intercuartílico y el test de Grubbs. Mientras que para el análisis de datos influyentes se considera el cálculo de la distancia de Cook y la diferencia en beta (DFBETA).

Rango intercuartílico: el análisis mediante el rango intercuartílico es visualmente apreciable en el diagrama de cajas como se muestra en la *Figura 3.12*, donde la caja corresponde a todos los valores entre el cuartil 1 y 3, lo cual representa el 50% de los datos. La distancia entre los valores límites de esa caja corresponde al rango

intercuartílico, y se fijan los límites extremos como 1,5 veces el valor del rango intercuartílico desde cada cara de la caja, todos los valores fuera de esos rangos se consideran valores atípicos.

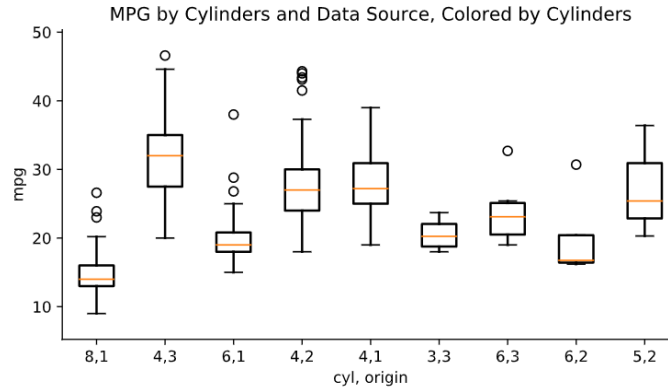


Figura 3.12: diagrama de cajas demostrativo. Fuente: extraído de la referencia [10].

Test de Grubbs: es una herramienta complementaria al rango intercuartílico, pero esta se encarga de comprobar si un valor analizado es atípico o no mediante un test de hipótesis, por lo cual se encarga de testear valor por valor, a diferencia del método anterior que sirve para encontrar un grupo de valores atípicos de forma inmediata. En este caso la hipótesis nula es que el dato analizado no es un valor atípico, mientras que la hipótesis alternativa es que si lo es, el estadístico se muestra en la ecuación (3.38) el cual corresponde a una relación entre la resta del valor extremo analizado con la media y la desviación estándar. Información obtenida de la referencia [12].

$$G = \left| \frac{\bar{x} - x_{min}}{S_x} \right| \text{ ó } \left| \frac{x_{max} - \bar{x}}{S_x} \right| \quad (3.38)$$

Dicho valor es comparado con el valor crítico mostrado en la ecuación (3.39), donde t corresponde a la distribución t student y α corresponde al nivel de significancia. En caso de que el valor del estadístico sea mayor que el valor crítico, entonces se rechaza la hipótesis nula, con lo cual se puede asumir que el valor estudiado es un dato atípico.

$$G_{crit} = \frac{N-1}{\sqrt{N}} \sqrt{\frac{t_{\left[\frac{\alpha}{2N}; N-2\right]}^2}{N-2 + t_{\left[\frac{\alpha}{2N}; N-2\right]}^2}} \quad (3.39)$$

Distancia de Cook: A diferencia de las herramientas anteriores, esta y la siguiente buscan encontrar aquellos puntos influyentes en el modelo, aquellos que alteran de forma importante el resultado de su estimación. Esta herramienta lo que busca determinar es cuanto varía la estimación al retirar un valor, lo cual se calcula gracias a la ecuación (3.40) obtenida de la referencia [5].

$$D_i = \frac{\sum_{k=1}^N (\hat{y}_{k(i)} - \hat{y}_k)}{p S_e^2} \quad (3.40)$$

Donde $\hat{y}_{k(i)}$ hace referencia al valor estimado en la posición k habiendo removido el valor i, p corresponde al número de estimadores, que en estos casos es 1, S_e^2 corresponde a la varianza del error (diferencia entre las dos estimaciones).

Diferencia en beta (DFBETA): sigue la misma lógica que la distancia de Cook, pero en vez de analizar la influencia de un valor sobre una estimación mediante el cálculo de las diferencias entre las estimaciones, en este caso se analiza la diferencia de los estimadores, para lo cual se aplica la ecuación (3.41) obtenida de la referencia [5].

$$DFBETA_{k,i} = \frac{\hat{\beta}_k - \hat{\beta}_{k(i)}}{S_{\hat{\beta}_{k(i)}} \sqrt{(1 - h_{ii(i)})}} \quad (3.41)$$

Donde se introducen dos nuevos términos, $\hat{\beta}_k$ corresponde al estimador del modelo para todos los datos, mientras que al poseer el subíndice i hace referencia al valor del mismo estimador pero habiendo retirado el dato i. El otro término es h_{ii} el cual corresponde a la ecuación (3.42).

$$h_{ii} = \frac{1}{N} + \frac{1}{N-1} \left(\frac{x_i - \bar{x}}{S_x} \right)^2 \quad (3.42)$$

3.2.4 Modelo estadístico y Resultados.

En la etapa del modelo estadístico se incluyen todos los pasos necesarios para darle forma al modelo que permitirá transformar los datos ya limpios en la información que se requiere. Dependiendo del nivel y tipo de información que se maneja, así mismo como del enfoque de los resultados esperados, este modelo puede ser enfocado hacia el Machine Learning, lo cual es muy común en análisis predictivos, pero para el caso de esta

memoria se aplicarán modelos estadísticos como la estimación de Weibull. Un punto importante de la etapa del modelo estadístico es que debe existir una retroalimentación con la etapa de limpieza de datos, verificando la calidad de aquellos que ya pasaron por las etapas anteriores y volviendo a limpiarlos.

Los fundamentos de estas dos etapas están fuertemente definidos según los conceptos de mantenimiento descritos en la sección 3.1, a continuación se deben juntar ambos pilares (análisis de datos y mantenimiento) para darle forma a la metodología, lo cual se realizará en el siguiente capítulo.

4. Metodología

4.1 Etapas de la metodología

Con lo descrito anteriormente, ahora se comienza a unificar la información de mantenimiento y de ciencia de datos en una metodología de trabajo, tomando la **Figura 3.11** como base de la metodología.

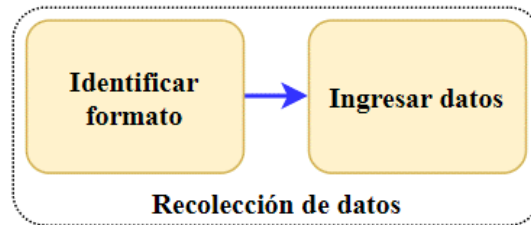


Figura 4.1: desglose sección recolección de datos. Fuente: elaboración propia.

El primer punto que se desglosa en la **Figura 4.1** es la **recolección de datos**, como se puede observar consiste en dos etapas principalmente, la primera etapa es la identificación del formato de los datos, lo cual condiciona la etapa de reprocesamiento, y la segunda etapa es el ingreso de los datos al flujo de trabajo en el formato requerido para las siguientes etapas.

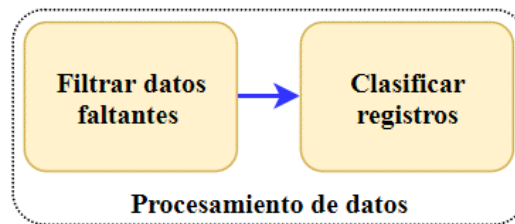


Figura 4.2: desglose sección procesamiento de datos. Fuente: elaboración propia.

Para la siguiente etapa que se desglosa en la **Figura 4.2**, el **preprocesamiento de los datos**, se realizan todas las tareas que son necesarias para pasar del formato de los datos ingresados desde la etapa anterior al formato que se considera ideal para los trabajos requeridos en las siguientes etapas.

El primer paso corresponde a filtrar datos faltantes, para lo cual se determinan que datos no pueden ser rescatados y condicionan el registro a ser inútil. Observando el registro que se considera ideal en la **Tabla 3.1**, en general las fechas y costos son elementos que no son registrados en la entrada de texto, por lo cual si no se encuentran explícitamente entonces el registro no se tomará en cuenta para la siguiente etapa.

El segundo paso es la clasificación de los registros según equipo y componente, lo cual es fundamental para generar políticas adecuadas de mantenimiento, ya que si se realiza un análisis de un equipo de forma global (realizando un ajuste de curva único para todos los datos del equipo), los resultados del análisis no podrán tener un nivel de especificación necesario como para tomar acciones concretas de mantenimiento.

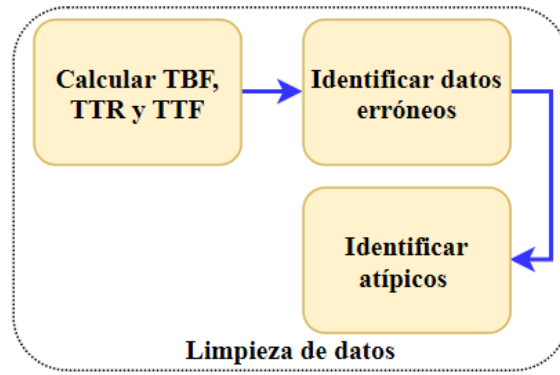


Figura 4.3: desglose sección limpieza de datos. Fuente: elaboración propia.

La siguiente etapa corresponde a la **limpieza de los datos**, mostrando su desglose en la **Figura 4.3**, el primer paso de esta etapa es el cálculo de los tiempos de interés (TBF, TTF, TTR), con lo cual se pueden identificar problemas con las fechas que son indicadores de registros defectuosos, como por ejemplo que uno de los cálculos de origen a un resultado menor que cero.

El siguiente paso corresponde a la identificación de datos atípicos en la distribución de TBF, lo cual se realiza en base (principalmente) a dos herramientas, el rango intercuartílico y el test de Grubbs.

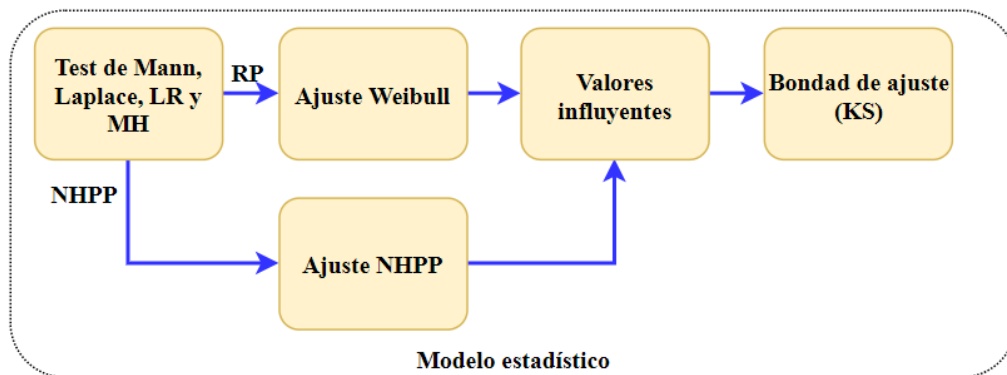


Figura 4.4: desglose sección modelo estadístico. Fuente: elaboración propia.

Como penúltima etapa está el **modelo estadístico**, el cual al desglosarlo según lo descrito en la sección 3.1 queda lo que se muestra en la **Figura 4.4**, en donde se puede apreciar que el primer paso es realizar los test para

verificar si se debe analizar un modelo RP o uno NHPP, en ambos casos se debe realizar su respectivo ajuste de curva como siguiente paso. Luego de tener el ajuste listo se analizan los puntos influyentes en la estimación, análisis que va de la mano con la identificación de valores atípicos, es fundamental que en este punto exista una retroalimentación con la etapa de limpieza de datos para verificar aquellos que deban eliminarse del análisis.

Luego de lo anterior, se realiza una prueba de bondad de ajuste, la cual es un indicador de resultado de gran relevancia para determinar si el ajuste que se está realizando, y por ende el modelo que se está analizando, puede ser considerado o se está forzando un ajuste que no sigue el comportamiento real de los datos.



Figura 4.5: desglose sección resultado. Fuente: elaboración propia.

Por último, se encuentra la etapa de **los resultados**, aquí se realiza la última transformación de los datos en la información relevante de salida, generando el tiempo óptimo de intervención, los gráficos de las curvas relevantes (Confiabilidad, probabilidad de fallas, tasa de fallas), la disponibilidad.

4.2 Metodología unificada

Tomando en conjunto todos los fragmentos de la metodología que se fueron detallando anteriormente, se genera la **Figura 4.6**, la cual corresponde a la metodología completa que se utilizará para el análisis de los historiales de falla. Es importante recalcar que esta metodología busca ser lo más específica posible sin llegar a un punto que le reste versatilidad, ya que dependiendo del caso de estudio sobre el cual se aplique se requerirán determinadas labores específicas a ser realizadas.

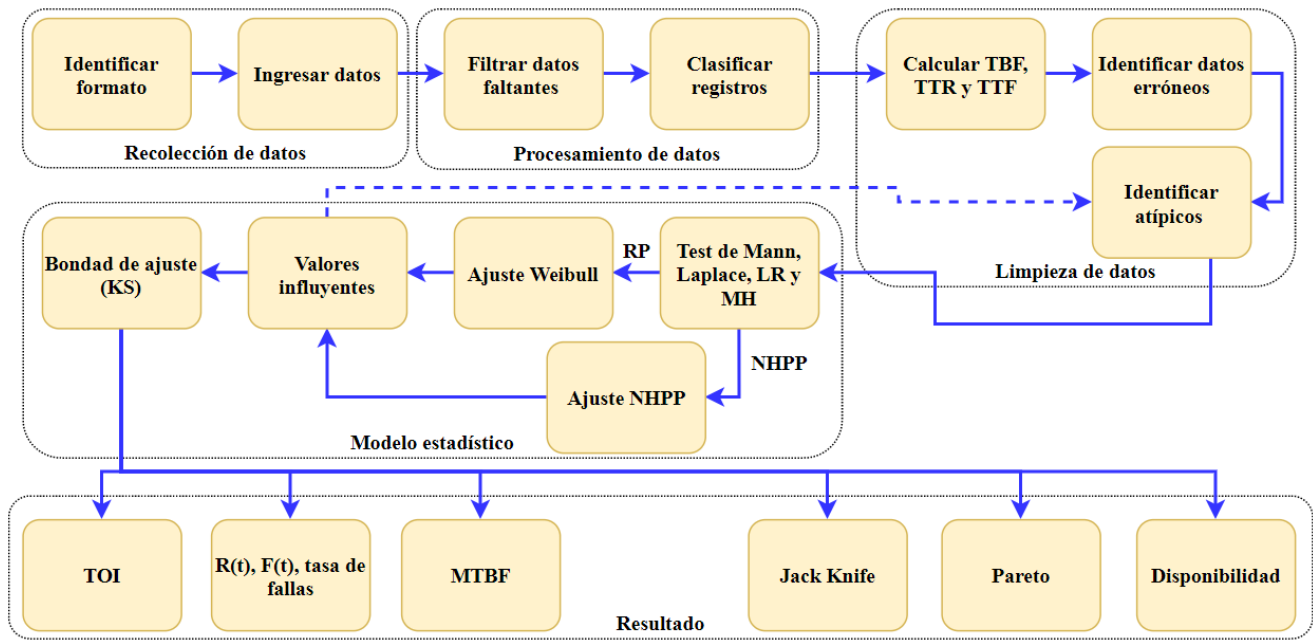


Figura 4.6: Diagrama de la metodología unificada para el análisis de historiales de falla. Fuente: elaboración propia.

5. Caso de estudio

Luego de haber estructurado la metodología necesaria para analizar un historial de fallas, es necesario probar dicha metodología en un caso de estudio para verificar cada etapa de la misma, para ello se analizará un historial de fallas entregado por CODELCO en labores efectuadas durante un proyecto para la empresa JRI, los historiales en detalle no se mostrarán, pero los análisis requeridos para esta memoria sí. Al momento de incluir detalles de costos es importante recalcar que las cifras se encuentran modificadas (amplificadas o reducidas por un factor) para respaldar la confidencialidad.

En las secciones siguientes se irán desglosando una a una las etapas de la metodología, especificando en cada paso la entrada de datos, la salida de datos y los pasos necesarios para conseguir ese flujo.

5.1 Recolección de datos: Identificar formato

Los datos obtenidos se encuentran separados en dos archivos, el primer archivo corresponde directamente al historial de fallas y presenta el formato mostrado en la **Tabla 5.1**, mientras que el segundo archivo agrega información sobre los equipos, del cual lo más relevante para el análisis es el emplazamiento donde se encuentra operando el equipo, aunque no indica en detalle ni tiempos de operación ni cambios de emplazamiento.

Tabla 5.1: Extracto de los datos presentes en el historial de fallas a analizar. Fuente: base de datos de CODELCO.

Clase de orden	Denominación	Orden	Texto breve	Cst.tot.reales	Fe.inicio real	Hora in.real	Fecha fin real	Fin real (hora)	Ubicac.técnica
MC	AGITADORES DE CAL	91957363	cambio de valvula drenaje	1.956,99	09-09-2002	8:00:00	26-09-2002	14:40:00	TCCN-CAL-AGI
MC	AGITADORES DE CAL	91540030	D.ROMERO 3.25.1 Fab. Piñón de Alta	203,7		0:00:00		0:00:00	TCCN-CAL-AGI
MC	AGITADORES DE CAL	91538107	cambio valvulas	3.969,02	04-10-2001	8:00:00	30-10-2001	14:44:16	TCCN-CAL-AGI

Detallando los datos de la **Tabla 5.1**, se puede observar que no se presenta una división por componentes, solamente se indica el código referente al equipo y una descripción de este en la columna “Denominación”, pero existe información relevante en la entrada de “Texto breve” para poder realizar la clasificación pertinente de los componentes del equipo.

Además, se aprecia una codificación de diferentes clases de orden, pudiéndose distinguir mantenimientos planificados y mantenimientos no planificados. También se aprecia la existencia de información relacionada a los costos y las fechas y horas tanto de inicio como de término de las tareas registradas.

Otro punto importante para recalcar es el formato de ingreso de cada tipo de dato, los costos se encuentran en dólares, las fechas tienen el formato día/mes/año y cada orden ingresada tiene un identificador único mostrado en la columna Orden. Teniendo todo esto en cuenta, el principal cambio de formato que se requiere para pasar al formato de ingreso mostrado en la *Tabla 3.1* es la clasificación de los componentes, trabajo que será abordado en la sección 5.4.

5.2 Recolección de datos: Ingresar datos

Para este paso ya comienza el traslado de los datos a la herramienta de preferencia, en el caso de este trabajo se utilizará Python debido a que requiere desarrollar un programa para la empresa JRI, el cual permita facilitar todo este proceso, por esto mismo el código permanecerá confidencial.

El primer paso es cambiar los títulos de las columnas por otros más genéricos y entendibles, como se enlistan a continuación bajo el formato “Título nuevo” (título antiguo):

- Clase de orden (Clase de orden)
- Denominación (Denominación)
- Orden (Orden)
- Texto (Texto breve)
- Costo (Cst.tot.reales)
- Fecha inicio (Fe.inicio real)
- Hora inicio (Hora in.real)
- Fecha termino (Fecha fin real)
- Hora termino (Fin real (hora))
- ID equipo (Ubicac.técnica)

Con eso en mente, se debe definir el estándar para ingresar cada uno de los datos, por lo cual se abordarán todos los ítems de las columnas y se definirán dichos criterios. Para el caso del tipo de orden se usará la codificación de “falla” para todos los eventos no planificados y “detención” para todos los eventos planificados.

En los casos de Denominación, Orden, Texto breve e ID equipo, se mantendrá el formato exacto en el cual se reciban los datos, ingresándolos como entradas de texto al sistema.

Para el caso de costos se ingresarán los valores como números decimales, pero en caso de que el valor a ingresar sea igual a cero, se ingresará como un dato nulo para no alterar los cálculos posteriores.

Por último, para las fechas y horas, se ingresarán como valores nulos aquellos registros en los que no existe la fecha, o que el formato es distinto al esperado. Para el caso específico de las horas existe un problema de formato en el cual existe el ingreso “0:00:00” haciendo referencia al inicio del día, pero también existe el ingreso “24:00:00”, por lo cual se decide reemplazar este último por “23:59:59”.

5.3 Procesamiento de datos: Filtrar datos faltantes

Luego de realizar todos los pasos anteriores se tiene un total de 341.173 registros ingresados, de los cuales 98.404 (28,84 %) no presentan las fechas ingresadas adecuadamente, 184 (0,05 %) no presentan bien ingresado los costos, 68.331 (20,03 %) presentan un ingreso de costos igual a cero.

Para poder presentar de forma adecuada los trabajos que corresponden a las siguientes secciones se debe realizar una priorización entre los 2.558 equipos presentes en el total de registros. Para realizar esta priorización se utilizan las herramientas de Jack Knife y Pareto. Para esto se debe calcular el TTR, lo cual simplemente es la diferencia de tiempo entre el inicio de la falla y el término de esta.

Para el grafico de Pareto solo es necesario conocer el número de fallas registradas para cada equipo, mientras que para Jack Knife se requiere además el MTTR, teniendo estos registros calculados se genera primero el grafico de Pareto, debido a la gran cantidad de equipos en la mostrado en la **Figura 5.1** se muestran los 20 equipos que presentan más fallas.

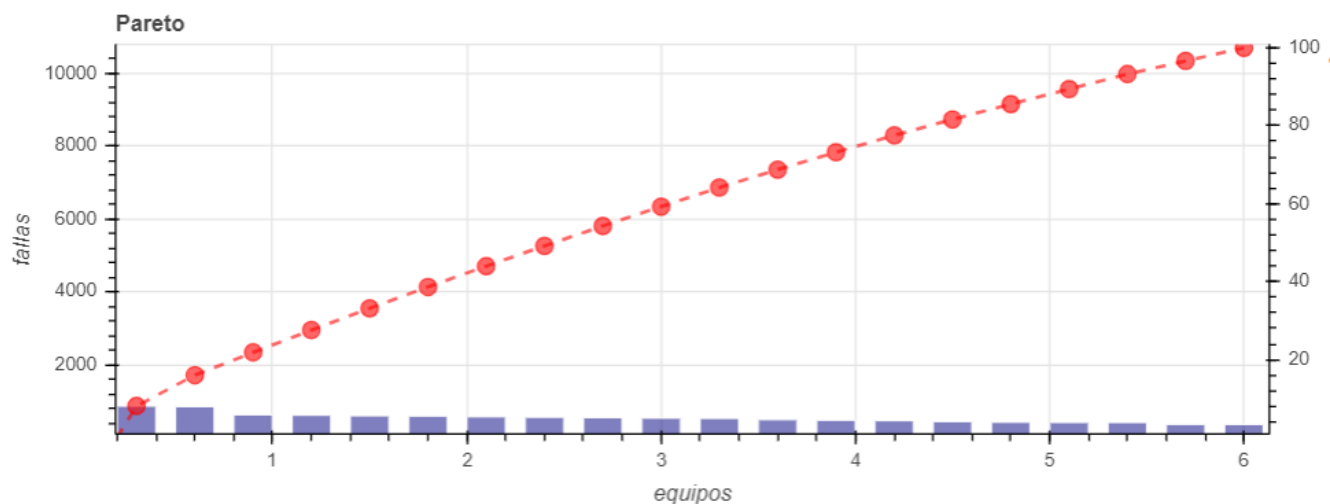


Figura 5.1: Diagrama de Pareto de los 20 equipos con más fallas del listado de equipos completo. Fuente: elaboración propia.

Para el caso del diagrama de Jack Knife se grafican todos los equipos, tal como se muestra en la **Figura 5.2**, en la cual se muestra enmarcado el grupo de equipos críticos, de los cuales se selecciona un Jumbo de perforación frontal (TM1R).

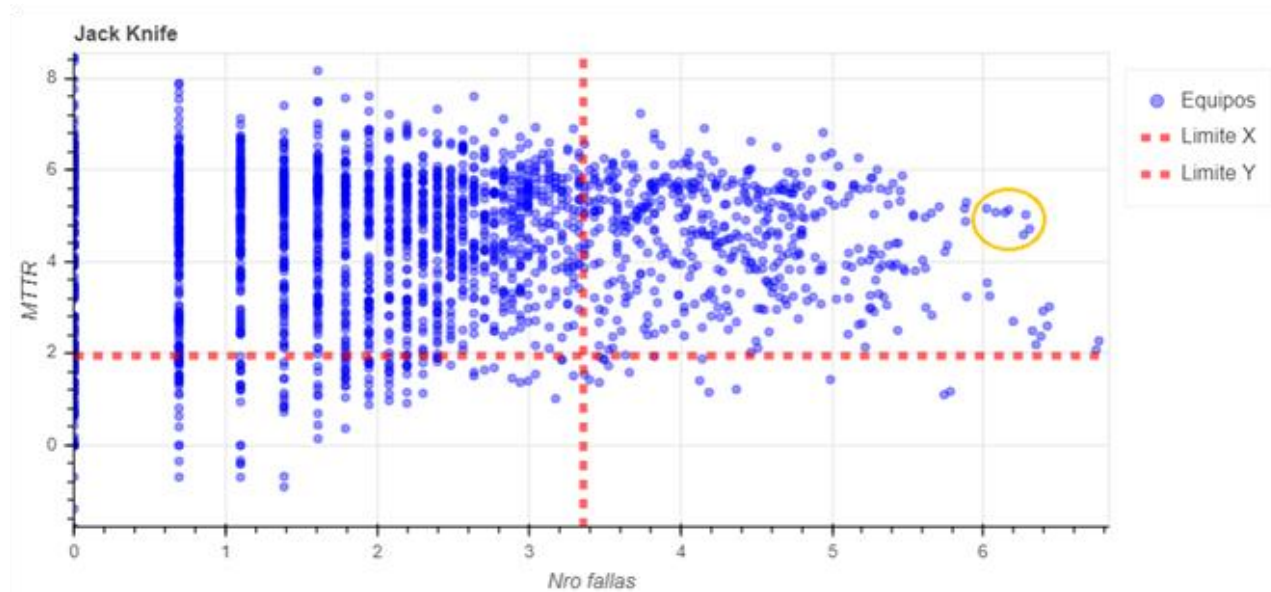


Figura 5.2: Diagrama de Jack-knife del listado completo de equipos. Fuente: elaboración propia.

Habiendo seleccionado ya el equipo en el cual se centrará la aplicación de esta metodología se tiene que para dicho equipo existen 1.040 registros, de los cuales 941 (90,48%) presentan el formato adecuado de ingreso. Además, se tiene que 534 de los registros que presentan el formato adecuado corresponden a registros de fallas, mientras que 407 corresponden a detenciones.

5.4 Procesamiento de datos: Clasificar registros

Especialmente para este caso de estudio se requiere trabajar exhaustivamente esta etapa, ya que no viene una clasificación previa por componente, sino que dicha clasificación debe extraerse de las entradas de texto. En este punto es importante recordar una expresión muy común en la ciencia de datos “garbage in, garbage out”, lo cual quiere decir que si se ingresa basura, lo único que se obtendrá de la salida del modelo es basura, por esto, aunque se le pueda sacar más provecho a datos con un ingreso deficiente, siempre debe mantenerse el foco de mejorar los sistemas de registro de datos.

Recordando el formato de ingreso de los registros mostrado en la **Tabla 5.1**, se puede apreciar una entrada de texto limitada en caracteres, con errores ortográficos y símbolos anotados. Considerando todo esto se requieren los siguientes pasos para trabajar con estos datos:

1. Generar un duplicado de la entrada de texto para poder modificarla sin perder el registro original.
2. Eliminar caracteres y mayúsculas del duplicado.
3. Corregir la ortografía.
4. Identificar palabras clave (nombres de componentes).
5. Generar clasificación de componentes.

Considerando cada uno de los pasos descritos anteriormente se generan registros con el formato mostrado en la *Tabla 5.2*, pudiéndose identificar los componentes deseados, para realizar esta tarea existen diversos métodos, uno de ellos es la aplicación de un modelo de procesamiento de lenguaje natural (NLP), pero el entrenamiento de este modelo requiere información previamente clasificada, además de presentar el problema de no ser tan versátil como se desea a la hora de enfrentarse a las diferentes culturas que se van formando en cada uno de los grupos de mantenimiento (las formas de referirse a determinados componentes varían de emplazamiento en emplazamiento, incluso de turno en turno).

Teniendo en cuenta todo lo descrito se decidió aplicar el método más sencillo y versátil, correspondiendo a un cruce entre las palabras clave de las que se puede hablar, considerando orden de prioridad en dichas palabras, esto debido a que algunos registros son más completos que otros, por ejemplo, existen varios registros sobre el cilindro de avance, pero a su vez existen diversos ingresos que solo mencionan un cilindro, sin indicar sobre cual se está hablando, siendo relevante priorizar los registros más específicos con respecto a los más generales.

Tabla 5.2: Extracto de los registros clasificados. Fuente: elaboración propia.

Clase_orden	ID_equipo	Texto	Texto_limpio	Clasificador
detención	TM1R	COMPRA ALTERNADOR N/P 01180005 ATLAS	compra alternador n p atlas	alternador
detención	TM1R	Control equipo en garantía	control equipo en garantía	
falla	TM1R	tp acumul cop 1838 n°	acumulador n	acumulador
detención	TM1R	MANT.15 JUMBO EMPERADOR # 371 BOOMER	mantenimiento jumbo emperador boomer	
falla	TM1R	tp cabezal 1838 cop n°	cabezal n	cabezal

También se observan diversos registros que son imposibles de clasificar dada la poca información entregada en la entrada de texto, o por hacer referencia a trabajos muy generales, sin indicar sobre que componente en particular se están realizando.

Pese a todo esto, se logran clasificar los componentes mostrados en la **Tabla 5.3**, lamentablemente en la ID “None” se muestran todos los registros que no fue posible clasificar teniendo en mente los problemas descritos anteriormente en torno a la deficiencia del ingreso de los registros de las entradas de texto, o la falta de un método más riguroso para indicar el componente específico al cual se le realiza la tarea registrada.

Con esta información ya se puede avanzar a la siguiente etapa de la metodología, correspondiente al cálculo de los tiempos de interés (TBF y TTR).

Tabla 5.3: Resumen de numero de fallas y detenciones de los componentes clasificados. Extracto de la Tabla 8.1 (anexo 8.1).

ID componente	Fallas	Detenciones
None	115	256
cabezal	75	44
hco	57	0
acumulador	54	40
perforadora	18	11

5.5 Limpieza de datos: Calcular TBF y TTR

Esta sección no requiere una explicación muy detallada, solo se deben calcular los tiempos entre fallas y los tiempos para reparar. El TTR ya fue calculado previamente para poder priorizar un equipo, sobre el cual se han estado realizando todos los cálculos y análisis hasta ahora. Solo quedan dos puntos importantes a mencionar.

El primer punto que se debe tener en cuenta es que se añade el cálculo del TBI (tiempo entre intervenciones), el cual obedece a la misma lógica que el TBF (mostrada también en la **Figura 3.2**), pero incluye tanto las fallas como las detenciones. Se incluye este tiempo dado que se utilizará el estimador de confiabilidad de Lewis, el cual requiere los tiempos entre intervenciones dado que considera en su estimación de confiabilidad tanto las fallas como las detenciones.

El segundo punto corresponde al criterio a adoptar ante mantenimientos sobrelapados, es decir, eventos de mantenimiento que comienzan mientras ya existe uno en curso, esto se puede deber a diversos factores como por ejemplo, mantenimiento oportunista (aprovechar que ya se está solucionando un problema en el activo físico para solucionar otros problemas o potenciales problemas), ingresos erróneos o una clasificación errónea. Para este caso de estudio se tomó la decisión de no incluir en el análisis los eventos sobrelapados.

5.6 Limpieza de datos: Identificar datos erróneos

Teniendo en cuenta el criterio establecido en la sección anterior, todos los registros en los que el TBF o TBI sea negativo o igual a 0 se consideran como datos erróneos, con lo cual se obtienen las cantidades de datos ingresados y datos validos mostradas en la

Tabla 5.4: Extracto de la Tabla 8.2: Número de registros validos luego del cálculo de TTR, incluyendo su respectivo porcentaje en relación al total de datos.

Componente	Fallas ingresadas	Fallas válidas	% fallas	Det. ingresadas	Det. válidas	% detenciones
cabezal	75	56	75	119	89	75
hco	57	57	100	57	57	100
acumulador	54	48	89	94	79	84
perforadora	18	18	100	29	28	97
sistema avance	17	17	100	17	17	100
cilindro avance	14	14	100	27	25	93
sistema eléctrico	14	14	100	14	14	100
peju	14	11	79	27	21	78
cilindro	12	12	100	16	15	94
cable	10	8	80	10	8	80

5.7 Limpieza de datos: Identificar datos atípicos

Teniendo los datos ya calculados y habiendo revisado aquellos erróneos, ahora es un buen momento para identificar los valores atípicos, para lo cual se deben aplicar las herramientas descritas en la sección 3.2.3. Como primera aproximación se tiene el diagrama de cajas mostrado en la **Figura 5.3**, el cual se realizó teniendo la consideración de clasificar como datos atípicos aquellos que se encuentran a una distancia de 3 RIC (rango intercuartílico).

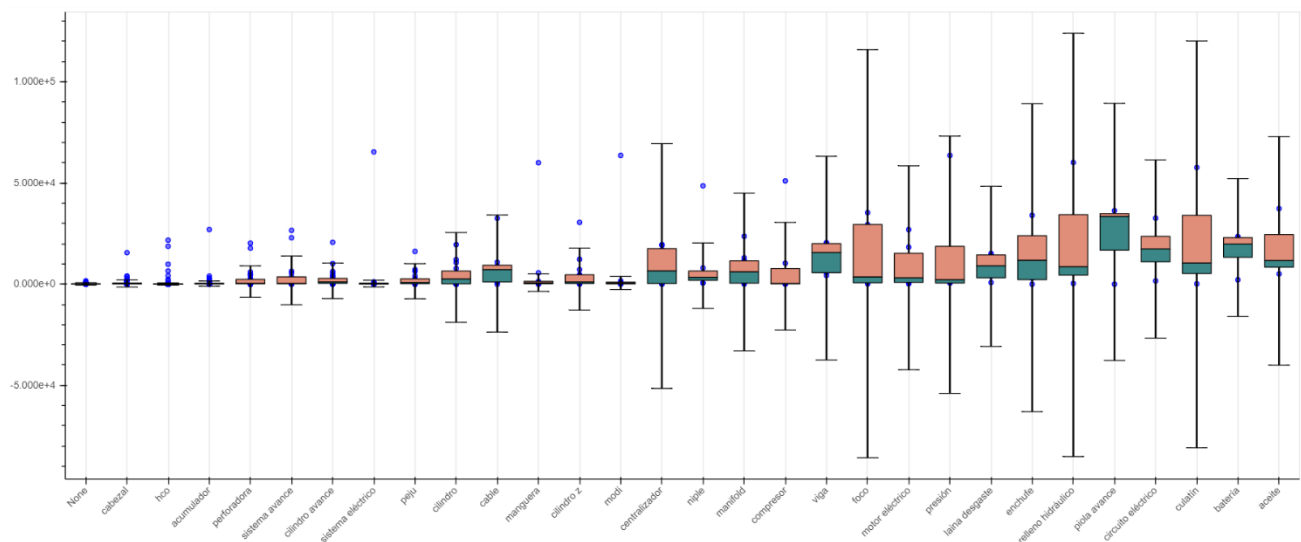


Figura 5.3: Diagrama de cajas para los 30 componentes Fuente: elaboración propia.

Continuando con lo descrito en la sección del marco teórico, se realizó el test de Grubbs mediante las ecuaciones (3.38) y (3.39). Considerando ambos criterios se determinan las cantidades de valores atípicos presentes para cada set de datos correspondientes a los registros clasificados por componente, mostrándose el resumen de estos datos en la **Tabla 5.5**.

Tabla 5.5: Extracto de la Tabla 8.3: Numero de datos atípicos y el impacto de la eliminación de un dato con respecto al total de datos clasificados por componente.

Componente	atípicos	% atípicos	% por dato
cabezal	4	4,49	1,12
hco	7	12,28	1,75
acumulador	6	7,59	1,27
perforadora	2	7,14	3,57
sistema avance	2	11,76	5,88
cilindro avance	1	4	4
sistema eléctrico	1	7,14	7,14
peju	1	4,76	4,76
cilindro	0	0	6,67
cable	0	0	12,5

Un dato extra a considerar que se muestra en la tabla reciente es el % por dato, lo cual corresponde a cuanto es el porcentaje de un dato respecto al total de datos, esto toma relevancia al considerar que ante la decisión de eliminar valores considerados atípicos, estos no deben superar el 5% siendo esta una consideración, no una ley, pero para este trabajo se tomará como si lo fuese.

En esta etapa el foco estaba en identificar los datos atípicos, no en eliminarlos. En situaciones normales nunca debe eliminarse un dato, si se identifican registros dudosos se debe hacer una investigación de las posibles causas que llevaron a generar dicho registro y corregirlo, pero dado que aquí se está trabajando con una base de datos facilitada por un tercero es imposible realizar dicho seguimiento, por lo cual con este análisis de valores atípicos y con el posterior análisis de la influencia de dichos valores en el modelo se establecerá un criterio para la eliminación de los valores más problemáticos, y uno de los criterios a considerar es que los datos eliminados no pueden superar el 5% del total de datos.

5.8 Modelo estadístico: Test de Mann, Laplace, LR y MH

Si bien, según lo descrito en los objetivos de este trabajo, el modelo NHPP se escapa de los alcances de esta memoria, la metodología generada considera la entrega de información que permita determinar si debiese considerarse la estimación de dicho modelo al no poder trabajar bajo el supuesto de un proceso renovable.

Para entregar dicha información se realizaron los test de Mann, de Laplace, de Lewis-Robinson y el MHT, utilizando las ecuaciones (3.9), (3.10), (3.11) y (3.12). Con lo cual se obtuvieron los resultados registrados en la **Tabla 5.6**.

Tabla 5.6: Extracto de la Tabla 8.4: Resultados de los test para determinar si aplicar un modelo para un proceso renovable o un proceso no homogéneo.

Componente	Mann	Mann_c	Mann dif%	L	LR	MH
cabecal	701	715,5	2,03	-1415,5	-541,83	216,33
hco	689	798	13,66	-702,0	-211,17	11,21
acumulador	441	473	6,77	-1178,9	-308,66	173,38
perforadora	56	76,5	26,8	-229,5	-113,64	22,97
sistema avance	35	68	48,5	-98,4	-49,83	4,43
cilindro avance	50	39	28,2	-197,4	-118,05	64,56
sistema eléctrico	36	45,5	20,8	-69,1	-19,71	0,92
peju	37	27,5	34,5	-149,6	-98,05	57,89

Componente	Mann	Mann_c	Mann dif%	L	LR	MH
cilindro	20	33	39,3	-82,8	-63,92	19,25
cable	17	14	21,4	-28,9	-23,71	17,5

Se puede apreciar como para el cabezal y el acumulador el test de Mann está por debajo del criterio que previamente se estipuló en la sección del Marco Teorice, pero para los demás componentes, sin ser una diferencia mayor al 50%, existe una diferencia importante entre el valor crítico del test y el estadístico, por lo cual sería interesante observar un contraste en todo el análisis que se realiza en las siguientes sesiones al aplicar un modelo NHPP, tanto para los componentes que cumplen con la hipótesis nula del test de Mann, como para aquellos que no lo hacen. Los demás test indican la presencia de una tendencia en los datos, siendo otro punto a considerar a la hora de optar por el ajuste de un modelo NHPP.

Un tema a considerar luego de observar estos resultados es la sensibilidad de ellos ante los datos faltantes, ya que un registro errático (no riguroso, con información faltante) puede generar una falsa perspectiva de una tendencia no homogénea, lo cual se convierte en otra interrogante, para la cual se requiere trabajar con muestras de datos más extensas y detalladas, en las cuales se pueda ir alterando dichos registros para observar la sensibilidad de estos análisis a la usencia de algunos registros.

5.9 Modelo estadístico: Ajuste Weibull

En esta sección de la metodología estipulada ya es momento de realizar el ajuste de curva, para ello el primer paso es estimar la confiabilidad, considerando la existencia de datos de detenciones y de fallas en los registros del caso de estudio se opta por utilizar el estimador de Lewis, calculado con la ecuación (3.17).

Para estimar la confiabilidad de esta forma primero se aplica un criterio descrito en la referencia [1], el cual determina que al momento de estimar la confiabilidad o la probabilidad de fallas en función de los tiempos entre intervenciones o entre fallas, primero se debe asegurar que no existan tiempos duplicados, ya que se estaría asignando una probabilidad distinta para dos tiempos iguales.

Habiendo realizado la estimación de la confiabilidad, se realiza el ajuste de curva de una Weibull de dos parámetros (alpha y beta). El ajuste de curva se realiza con el método de los mínimos cuadrados presentado en la referencia [1], lo cual es realizado para cada uno de los componentes, obteniéndose los resultados mostrados en la **Tabla 5.7**.

Tabla 5.7: extracto de la Tabla 8.5: Resultados de la estimación de un modelo de Weibull de dos parámetros.

Componente	R^2	α (alpha)	β (beta)
cabezal	0,96	913,84	0,74
hco	0,81	256,22	0,52
acumulador	0,93	829,02	1,04
perforadora	0,96	3.230,17	0,57
sistema avance	0,81	2.303,83	0,52
cilindro avance	0,98	5.998,73	0,69
sistema eléctrico	0,81	890,23	0,44
peju	0,90	4.295,45	0,80
cilindro	0,97	6.133,73	0,51
cable	0,91	10.920,17	0,73

Se puede observar una buena correlación en la mayoría de los ajustes, pero en casi todos los equipos el valor de beta está por debajo de 1, lo cual debe tenerse en cuenta dada la relevancia de dicho valor en el análisis del comportamiento del componente.

A continuación se muestran los gráficos de interés generados para uno de los componentes que se están analizando, esto se debe simplemente a una consideración de la extensión del documento y la relevancia de dicha información para los fines de este trabajo, ya que el foco está en la metodología y no en los datos en sí.

En la **Figura 5.4** se muestra el gráfico de las curvas confiabilidad y probabilidad de falla generada luego de la obtención de los parámetros del modelo Weibull de dos parámetros para el cabezal, además se encuentran graficados los puntos de confiabilidad estimados mediante Lewis.

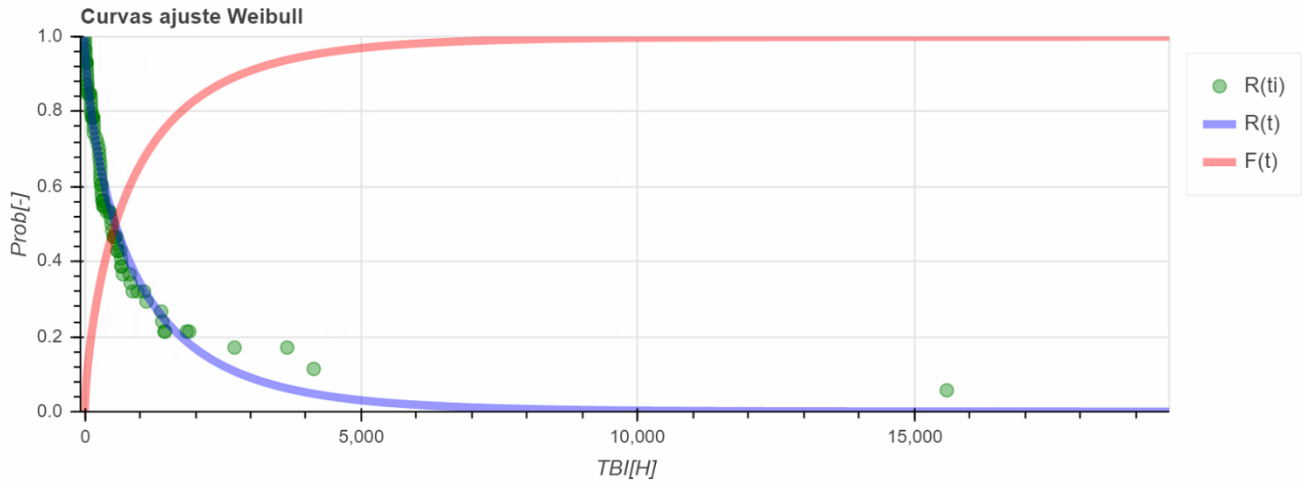


Figura 5.4: Curvas de confiabilidad y probabilidad de fallas para el componente clasificado como cabezal. Fuente: elaboración propia.

A simple vista puede observarse la presencia de un dato muy alejado del resto (más de 3 veces el valor del dato más cercano), lo cual se debe a que aún no se han removido los datos considerados atípicos. Antes de considerar remover dichos datos es importante determinar si existe un sesgo en la estimación, lo cual se comprueba mediante el análisis del residuo entre el ajuste de curva y los valores específicos de confiabilidad estimados con Lewis, dicho análisis se puede realizar mediante un histograma del cálculo mencionado, en caso de existir un sesgo entonces los valores graficados en el histograma no se comportarán como una normal con media igual a 0, lo cual se ve claramente mostrado en la **Figura 5.5**, donde se aprecia un importante desplazamiento de los residuos al cuadrante negativo.

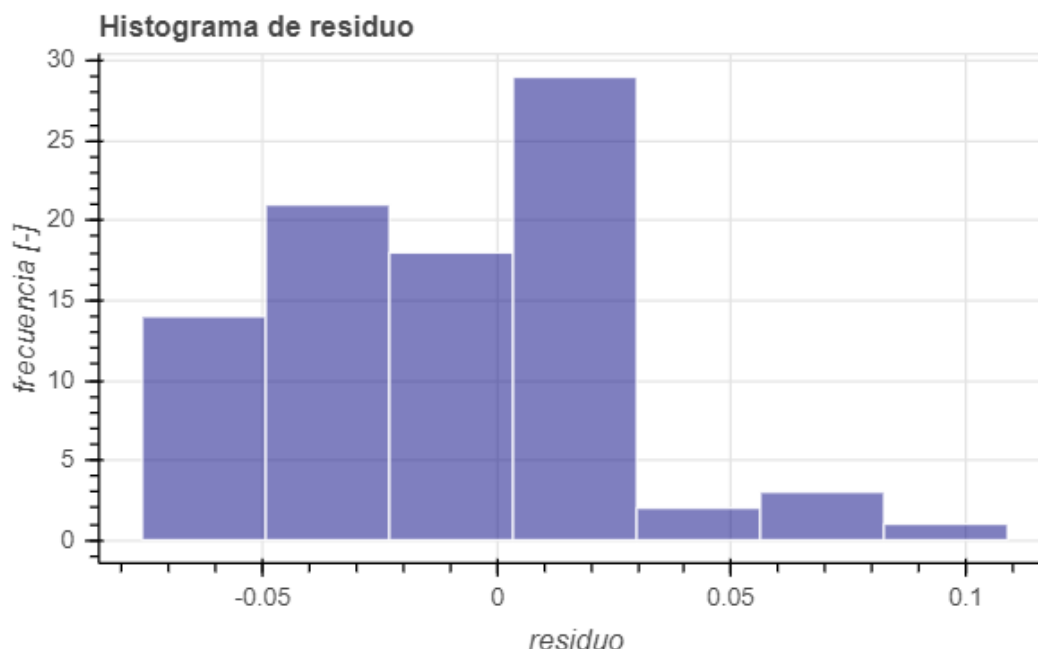


Figura 5.5: Histograma de residuos sin eliminar valores atípicos. Fuente: elaboración propia.

5.10 Modelo estadístico: Valores influyentes

Teniendo el primer ajuste realizado, en esta sección se determina la existencia de los valores influyentes, lo cual corresponde a valores que tienen un gran impacto en el resultado del ajuste del modelo. Para ello se utilizan las ecuaciones (3.40) y (3.41) para estimar la distancia de Cook y la DFBETA respectivamente.

Con este cálculo realizado se opta por eliminar los datos que estén a más de 3 RIC alejados del 50% de los datos, la DFBETA del dato en si este alejada a más de 1,5 RIC de los valores de DFBETA de los demás datos, siendo criterios concordantes con la distancia de Cook y el test de Grubbs. Además recordando que los datos eliminados no pueden superar el 5% del total de datos. El registro de los datos eliminados, junto con los números totales de datos, la cantidad de atípicos y el porcentaje que significa un dato con respecto al total se muestra en la **Tabla 5.8**.

Tabla 5.8: Resumen de datos eliminados para cada componente.

Componente	Fallas válidas	Detenciones válidas	atípicos	% por dato	eliminados
cabecal	56,00	89,00	4,00	1,12	1,00
hco	57,00	57,00	7,00	1,75	2,00
acumulador	48,00	79,00	6,00	1,27	3,00
perforadora	18,00	28,00	2,00	3,57	1,00

Componente	Fallas válidas	Detenciones válidas	atípicos	% por dato	eliminados
cilindro avance	14,00	25,00	1,00	4,00	1,00
peju	11,00	21,00	1,00	4,76	1,00

Con los datos atípicos removidos se vuelve a analizar el sesgo presente en la estimación del modelo, tomando el mismo componente (cabezal), se obtiene el histograma mostrado en la **Figura 5.6**, en el cual se puede apreciar una disminución en el sesgo apreciado en la **Figura 5.5**, pese a no haberse eliminado completamente, lo cual también obedece a los rigurosos criterios establecidos para la eliminación de datos atípicos.



Figura 5.6: Histograma de residuos luego de eliminar valores atípicos. Fuente: elaboración propia.

Solo como prueba, para determinar el impacto del criterio DFBETA en la eliminación del sesgo de la estimación, considerando el mismo componente se decide remover todos los datos que cumplan con los demás criterios, con lo cual se genera la **Figura 5.7**, pudiendo apreciarse una mayor disminución del sesgo, centrándose aún más los residuos en torno al cero.

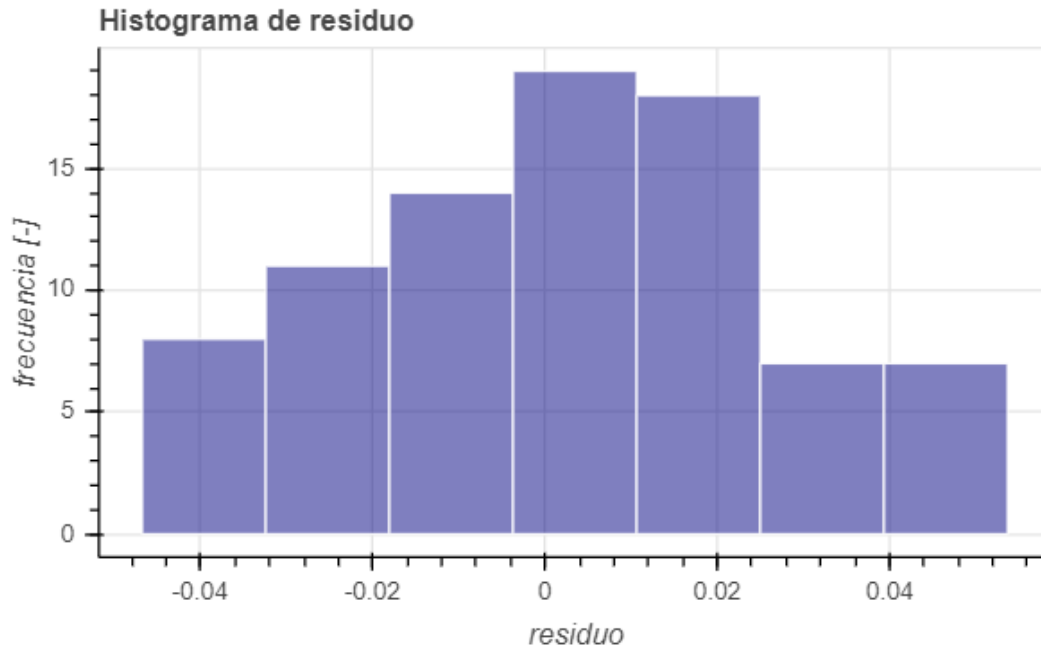


Figura 5.7: Histograma de residuos eliminando el máximo de valores sin considerar criterio de DFBETA. Fuente: elaboración propia.

5.11 Modelo estadístico: Bondad de ajuste

Como último paso de la verificación del modelo estadístico elegido se utiliza el test de kolmogorov-smirnov (KS) para comprobar que los datos se ajustan al modelo que se les está asignando, pero además de hacer esta prueba con el modelo de Weibull de dos parámetros también se realiza con el modelo de tres parámetros dado que el proceso de limpieza de los datos ya está terminado y es un buen momento para verificar diversos ajustes de curvas que no requieran unas consideraciones particulares en su limpieza de datos.

Los resultados de estos cálculos se muestran en la **Tabla 5.9**, en donde las primeras dos columnas corresponden a los parámetros de la Weibull de dos parámetros, mientras que las dos columnas siguientes corresponden al estadístico del test de KS y el valor crítico para dicho test respectivamente, repitiéndose el esquema en las siguientes columnas, pero ahora para la curva de Weibull de tres parámetros.

Tabla 5.9: extracto de la Tabla 8.6: resultados de las pruebas de bondad de ajuste tanto para el modelo de dos parámetros como para el de tres.

Componente	alpha	beta	KS	Kscritico	alpha	beta	t_0	KS	Kscritico
cabezal	785,01	0,79	0,08	0,19	778,48	0,81	-2,58	0,08	0,19
hco	176,76	0,61	0,14	0,19	180,47	0,47	2,50	0,16	0,19
acumulador	664,82	1,16	0,11	0,21	656,92	0,91	30,45	0,09	0,21
perforadora	2552,97	0,60	0,10	0,33	2606,61	0,64	-20,30	0,12	0,33
sistema avance	2303,83	0,52	0,16	0,33	2107,60	0,49	37,24	0,16	0,33
cilindro avance	4677,78	0,74	0,08	0,39	4354,10	0,86	-54,41	0,05	0,39
sistema eléctrico	890,23	0,44	0,21	0,36	1144,85	0,26	6,49	0,22	0,36
peju	2833,62	1,00	0,08	0,41	3012,33	0,71	182,68	0,06	0,41
cilindro	6133,73	0,51	0,11	0,39	7048,27	0,76	-435,98	0,09	0,39
cable	10920,17	0,73	0,12	0,48	11486,74	0,80	-384,08	0,12	0,48

Considerando estos valores se observa que todos los ajustes de curva cumplen con los criterios de bondad de ajuste del test KS, además de ya encontrarse aplicados los criterios de limpieza de datos establecidos en la metodología, por lo cual se considera oportuno pasar a la última etapa, correspondiente a la entrega de resultados.

5.12 Resultados

Para los modelos de disponibilidad de cada componente se obtienen dos indicadores, el primero corresponde a la disponibilidad genérica y el segundo corresponde a la disponibilidad inherente. Mientras que para el análisis de los tiempos óptimos de intervención se obtiene el tiempo optimo en si (T_p) correspondiente al mínimo de la curva EC presentada en la sección 3.1.11, pero en caso de que dicha curva no posea un mínimo se indica la política de correr a la falla, además se obtiene el mínimo y la asíntota a la que tiende la curva EC. También se obtiene el tiempo en el cual la confiabilidad baja al 80% y el MTBF. Todos estos resultados se presentan en la **Tabla 5.10**

Tabla 5.10: Resumen resultados finales TOI, cálculos de disponibilidad y MTBF por componente. Extracto de la Tabla 8.7

Componente	T_p	EC min	EC asíntota	MTBF	TOI_R0,8	A_G	A_I
cabezal	Correr a la falla	0,8560	0,7349	868,1651	120,9493	0,8123	0,8815
hco	Correr a la falla	2,1967	1,0520	406,1681	10,0230	0,9954	0,9940
acumulador	Correr a la falla	0,5122	0,4972	715,6596	157,9008	0,8175	0,8905
perforadora	Correr a la falla	0,6037	0,4111	3621,8546	227,3636	0,9886	0,9941
sistema avance	Correr a la falla	0,9861	0,6503	4350,6634	138,3497	0,9896	0,9902
cilindro avance	Correr a la falla	0,1372	0,1087	4657,2004	702,7258	0,9406	0,9746

Componente	Tp	EC min	EC asintota	MTBF	TOI_R0,8	A_G	A_I
sistema eléctrico	Correr a la falla	0,2505	0,0732	23899,9245	9,7589	0,9910	0,9981
peju	Correr a la falla	0,4585	0,2662	3964,1993	542,8081	0,9014	0,9512
cilindro	Correr a la falla	0,1684	0,1309	7829,6300	554,1488	0,9819	0,9896
cable	Correr a la falla	0,1344	0,1109	12599,0631	1388,8031	0,9904	0,9934

Con todos estos resultados ya se tienen los insumos necesarios para gestionar una política de mantenimiento que sea respaldada gracias a una información más robusta. Además, se puede observar que la información obtenida es consistente con equipos que se encuentran sobre exigidos, con una mortalidad prematura, lo cual es coherente con la época de la minería a la cual pertenecen los datos, siendo una época de grandes niveles de producción y ganancias para dicho rubro.

6. Conclusiones

Habiendo partido de una elaboración teórica de la metodología para el procesamiento de la información, al aplicarla a un caso de estudio en particular se determinó el ingreso de información mínimo para los análisis estipulados permite realizar cada una de las etapas requeridas para agregarle valor al proceso de transformación de los datos en información, obteniendo la disponibilidad, la confiabilidad y otros indicadores (como el tiempo óptimo de intervención) para variados componentes que son rescatados de un conjunto de registros que no estaban previamente clasificados. Pero a su vez se puede observar como la ausencia de un registro riguroso de la información da lugar a una mala calidad de la información obtenida, ya que pese a las mejoras que se puedan realizar en el procesamiento de esta, la calidad del resultado depende en gran medida de la calidad del ingreso.

Con esto en mente se consideran dos escenarios, el mínimo de información requerida para obtener los modelos estadísticos es justamente el caso estudiado, en el cual es posible realizar una segmentación de los componentes gracias a una entrada de texto, pero el caso ideal es que dicha segmentación no sea necesario forzarla, sino que cada registro se realice al momento de ingresar el evento considerando el componente específico al cual se le está realizando la intervención.

Considerando aquello, tal como se muestra en la sección 5.12, fue posible obtener el modelo de confiabilidad y disponibilidad de cada uno de los componentes clasificado y que presentaba un número adecuado de eventos registrados, aunque lamentablemente en las clasificaciones de dichos componentes fue imposible distinguir en los registros aquellos componentes que se encuentran más de una vez en el equipo, es decir que al momento de hablar de neumáticos por ejemplo, no se especifica de cuál de todos se está hablando.

También fue posible identificar qué tipo de política de intervención a edad constante se ve reflejada en los datos, lo cual corresponde en su mayoría a una política de correr a la falla para el caso de estudio analizado. Con lo cual, teniendo en mente los insumos que se desea generar con la metodología, establecidos en el último paso de ésta, se pudo obtener una buena base de insumos para una gestión de activos fijos más robusta.

7. Referencias

- [1] Mora, L. A. (2009). Mantenimiento - Planeación, ejecución y control. Alfaomega Grupo Editor.
- [2] T. Grubessich F. «Ingeniería de pantas industriales – ICN342», p. 77.
- [3] E. E. Lewis (1987). Introduction to reliability engineering (No. 04; TA169, L4.).
- [4] R. Pascual (2005). El arte de mantener.
- [5] «Online Statistics Course Materials». <http://lithoguru.com/scientist/statistics/course.html> (accedido may 24, 2020).
- [6] Schutt, R., & O'Neil, C. (2014). Doing data science: Straight talk from the frontline. O'Reilly.
- [7] Létourneau, S., Yang, C., Drummond, C., Scarlett, E., Valdés, J., & Zaluski, M. (2005). A domain independent data mining methodology for prognostics. In Essential technologies for successful prognostics: proceedings of the 59th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, Virginia, April 18-21, 2005.
- [8] D. M. Louit, R. Pascual, y A. K. S. Jardine, «A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data», Reliab. Eng. Syst. Saf., vol. 94, n.o 10, pp. 1618-1628, oct. 2009, doi: 10.1016/j.ress.2009.04.001.
- [9] A. J. Cárdenas, «Sr. JUAN CARLOS FIGUEROA BARRA», p. 111.
- [10] «Boxplot chart — HoloViews 1.13.2 documentation». http://holoviews.org/gallery/demos/matplotlib/boxplot_chart.html (accedido may 24, 2020).
- [11] Machine Learning Approaches for Failure Type Detection and Predictive Maintenance - Jahnke Patrick 2015.
- [12] «BioMates: Contraste de Grubbs». https://www.sgapeio.es/INFORMEST/VICongreso/taller/applets/biomates/explora/explora_grubbs/explora_grubbs.htm (accedido may 24, 2020).
- [13] E. F. Alsina, M. Chica, K. Trawiński, and A. Regattieri, «On the use of machine learning methods to predict component reliability from data-driven industrial case studies, » *Int. J. Adv. Manuf. Technol.*, vol. 94, no. 5–8, pp. 2419–2433, 2018, doi: 10.1007/s00170-017-1039-x.
- [14] A. Famili, W. M. Shen, R. Weber, and E. Simoudis, «Data preprocessing and intelligent data analysis», *Intell. Data Anal.*, vol. 1, no. 1, pp. 3–23, 1997, doi: 10.3233/IDA-1997-1102.
- [15] F. Gürbüz, L. Özbakir, and H. Yapici, «Data mining and preprocessing application on component reports of an airline company in Turkey», *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6618–6626, 2011, doi: 10.1016/j.eswa.2010.11.076.
- [16] F. Herrera, «Big Data: Preprocesamiento y calidad de datos», *Novática*, no. 237, p. 17, 2016.
- [17] R. B. Patil, B. S. Kothavale, and L. Y. Waghmode, «Selection of time-to-failure model for computerized numerical control turning center based on the assessment of trends in maintenance data», *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 233, no. 2, pp. 105–117, 2019, doi: 10.1177/1748006X18759124.
- [18] A. Tam and I. Kwan, «Data quality in asset management—Creating and maintaining a foundation for data analytics». Springer International Publishing, 2019.
- [19] P. Tsarouhas, «Statistical analysis of failure data for estimating reliability, availability and maintainability of an automated croissant production line», *J. Qual. Maint. Eng.*, vol. 25, no. 3, pp. 452–475, 2019, doi: 10.1108/JQME-04-2018-0029.

- [20] L. Y. Zhai, W. F. Lu, Y. Liu, X. Li, and G. Vachtsevanos, «Analysis of time-to-failure data with weibull model in product life cycle management», *Re-Engineering Manuf. Sustain. - Proc. 20th CIRP Int. Conf. Life Cycle Eng.*, pp. 699–703, 2013.
- [21] Martínez Laura, «Métodos De Inferencia Para La Distribución Weibull: Aplicación En Fiabilidad Industrial», p. 68, 2011, [Online]. Available: http://eio.usc.es/pub/mte/descargas/proyectosfinmaster/proyecto_613.pdf.
- [22] «Confidence Bounds - ReliaWiki». http://reliawiki.org/index.php/Confidence_Bounds (accedido jun 24, 2020).
- [23] F. S. C. Reli and M. I. L-hobk-, «MIL-HDBK- MILITARY HANDBOOK RELIABILITY GROWTH MANAGEMENT», no. I.

8. Anexo

8.1 Registros clasificados

Tabla 8.1: cantidad de registros de fallas y detenciones por componente clasificado.

ID componente	Fallas	Detenciones
None	115	256
cabezal	75	44
hco	57	0
acumulador	54	40
perforadora	18	11
sistema avance	17	0
cilindro avance	14	13
sistema eléctrico	14	0
peju	14	13
cilindro	12	4
cable	10	0
manguera	9	0
cilindro z	9	7
modi	8	0
centralizador	8	0
niple	7	0
manifold	7	2
compresor	6	2
viga	5	0
foco	5	0
motor eléctrico	4	4
presión	4	0
lana desgaste	4	0
enchufe	4	0
relleno hidráulico	3	0
piola avance	3	0

ID componente	Fallas	Detenciones
circuito eléctrico	3	1
culatín	3	0
batería	3	1
aceite	3	0
cilindro gato	2	1
comando	2	0
enfriador	2	1
cojinete	2	0
válvula	2	0
neumático	2	0
piola retroceso	2	0
bomba	2	1
tirante	2	0
cilindro telescopio	2	0
pernos anclaje	2	0
pernos	1	0
filtro	1	0
flotador	1	0
mordazas	1	0
cilindro basculación	1	0
relé	1	0
carrete	1	0
estanque	1	0
mesa	1	0
sistema partida	1	0
sistema perforación	1	0
switch	1	0
transmisión	1	0
sistema inyección	1	0
unidad rotación	0	3
buje	0	2

ID componente	Fallas	Detenciones
alternador	0	1

8.2 Conteo ingresos validos e inválidos.

Tabla 8.2: Número de registros validos luego del cálculo de TTR, incluyendo su respectivo porcentaje en relación al total de datos.

Componente	Fallas ingresadas	Fallas válidas	% fallas	Det. ingresadas	Det. válidas	% detenciones
None	115	87	76	371	271	73
cabezal	75	56	75	119	89	75
hco	57	57	100	57	57	100
acumulador	54	48	89	94	79	84
perforadora	18	18	100	29	28	97
sistema avance	17	17	100	17	17	100
cilindro avance	14	14	100	27	25	93
sistema eléctrico	14	14	100	14	14	100
peju	14	11	79	27	21	78
cilindro	12	12	100	16	15	94
cable	10	8	80	10	8	80
manguera	9	9	100	9	9	100
cilindro z	9	7	78	16	13	81
modi	8	8	100	8	8	100
centralizador	8	8	100	8	8	100
niple	7	7	100	7	7	100
manifold	7	7	100	9	9	100
compresor	6	5	83	8	7	88
viga	5	5	100	5	5	100
foco	5	5	100	5	5	100
motor eléctrico	4	4	100	8	7	88
presión	4	4	100	4	4	100
lana desgaste	4	4	100	4	4	100
enchufe	4	4	100	4	4	100
relleno hidráulico	3	3	100	3	3	100

Componente	Fallas ingresadas	Fallas válidas	% fallas	Det. ingresadas	Det. válidas	% detenciones
piola avance	3	3	100	3	3	100
circuito eléctrico	3	3	100	4	4	100
culatín	3	3	100	3	3	100
batería	3	3	100	4	4	100
aceite	3	3	100	3	3	100

8.3 Conteo valores atípicos

Tabla 8.3: Numero de datos atípicos y el impacto de la eliminación de un dato con respecto al total de datos clasificados por componente.

Componente	Detenciones válidas	% detenciones	atípicos	% atípicos	% por dato
None	271	73,05	8	2,95	0,37
cabezal	89	74,79	4	4,49	1,12
hco	57	100	7	12,28	1,75
acumulador	79	84,04	6	7,59	1,27
perforadora	28	96,55	2	7,14	3,57
sistema avance	17	100	2	11,76	5,88
cilindro avance	25	92,59	1	4	4
sistema eléctrico	14	100	1	7,14	7,14
peju	21	77,78	1	4,76	4,76
cilindro	15	93,75	0	0	6,67
cable	8	80	0	0	12,5
manguera	9	100	2	22,22	11,11
cilindro z	13	81,25	1	7,69	7,69
modi	8	100	1	12,5	12,5
centralizador	8	100	0	0	12,5
niple	7	100	1	14,29	14,29
manifold	9	100	0	0	11,11
compresor	7	87,5	1	14,29	14,29
viga	5	100	0	0	20
foco	5	100	0	0	20
motor eléctrico	7	87,5	0	0	14,29

Componente	Detenciones válidas	% detenciones	atípicos	% atípicos	% por dato
presión	4	100	0	0	25
lana desgaste	4	100	0	0	25
enchufe	4	100	0	0	25
relleno hidráulico	3	100	0	0	33,33
piola avance	3	100	0	0	33,33
circuito eléctrico	4	100	0	0	25
culatín	3	100	0	0	33,33
batería	4	100	0	0	25
aceite	3	100	0	0	33,33

8.4 Calculo indicadores test para determinar modelo optimo a ser aplicado (RP o NHPP)

Tabla 8.4: Resultados de los test para determinar si aplicar un modelo para un proceso renovable o un proceso no homogéneo.

Componente	Mann	Mann_c	Mann dif%	L	LR	MH
None	2394	1958	22,27	-7662,82	-6926,29	779,93
cabezal	701	715,5	2,03	-1415,56	-541,83	216,33
hco	689	798	13,66	-702,07	-211,17	11,21
acumulador	441	473	6,77	-1178,92	-308,66	173,38
perforadora	56	76,5	26,8	-229,51	-113,64	22,97
sistema avance	35	68	48,53	-98,45	-49,83	4,43
cilindro avance	50	39	28,21	-197,44	-118,05	64,56
sistema eléctrico	36	45,5	20,88	-69,13	-19,71	0,92
peju	37	27,5	34,55	-149,69	-98,05	57,89
cilindro	20	33	39,39	-82,88	-63,92	19,25
cable	17	14	21,43	-28,97	-23,71	17,5
manguera	11	18	38,89	-29,89	-11,64	0,84
cilindro z	10	10,5	4,76	-68,73	-40,21	33,51
modi	12	14	14,29	-23,21	-8,75	0,47
centralizador	6	14	57,14	-25,45	-24,8	5,7
niple	3	10,5	71,43	-18,11	-10,4	1,81
manifold	1	10,5	90,48	-31,69	-30,61	5,16

Componente	Mann	Mann_c	Mann dif%	L	LR	MH
compresor	4	5	20	-19,41	-9,98	8,16
viga	3	5	40	-9,35	-15,94	4,13
foco	1	5	80	-8,56	-6,93	2,53
motor eléctrico	2	3	33,33	-20,8	-17,73	8,41
presión	1	3	66,67	-3,19	-1,76	0,2
lana desgaste	3	3	0	-6	-7,1	4,21
enchufe	1	3	66,67	-4,53	-4,1	1,94
relleno hidráulico	0	1,5	100	-0,33		0,29
piola avance	2	1,5	33,33	-2,55		2,94
circuito eléctrico	1	1,5	33,33	-5,66	-7,57	3,61
culatín	0	1,5	100	-0,39		0,34
batería	1	1,5	33,33	-5,33	-8,81	3,77
aceite	3	1,5	100	-3,91		7,07

8.5 Parámetros estimación Weibull

Tabla 8.5: Resultados de la estimación de un modelo de Weibull de dos parámetros.

Componente	R^2	α (alpha)	β (beta)
None	0,98	590,85	0,97
cabezal	0,96	913,84	0,74
hco	0,81	256,22	0,52
acumulador	0,93	829,02	1,04
perforadora	0,96	3.230,17	0,57
sistema avance	0,81	2.303,83	0,52
cilindro avance	0,98	5.998,73	0,69
sistema eléctrico	0,81	890,23	0,44
peju	0,90	4.295,45	0,80
cilindro	0,97	6.133,73	0,51
cable	0,91	10.920,17	0,73
manguera	0,85	3.155,97	0,42
cilindro z	0,85	11.302,63	0,51

Componente	R^2	α (alpha)	β (beta)
modi	0,75	2.812,86	0,37
centralizador	0,88	11.048,62	0,47
niple	0,89	10.486,96	0,61
manifold	0,92	14.448,09	0,58
compresor	0,97	19.151,33	0,38
viga	0,84	19.625,52	1,28
foco	0,94	19.694,81	0,44
motor eléctrico	0,97	17.535,09	0,64
presión	0,97	19.096,56	0,33
lana desgaste	0,84	14.902,79	0,95
enchufe	0,93	26.374,85	0,59
relleno hidráulico	1,00	50.061,66	0,51
piola avance	1,00	36.076,85	12,09
circuito eléctrico	0,98	26.700,05	1,88
culatín	1,00	49.045,01	0,58
batería	1,00	23.522,01	29,95
aceite	1,00	33.512,20	0,86

8.6 Parámetros Weibull de dos y tres parámetros, junto con pruebas de bondad de ajuste

Tabla 8.6: resultados de las pruebas de bondad de ajuste tanto para el modelo de dos parámetros como para el de tres.

Componente	alpha	beta	KS	Kscritico	alpha	beta	t_0	KS	Kscritico
None	574,30	0,98	0,05	0,15	564,73	1,00	-1,01	0,05	0,15
cabezal	785,01	0,79	0,08	0,19	778,48	0,81	-2,58	0,08	0,19
hco	176,76	0,61	0,14	0,19	180,47	0,47	2,50	0,16	0,19
acumulador	664,82	1,16	0,11	0,21	656,92	0,91	30,45	0,09	0,21
perforadora	2552,97	0,60	0,10	0,33	2606,61	0,64	-20,30	0,12	0,33
sistema avance	2303,83	0,52	0,16	0,33	2107,60	0,49	37,24	0,16	0,33
cilindro avance	4677,78	0,74	0,08	0,39	4354,10	0,86	-54,41	0,05	0,39
sistema eléctrico	890,23	0,44	0,21	0,36	1144,85	0,26	6,49	0,22	0,36
peju	2833,62	1,00	0,08	0,41	3012,33	0,71	182,68	0,06	0,41

cilindro	6133,73	0,51	0,11	0,39	7048,27	0,76	-435,98	0,09	0,39
cable	10920,17	0,73	0,12	0,48	11486,74	0,80	-384,08	0,12	0,48
manguera	3155,97	0,42	0,15	0,45	3232,38	0,23	68,43	0,16	0,45
cilindro z	11302,63	0,51	0,09	0,52	10923,46	0,62	-298,02	0,09	0,52
modi	2812,86	0,37	0,19	0,48	1806,77	0,22	100,90	0,13	0,48
centralizador	11048,62	0,47	0,15	0,48	36864,31	3,15	-22793,48	0,11	0,48
niple	10486,96	0,61	0,14	0,52	9731,39	0,50	408,11	0,13	0,52
manifold	14448,09	0,58	0,11	0,52	16466,40	1,30	-3108,29	0,06	0,52
compresor	19151,33	0,38	0,08	0,62	19341,83	0,49	-410,55	0,06	0,62
viga	19625,52	1,28	0,15	0,62	44232,16	4,30	-24860,58	0,13	0,62
foco	19694,81	0,44	0,10	0,62	19050,47	0,39	272,95	0,09	0,62
motor eléctrico	17535,09	0,64	0,05	0,62	17278,24	0,61	227,83	0,05	0,62
presión	19096,56	0,33	0,06	0,71	16395,83	0,26	414,62	0,02	0,71
lana desgaste	14902,79	0,95	0,12	0,71	33177,95	3,20	-18366,97	0,11	0,71
enchufe	26374,85	0,59	0,08	0,71	67290,76	2,86	-40141,27	0,03	0,71
relleno hidráulico	50061,66	0,51	0,00	0,84	47983,62	0,47	1548,25	0,00	0,84
piola avance	36076,85	12,09	0,00	0,84	6133,25	1,70	29876,26	0,00	0,84
circuito eléctrico	26700,05	1,88	0,04	0,71	15474,88	0,90	10376,11	0,00	0,71
culatín	49045,01	0,58	0,00	0,84	21352,30	0,12	10381,61	0,00	0,84
batería	23522,01	29,95	0,00	0,84	568,90	0,28	22800,68	0,00	0,84
aceite	33512,20	0,86	0,00	0,84	38858,82	1,07	-5031,18	0,00	0,84

8.7 Resumen de resultados

Tabla 8.7: Resumen resultados finales TOI, cálculos de disponibilidad y MTBF por componente.

Componente	Tp	EC min	EC asíntota	MTBF	TOI_R0,8	A_G	A_I
None	Correr a la falla	3,3079	5,6135	563,5390	125,1566	0,5509	0,7679
cabecal	Correr a la falla	0,8560	0,7349	868,1651	120,9493	0,8123	0,8815
hco	Correr a la falla	2,1967	1,0520	406,1681	10,0230	0,9954	0,9940
acumulador	Correr a la falla	0,5122	0,4972	715,6596	157,9008	0,8175	0,8905
perforadora	Correr a la falla	0,6037	0,4111	3621,8546	227,3636	0,9886	0,9941
sistema avance	Correr a la falla	0,9861	0,6503	4350,6634	138,3497	0,9896	0,9902
cilindro avance	Correr a la falla	0,1372	0,1087	4657,2004	702,7258	0,9406	0,9746
sistema eléctrico	Correr a la falla	0,2505	0,0732	23899,9245	9,7589	0,9910	0,9981

Componente	Tp	EC min	EC asíntota	MTBF	TOI_R0,8	A_G	A_I
peju	Correr a la falla	0,4585	0,2662	3964,1993	542,8081	0,9014	0,9512
cilindro	Correr a la falla	0,1684	0,1309	7829,6300	554,1488	0,9819	0,9896
cable	Correr a la falla	0,1344	0,1109	12599,0631	1388,8031	0,9904	0,9934
manguera	Correr a la falla	0,0177	0,0039	148273,4493	72,7168	0,9995	1,0000
cilindro z	Correr a la falla	0,0511	0,0345	15349,0842	689,1634	0,9519	0,9839
modi	Correr a la falla	0,0308	0,0067	95485,8518	102,9983	0,9969	0,9997
centralizador	Correr a la falla	0,1454	0,1559	10196,8933	88,7294	0,9974	0,9978
niple	Correr a la falla	0,0375	0,0185	19857,0662	893,2277	0,9998	0,9999
manifold	Correr a la falla	0,1481	0,1440	12088,0390	2108,7261	0,9944	0,9964
compresor	Correr a la falla	0,1234	0,0702	39413,3817	508,4034	0,9788	0,9947
viga	16900	0,1755	0,2059	15398,7573	6348,0981	0,9923	0,9934
foco	Correr a la falla	0,0170	0,0073	67015,4698	690,2262	0,9998	1,0000
motor eléctrico	Correr a la falla	0,0110	0,0050	25927,4845	1677,9873	0,9822	0,9938
presión	Correr a la falla	0,0058	0,0015	312963,8716	465,8929	0,9999	1,0000
lana desgaste	Correr a la falla	0,1402	0,1480	11349,0169	2395,8740	0,9988	0,9991
enchufe	Correr a la falla	0,0331	0,0284	19828,3401	-305,0154	0,9998	0,9998
relleno hidráulico	Correr a la falla	0,0032	0,0008	109883,7648	3520,3478	1,0000	1,0000
piola avance	Correr a la falla	0,0131	0,0131	35348,7478	32413,5554	0,9999	0,9999
circuito eléctrico	Correr a la falla	0,0829	0,0600	26698,0175	13277,6900	0,9912	0,9943
culatín	Correr a la falla	0,0010	0,0000	2281821327,2	10381,6748	1,0000	1,0000
batería	22900	0,0289	0,0220	30021,2529	22803,3954	0,9996	0,9998
aceite	Correr a la falla	0,0286	0,0239	32778,0520	4576,5394	0,9738	0,9854