

UNIVERSIDAD POLITECNICA DE VALENCIA
Departamento de Estadística e Investigación Operativa
Aplicadas y Calidad



**ANALISIS DE SUPERVIVENCIA EN FIABILIDAD.
PREDICCIÓN EN CONDICIONES DE ALTA CENSURA Y
TRUNCAMIENTO: EL CASO DE LAS REDES DE
SUMINISTRO DE AGUA POTABLE**

TESIS DOCTORAL

Presentada por:

Hernando Solano Hurtado

Directores:

Dr. Andrés Carrión García

Dra. Ana Debón Aucejo

Valencia, Octubre de 2008

A la Memoria de mi madre

A mis hijos

Lina María, Andrés Felipe y Carlos Hernando

Agradecimientos

Al parafrasear a Mario Capecchi (genetista molecular italoamericano, premio Nobel de Medicina 2007), orgulloso me encantaría poder decir a todo pulmón, “*Tengo 63 años y cuanto más logro estudiar, menos sé, pero más me divierto.....*”. Y me divierto porque la vida no es una carrera, sino un viaje que debe ser disfrutado a cada paso y en cada instante. En mi caso, se traduce en el disfrute no sólo por la adquisición del conocimiento como forma de vida, sino también, por la inmensa satisfacción personal de haberme sentido acogido y acompañado en buena parte de este proceso, por el selecto grupo de personas que tuve la suerte de conocer, dentro y fuera de la Universidad. A todos ellos mis agradecimientos, particularmente por sus distintas expresiones de colaboración y ayuda, sobre todo por permitirme convertir esta utopía en realidad.

Comenzaré por agradecer en primer lugar a Andrés Carrión García, por haber sido no sólo mi mentor, director y amigo, sino también por haber creído en mi y permitirme trabajar en el tema que más quería, la *Fiabilidad*; y sobre todo por haberme enseñado con su ejemplo, el concepto más importante del ser humano: *La calidad humana*. De verdad, ¡Gracias Andrés!. También agradezco a Ana Debón A., por haber respondido incondicionalmente en todo momento a mis expectativas como codirectora, y como persona, para contribuir significativamente a la feliz culminación de este trabajo.

Un reconocimiento especial a María Luz Gámiz (Universidad de Granada), por su constante retroalimentación académica (por las luces que me dio, por el estímulo, por los materiales que me proporcionó, por un largo etc, etc.) y sobre todo por su adhesión incondicional y amistad, especialmente en los momentos más duros y difíciles de mi estadía en su acogedora ciudad. ¡Gracias!

También tengo mucho que agradecer, al Instituto Tecnológico del Agua de la UPV (ITA), donde se gestó este proyecto, en particular a su director don Enrique Cabrera, por la beca concedida para la realización de esta memoria.

Y como me voy a olvidar de Gustavo Vivas y Francisco Morán por su decidido y efectivo apoyo, quienes desde la Vicerrectoría de Cooperación (VCPD) hicieron posible en diferentes oportunidades las ayudas para mis estancias en la UPV.

A mis profesores de doctorado por abrirme nuevas perspectivas, Rafael Romero y Luisa Zúnica, por su especial deferencia conmigo. A Nieves Martínez, por acrecentar mi interés por la Fiabilidad. A Alberto Ferrer, por su rigor y exigencia académica. A Francisco Aparisi, por su simpatía y permanente *cachondeo*. A Gonzalo Clemente y José Saénz por sus acertados análisis de encuestas y a José Domingo Bermúdez por su particular pedagogía en la Simulación. Finalmente a Bernardo Richart, a Julia Durba y a Ma. Ángeles Marín por su apreciado espíritu de cooperación y servicio. Igualmente a Miguel Granell desde Alumnado, por su sentido aprecio y reiterados gestos de colaboración y asistencia. Gracias a todos y a todas.

A mis compañeros de despacho, *Maneco*, (mi consultor de cabecera en ofimática), por sus reiteradas y casi siempre acertadas soluciones y su particular sentido del humor. A Marco Antonio, que a pesar de su seriedad siempre estaba disponible para aclararnos dudas en todo momento; igualmente a José María, todo un personaje que desde ya, transmite credibilidad a pesar de su juventud y a Gustavo Andrés por compartirme algunas de sus confidencias. A todos ellos gracias, por la compañía, por los buenos (y no tan buenos) momentos, por la amistad y por el cariño de siempre.

A mis hijos Andrés Felipe y Carlos Hernando, gracias por haberme permitido *sustraerles* buena parte del tiempo (varios años) que debería haberles dedicado y que utilicé en forma egoísta en el desarrollo de este proyecto. Gracias hijos.

A Adriana mi mujer, por haber soportado con estoicismo y paciencia mis altibajos emocionales, por su apoyo físico y espiritual durante su estancia en España, especialmente en los momentos clave y de mayor riesgo en la concreción de este ideal. Mil y mil gracias.

Contenido

1. Justificación	1
2. Objetivos	5
3. Metodología y contenidos de la Tesis	9
3.1 Metodología	9
Recopilación de datos.....	10
Depuración.	10
Base de datos depurada.....	10
Revisión de antecedentes y selección de los modelos.	10
Modelos Considerados.	11
Comparación de los Modelos estudiados.	11
Conclusiones y Futuras líneas de investigación	12
3.2 Contenidos de la Tesis.....	13
4. Antecedentes	15
4.1 Introducción al estudio de la Fiabilidad	16
4.2 Tipos de Censura.....	19
4.2.1 Censura por la derecha	19
4.2.2 Censura por la izquierda.....	19
4.2.3 Censura doble.....	20
4.2.4 Censura tipo I	20
4.2.5 Censura tipo II	20
4.2.6 Censura Aleatoria.	20
4.2.7 Censura Múltiple.	21
4.2.8 Datos Truncados por la izquierda	22
4.2.9 Modelo de truncamiento por la izquierda y censura por la derecha..	23
4.3 Estructura usual de los datos en una Red de agua.....	24
4.4 Funciones Básicas en Fiabilidad.....	25
4.4.1 La Función de Fiabilidad o de Supervivencia	25
4.4.2 La función de Riesgo	26
4.4.3 La Tasa de riesgo acumulada.....	28
4.5 Algunos modelos de fiabilidad básicos	31
4.5.1 Modelos Paramétricos	31
4.5.1.1 Distribución Exponencial	33
4.5.1.2 Distribución Weibull	34
4.5.1.3 Distribución Log-Normal	37

4.5.1.4 Distribución de Herz.	41
4.5.1.5 Representación gráfica de los modelos paramétricos.....	43
4.5.2 Modelos no Paramétricos	45
4.5.2.1 Tablas de vida	45
4.5.2.2 El estimador de Kaplan- Meier (KM) de la función de fiabilidad.	50
4.5.2.3 El estimador de Kaplan- Meier ponderado (KM _P).....	53
4.5.2.4 El estimador Nelson-Aalen (NA) de la función de riesgo acumulado	56
4.5.2.5 Estimadores no paramétricos de la función de fiabilidad para datos truncados a la izquierda y censurados a la derecha	57
El estimador de Turnbull	59
Estimador de Nelson-Aalen extendido (NAE)	60
4.5.2.6 Comparación de funciones de fiabilidad	62
4.5.3 Modelos Semiparamétricos	65
4.5.3.1 Modelo de riesgo proporcional de Cox	65
4.5.3.2 Modelos de vida acelerada.....	70
4.5.3.3 Residuos de Cox-Snell	73
4.6. Modelos de fiabilidad en redes de suministro de agua potable.....	75
4.6.1 Modelos para el análisis del estado de deterioro de una red de agua.	82
4.6.1.1. Modelos exponenciales y lineales en el tiempo.	83
4.6.1.2. Modelos de riesgo proporcional y de vida acelerada.	88
4.6.2 Tablas resumen.	92
5. Análisis de fiabilidad en una red de suministro de agua potable.	95
5.1 Caracterización de la red	95
5.1.1 Estructura original de la base de datos disponible para el estudio ...	96
5.2 Características de la base de datos del estudio.....	99
5.2.1 Presentación y descripción general de la red estudiada.....	102
5.3 Evaluación de la fiabilidad en la red.....	117
6. Resultados y análisis	119
6.1 Evaluación de la fiabilidad en la red con modelos no paramétricos	119
6.1.1 Tablas de vida	119
6.1.2 Estimaciones con el enfoque de Kaplan- Meier	136
6.1.3 Estimación de la fiabilidad de la red, con datos truncados por la izquierda y con censura por la derecha, utilizando el método de NAE	147
6.1.3.1 Estimación global de la fiabilidad.....	148
6.1.3.2 Estimación estratificada de la fiabilidad por tipo de material	153
6.1.3.3 Estimación estratificada de la fiabilidad por tráfico	157
6.2 Evaluación de la fiabilidad en la red con modelos semiparamétricos ...	160

6.2.1 Validación de los supuestos de Proporcionalidad de Cox	163
6.3 Evaluación de la fiabilidad en la red con modelos paramétricos	174
6.3.1 Identificación del Mejor Modelo para los Datos	175
6.3.2 Estimación de los Parámetros del Modelo	177
6.4 Comparación de las estimaciones de $R(t)$ con los cuatro métodos.....	183
7. Necesidades de información para la modelización estadística	187
7.1 Generalidades.....	187
7.2 Bases de datos en fiabilidad	189
7.3 Variables de mayor influencia en el fallo	193
7.4 Problemática de la base de datos analizada.....	195
7.5 Contenido y características recomendables en una base de datos de vida de redes de suministro de agua.....	197
7.5.1. Características de la red	198
7.5.2. Características de los tramos	199
7.5.3. Información sobre fallos y averías.	201
7.6 Indicadores estadísticos para la gestión	203
7.7 Explotación de la base de datos	204
8. Conclusiones y sugerencias	207
8.1 Conclusiones sobre la base de datos utilizada	207
8.2 Conclusiones sobre los modelos no paramétricos	209
8.3 Conclusiones sobre los modelos Semiparamétricos: factores influyentes en la supervivencia	211
8.4 Conclusiones sobre los modelos paramétricos	212
8.5 Comparación de los modelos.....	213
8.6 Recomendaciones para una base de datos de supervivencia en redes de suministro de agua potable.	214
8.7 Nuevas líneas de investigación.....	216
9. Referencias	219
Anexos	231

Resumen

Los sistemas de abastecimiento de agua potable en áreas urbanas y áreas rurales relativamente densas, están constituidos fundamentalmente por redes de tuberías usualmente instaladas bajo tierra, cuyas estructuras son cambiantes en el tiempo. Ello dificulta en gran medida las tareas de mantenimiento y el llevar un registro exacto de algunas actuaciones básicas complementarias (inventarios, sustitución y/o reparación de tuberías, gestión de averías y acometidas, gestión de demanda del servicio, etc.) necesarias no sólo para la adecuada conservación de su infraestructura sino también para una gestión integral y eficaz del servicio.

En virtud de lo anterior y con el fin de contribuir a realizar de manera eficiente las labores de planeación, diseño, construcción, operación y mantenimiento en las redes de agua potable, es necesario contar con un sistema de información actualizado de la infraestructura hidráulica disponible y del estado técnico y de deterioro de la misma. Se podrá así abordar no sólo la gestión y planificación de las actividades anteriormente mencionadas (mantenimiento, reparación, renovación y/o sustitución de las tuberías y otros elementos básicos de estos sistemas), sino también el registro y cuantificación exacta mediante la modelización estadística del verdadero estado de deterioro estructural y fiabilidad del sistema de abastecimiento en general.

En esta tesis se hace una revisión de los métodos existentes para la modelización de fallos en tramos individuales de tubería de redes de suministro de agua y se valora la suficiencia de los datos en calidad y cantidad para la retroalimentación de estos modelos. Se describen también, los métodos estadísticos más apropiados para la modelización de estos fallos en los conductos y se aplican los más adecuados, al caso de una red del mediterráneo Valenciano, que se caracteriza por su corto historial de roturas y su alta censura.

Dentro del marco general del objetivo de este trabajo (estudio, evaluación, adaptación y aplicación de modelos de supervivencia en ingeniería), se aborda

en primer lugar el problema de identificar el patrón de fiabilidad y supervivencia de una red de agua y sus tuberías, como un primer acercamiento con técnicas de libre distribución, tales como, Kaplan Meier, tablas de vida y Nelson-Aalen extendido, con resultados de estas técnicas altamente alentadores, que refuerzan la confianza en la coherencia y consistencia de estos análisis.

Por otra parte se aplican modelos de regresión de Cox (después de verificar supuestos, estimar parámetros por máxima verosimilitud y revisar el ajuste a los datos) para identificar los factores de riesgo (físico-ambientales) más influyentes de la red y sus tuberías, tratando de modelar cuantitativa y unificadamente la relación entre la tasa de fallo y el entorno del sistema de distribución. En otra fase de este trabajo, se identifica la familia paramétrica (con gráficos de probabilidad) entre más de diez modelos posibles de la distribución de vida de las tuberías; se estiman los parámetros del modelo escogido por mínimos cuadrados y máxima verosimilitud, (con resultados muy similares de estas estimaciones) y finalmente se valora el mejor ajuste a los datos (con los indicadores de Anderson-Darling y correlación de Pearson).

En el estudio se observa una ligera tendencia a sobrestimar la fiabilidad de los conductos de la red por parte del método de Cox, comparado con las estimaciones de Kaplan-Meier, las de Nelson-Aalen extendido y las del modelo paramétrico seleccionado (modelo lognormal), a partir de los 35-40 años de operación de la red. El estudio también reveló, que los factores influyentes en la supervivencia y en el riesgo de fallo van más allá de la edad, incluyendo como tales al material constituyente de la tubería y a las condiciones de trabajo de la red.

Resum

Els sistemes d'abastiment d'aigua potable en àrees urbanes i àrees rurals relativament denses, estan constituïts fonamentalment per xarxes de canonades usualment instal·lades baix terra, les estructures dels quals són canvians en el temps. Això dificulta en gran manera les tasques de manteniment i el portar un registre exacte d'algunes actuacions bàsiques complementàries (inventaris, substitució i/o reparació de canonades, gestió d'averies i connexions, gestió de demanda del servei, etc.) necessàries no sols per a l'adequada conservació de la seua infraestructura sinó també per a una gestió integral i eficaç del servei.

En virtut de l'anterior i a fi de contribuir a realitzar de manera eficient les labors de planeació, disseny, construcció, operació i manteniment en les xarxes d'aigua potable, és necessari comptar amb un sistema d'informació actualitzat de la infraestructura hidràulica disponible i de l'estat tècnic i de deteriorament de la mateixa. Es podrà així abordar no sols la gestió i planificació de les activitats anteriorment mencionades (manteniment, reparació, renovació i/o substitució de les canonades i altres elements bàsics d'estos sistemes), sinó també el registre i quantificació exacta per mitjà de la modelització estadística del verdader estat de deteriorament estructural i fiabilitat del sistema d'abastiment en general.

En esta tesi es fa una revisió dels mètodes existents per a la modelització de fallades en trams individuals de canonada de xarxes de subministrament aigua i es valora la suficiència de les dades en qualitat i quantitat per a la retroalimentació d'estos models. Es descriuen també, els mètodes estadístics més apropiats per a la modelització d'estes fallades en els conductes i s'apliquen els més adequats, al cas d'una xarxa del mediterrani Valencià, que es caracteritza pel seu curt historial de ruptures i la seua alta censura.

Dins del marc general de l'objectiu d'este treball (estudi, avaluació, adaptació i aplicació de models de supervivència en enginyeria), s'aborda en primer lloc el problema d'identificar el patró de fiabilitat i supervivència d'una

xarxa d'aigua i les seues canonades, com un primer acostament amb tècniques de lliure distribució, com ara, Kaplan Meier i taules de vida, amb resultats d'estes tècniques altament coincidents, que reforcen la confiança en la coherència i consistència d'estes anàlisis.

D'altra banda s'apliquen models de regressió de Cox (després de verificar supòsits, estimar paràmetres per màxima versemblança i revisar l'ajust a les dades) per a identificar els factors de risc (físicoambientals) més influents de la xarxa i les seues canonades, tractant de modelar quantitativa i unificadament la relació entre la taxa de fallada i l'entorn del sistema de distribució. En una altra fase d'este treball, s'identifica la família paramètrica (amb gràfics de probabilitat) entre més de deu models possibles de la distribució de vida de les canonades; s'estimen els paràmetres del model triat per mínims quadrats i màxima versemblança, (amb resultats molt semblants d'estes estimacions) i finalment es valora el millor ajust a les dades (amb els indicadors d'Anderson-Darling i correlació de Pearson).

En l'estudi s'observa una lleugera tendència a sobreestimar la fiabilitat dels conductes de la xarxa el mètode de Cox, comparat amb les estimacions de Kaplan-Meier i amb les estimacions del model paramètric seleccionat (model lognormal), a partir dels 35-40 anys d'operació de la xarxa. L'estudi també va revelar, que els factors influents en la supervivència i en el risc de fallada van més enllà de l'edat, incloent com a tals al material constituent de la canonada i a les condicions de treball de la xarxa.

Summary

Drinking water supply systems in relatively dense urban and rural areas consist mainly of pipes usually installed underground, whose structures are changing over time. This greatly hinders the work of maintaining and carrying an accurate record of some key actions complementary (inventories, replacement and/or repair of pipelines, failures and damages management, demand management service, etc.) necessary not only for proper infrastructure maintenance, but also for a comprehensive and effective management of the service.

In reference to the above mentioned, and in order to contribute to perform efficiently the work of planning, design, construction, operation and maintenance of water supply networks, it is necessary to count with an updated information system of the hydraulic infrastructure, its technical state and its deterioration. It may well address not only the management and planning of the network, but also recording and precise quantification of the real state of structural deterioration and reliability of supply system in general, allowing statistical modeling.

In this thesis, a review of failure models in individual sections of pipe water supply networks is conducted. Additionally, it assesses the adequacy of the data quality and quantity for feedback from these models. It also describes the most appropriate statistical methods to model these pipes failures and the most suitable are applied to the case of a network in a city of the Spanish Mediterranean coast. This database is characterized by its short history of breakages and high censorship.

The overall objective of this study (review, evaluation, adaptation and application of survival models in engineering), primarily addressed the problem of identifying the pattern of reliability and survival of water pipes and its networks as a first approach with free distribution techniques such as Kaplan Meier and life-tables, obtaining highly coincidental results that reinforce confidence in the coherence and consistency of these analyses.

Moreover, Cox regression models are applied (after verifying assumptions, estimating parameters for maximum likelihood and revise the

adjustment to the data) to identify most influential risk factors (physical and environmental). The aim is to model the relationship between failure rate and environment of the distribution system. In another phase of this work, the best fitting parametric family is identified, among more than ten models of possible distributions of pipe's life. Model parameters are estimated by least squares and maximum likelihood, with very similar results, and finally selected the best fit to the data (using Anderson-Darling indicator and Pearson correlation).

The study notes a slight tendency to overestimate pipes reliability in the Cox method, compared with the Kaplan-Meier estimates and the parametric model estimates (lognormal model), from 35-40 years of the network operation. The study also revealed that survival and the failure risk influencing factors goes beyond age, including such as material component of the pipeline and working conditions.

1. Justificación

En cualquier sistema de abastecimiento de agua potable, la red de distribución es el componente más caro de todo el sistema (de hecho, las ciudades Europeas están gastando del orden de *1100 millones de euros anuales en la rehabilitación de sus redes de abastecimiento de agua*, Rostum (2000)) con costes que tienden a crecer en los próximos años por el envejecimiento y paulatina degradación de sus tuberías. Ello llega a representar incluso hasta el 80% del total de los gastos involucrados en el sostenimiento de todo el sistema. El deterioro de estas infraestructuras se traduce en altos costes de mantenimiento, aumento en la tasa de roturas, disminución de la capacidad hidráulica de la red, reducción de la calidad físico-química del líquido, reducción en la calidad de la prestación del servicio (suspensión frecuente y prolongada del servicio) y en cuantiosas pérdidas por fugas y escapes del líquido, aparte de la interrupción de las actividades normales de las personas en las vías públicas, generando problemas sociales de insospechadas proporciones.

Una adecuada planificación en la rehabilitación y renovación de las *redes de agua y/o* de algunos de sus componentes, es el *ingrediente* imprescindible para cumplir adecuadamente con los objetivos de calidad en la prestación del servicio de abastecimiento de agua potable, de ahí la importancia de comprender y cuantificar los mecanismos de deterioro de las tuberías, por ser una parte *esencial* del proceso de gestión integral de la red y su infraestructura.

La estrategia consiste, en aprovechar en toda su extensión la vida útil de cada tubo, y al mismo tiempo abordar cuestiones de seguridad, fiabilidad, calidad del agua y la eficiencia económica del sistema.

El deterioro de las tuberías puede clasificarse en dos categorías. La primera es un deterioro estructural, que disminuye la resistencia de la estructura de las tuberías y su capacidad para resistir los distintos tipos de tensión (sobresfuerzo) que se le aplique. El segundo, es el deterioro de la superficie interna de las tuberías como consecuencia de la disminución en la

capacidad hidráulica (obstrucción), la degradación de la calidad físico-química del agua y la reducción de la adaptación al cambio estructural, en los casos severos de corrosión interna. El deterioro de la superficie interna de las tuberías, sólo se menciona como un factor más que contribuye al deterioro estructural de la misma. Ambas categorías de deterioro, contribuyen sustancialmente a disminuir la fiabilidad de la red de distribución.

Puesto que los materiales de las tuberías también son determinantes en el proceso de deterioro de una red, es oportuno mencionar a modo de comentario, los resultados de dos importantes estudios realizados en los USA y Canadá, (y que coinciden plenamente), acerca de la forma como se distribuyen usualmente (en tanto por ciento) los distintos materiales en estas redes. De un lado Kirmeyer y otros (1994) estimaban en un estudio realizado a comienzos de 1992, en los Estados Unidos, que más de dos tercios de las tuberías de agua eran metálicas (alrededor de 48% de hierro fundido y un 19% hierro dúctil), alrededor del 15% son de asbesto cemento (Uralita) y el 18% restante eran de plástico (polietileno), hormigón y otros materiales. En contraste, de las nuevas tuberías que se están instalando alrededor de 48% son de fundición (hierro) dúctil, 39% de PVC y el 12,5% de concreto a presión

De otra parte, Rajani y McDonald (1995), en un estudio similar que realizaron con una cobertura de 21 ciudades canadienses (alrededor del 11% de la población del Canadá) observaron una distribución porcentual muy similar de los distintos tipos de material de las tuberías de las redes de abastecimiento, Kleiner y Rajani (2000).

Los mecanismos físicos que llevan a la rotura de tuberías a menudo son muy complejos y no es fácil entenderlos completamente. El hecho de que la mayor parte de las tuberías estén bajo tierra, y el que relativamente son pocos los datos disponibles acerca de sus modos de rotura (debido a la histórica falta de conciencia acerca de la importancia de la recopilación de estos datos en los servicios públicos del agua), contribuye en gran medida a la falta de un mayor conocimiento sobre el tema.

El enfoque que se sigue para el estudio de la modelación estadística con este propósito, es el de Kleiner y Rajani (2000), que recoge básicamente cuatro aspectos claves de la modelización: una breve descripción del modelo, la formulación de su estructura matemática, sus ventajas y desventajas y la caracterización de la información mínima que se requiere para su implementación, como se verá posteriormente. La intención de este formato de valoración es facilitar la identificación de las principales características de los modelos estudiados con sus correspondientes limitaciones y requerimientos de información.

2. Objetivos

De acuerdo a lo anteriormente expuesto, el objetivo principal de este trabajo es establecer y aplicar una metodología de análisis, que permita estudiar, evaluar, adaptar y aplicar algunos modelos estadísticos, que serán utilizados en la predicción de los fallos y averías de las tuberías y en la cuantificación del estado de deterioro y degradación de las redes de distribución de agua potable.

Los modelos considerados aquí para la descripción del *estado de las tuberías y del grado de deterioro actual y futuro de la infraestructura de estos sistemas*, serán el principal instrumento para la planificación del mantenimiento de la red. Estos modelos se han aplicado a una base de datos, que se caracteriza por la alta censura y el truncamiento en la información (de fallos y averías) de las tuberías de la red de agua potable, de una importante zona mediterránea de la Comunidad Valenciana, como anteriormente se mencionó. El fin buscado es que los gestores y planificadores de estas redes, puedan predecir y determinar con la máxima confianza y el mínimo riesgo, las prioridades en materia de reparación y/o sustitución y mantenimiento de los distintos tramos de tubería de una red en general.

Varios son los interrogantes, que en principio nos planteamos en esta tesis y cuyas respuestas, tratamos alcanzar con el desarrollo de este trabajo. Para ello, se ha procedido a estructurar el estudio en cuatro etapas o fases, que se realizan con objetivos y metodologías muy particulares. Veamos:

- *Un primer objetivo, aborda el problema de valorar la probabilidad de supervivencia, fruto del estado de deterioro de la red y sus tuberías, con técnicas de fiabilidad no paramétricas o de libre distribución. Se trabajará en forma general y estratificada (de acuerdo a ciertas características físicas y medioambientales de instalación de las tuberías). Se pretende conocer, en una primera aproximación, una predicción del patrón de comportamiento de los riesgos de fallo y/o*

averías de un tramo de tubería, establecer cuál es la predisposición al fallo y la rapidez con que estos ocurren, todo ello según el tipo de material, longitud, diámetro, zona de tráfico, etc. de los diferentes conductos del sistema de suministro de agua potable.

- *Otro objetivo nuestro, se centra en identificar qué factores o características de las tuberías y de su entorno influyen decididamente y en qué medida lo hacen, sobre la duración y/o riesgo de fallo de los distintos conductos de la red, es decir, influyen en la degradación y deterioro de los sistemas de suministro de agua potable. Se trata, de valorar no solamente si estos factores afectan o no el riesgo de fallo de los distintos conductos de la red (y por tanto a toda la red), sino también, en qué sentido y en qué medida lo afectan. Se trata también de establecer, mediante la modelización estadística semiparamétrica, una representación más unificada que mida la relación entre el entorno del sistema de distribución y el proceso de fallo del mismo.*

- *En la tercera etapa, se plantea como objetivo, ajustar con los datos disponibles, la edad hasta el fallo de un tramo de tubería (con técnicas paramétricas), para identificar entre más de diez familias de modelos, el o los modelos óptimo(s) de mejor ajuste, de modo que, podamos describir y caracterizar la distribución de vida de las tuberías, a fin de poder realizarles estimaciones confiables de su riesgo de fallo a cualquier edad, incluso más allá de la información contenida en la muestra. Así mismo se busca hacer estimaciones sobre la edad media hasta el fallo (MTTF), por desgaste y envejecimiento natural de un tramo de tubería, estimaciones que de otra manera son imposibles de realizar con los métodos no paramétricos y semiparamétricos, debido a las limitaciones de estos métodos en presencia de datos con censura y truncamiento, que como se mencionó al comienzo de este apartado, es el caso de nuestra información.*

- *Estos objetivos se complementan y se concretan finalmente con el análisis y la comparación, de los resultados que se obtienen con los tres tipos de modelos anteriormente referenciados.*

- *En la última etapa se plantea como objetivo, estudiar en profundidad, el tema de la necesidad de información adecuada para la modelación estadística, a fin de contribuir, con mejores prácticas para el diseño y desarrollo, de idóneos sistemas de información (captura, almacenamiento y gestión de los datos) para la mejor gestión de las redes de suministro de agua potable. Se trata de determinar, si la naturaleza de los datos, que normalmente recogen y registran las entidades gestoras de este servicio, son lo suficientemente idóneas (en calidad y cantidad), para facilitar la adecuada retroalimentación y óptima utilización de los modelos de predicción, en el diseño de buenos indicadores de gestión para el control de abastecimientos. En síntesis, se trata de responder con este objetivo a la pregunta: ¿Qué tipo de información clave (indicadores y características asociadas a las redes) permite avanzar hacia la gestión proactiva de su infraestructura, desde la visión técnico operativa y de calidad en el servicio?*

En resumen, estos objetivos específicos se sintetizan en:

1. Identificar, el patrón de comportamiento de los fallos de los diferentes conductos de una red de suministro de agua potable, con métodos estadísticos no paramétricos.
2. Adaptar y aplicar, los modelos semiparamétricos de Cox, en la identificación de los factores de riesgo más influyentes, sobre el tiempo de vida de las tuberías de los sistemas de distribución de agua potable.
3. Identificar, con un enfoque paramétrico, la familia de modelos de mejor ajuste a los datos disponibles para superar las limitaciones impuestas por la censura (en los modelos de Cox y en los no paramétricos) en la predicción del riesgo de fallo de las tuberías a cualquier edad, incluso más allá de la información contenida en la muestra.
4. Estudiar y analizar el tipo de información requerida, en los sistemas de abastecimiento de agua potable, para garantizar sistemas idóneos (en calidad, cantidad y fiabilidad del dato), a fin de avanzar hacia la gestión proactiva de su infraestructura.

3. Metodología y contenidos de la Tesis

3.1 Metodología

El siguiente esquema, muestra la estructura del enfoque metodológico, seguido en este estudio.

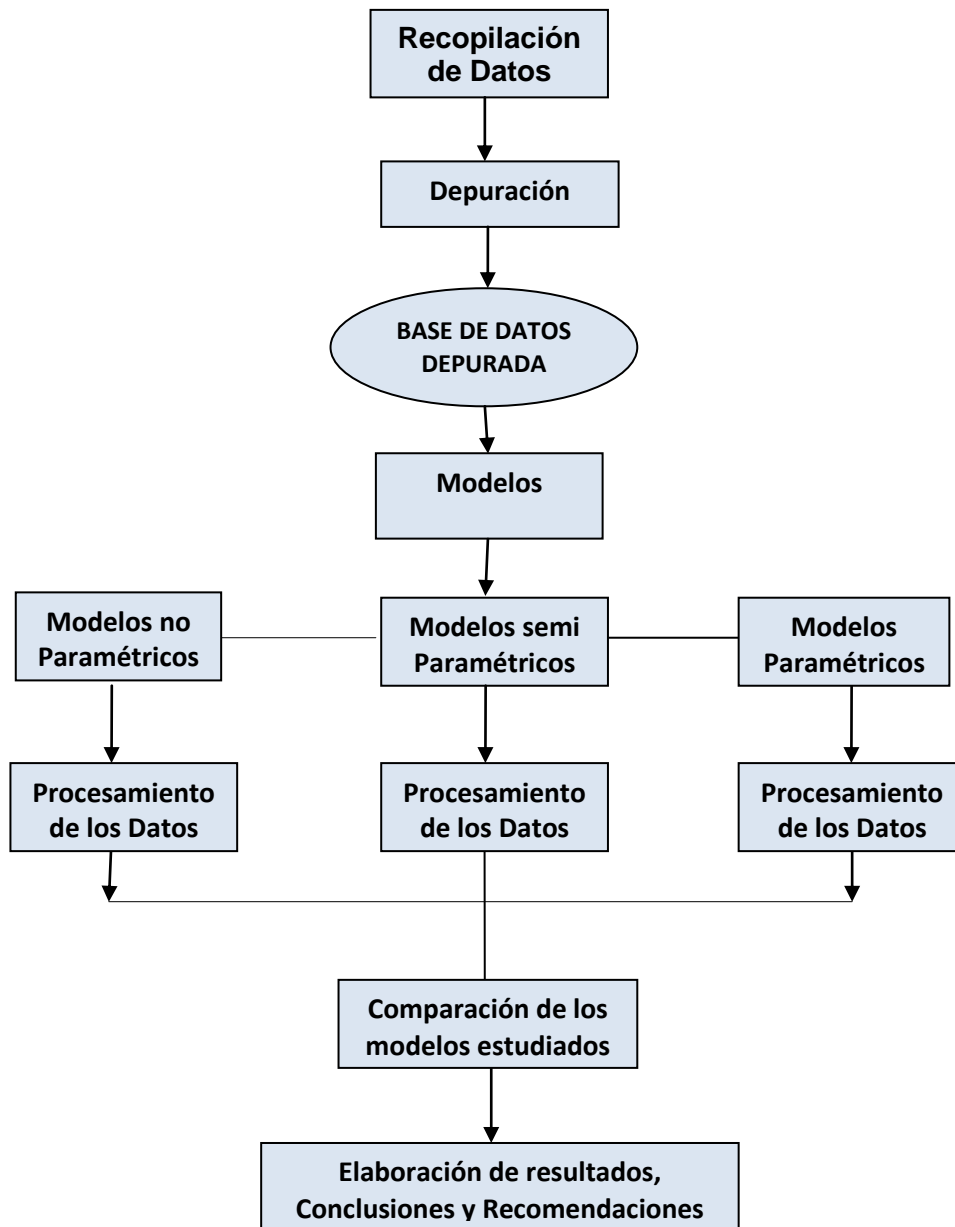


Figura 3.1 Estructura de la Metodología del estudio

Recopilación de datos.

Como se ha venido mencionando desde un comienzo, los datos para la realización de este estudio, fueron proporcionados por una empresa prestadora del servicio de suministro de agua potable en una zona del mediterráneo español. Los datos recopilados recogen información básica sobre las principales características de los distintos tramos de tubería de la red, esto es, datos en la red sobre mantenimiento, características ambientales y de estructura de los tubos.

Depuración.

Debido a la múltiples y variadas inconsistencias de la información suministrada, se sometió ésta, desde un comienzo, a un riguroso proceso de depuración, consistente de una parte, en identificar y eliminar registros con mediciones de fallos y características nulas y/o negativas, y de otra en discriminar entre averías provocadas y averías naturales (las que son consecuencia del paso del tiempo, el uso, etc.).

Base de datos depurada.

La depuración de los datos originales, redujo el volumen de información aproximadamente en un 17,6%. Los datos restantes se han caracterizado por un corto historial de fallos, con un altísimo porcentaje de censura y truncamiento. El resto de la información corresponde a tuberías averiadas por el deterioro natural y desgaste propio del paso del tiempo. En el capítulo 5, se describe en detalle, las principales características de estos datos depurados.

Revisión de antecedentes y selección de los modelos.

Esta etapa, se inicia revisando diversos antecedentes de la literatura científica existente, sobre la modelación estadística de los fallos en los conductos de las redes de agua potable; se destaca en forma resumida, los alcances y limitaciones de cada modelo, los requerimientos de información, una breve descripción del mismo y la formulación de su estructura matemática. Este enfoque (Kleiner y Rajani, 2000), facilita de una parte, la importante recopilación de información básica para el diseño de bases de datos idóneas para la modelación estadística y de otra parte, permite la reformulación de

indicadores de gestión (vitales en la planificación de la rehabilitación de las redes de abastecimiento).

Modelos Considerados.

Para responder a los objetivos planteados en este estudio, de fiabilidad en redes de suministro de agua potable, con cortos historiales de rupturas y alta censura múltiple, se consideraron tres grandes categorías de modelos, que son presentados y caracterizados en el capítulo 4 de esta memoria.

- De una parte, los denominados modelos *no Paramétricos*; en esta categoría se estudiaron, se valoraron y se aplicaron los modelos de: *Kaplan-Meier*; *Kaplan- Meier Ponderado*, *Actuarial (tablas de Vida)*, modelos que fueron aplicados al caso del esquema muestral con censura por la derecha únicamente y el modelo de *Nelson- Aalen, extendido*, que fue aplicado al caso de datos truncados a la izquierda y censurados a la derecha.

- En la categoría de modelos *semi Paramétricos*, se revisaron los modelos de Vida acelerada y el de Riesgo proporcional de Cox, que fue el que finalmente se aplicó con nuestros datos depurados, para identificar, factores de riesgo y de mayor influencia en el deterioro de las tuberías.

- En lo que respecta a la modelación *Paramétrica* (para responder al tercer objetivo), se probaron más de una decena de modelos (Weibull, Valores Extremos, Logístico, Log-normal, etc.), a los que se les analizó el ajuste a los datos disponibles, con los indicadores de Anderson-Darling y con la Correlación de Pearson. La estimación de los parámetros se hizo con métodos robustos de Máxima Versosimilitud y con el método de los Mínimos Cuadrados. Los resultados se presentan en el capítulo 6.

Comparación de los Modelos estudiados.

Se realizaron dos tipos de comparaciones. Una hace referencia, a la comparación de las estimaciones de la fiabilidad de los conductos de la red, con cada uno de los modelos: Kaplan-Meier, Nelson-Aalen extendido, Cox y Log-normal; que están representando a las categorías ya referenciadas (*semi no Paramétricas* y *Paramétricas*). Se trata en realidad de apreciar, si hay

diferencias importantes en la estimación de la fiabilidad de la red, entre cada uno de estos modelos. Los resultados, se representan gráficamente y se analizan igualmente en el capítulo 6.

El otro tipo de comparación, es el que se realiza exclusivamente con los modelos *Paramétricos*, y se refiere dentro de esta categoría a la identificación de la mejor familia (utilizando gráficos de Probabilidad); y el mejor ajuste de cada modelo, respecto a los datos depurados, utilizando también los indicadores de Anderson-Darling y de correlación de Pearson. Los resultados también se analizan en el capítulo 6 de esta memoria.

Conclusiones y Futuras líneas de investigación

En esta última fase de la metodología, se formulan las conclusiones generales del estudio y se presentan sugerencias respecto a los diferentes temas objetos de interés. Se proponen algunas recomendaciones para las bases de datos de supervivencia en redes de suministro de agua potable. También se presentan sugerencias sobre posibles líneas de investigación. Esta fase corresponde al capítulo 8 de la tesis.

3.2 Contenidos de la Tesis

En el contenido de esta memoria, se distinguen claramente dos partes. Los cuatro primeros capítulos, conforman la primera parte y los tres restantes la segunda.

En los capítulos 1 y 2 (**Justificación y Objetivos** respectivamente), se presenta un planteamiento general de la problemática a investigar es decir, se presenta “*el escenario del problema*”, conjuntamente con las razones que explican su estudio y además se formulan los objetivos que se espera alcanzar en el desarrollo de este trabajo.

El capítulo 3 (**Metodología y contenidos de la tesis**), recoge la estructura de la metodología seguida en el desarrollo de este estudio, con algunas explicaciones en cada una de sus etapas.

En el capítulo 4 (**Revisión y Antecedentes**) se presentan los antecedentes bibliográficos, que hacen referencia al problema a investigar. Se resaltan aquí tres aspectos. En primer lugar, la revisión y presentación de los antecedentes básicos para la modelación estadística de los fallos de las tuberías de una red de suministro de agua potable; esta parte, hace referencia esencialmente a los conceptos básicos de Análisis de Fiabilidad y Supervivencia. Luego se hace una revisión de lo esencial de la literatura sobre modelación de fallos de tuberías en redes de agua, presentando algunos modelos existentes, acompañados de comentarios sobre las limitaciones de cada uno de ellos, el tipo de información que requieren, una pequeña descripción del modelo y la formulación de su estructura matemática. Finalmente se describe un tercer aspecto, relacionado con los modelos estadísticos más relevantes (paramétricos, semiparamétricos y no paramétricos), del análisis de fiabilidad y supervivencia, que fueron aplicados en el desarrollo de este trabajo.

A partir de aquí, la segunda parte de la memoria la constituyen los capítulos 5, 6, y 7 cuyos contenidos se resumen a continuación.

El capítulo 5 (**Análisis de fiabilidad en una red de suministro de agua potable**) Se inicia con una presentación general, de las características básicas

(factores de uso y ambientales) de toda la información asociada a las tuberías de una red de suministro de agua potable de la comunidad Valenciana. Esta información se constituye en la columna vertebral de la aplicación de los modelos estadísticos seleccionados en este estudio, para describir la fiabilidad de las redes de abastecimiento de agua.

El capítulo 6 (**Resultados y Análisis**) recoge los resultados y la discusión de los distintos enfoques (no paramétrico, semiparamétrico y paramétrico) empleados en la valoración del estado de deterioro de los conductos de la red. En el primer apartado de este capítulo, además del análisis exploratorio de los datos, se presentan y analizan los resultados de tres métodos no paramétricos diferentes, utilizados en la estimación de la supervivencia de los conductos de la red: *Kaplan Meier*, *Tablas de vida (Método Actuarial)* y *Nelson Aalen extendido*. Se comparan además los resultados de estas 3 metodologías. En el apartado siguiente, se identifican los factores de riesgo de las tuberías a nivel general y por estratos, según tipo de material y nivel de tráfico de los conductos de la red. En el apartado final, se identifica el modelo paramétrico de mejor ajuste a los datos, que es utilizado para la predicción a largo plazo del estado de deterioro de la red.

El capítulo 7 (**Necesidades de información para la modelación estadística**) Revisa cuáles son las variables que con más frecuencia se emplean en los estudios sobre fiabilidad de redes de suministro, y junto con los resultados de nuestro estudio, se propone lo que sería una estructura recomendable para bases de datos a utilizar en la modelación de la supervivencia de redes de suministro de agua potable.

Finalmente, en el capítulo 8 (**Conclusiones y Recomendaciones**) se enumeran las principales conclusiones obtenidas del presente estudio (por capítulo) y se plantean algunas sugerencias para la mejor gestión de la información y explotación de la misma y se sugieren finalmente (de cara al futuro) posibles líneas de investigación en esta temática

4. Antecedentes

El deterioro de las tuberías existente en las redes de abastecimiento de agua varía de acuerdo a distintos tipos de factores esencialmente ambientales, dando lugar a una gran variabilidad en su patrón de fallos, tanto dentro del sistema de abastecimiento en cada ciudad, como entre ciudades. Es importante por lo tanto, acercarse a los modelos y métodos estadísticos más adecuados para describir, no sólo las características del fallo, sino también la predicción y análisis del comportamiento actual y futuro de los mismos (en cada tubería del sistema). Así pues, los modelos estadísticos de predicción de fallos en los conductos, se convierten en el paso esencial para el análisis de la fiabilidad de la red y para darle soporte a la toma de decisiones, al amparo de exigentes criterios económicos, en lo que concierne al mantenimiento y rehabilitación de las redes de suministro de agua potable.

Por tanto, en los sistemas de distribución de agua para consumo humano, son muchas las razones que nos permiten afirmar que el estudio de su fiabilidad es en gran medida un estudio estadístico. Entre otras:

- Los modelos empleados para representar la vida hasta el fallo de los distintos componentes básicos de las redes (tuberías, válvulas, tanques de almacenamiento y estaciones de bombeo) son modelos estadísticos; en donde las tuberías (es decir, la red de tuberías) juegan el papel más importante de todo el sistema.
- La estimación de los parámetros, de los diferentes modelos de predicción, que se han desarrollado para evaluar el estado estructural de las redes de tuberías (O'Day et al. 1986; Elnaboulsi y Alexandre 1996; Karaa y Marks 1990) se realiza estadísticamente, particularmente cuando se trata de identificar factores ambientales y estructurales de las redes, de mayor la influencia sobre el tiempo de vida de las tuberías.

Siendo la metodología estadística fundamental en el análisis de la fiabilidad, se presenta una breve revisión de aquellos conceptos que han sido utilizados y aplicados en este estudio.

4.1 Introducción al estudio de la Fiabilidad

La Fiabilidad, en la ingeniería se define como “*la probabilidad de que un dispositivo realice su función bajo determinadas condiciones de funcionamiento, durante un período de tiempo establecido*”; su estudio se ha orientado a los *tiempos hasta el fallo* de dispositivos y componentes de diferente naturaleza con el fin de predecir, la probable ocurrencia de las averías, la rapidez con que se presentan y el instante en que ocurrirán. De esta manera se pueden determinar políticas de mantenimiento y rehabilitación, predicción de costes en mantenimiento de redes de suministro de servicios públicos y en general plazos de garantías en productos y servicios de abastecimiento. Algunos investigadores del tema, han adaptado estos análisis a la ingeniería y particularmente al estudio de fallos de tuberías en redes de agua para el consumo humano (Clark et al. 1982, 1988; Andreou et al. 1987), ya que estas mismas técnicas, tradicionalmente han sido utilizadas particularmente en el campo de la biomedicina, de donde provienen y donde se conocen con el nombre de *Análisis de Supervivencia*.

El análisis de Fiabilidad incorpora el hecho de que, mientras algunas tuberías fallan (por envejecimiento natural o por una causa provocada) otras no, todo este acopio de información se considera de vital importancia y de gran impacto a la hora de analizar los fallos en los sistemas de abastecimiento de agua, pues las tuberías pueden fallar muchas veces durante su vida útil. Cada vez que se observa un fallo, se necesita intervenir inmediatamente la red. Algunos investigadores (Andreou et al. 1987; Eisenbis 1994; Guftafson Clancy 1999) han demostrado que el patrón de ruptura depende fuertemente del *número de rupturas previas que las tuberías han sufrido en el pasado*. De hecho, *el número de rupturas previas es a menudo reportado como el factor más importante al predecir futuros fallos*. El análisis de Fiabilidad es particularmente útil en este campo, cuando los registros de los fallos se han conservado en buena proporción, en relación al historial de averías y rupturas de la red. Sin embargo, tratar de realizar un análisis de Fiabilidad “Textbook” (textualmente o en forma estrictamente rigurosa), no es fácil en un estudio de esta naturaleza, porque los registros históricos que recogen las empresas

gestoras del servicio sobre sus rupturas son muy cortos, tienen menos de diez años y además son registros incompletos en el tiempo (contienen altos porcentajes de información parcial sobre las averías de los conductos, que dificulta el conocer el historial completo de fallos en la red) y esto obviamente es un obstáculo para la modelización estadística de los fallos.

El Análisis de Supervivencia o Análisis de Fiabilidad es el conjunto de técnicas que se emplean para analizar los datos en los que la variable de interés es el tiempo que transcurre desde un instante inicial bien definido, como puede ser la instalación de una tubería, hasta la ocurrencia de un determinado suceso o instante final, como puede ser la rotura o avería de la misma. Para una revisión en profundidad del Análisis de Supervivencia, véanse los textos de Klein y Moeschberger (1997), Andersen, Borgan, Gill y Keiding (1993), Cox y Oakes (1984), Lawless (1982), Kalbfleish y Prentice (1980), entre otros.

La principal razón por la que en este tipo de estudios no se aplican las técnicas de análisis de datos habituales (como son los modelos clásicos de regresión multivariados), es porque la variable tiempo de supervivencia no sigue una distribución Normal, más bien suele tener una distribución asimétrica con una larga cola a la derecha. Aunque podrían intentarse transformaciones para su normalización, la segunda dificultad y característica principal de los datos de supervivencia, es que estos tiempos de supervivencia son frecuentemente, *censurados*. Se dice que, el tiempo de supervivencia de un individuo es censurado, cuando el suceso de interés no ha sido observado en ese individuo. En nuestro estudio serán censurados todos los datos correspondientes a tuberías que durante el periodo de observación no habían presentado ninguna rotura.

La variable de interés es T : *tiempo que tarda una tubería en romperse desde su instalación* y por tanto, es el tiempo durante el cual la tubería ha estado en condiciones de funcionamiento. Asociadas a dicha variable aparecen dos funciones de fundamental interés, que describen su comportamiento o distribución.

Por una parte la *función de supervivencia* o *función de fiabilidad*, $R(t) = P(T > t)$, Kalbfleisch (1980) y Crowder et al. (1991), proporciona la probabilidad de que *el tiempo de supervivencia* sea mayor que t . En nuestro estudio, es la probabilidad que tarda una tubería en romperse por encima de un instante t y por tanto es favorable tener probabilidades altas en valores altos de t .

La *función de riesgo* o *tasa de fallo* $h(t)$ se utiliza para expresar porcentualmente el riesgo o tasa de fallo (condicional) de una unidad en un instante de tiempo t . En el caso de los conductos de una red, es la probabilidad de que una tubería falle en el instante t condicionado a que esa tubería no ha fallado (y por tanto ha sobrevivido) hasta ese preciso instante. La definición formal para estas y otras funciones básicas por su importancia, se presentan en los siguientes apartados.

4.2 Tipos de Censura

Son múltiples las ocasiones en las que solo se tienen datos completos de la variable de interés T (tiempo de ocurrencia de un suceso, igualmente de interés) en una parte (que usualmente es pequeña) de las n unidades o individuos de la muestra que se desea analizar, mientras que del resto solamente se tiene información parcial o incompleta. Esta particularidad, es la que dificulta el análisis estadístico en los estudios de Fiabilidad y de Análisis de Supervivencia, pues se dice que buena parte de los datos de la muestra vienen con *censura*.

Atendiendo a las causas que dan lugar a la censura, (y que pueden ser *aleatorias* o *controladas*) se distinguen los siguientes tipos de censura:

4.2.1 Censura por la derecha Una observación T_i se dice *censurada a la derecha* de un cierto valor conocido Y , si se desconoce el valor exacto de la observación T_i y solo se sabe que ésta es mayor que Y . Convencionalmente esta situación se representa con n pares de variables aleatorias (t_i, δ_i) , donde

$$t_i = \min(T_i, Y) \quad y \quad \delta_i = \begin{cases} 1 & \text{si } T_i \leq Y \\ 0 & \text{si } T_i > Y \end{cases}$$

La variable δ_i (*indicador de censura*), indica si la observación T_i es censurada o no, y $t_i = T_i$, si es un dato completo u observado y si no lo es $t_i = Y$.

4.2.2 Censura por la izquierda. Análogamente, una observación T_i se dice *censurada a la izquierda* de un cierto valor conocido X , si se desconoce el valor exacto de la observación T_i y solo se sabe que ésta es menor que X ; el momento exacto en el que ocurrió el fallo es desconocido, sólo se sabe que ha ocurrido antes de que el individuo se incluya en el estudio. Es posible por ejemplo, encontrarse en la muestra con unidades que ya han fallado antes del comienzo del estudio, sin saber exactamente cuándo.

4.2.3 Censura doble. Cuando existen algunos datos censurados tanto por la izquierda como por la derecha, se dice simplemente que *la Censura de los datos es doble*.

Situaciones prácticas con muestras doblemente censuradas han sido estudiadas entre otros por, Gehan (1965), Mantel (1967), Peto (1973), Leiderman et al.(1973), Turnbull (1974), Turnbull y Weiss (1978), Morales et al. (1991), Tang et al. (1995).

4.2.4 Censura tipo I (censura única por tiempo): El suceso de interés (fallo, avería o ruptura de la tubería) se observa si ocurre antes de un instante de tiempo fijo predeterminado C . En este caso, C es una constante (de censura) prefijada por el investigador para todas las unidades muestrales. Este tipo de censura es común cuando, por diversas causas, el investigador finaliza el estudio antes de que todos los individuos hayan experimentado el suceso de interés. En el campo de los test de fiabilidad se relaciona con los ensayos limitados por tiempo.

4.2.5 Censura tipo II (censura única por número de fallos): Este tipo de censura surge cuando se fija al final del estudio en el momento en que un número $r < n$ predeterminado de individuos falla. Los tiempos de vida observados son los r menores valores de la muestra, de forma que C ahora, se convierte en una variable aleatoria $C = T(r)$. Este tipo de censura se usa con frecuencia en los experimentos industriales y es más fácil de analizar desde el punto de vista estadístico. Estos experimentos se denominan ensayos limitados o censurados por número de fallos.

Es importante señalar, que el valor de C en el esquema de tipo I y el valor de r (o la fracción r/n) que indica la tasa de censura en el esquema de tipo II, deben establecerse antes de iniciar el experimento y no durante el transcurso del mismo, dependiendo de los resultados que se observen. La necesidad de que el mecanismo de censura sea independiente de la observación del fenómeno es un requisito imprescindible para la validez de las conclusiones.

4.2.6 Censura Aleatoria. Generalmente en los estudios de fiabilidad se fija de antemano su duración y las unidades entran a formar parte de la muestra a lo

largo de ese período. En las unidades que no se ha experimentado el suceso de interés al final del estudio, la censura de sus tiempos de vida es semejante a la de tipo I. En ocasiones durante la duración del estudio, algunas unidades experimentan otros sucesos independientes del de interés que provocan su eliminación del mismo. Esta situación se denomina *Censura Aleatoria*. Es el caso del abandono de pacientes en estudios de supervivencia en biomedicina; las averías y roturas en los conductos de las redes, ocasionadas (en plena duración del estudio) en forma imprevista por causas externas, diferentes al envejecimiento natural de la tubería, etc. En este tipo de censura, C es una variable aleatoria que se supone independiente de la variable de interés T .

4.2.7 Censura Múltiple. La censura múltiple se produce cuando *no existe un único tiempo de censura* y puede ser por alguna de las siguientes causas:

- Las unidades entran en servicio en diferentes instantes de tiempo. Es el caso de nuestros datos, donde las tuberías fueron instaladas en diferentes años.
- Puede ocurrir que existan distintos equipos de prueba o ensayo de las unidades y algunos de estos equipos se averían mientras tienen lugar las pruebas. Otra situación sería el caso en que se “pierda” el seguimiento del individuo o unidad por causas distintas a lo que es objeto de nuestro estudio; tal es el caso de los daños causados por obras realizadas en las vías públicas, que pueden afectar a redes de abastecimiento de agua causando fugas, averías y desperfectos en general en los conductos de estas redes; en este caso la tubería ha fallado por una causa externa provocada y no precisamente por el uso y el paso del tiempo. Generalmente, este tipo de observaciones se representan mediante un par de variables (T, δ) , donde T es el tiempo transcurrido desde la entrada del individuo al ensayo hasta su salida del mismo y δ es una variable binaria indicadora del tipo de observación, que toma el valor 1 si se ha observado el fallo y el valor 0 si se trata de una observación censurada.

Para una descripción detallada de los mecanismos que originan los distintos tipos de censura puede consultarse a Hill et al. (1990), Miller (1981), Lawless (1982), Kalbleisch y Prentice (1980).

4.2.8 Datos Truncados por la izquierda

En un esquema muestral con truncamiento, sólo aquellos individuos que verifican cierta condición definida de antemano son observados por el investigador.

En estudios de supervivencia, el tipo más común de truncamiento por la izquierda ocurre cuando los sujetos comienzan a ser observados a edades aleatorias, esto es, el origen del tiempo de vida precede al origen del estudio. En tal caso, aquellos sujetos en los que el fallo tiene lugar antes del inicio del estudio serán ignorados por el investigador. Como consecuencia, si T es el tiempo de fallo, y X es el tiempo en que éste se incorpora al estudio, un sujeto formará parte de la muestra únicamente si $T \geq X$. De modo que, bajo un esquema muestral con truncamiento a la izquierda la inferencia no se realiza sobre la variable T , sino sobre la variable T condicionada a la ocurrencia del suceso o condición $T \geq X$. Es el caso bastante frecuente que ocurre, cuando los datos al inicio de un estudio, no se recogen desde el mismo instante en que las unidades o sistemas (a estudiar) entran en operación, pero sí en algún otro momento posterior.

El tipo de truncamiento a la izquierda más habitual ocurre cuando los sujetos entran a estudio a edades aleatorias (no necesariamente cuando se inicia el proceso para la ocurrencia del suceso de interés) y son seguidos a partir de este tiempo de entrada con retraso hasta que el suceso ocurre o hasta que el sujeto es censurado a la derecha. En este caso, todos los sujetos que experimentan el suceso de interés antes de la entrada en estudio no serán conocidos por el investigador, para el truncamiento a la izquierda, estos individuos no son considerados en el estudio.

Por ejemplo, supongamos que la variable de interés T , es el tiempo de duración de una determinada enfermedad y que otra variable X , mide el tiempo desde que la enfermedad comienza en el individuo hasta que éste entra en estudio. Únicamente se dispondrá de los datos correspondientes a los pacientes que entran en el estudio, es decir, aquellos individuos para los que se cumpla que $X_i \leq T_i$.

4.2.9 Modelo de truncamiento por la izquierda y censura por la derecha.

En muchos problemas de Análisis de Supervivencia y Fiabilidad, hay una fuerte presencia de datos que son a la vez, truncados y censurados y que requieren de un nuevo modelo para su estudio.

El modelo teórico que recoge esta situación, denominado *modelo de truncamiento aleatorio por la izquierda y censura aleatoria por la derecha*, es el siguiente:

Sea (X, T, C) un vector aleatorio, donde X es el tiempo de truncamiento; T es el tiempo de fallo y C es el tiempo de censura. Se supone que Bajo un modelo de truncamiento por la izquierda y censura a la derecha, las observaciones muestrales serán observaciones del vector (X, Y, δ) , donde $Y = \min\{T, C\}$ y δ es el indicador de censura es decir $\delta = 1_{\{T \leq C\}}$ si T es un tiempo de fallo o 0 si T es un tiempo de censura. Si $Y < X$ no hay observación. Es decir que la muestra está constituida por n observaciones de tipo (x_i, y_i, δ_i) en los que $x_i \leq y_i$ para todo $i = 1, 2, \dots, n$.

Las condiciones habituales en este modelo son :

T es independiente de (X, C)

X, T y C son mutuamente independientes

Sobra decir que la segunda condición implica la primera.

4.3 Estructura usual de los datos en una Red de agua

La Figura 4.1 muestra la disposición más o menos típica de los datos en una red de abastecimiento de agua (Rostum, 2000). Se observan básicamente tres zonas: La del centro es la ventana de los fallos disponibles (que coincide con el período del estudio), esto es, los instantes de las rupturas y averías que se observan en las tuberías. Esta zona puede incluir también datos de instalación de tuberías. Los tiempos t_i y t_f son los instantes de inicio y finalización del estudio. A la izquierda de la ventana de fallos disponibles, están los fallos y/o las censuras que pudieron haber ocurrido y que no son conocidos (porque ni siquiera fueron registrados) o lo fueron parcialmente, estos últimos se suelen considerar datos *censurados por la izquierda* o bien tratarse de datos *truncados por la izquierda* (la diferencia se verá más adelante); esta zona también contiene buena parte de los registros sobre el año de instalación de las tuberías.

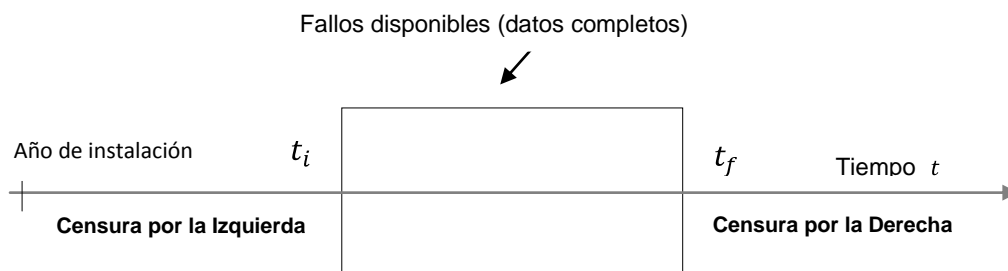


Figura 4.1 Estructura de Datos sobre fallos disponibles en las redes de agua

A la derecha de la ventana central, quedan registradas parcialmente con el tiempo t_f^+ (tiempo final con censura), todas las tuberías que no fallaron al finalizar el período $[t_i; t_f]$, es la zona que corresponde a los *datos censurados por la derecha*.

4.4 Funciones Básicas en Fiabilidad.

En el análisis de las probabilidades, se sabe que una variable aleatoria continua queda completamente caracterizada, cuando se conoce su función de densidad ($f(t)$) o su función de distribución ($F(t)$). Sin embargo, en el análisis de Fiabilidad o de Supervivencia, es usual considerar otras funciones, que sirven para enriquecer y caracterizar con otra visión la misma variable. En nuestro caso la variable de interés T es el *tiempo hasta el fallo* (o *tiempo de supervivencia*) de un tramo de tubería en una red de suministro de agua y que se interpreta como el tiempo que transcurre, desde la instalación de la tubería hasta el preciso instante de su fallo. En la mayoría de los casos para efectos prácticos, la variable T así definida se considera continua. Las más importantes funciones son *la función de fiabilidad o supervivencia* y *la función de riesgo*.

4.4.1 La Función de Fiabilidad o de Supervivencia

Se utiliza para cuantificar el grado de seguridad de funcionamiento de un dispositivo, es decir, la capacidad que tiene para cumplir con éxito una función o misión determinada. En general, se define por la probabilidad de que el dispositivo sobreviva más allá del instante t , esto es:

$$R(t) = P(T > t) \quad 0 \leq t < \infty$$

Por lo tanto, la función de Fiabilidad da la probabilidad complementaria de la habitual función de distribución acumulativa, $F(t) = 1 - R(t)$. Como se mencionó anteriormente, es la definición utilizada por autores clásicos como Kalbfleisch (1980) y Crowder et al. (1991). Definida de esta forma, la función de fiabilidad es una función monótona, decreciente, continua por la izquierda, con

$$R(0) = 1 \quad \text{y} \quad \lim_{t \rightarrow \infty} R(t) = 0$$

Si T es una variable continua, la función de Fiabilidad, $R(t)$, se puede obtener por integración de la función de densidad $f(x)$, es decir,

$$R(t) = P(T > t) = \int_t^{\infty} f(x)dx \quad 0 \leq t < \infty$$

En la Figura 4.2 se aprecia la relación entre las funciones $R(t)$ y $F(t)$.

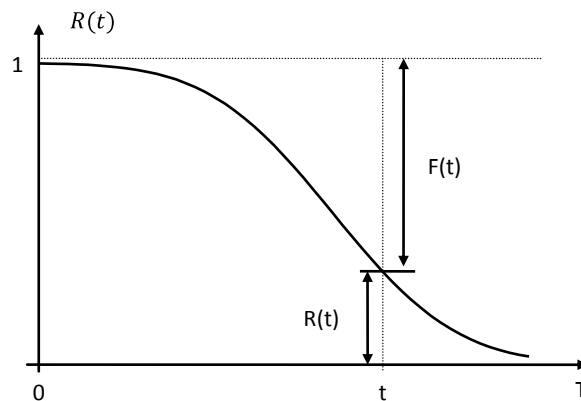


Figura 4.2. Función de Supervivencia

4.4.2 La función de Riesgo

Conocida también como *fuerza de mortalidad (FOM)*, *tasa de fallo o riesgo de fallo*, indica la razón (instantánea) de fallo en el instante $T = t$ condicionada a la supervivencia del dispositivo, hasta el inicio de este instante t . Es decir, mide la variación relativa del número de supervivientes en el instante t y se relaciona con el número de fallos por unidad de tiempo; se define también como *la velocidad de extinción o degradación del dispositivo*; es la intensidad, *la fuerza con que se presentan los fallos en el dispositivo, justo en el instante t* . Representa la evolución de la probabilidad de fallo en relación con la edad de los individuos.

Si T es una variable continua la función de riesgo en t , que con frecuencia se denota por $h(t)$ o por $\lambda(t)$, se define como:

$$h(t) = \lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{R(t)}$$

La tasa de fallo, $h(t)$, de casi cualquier tipo de dispositivo, (incluso los conductos de una red), varía en función del tiempo, y su forma más frecuente se asemeja a la de la figura 4.3, que por su forma, se conoce como *Curva de la Bañera*; y que se caracteriza, porque durante el primer periodo de vida de las unidades, la tasa de fallo es decreciente (periodo de *fallos precoces*) hasta que se alcanza un valor en el cual se mantiene sensiblemente constante (periodo de *fallos accidentales*) y que es la zona llamada de *vida útil* del producto.

Finalmente, a partir de un determinado instante de tiempo, la tasa de fallo crece suavemente al comienzo, pero luego con el tiempo de un modo más rápido (periodo de fallos por *envejecimiento*). Si la función de riesgo de una tubería fuese constante, la $R(t)$ se reduciría a la función de supervivencia del modelo exponencial, que será caracterizada posteriormente. La función $\lambda(t)$ presenta diferentes tipologías en su forma geométrica, no solamente adopta la forma de curva de bañera, puede ser, creciente, decreciente, constante, en forma de joroba, etc. Es de anotar, que en la literatura se conocen dos tipos de curvas de bañeras. Una para *sistemas reparables*, donde el sistema o dispositivo puede fallar varias veces y la otra para *sistemas No reparables*; estos últimos incluyen elementos que solamente tienen una vida útil o son sistemas que después de reparados, vuelven a funcionar perfectamente, como cuando eran nuevos.

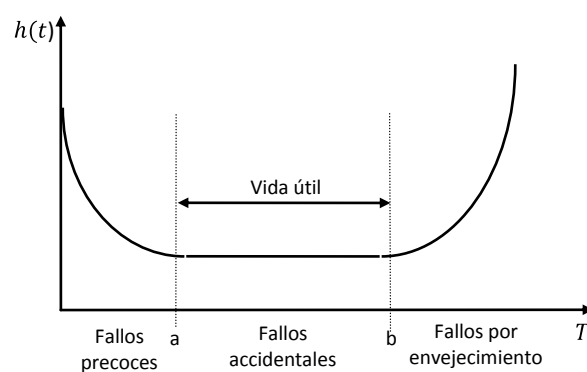


Figura 4.3. Evolución de la Tasa de fallo $h(t)$

4.4.3 La Tasa de riesgo acumulada (*función de Riesgo acumulado o función de Impacto*).

Para el caso de la variable continua T de este estudio, y en general para toda variable continua, la función de impacto $H(t) = \Lambda(t)$, se define por

$$H(t) = \Lambda(t) = \int_0^t h(u)du = -\log R(t)$$

La función $H(t)$ así definida, también permite caracterizar la distribución de probabilidades de T y además, es importante en la medición de la rapidez con que ocurren las rupturas en el tiempo, en la construcción de papeles probabilísticos y en el análisis de residuos.

Hoyland, A. y Rausand M. (1994), proponen utilizar los gráficos de la función $H(t)$ estimada cuando se trata de interpretar, si una unidad tiene una tasa de fallo creciente (IFR), decreciente (DFR), constante (CFR) o en forma de curva de bañera. El esquema de los gráficos se presenta en las Figuras 4.4 a (IFR), 4.4b (DFR) y 4.4c (curva en bañera)

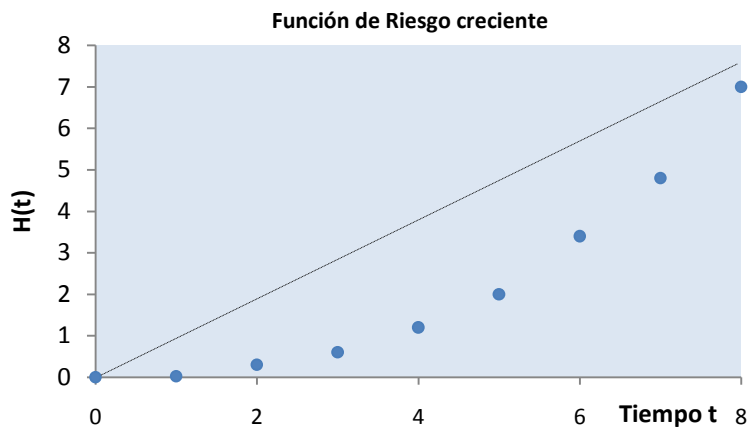


Figura 4.4.a. Función de impacto $H(t)$ indicando IFR

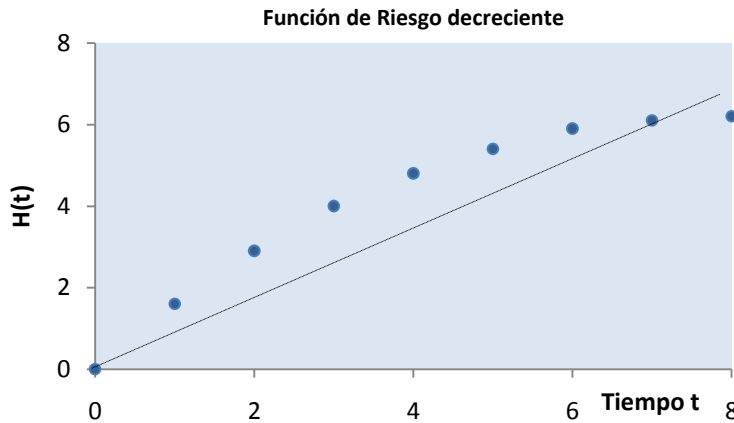


Figura 4.4 b. Función de impacto $H(t)$ indicando DFR

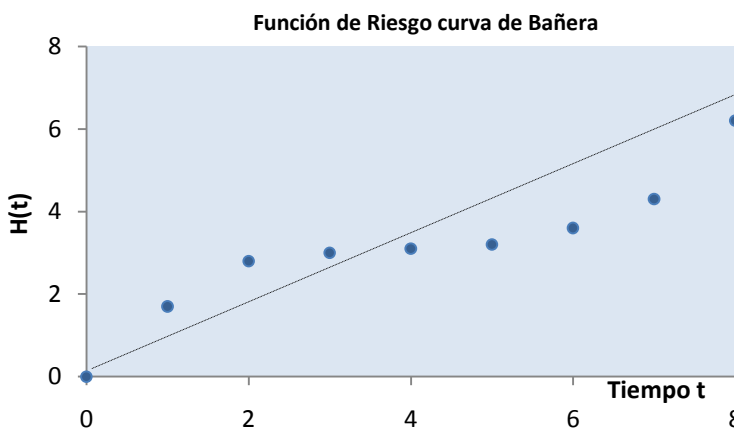


Figura 4.4.c. Función de impacto $H(t)$ indicando curva en forma de Bañera

El seguimiento a los cambios en la pendiente de la tangente en cada punto de estas curvas, es un indicador para valorar la evolución de $h(t)$ en el tiempo y poder así, identificar zonas con IFR, DFR, CFR y con curva de bañera. En particular en la Figura 4.4 c, se distinguen claramente las 3 fases de la curva de la bañera: inicialmente, un decrecimiento de $h(t)$ (*mortalidad infantil*) en el intervalo $(0;3)$; una relativa estabilidad en $(3;5)$ (*vida útil*) y un aumento en el intervalo $(5;8)$ (*envejecimiento*).

Todas estas funciones $f(t)$, $F(t)$, $R(t)$, $h(t)$ y $H(t)$ sirven para caracterizar la distribución de probabilidad de T y además son equivalentes entre sí, de modo

que conocida una de ellas, las restantes son fáciles de obtener a través de alguna de las siguientes relaciones,

$$f(t) = -\frac{d}{dt}R(t)$$

$$h(t) = -\frac{d}{dt}\ln R(t)$$

$$R(t) = e^{-\left[\int_0^t h(x)dx\right]} = e^{-H(t)}$$

4.5 Algunos modelos de fiabilidad básicos

En esta sección vamos a revisar algunas de las distribuciones de probabilidad más empleadas en Fiabilidad y Análisis de Supervivencia. En principio, cualquier distribución no negativa se puede utilizar para modelizar una variable tiempo de vida T . El objetivo es disponer de un conjunto de distribuciones lo suficientemente flexibles para adaptarse a los distintos tipos de datos y que sean lo más sencillas posibles para facilitar su análisis (Abaurrea, J. y Cebrián, A.1998).

Los métodos tradicionalmente utilizados en análisis de supervivencia, pueden dividirse en tres grupos: paramétricos, no paramétricos y semiparamétricos. Los más extendidos, son los semiparamétricos y paramétricos; y las estimaciones que se obtienen con estos métodos, sirven de gran apoyo a posteriores análisis estadísticos más detallados y elaborados.

Los métodos no paramétricos tienden a ser más sencillos. Éstos métodos son menos eficientes que los métodos paramétricos, pero resultan de gran utilidad cuando no se conoce ningún modelo paramétrico que se ajuste adecuadamente a los datos.

Con mucha frecuencia, se han empleado en la literatura modelos paramétricos, para estimar funciones de fiabilidad y realizar contrastes de hipótesis sobre las mismas. En general estos modelos son usados en el análisis del tiempo de vida y en problemas relacionados con la modelización del envejecimiento y el proceso de fallo.

4.5.1 Modelos Paramétricos

Unas cuantas distribuciones paramétricas univariantes han venido jugando un importante papel en la modelización de un amplio rango de situaciones, que incluyen la modelación de fallos en conductos de una red. Dentro de estos modelos, los más frecuentemente utilizados son:

- *Exponencial*, Davis (1952), Epstein y Tsao (1953), Epstein y Sobel, (1953, 1954), Bartholomew (1957, 1963), Epstein (1958), Zelen y Dannemiller (1961).

- *Weibull*, Weibull (1951), Lieblein y Zelen (1956), Kao (1959), Cohen (1965), Harter y Moore (1965).

- *Normal y log-Normal*, Dixon (1960), Feinleib, (1960), Cohen, (1961), Hill, (1963).

- *Gamma*, Gupta y Groll, (1961), Wilk et al. (1962a, 1962b); Stacy y Mihram (1965), Harter (1967).

Para la utilización de estos métodos, se asumen ciertas funciones de probabilidad para la variable tiempo de vida. Si estos supuestos son razonables, (y para averiguarlo antiguamente se realizaban procedimientos gráficos manualmente, basados en papel de probabilidad (Chernoff y Lieberman (1954); Kimball (1960) y Nelson (1969, 1972, 1979, 1982, 1990), hoy en día, tales procedimientos se realizan automáticamente en la mayor parte del software estadístico existente) se procede luego, con la estimación de los parámetros de la familia elegida. El método consiste en estimar, por métodos robustos (máxima verosimilitud o mínimos cuadrados), los parámetros característicos de la distribución, y usar su normalidad asintótica para realizar la estimación por intervalos y los contrastes de hipótesis del caso.

Las pruebas de bondad de ajuste se realizan con la habitual prueba basada en la ji-cuadrado, con la del logaritmo del cociente de verosimilitudes o con gráficos de probabilidad como se mencionó anteriormente. Una útil referencia para profundizar en estos temas, es el trabajo realizado por Johnson y Kotz (1970), quienes tratan extensamente catálogos matemáticos y las propiedades estadísticas de las distribuciones tratadas en este documento y también proveen referencias concernientes a sus áreas de aplicación. El software disponible hoy en día en el mercado (Minitab 15, entre otros) reduce el riesgo que existía antes, de seleccionar equivocadamente el modelo subyacente de mejor ajuste a los datos, Bañ y Englehart, (1980), para evitar que las estimaciones máximo verosímiles, conduzcan a resultados inapropiados y asintóticamente sesgados. Los métodos paramétricos en

general se consideran más potentes que los no paramétricos (Kaplan-Meier, Actuarial, etc.).

Algunas de las distribuciones anteriormente referenciadas, se presentan a continuación.

4.5.1.1 Distribución Exponencial

Se utiliza para modelar el tiempo transcurrido entre dos sucesos aleatorios no muy frecuentes cuando la tasa de ocurrencia, λ , se supone constante.

En fiabilidad se usa para describir los tiempos de fallo de un dispositivo durante su etapa de vida útil, en la cual la tasa de fallo es (aproximadamente) constante, $h(t) = \lambda$. Una tasa de fallo constante significa que, para un dispositivo que no haya fallado con anterioridad, la probabilidad de fallar en el siguiente intervalo infinitesimal es independiente de la edad del dispositivo.

La tasa de fallo λ es el parámetro que caracteriza a esta distribución. Este valor es la inversa del tiempo medio que transcurre hasta el fallo (o entre dos fallos consecutivos, MTBF, si el dispositivo sigue funcionando) $MTTF = 1/\lambda$.

La expresión de la función de densidad cuando la vida de la unidad (un tramo de tubería) sigue una distribución exponencial es

$$f(t) = \lambda e^{-\lambda t} \quad t \geq 0$$

donde λ es una constante positiva ($\lambda > 0$).

La función de distribución, ó, la desfiabilidad es,

$$F(t) = 1 - e^{-\lambda t} \quad t \geq 0$$

La función de fiabilidad, probabilidad de supervivencia a una duración t , queda como

$$R(t) = 1 - F(t) = e^{-\lambda t} \quad t \geq 0$$

La tasa de fallo es:

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \quad t \geq 0$$

Puede observarse que $h(t)$ no depende de t , en otras palabras, la tasa de fallo es constante, por ello emplearemos la distribución exponencial durante el período de vida útil del producto.

Es frecuente representar al parámetro $1/\lambda$ por θ , vida media, así pues, las fórmulas anteriores quedarían como:

$$f(t) = \frac{e^{-t/\theta}}{\theta}, F(t) = 1 - e^{-t/\theta}, R(t) = e^{-t/\theta}, \mu = \theta, \sigma = \theta$$

Si nos fijamos en las expresiones de las funciones de densidad, distribución y fiabilidad (la que es aquí objeto de estudio), observamos que es la relación entre la duración de la misión y la vida media, (o sea el ratio t/θ), el que define el valor de las mismas y no tanto el valor exclusivo de t .

La distribución exponencial tiene las siguientes características:

- Es la distribución más sencilla utilizada en el análisis de datos de fiabilidad. Su función de riesgo es constante.
- Utilizada como distribución de tiempo hasta el fallo de componentes electrónicos. No si éstos tienen fallos debidos a defectos de calidad.
- Descripción de tiempo de vida para componentes *durante su periodo de vida útil*, ya que la tasa de fallo es constante.

4.5.1.2 Distribución Weibull

La distribución Weibull es quizás la más usada para modelar la distribución del tiempo de vida. La utilidad de las aplicaciones de esta distribución han sido ampliamente definidas por diversos autores Weibull, (1951); Berretoni, (1964), y utilizadas como modelos de diversos tipos de productos como tubos de vacío Kao, (1959), aislamiento eléctrico Nelson,

(1972) así como también en aplicaciones biomédicas, tales como, estudios del tiempo hasta la aparición de un tumor en poblaciones humanas, Altshuler, (1970) o aplicaciones en el laboratorio con animales, Pike (1966), Peto et al., (1972) y en muchas otras situaciones.

Se ha comentado antes que la distribución exponencial es utilizada a menudo para modelar los tiempos de fallo cuando la tasa de riesgo es constante. Si por el contrario, la probabilidad de fallo varía con el tiempo resulta más apropiada una Weibull (de hecho la exponencial puede verse como un caso particular de la Weibull). La distribución Weibull es tan flexible que, eligiendo adecuadamente sus parámetros, permite describir las tres etapas de la función tasa de fallos (curva de la bañera).

Una variable aleatoria continua no negativa, que habitualmente interpretaremos como la vida de la unidad estudiada sigue una *distribución de Weibull* si su función de densidad es:

$$f(t) = \beta(t - \delta)^{-1} \left(\frac{t - \delta}{\theta - \delta} \right)^{\beta} e^{-\left(\frac{t - \delta}{\theta - \delta} \right)^{\beta}} \quad t > \delta$$

expresión en la cual:

δ es la vida mínima o umbral (*threshold*) de las unidades estudiadas ($\delta \geq 0$).

θ es la vida característica (*scale*) de esas unidades ($\theta > \delta$)

β es el parámetro de forma (*shape*) o pendiente de Weibull ($\beta > 0$)

La vida mínima δ es una edad que con seguridad van a alcanzar las unidades estudiadas, y puede hacer referencia al hecho de que estas unidades no sean nuevas en el momento de iniciar el estudio. Con frecuencia, como luego se comenta, la vida mínima toma el valor cero.

La vida característica θ es una edad tal que la probabilidad de que sea superada esa duración es el 36.79%, o, lo que es equivalente, tal que un 63.21% de las unidades fallan antes de alcanzarla. Aunque la vida

característica no es la media de la distribución de Weibull, puede interpretarse como un indicador aproximado de posición (recuérdese que en la distribución exponencial la probabilidad de que la media sea superada es precisamente el 36.79%).

Por último, el parámetro de forma β describe la forma de la distribución, y como se verá, es clave para entender el comportamiento de la variable vida útil o duración de las unidades estudiadas.

La función de distribución será:

$$F(t) = 1 - e^{-\left(\frac{t-\delta}{\theta-\delta}\right)^\beta}$$

Como ya se ha comentado, con frecuencia la vida mínima δ toma el valor cero, con lo cual las expresiones anteriores se simplificarían, quedando la denominada *distribución reducida de Weibull*, cuya función de densidad es:

$$f(t) = \beta t^{-1} \left(\frac{t}{\theta}\right)^\beta e^{-\left(\frac{t}{\theta}\right)^\beta}$$

y la función de distribución:

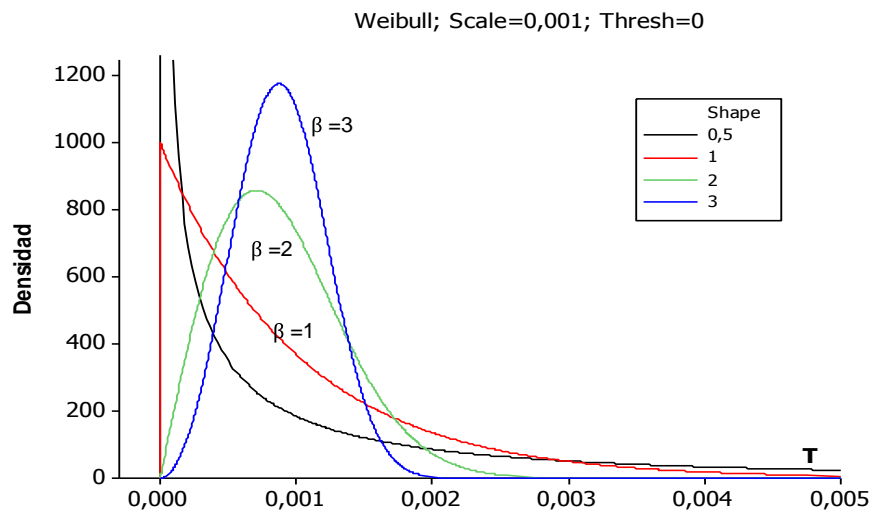
$$F(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^\beta}$$

En esta distribución reducida de Weibull, el valor medio resulta ser:

$$E(T) = \theta \left(\Gamma \left(1 + \frac{1}{\beta} \right) \right)$$

Donde Γ es la *función gamma* tabulada (o bien obtenida por integración numérica).

El aspecto que tiene la función de densidad de esta variable de Weibull depende del valor de sus parámetros. En la figura 4.5 se recoge la forma de la función de densidad para diferentes valores de β (0,5; 1; 2; 3), con umbral $Tresh = \delta = 0$ y con escala $\theta = 0,001$ fijos.



Obsérvese que para valores elevados de β la forma de la distribución se asemeja a la campana de Gauss, es decir, a la distribución normal. En la práctica a partir de $\beta = 3.2$ se aproxima la distribución de Weibull a la normal.

4.5.1.3 Distribución Log-Normal

La distribución *Log-normal* se utiliza con mucha frecuencia en el cálculo actuarial para ajustar datos relativos al coste de siniestros en las pólizas de seguros, se reportan también aplicaciones en geología, especialmente con datos relativos a la explotación de yacimientos de oro (en algunas regiones de Africa del Sur), Kock y Link, (1970); Agterberg, (1974).

Se dice que una variable aleatoria sigue una distribución *log-normal* (en base e) o logarítmico normal cuando su transformada logarítmica (o más generalmente, la transformación logarítmica de una transformación lineal de la variable original) tiene una distribución normal.

Formalmente, *una variable aleatoria T tiene un comportamiento log-normal, si su logaritmo es normal, es decir, si:*

$$\ln(T) \sim N(\mu, \sigma^2)$$

Si T sigue una distribución log-normal se representa por $T \sim LN(\mu, \sigma^2)$

Donde μ y σ son los parámetros de posición y dispersión de la distribución de $\ln(T)$ y que toman diferentes nombres, como se verá posteriormente.

Y sus funciones de densidad y de distribución vienen dadas respectivamente por:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-(\ln t - \mu)^2 / 2\sigma^2} \quad t \geq 0$$

$$F(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad t \geq 0 \quad (4.1)$$

Donde $\Phi(z)$ representa la función de distribución de una normal estándar, cuyo cálculo se obtiene con la integral indefinida

$$\Phi(z) = \int_{-\infty}^z \varphi(u) du$$

Donde, $\varphi(u)$ es la f.d.p de una normal (0,1).

La función de fiabilidad $R(t)$ y la función de riesgo $h(t)$ dependen también de $\Phi(z)$, como se puede ver en la siguiente expresión

$$R(t) = P(T > t) = 1 - F(t) = P\left(\frac{\ln T - \mu}{\sigma} > \frac{\ln t - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)$$

La función de riesgo $h(t) = \frac{f(t)}{R(t)}$ tiene valor cero en $t = 0$, es creciente hasta un máximo y después decrece aproximándose a cero cuando t aumenta. El modelo log-normal es muy fácil de utilizar si no hay censura, pero con censura los cálculos y operaciones se complican.

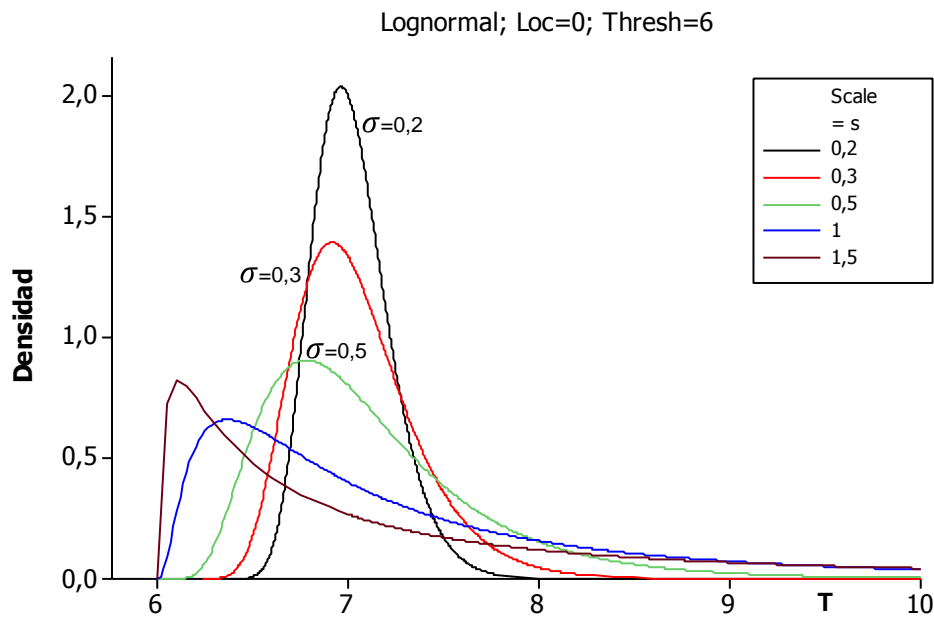


Figura 4.6 Funciones de densidad de varias distribuciones lognormales

Características de la log-normal con dos parámetros.

Si $\ln(T) \sim N(\mu, \sigma^2)$, entonces, la media y la varianza de T son, respectivamente:

$$E(T) = e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$

$$Var(T) = e^{(2\mu + \sigma^2)}(e^{\sigma^2} - 1)$$

A partir de (4.1) se obtiene el cuantil de orden p ,

$$t_p = e^{(\mu + z_p \sigma)}$$

Donde z_p representa el cuantil de orden p de una variable normal estándar (0,1). En particular, la mediana es:

$$t_{0,5} = e^{(\mu + z_{0,5} \sigma)} = e^\mu$$

y la moda es, $M_o = e^{(\mu - \sigma^2)}$

Distribución log-normal con tres parámetros

En algunas ocasiones ciertas variables no están definidas sino a partir de cierto valor límite δ , esto es $t > \delta$. Este límite en realidad es un parámetro, que al ser incorporado al modelo log-normal básico da lugar a la distribución log-normal con tres parámetros Sarabia J.M et al. (2007), cuya función de densidad es:

$$f(t) = \frac{1}{(t - \delta)\sigma\sqrt{2\pi}} e^{-\frac{(\ln(t-\delta) - \mu)^2}{2\sigma^2}} \quad t > \delta$$

Donde $\mu \in \mathbb{R}$ es un parámetro de localización (location), $\sigma > 0$ de escala (scale) y $\delta \in \mathbb{R}$ es la vida mínima o umbral (threshold) de las unidades en estudio. La función de distribución es ahora,

$$F(t) = \Phi\left(\frac{\ln(t - \delta) - \mu}{\sigma}\right) \quad t > \delta$$

La media de esta distribución viene dada por,

$$E(T) = \delta + e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$

La varianza de la distribución log-normal con tres parámetros, es la misma varianza de la distribución log-normal con dos parámetros. Lo que si cambia es el cuantíl de orden p ($0 < p < 1$), definido por,

$$t_p = \delta + e^{(\mu + z_p\sigma)}$$

Se debe tener alguna precaución al utilizar la distribución log-normal con tres parámetros en algunas aplicaciones, por la dificultad que a veces se presenta al estimar δ cuando la muestra es pequeña, Azorín F y Palacios M.J. (1975). En general una subestimación de δ conduce a varios sesgos en la estimación de la media de T, Koch y Link (1970), pero su sobreestimación puede dar lugar a límites de confianza demasiado próximos, por ser demasiado pequeña la varianza de los logaritmos.

4.5.1.4 Distribución de Herz.

La distribución estadística de Herz, fue introducida por Herz (1996, 1997, 1998) y utilizada por Trujillo (1997) para describir el envejecimiento de las tuberías de una red de agua. El modelo se basa en un modelo matemático de supervivencia desarrollado en la Universidad de Karlsruhe (Alemania).

En este modelo la red se divide en cohortes de tuberías, es decir, en grupos de tuberías instaladas el mismo año, del mismo material y otras características comunes que afectan a su rendimiento con el paso del tiempo. La distribución Herz fue desarrollada específicamente para valorar el envejecimiento de los distintos elementos básicos de la infraestructura de una red (conductos, válvulas, tanques de almacenamiento y estaciones de bombeo) y tiene la característica de que la tasa de fallo / renovación, aumenta con la edad cada vez más, antes que en forma gradual, tendiendo asintóticamente a un valor frontera. Lo que el autor llama la *tasa de fallo o de renovación*, en términos estadísticos es la función de riesgo para el tiempo de servicio de una tubería. La tubería se sustituirá cuando su vida útil haya caducado. La tasa de renovación está dado por:

$$h(t) = \frac{be^{b(t-c)}}{a+e^{b(t-c)}} \quad \text{Para } t \geq c$$

donde los valores a, b y c son los 3 parámetros del modelo (factor de edad a, factor de fallo b y el tiempo de resistencia c) que pueden ser estimados empíricamente durante los últimos períodos de instalación y por tipos específicos de tubería. Cuando se utiliza este modelo, en la estimación de la supervivencia de los tubos, se debe tener en cuenta la opinión de los expertos, es decir, la de los gestores e ingenieros de la red (Herz, 1996). La función de envejecimiento o de fiabilidad de los tubos, se debe establecer para cada grupo de tuberías. El modelo también predice la vida residual (es decir, vida útil restante del tubo) para cada cohorte de tuberías, que se puede utilizar para estimar las necesidades de rehabilitación de las tuberías.

Varias de las principales ciudades europeas (no sólo en Alemania) han utilizado la distribución Herz para la planificación de la renovación y la rehabilitación de las tuberías. El modelo se recoge en un software de fácil uso, denominado KANEW que es un proyecto de investigación patrocinado por la American Water Works Association Research Foundation (AWWARF), con la autoría de Deb et al. (1998), quienes lo implementaron (como parte del proyecto en un estudio piloto), con cinco (5) casos en los EE.UU. Los autores llegaron a la conclusión de que KANEW es útil para evaluar las futuras necesidades de rehabilitación de los conductos, pero algunos aspectos del modelo se deben mejorar mediante el desarrollo de mejores métodos de estimación de los parámetros de la función de supervivencia, utilizando registros tanto de la operación (funcionamiento) de la red y sus tuberías como de su mantenimiento.

La distribución Herz (1996), está caracterizada por las siguientes funciones dependientes de la edad que están matemáticamente relacionadas, Función de densidad del tiempo de vida $f(t)$, de las tuberías

$$f(t) = \frac{(a + 1)e^{b(t-c)}}{[a + e^{b(t-c)}]^2} \quad \text{si } t > c$$

Función supervivencia

$$R(t) = 1 - F(t) = \frac{a + 1}{a + e^{b(t-c)}} \quad \text{si } t > c$$

Estas funciones, igual que la tasa $h(t)$ de renovación o de fallo de la distribución de Herz, también dependen de los mismos 3 parámetros a , b y c , anteriormente definidos. Si $a = 0$, la distribución de Herz se convierte en una distribución exponencial con ratios de fallo constantes a lo largo del tiempo. El factor de fallo b es el ratio final de fallo a edades muy elevadas. Por encima de un tiempo máximo de resistencia, no hay rehabilitación. Estos parámetros dependientes de la edad, deben de ser calculados para situaciones específicas.

La distribución de Herz, tiene enormes ventajas computacionales y aparece como el modelo más apropiado para el estudio del deterioro de la infraestructura de los elementos básicos de una red de suministro de agua, según Baur y Herz (2002). Pero también tiene desventajas y básicamente son las siguientes:

- KANEW no permite el análisis individual de los tubos, ya que no se incluye la estructura de covariables en el modelo. La función de fiabilidad o de envejecimiento se especifica para cada tipo de tuberías (cohorte), no para tubos individuales. Esto implica que el modelo sólo se debe utilizar cuando se analizan las estrategias y las necesidades de rehabilitación para toda la red de distribución.

- KANEW no considera factores importantes como la capacidad hidráulica, la calidad del agua y el análisis de fiabilidad de la red dentro del análisis de las necesidades de rehabilitación / renovación de los tubos. El modelo asume que estas propiedades están en función de la edad (para cada cohorte), y que están explicadas por la función de densidad (de probabilidad) de su vida de servicio.

- Los parámetros en la distribución Herz se basan en las tasas históricas de renovación y no en las tasas históricas de roturas. Las tasas de renovación reflejan tanto las políticas de rehabilitación del pasado como las condiciones técnicas y económicas de la época. Cómo lo más probable es que las políticas de rehabilitación sean susceptibles de cambio; lo lógico es que los parámetros tengan que ser modificados con el fin de que reflejen estas futuras normas y políticas.

4.5.1.5 Representación gráfica de los modelos paramétricos

Numerosos autores han estudiado las técnicas gráficas para la selección de modelos paramétricos, que mejor se ajusten (adecúen) a la observaciones disponibles en un estudio de fiabilidad. Se destacan entre otros los estudios de Chernoff y Lieberman (1954), Kimball (1960) y Nelson (1969,1972,1979,1982,1990), este último es quien más ha trabajado sobre el tema.

Estas técnicas, que durante mucho tiempo se han utilizado en el análisis tanto de datos completos como en el análisis de datos con censura (múltiple y única), han demostrado ser herramientas empíricas simples, efectivas y de gran potencia. Se destacan *el papel probabilístico* y *el papel de riesgo*, que para su utilización requieren *de antemano de una estimación no paramétrica* (Kaplan-Meier o tablas de vida) de la función de fiabilidad de los tiempos de fallo de una unidad.

Un método general para desarrollar el papel probabilístico se puede establecer y es válido para aquellos modelos paramétricos que tienen la estructura

$$Y = \mu + \sigma W$$

Donde $Y = \ln T$ y W es una variable de distribución conocida (valores extremos, normal, logística).

Los modelos con la estructura anterior permiten establecer una relación lineal entre $\ln T$ y $W = F_W^{-1}[F_T(t)]$, ya que en estos modelos se verifica que

$$F_T(t) = F_W(w)$$

Martínez (1998), por lo que

$$F_W^{-1}[F_T(t)] = F_W^{-1}[F_W(w)] = W$$

Por lo tanto la estructura original $Y = \mu + \sigma W$, se convierte en la importante relación lineal,

$$\ln T = \mu + \sigma F_W^{-1}[F_T(t)]$$

que es el fundamento de la técnica de los gráficos de probabilidad.

Por lo tanto, siempre que podamos obtener la inversa de la función de distribución de W , podremos determinar mediante una comprobación gráfica, la linealidad existente entre $F_W^{-1}[F_T(t)]$ y $\ln T$.

Esta verificación, se realiza usualmente a partir de una estimación previa (no paramétrica) de la función de distribución $F_T(t)$, en los tiempos de fallo, esto es con $\hat{F}_T(t)$.

Formalmente, si $t_1 < \dots < t_k$ son los diferentes tiempos de fallo observados en una muestra de tamaño n de una población homogénea, el mejor modelo para los datos (el más adecuado), es aquel cuya nube de puntos de la forma

$$(Y, W) = (\ln t_j, F_W^{-1}[\hat{F}_T(t_j)]) \quad j = 1, \dots, k$$

se ajusta mejor a una línea recta.

4.5.2 Modelos no Paramétricos

En ocasiones puede resultar ventajoso, o incluso necesario, comenzar el análisis con métodos no paramétricos, pues éstos no requieren de grandes supuestos previos sobre el modelo de las observaciones. Los modelos no paramétricos, son métodos analíticos y gráficos que permiten interpretar los datos obtenidos, sin la distorsión que podría causar la elección de un modelo paramétrico subyacente no demasiado acertado.

En los estudios no paramétricos, no se asume ningún tipo concreto de modelo probabilístico para los tiempos de fallo y las funciones básicas (fiabilidad, riesgo) se estiman directamente de los datos. En algunos casos, estos métodos no paramétricos serán suficientes para realizar el análisis de los datos. Sin embargo, en otras circunstancias, son un paso intermedio hacia un modelo más estructurado (paramétrico), que permita profundizar más en el análisis de las observaciones.

En nuestro trabajo vamos utilizar tres de ellos, las *Tablas de Vida*, el estimador *Kaplan-Meier* y el método de *Nelson Aalen extendido (NAE)*.

4.5.2.1 Tablas de vida

Las tablas de vida, Bohmer, (1912) tienen como objetivo describir y establecer previsiones sobre la mortalidad, fiabilidad o supervivencia de una

población de interés, a partir de la consideración de una cohorte, o conjunto de datos procedentes de un estudio (o ensayo clínico), a los cuales se les hace un seguimiento en un período de tiempo determinado, comprobando si se registra en cada uno de sus miembros o elementos, la presencia o ausencia de una característica o evento de interés (Aprendizaje de un método, recuperación física de un paciente, mortalidad de un enfermo, fallo de un dispositivo, etc.) en la población. La validez de éste método exige que la distribución del tiempo de fallo de todos los individuos, censurados y no censurados, sea la misma.

Este es uno de los métodos más clásicos y directos para describir la fiabilidad de una muestra a través de la llamada *tabla de Supervivencia o Actuarial*, la cual no es más que una tabla de frecuencias mejorada y ampliada. A partir de ella, es posible hacer una primera estimación sobre los comportamientos de las funciones de supervivencia $R(t)$, distribución $F(t)$, densidad $f(t)$, y tasa de fallo $h(t)$.

La distribución de los tiempos de fallo se divide en un determinado número de intervalos que denotamos $(t_{i-1}, t_i]$. Para cada intervalo se registra el número de observaciones o dispositivos que han entrado en buen estado n_i (número que entra en el intervalo), el número de los que han fallado d_i (número de eventos terminales), y el número de observaciones perdidas o censuradas en r_i (número que sale en el intervalo). Se calcula a partir de ellos el número de expuestos al riesgo, asumiendo que las pérdidas se producen homogéneamente a lo largo del mismo, su número promedio es $n'_i = n_i - 0.5 r_i$. La probabilidad de fallo es la proporción $p_i = d_i/n'_i$, y la de supervivencia es $q_i = 1 - p_i$.

Casi todas las tablas de vida presentan una estructura más o menos estándar con la descripción detallada de las siguientes columnas, Abaurrea, J. y Cebrián, A. (1998):

$(t_{i-1}, t_i]$: identifica los extremos del intervalo de tiempo i -ésimo. El extremo inferior del primer intervalo, t_0 , suele ser 0; si el ensayo no termina hasta que se observe el fallo de todos los individuos, el extremo superior del último intervalo es $t_{s+1} = \infty$.

\bar{t}_i : punto medio del intervalo i -ésimo.

l_i : número de individuos que abandonan el experimento y cuyo tiempo de censura, pertenece al intervalo i -ésimo.

c_i : número de individuos cuya respuesta o fallo no se ha observado al finalizar el instante final t_i del intervalo i -ésimo.

El tratamiento en el proceso de estimación de ambos tipos de observaciones censuradas, l_i y c_i , es idéntico; por ello, en la mayoría de las tablas no se hace la distinción anterior y se define una única columna r_i ,

r_i : número de censuras en el intervalo i -ésimo. Desde luego, $r_i = l_i + c_i$

d_i : número de fallos en el intervalo $(t_{i-1}, t_i]$

n_i : número de individuos en riesgo al inicio del intervalo i -ésimo (que permanecen vivos y en el estudio al comienzo de t_{i-1}). La siguiente relación facilita su cálculo,

$$n_i = n_{i-1} - (d_{i-1} + r_{i-1})$$

n'_i : número estimado de individuos expuestos a riesgo durante el intervalo i -ésimo. Si en el intervalo no hay observaciones censuradas, $n_i = n'_i$; en otro caso, la hipótesis usual es suponer que las observaciones censuradas ocurridas durante el intervalo se distribuyen en el uniformemente. Por ello, en media, los individuos con observación censurada están expuestos a riesgo durante la mitad de la duración del intervalo, de modo que,

$$n'_i = n_i - \frac{1}{2} r_i$$

p_i : proporción de fallo en el intervalo i -ésimo. Este valor estima la probabilidad de fallo en ese intervalo, condicionada a que el individuo no había fallado al comienzo del mismo. Si en el intervalo no hay observaciones

censuradas, el estimador natural de esta probabilidad es d_i/n_i si las hay, este estimador tiende a subestimar dicha probabilidad, ya que es posible que alguno de los individuos censurados en el intervalo muera antes de finalizar el mismo. Por esta causa, es necesario realizar un tipo de ajuste; la alternativa habitual consiste en sustituir n_i por el número estimado de individuos expuestos a riesgo n'_i

$$p_i = \frac{d_i}{n'_i}$$

La validez de este ajuste, en cierta medida arbitrario, depende de las características de los procesos de censura y de fallo. Breslow y Crowley (1974) establecieron que, bajo el esquema de censura aleatoria, este estimador era inconsistente y sesgado. Sin embargo, cuando la proporción de censura no es excesiva y se distribuye de forma homogénea, si los intervalos no son muy amplios y los valores de n_i no son demasiado pequeños, el comportamiento de este estimador es aceptable.

q_i : proporción de supervivencia en el intervalo i -ésimo. Este valor estima la probabilidad de que un individuo sobreviva al instante t_i , dado que estaba vivo al inicio del intervalo i -ésimo. Se define,

$$q_i = 1 - p_i$$

$\hat{P}(t_i) = \hat{R}_i$: proporción de supervivencia en el instante t_i del ensayo. Este valor estima $R(t_i) = P(T > t_i)$. Si un individuo sobrevive hasta el inicio del intervalo $(i + 1)$ -ésimo, implica que dado que ha sobrevivido hasta el comienzo del intervalo i -ésimo, no falla durante el intervalo i -ésimo; así,

$$\hat{P}(t_i) = \hat{R}_i = q_i \hat{P}(t_{i-1})$$

Aplicando reiteradamente esta relación y dado que $\hat{P}(t_0)$ es igual a 1, se obtiene,

$$\hat{P}(t_i) = \hat{R}_i = q_i q_{i-1} \dots q_1 .$$

Este estimador de la función de supervivencia se llama *estimador actuarial*. En los instantes que no son extremo de un intervalo, es habitual calcular su valor mediante interpolación lineal entre los valores de los correspondientes extremos.

$f(\bar{t}_i)$: tasa o proporción de fallo en el intervalo i -ésimo por unidad de tiempo. Este valor estima la función de densidad de T en el punto medio del intervalo de anchura b_i .

$$f(\bar{t}_i) = \frac{\hat{P}(t_{i-1}) - \hat{P}(t_i)}{b_i} = \frac{\hat{P}(t_{i-1}) \cdot p_i}{b_i}$$

$h(\bar{t}_i)$: tasa instantánea condicional de fallo correspondiente al punto medio del intervalo i -ésimo. Con este valor se estima la función de riesgo en \bar{t}_i

$$h(\bar{t}_i) = f(\bar{t}_i) / \hat{P}(t_i) = f(\bar{t}_i) / \hat{R}_i$$

La estimación de la función de riesgo también puede expresarse como

$$h(\bar{t}_i) = \frac{2p_i}{b_i(1 + q_i)} = \frac{d_i}{b_i(n'_i - d_i/2)}$$

Precisión de las estimaciones.

Los valores de q_i , p_i y $\hat{P}(t_i) = \hat{R}_i$ son estimaciones sujetas a la variabilidad inherente al proceso de muestreo, por lo que deben completarse con información relativa a su precisión. Bajo determinadas hipótesis sobre los mecanismos de censura es posible, aunque complicado, deducir estimaciones de sus varianzas. Por esta razón, aunque la metodología de las tablas de vida clínicas es antigua, el estudio teórico de las propiedades estadísticas de sus estimadores es reciente y está aún por completar. En este apartado se presentan algunas de las propiedades y resultados más utilizados. La mayor parte de estos resultados se han obtenido para el caso de muestras completas, pero se suelen generalizar y aplicar también al caso de muestras censuradas.

La estimación mas empleada de la varianza de $\hat{P}(t_j) = \hat{R}_j$ es la propuesta por Greenwood (1926)

$$Var[\hat{R}_j] \approx \hat{R}_j^2 \sum_{t=1}^j \frac{p_t}{q_t n'_t} = \hat{R}_j^2 \cdot \sum_{t=1}^j \frac{d_t}{n'_t(n'_t - d_t)}$$

Esta estimación, resultado de una aproximación asintótica (método *Delta*, *Meeker and Escobar (1998)*) es *razonable* cuando el valor esperado de n_j no es demasiado pequeño y requiere, si la proporción de censura en la muestra es importante, que el número de intervalos considerados no sea muy pequeño. La formula de Greenwood tiende a subestimar la varianza de $\hat{P}(t_j)$, especialmente en los intervalos de la cola derecha de la distribución donde el valor esperado de n_j suele ser pequeño. No obstante en esos casos su cálculo no es adecuado ya que la distribución de $\hat{P}(t_j)$ suele ser muy sesgada y, en consecuencia, la varianza no es una buena medida de precisión de la estimación.

4.5.2.2 El estimador de Kaplan- Meier (KM) de la función de fiabilidad.

El impulso de las técnicas de estimación no paramétricas con datos censurados es relativamente reciente. Se inicia con los aportes de Kaplan y Meier en 1958, quienes publican algunos resultados obtenidos en ese momento para observaciones censuradas a la derecha y añaden un estudio de las propiedades básicas de un nuevo estimador, que se conocerá más tarde con el nombre de sus creadores. De los métodos no paramétricos, desarrollados para estimar la función de fiabilidad con datos no agrupados en presencia de censura, el más utilizado es *el estimador producto límite de Kaplan-Meier* (Kaplan and Meier, (1958)). Dicho método descompone la supervivencia o fiabilidad de una unidad al cabo de t años, en un producto de probabilidades condicionadas, que deben ser previamente estimadas, antes del cálculo del estimador.

La diferencia fundamental entre el método actuarial y el de KM (que también es una función escalonada), radica en que las estimaciones de Kaplan–Meier están basadas en tiempos de fiabilidad individuales, sin agrupar, mientras que en la estimación por tablas de vida los individuos han sido previamente agrupados en intervalos. La ventaja del método de KM respecto a las tablas de vida, es que las estimaciones resultantes por este método no dependen de cómo se agrupan los datos en los intervalos. De hecho, Kaplan–Meier se podría considerar como un caso particular del método actuarial. La estimación producto-límite es definido de la siguiente manera (Lawless,1982):

Suponga que hay observaciones de n unidades o individuos y K ($K \leq n$) distintos tiempos $t_1 < t_2 < \dots < t_k$ en los cuales ocurren fallos. En los n individuos se permite la posibilidad de que haya más de un fallo en t_j , denotamos por d_j el número de fallos en t_j . Adicional a los tiempos de vida t_1, \dots, t_k existen a su vez tiempos de censura t_i^+ , para aquellos individuos en los que el tiempo de vida no es observado. La estimación producto-límite de $R(t)$ para la duración t , es una función escalonada, que se calcula como el producto de uno menos el riesgo existente hasta el período t :

$$\hat{R}(t) = \prod_{j: t_j < t} \frac{n_j - d_j}{n_j}$$

Donde d_j representa el número de fallos ocurridos en el momento t_j y n_j es la población superviviente en el momento t_j ó el número de individuos en riesgo en t_j . Si se diese el caso de alguna observación censurada cuyo valor coincidiera con un tiempo de fallo, se hace la hipótesis de que la unidad censurada ocurre inmediatamente después del tiempo de fallo y, en consecuencia, las unidades censuradas en ese instante se contabilizan como unidades en riesgo.

El estimador KM de $R(t)$ da una estimación puntual o un único valor para esta función en cualquier instante t . Por lo tanto, si se desea tener una medida de la precisión de este estimador en diferentes instantes de tiempo o sobre

diferentes muestras, es necesario contar con un buen estimador de la varianza del estimador KM, el cual viene dado por la fórmula de Greenwood (1926),

$$Var [\hat{R}(t)] = \hat{R}(t)^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}$$

Utilizando la normalidad asintótica de $\hat{R}(t)$, se puede construir el siguiente intervalo de confianza aproximado para $\hat{R}(t)$, a un nivel del $100(1-\alpha)\%$

$$\hat{R}(t) \pm Z_{1-\frac{\alpha}{2}} ee(\hat{R}(t))$$

donde $Z_{1-\frac{\alpha}{2}}$ es el cuantil correspondiente a la distribución normal estándar y $ee(\hat{R}(t)) = \sqrt{Var(\hat{R}(t))}$ es el error estándar de estimación del estimador KM de $R(t)$, que se calcula con la anterior fórmula de Greenwood.

Algunas propiedades del estimador Kaplan-Meier

En la literatura estadística, las propiedades deseables en un buen estimador tienen que ver con que sea Insesgado, Consistente, Eficiente y Suficiente, y el estimador KM, tiene varias de estas propiedades, esto es: *Consistencia* (Efron, 1967) y *Eficiencia* (Wellner, 1982), aparte de ser un estimador de *máxima verosimilitud* para datos censurados, (Peterson 1977), (Johansen, 1978), (Oakes, 1986), (Wang, 1996), y además es normalmente asintótico (Breslow y Crowley (1974), han examinado propiedades asintóticas de este estimador), lo que se traduce en una gran facilidad de cálculo y en poder ser utilizado casi con exclusividad en problemas con datos censurados a la derecha; aparte de que cuando las estimaciones de $R(t)$ se hacen con datos completos (ausencia de censura), su expresión coincide con la del estimador no paramétrico de la función empírica de fiabilidad (FES).

Los años setenta y ochenta, fueron pródigos en la obtención de resultados relativos a distintas propiedades del estimador de Kaplan-Meier, (Miller, 1981).

Resumiendo, el estimador KM con grandes muestras, tiene muchas de las propiedades deseables en un estimador, pero estas mismas propiedades con pequeñas muestras ya no son tan robustas. En particular, es sesgado para muestras finitas y la magnitud del sesgo es inversamente proporcional al tamaño de la muestra, Gross y Clark (1975). De otra parte la eficiencia asintótica del estimador KM, es inferior a la del estimador paramétrico de máxima verosimilitud, si el nivel de censura es alto o la supervivencia está próxima a cero, Miller (1983).

De otra parte, cuando los datos están doblemente censurados en la estimación no paramétrica de $R(t)$, se emplea con alguna frecuencia el estimador propuesto por Turnbull (1974), el cual tiene propiedades similares a las de KM, entre ellas la autoconsistencia, Gu y Zhan (1993), Mykland y Ren, (1996). En 1976, el propio Turnbull presentó una versión más general de su estimador para datos arbitrariamente agrupados, censurados y truncados.

En cierta medida se obtuvieron propiedades paralelas a las de la distribución empírica, que ya estaban perfectamente desarrolladas en el libro de Billingsley (1986).

4.5.2.3 El estimador de Kaplan- Meier ponderado (KM_P)

El problema que se presenta con el estimador de Kaplan-Meier (KM) cuando la base de datos del estudio contiene una *fuerte censura* o altísimo porcentaje de observaciones con *censura*, como ocurre en este estudio, es que sus estimaciones por lo regular, no sólo tienen la tendencia a sobrestimar la fiabilidad o supervivencia de las unidades o individuos en estudio con un alto margen de sesgo, sino que van acompañadas de muy poca variabilidad de las estimaciones. Las estimaciones de KM obtenidas en realidad son estimaciones sesgadas (sobrestimaciones), porque el método parte del supuesto de que *los dispositivos con censura, se conservan vivos (sin averías) hasta el siguiente fallo* (incluso después de largos períodos de tiempo). Nada más lejos de la realidad, pues esto es como suponer que el paso del tiempo de un año al siguiente no tiene ningún efecto, ni acción sobre las observaciones (*personas*

en el caso de estudios y tratamiento de enfermedades y *tramos de tubería* en el caso que nos ocupa), razón por la cual, cobra importancia la necesidad de reducir el sesgo que producen las estimaciones de KM con datos censurados.

Para corregir en principio esta aparente debilidad, se propone una modificación al método original de estimación de KM y que consiste en ponderar o acompañarle a las observaciones con censura, un factor o *tasa de no censura*, propuesta por Bahrawar et al. (2005), quienes aplicaron esta metodología con “datos de Transplante de Corazón en Stanford”, donde la variable respuesta era el tiempo de supervivencia del paciente después del transplante. La tasa de censura de la base de datos de Stanford era del 27%.

El método de KM_p propuesto, responde a esta situación y considera la censura como una parte importantísima del análisis, que se involucra directamente en la fórmula de su factor de ponderación W_j , conocido como *tasa de no censura* y que está definido por:

$$W_j = \frac{n_j - c_j}{n_j} \quad \text{con } 0 \leq W_j \leq 1$$

De modo que si $W_j = 1$, no hay censura en el instante t_j (pues $c_j = 0$), pero si $W_j < 1$, en el instante t_j hay al menos una censura.

El estimador KM_p , se define entonces como:

$$\hat{R}(t) = \begin{cases} 1 & \text{si } t = 0 \\ \prod_{j: t_{(j)} \leq t} W_j \cdot \left(\frac{n_j - d_j}{n_j} \right) & \\ 0 & \text{si } t \geq t_n \end{cases}$$

En este caso las estimaciones $KM_p = \hat{R}(t)$ estarían definidas para todo $t \geq 0$, alcanzando el valor 0, aún en el caso de que la última estimación sea censurada, ya que Bahrawar et al. (2005), sugieren considerar la última observación censurada como un fallo, para garantizar su definición en todo

instante. Si no hay censura, $\hat{R}(t)$ coincide con la función empírica de supervivencia FES, ya que $W_j=1$ y $\hat{R}(t)$ coincidiría con KM, que a su vez se convierte en la FES, si todos los datos son completos.

La Función empírica de Fiabilidad (FES)

Si se tienen n tiempos de fallo ordenados, $t_1 \leq t_2 \leq t_3 \dots \leq t_n$, donde no hay censura, el número de unidades que sobreviven el instante t_i es $n - i$.

Por lo tanto, una plausible estimación no paramétrica de la función de fiabilidad $R(t)$ en t_i , sería simplemente la proporción de unidades que sobreviven en el instante t_i

$$\hat{R}(t_i) = \frac{n - i}{n} \quad i = 1, 2, \dots, n$$

En consecuencia habría una probabilidad cero de sobrevivir más allá de t_n . Como es improbable que ningún valor muestral alcance el tiempo hasta el fallo más alto, esta expresión tiende a subestimar la fiabilidad de la componente. En una muestra de datos sin censura, el estimador KM coincide con este estimador, usualmente conocido como *función empírica de supervivencia o fiabilidad (FES)*

4.5.2.4 El estimador Nelson-Aalen (NA) de la función de riesgo acumulado

Fue propuesto por primera vez en el ámbito de la fiabilidad por Nelson, W. (1969 y 1972), y redescubierto independientemente por Altschuler (1970), quien lo obtuvo utilizando técnicas de procesos de conteo con animales. Los detalles de estas técnicas pueden encontrarse en Andersen *et al.*(1993), Fleming y Harrington (1991), Lawless (1982).

Dado que la función de riesgo acumulado está definida por $H(t) = \Lambda(t) = -\ln R(t)$, un estimador natural de esta función se define también por

$$\hat{H}(t) = -\ln \hat{R}(t)$$

Donde $\hat{R}(t)$ es el estimador KM de la función $R(t)$. Otro posible estimador de esta misma función $H(t) = \Lambda(t)$, definido por Nelson (NA), es *la función empírica de riesgo acumulado*,

$$\tilde{H}(t) = \tilde{\Lambda}(t) = \sum_{j, t_j \leq t} \frac{d_j}{n_j}$$

Donde d_j representa el número de fallos ocurridos en el instante t_j y n_j es el número de individuos en riesgo en t_j .

El cociente d_j/n_j , proporciona una estimación de la probabilidad condicionada de que una unidad que sobrevive hasta justo antes del instante t_j , falle en el instante t_j . Esta es la cantidad básica a partir de la cual se construyen los estimadores de la función de fiabilidad $R(t)$ y la función de impacto $H(t)$. En realidad, $\tilde{H}(t)$ lo que hace es recoger (para todo $t_j \leq t$) las estimaciones puntuales d_j/n_j de la función de riesgo, $h(t_j) = h_j$ en cada instante de fallo t_j y acumularlas.

A partir de la relación logarítmica entre $\Lambda(t)$ y $R(t)$, se obtiene un estimador alternativo $\tilde{R}(t)$, de la función de supervivencia, conocido comúnmente como *el estimador de Nelson-Aalen*, de la función de supervivencia, cuya relación es,

$$\tilde{R}(t) = e^{-\tilde{\lambda}(t)} = e^{-\sum_{j:t_j \leq t} \left[\frac{d_j}{n_j} \right]}$$

Este estimador ha sido tema de discusión (para el caso de una variable continua) por parte de diferentes autores, entre ellos Altschuler (1979), Breslow y Crowley (1974), Nelson (1972b) y Efron (1977). La conclusión es que, en el caso de que T sea una v.a. continua, $\hat{H}(t)$ y $\tilde{\lambda}(t)$ son asintóticamente equivalentes y con la excepción de valores altos de t , donde las estimaciones son más inestables, la diferencia entre ambos será, por general pequeña. En la práctica $\tilde{\lambda}(t)$ es una aproximación lineal de primer orden de la función $\hat{H}(t)$. Aunque $\tilde{\lambda}(t)$ es más sencillo de calcular, no hay razones de fondo para preferir uno u otro.

Estas estimaciones son de gran utilidad en la construcción de *gráficas*, para evaluar la selección de una determinada familia paramétrica de distribuciones, cuando se trata de modelizar la distribución del tiempo de vida de una unidad y realizar unas primeras estimaciones de los parámetros del modelo seleccionado, Nelson (1982), Lawless (1982).

4.5.2.5 Estimadores no paramétricos de la función de fiabilidad para datos truncados a la izquierda y censurados a la derecha

En lo que sigue, nos planteamos el problema de obtener un estimador para la función de fiabilidad $R(t) = P(T > t)$. A partir de dicho estimador, es posible obtener estimaciones para otras funciones de interés como son la función de distribución $F(t) = 1 - R(t)$ o la función de impacto (riesgo acumulado) $H(t) = -\ln(R(t))$.

La primera referencia sobre un estimador no paramétrico de la función de distribución F para datos censurados y truncados se encuentra en la revista *Journal of the Royal Statistical Society* y está firmada por Turnbull (1976), quien en su trabajo utiliza una idea sobre autoconsistencia para construir un algoritmo que obtiene el estimador de máxima verosimilitud de F :

Las primeras propiedades del estimador de Turnbull, que resulta ser un estimador de tipo límite producto, al igual que los de Kaplan-Meier y Linden-Bell, fueron estudiadas por Tsai, Jewell y Wang (1987), quienes derivaron su comportamiento asintótico, y por Lai y Ying (1991), que estudiaron la consistencia fuerte uniforme y la convergencia débil del proceso para una versión modificada del citado estimador bajo hipótesis más generales.

En cuanto a las representaciones casi seguras que descomponen al estimador en una suma de variables i.i.d. más un término de error, se deben a Gijbels y Wang (1993), Zhou (1996) y Zhou y Yip (1999).

Los estimadores para la función de supervivencia y otras funciones, basados en muestras censuradas a la derecha, pueden modificarse para manipular muestras truncadas a la izquierda y censuradas a la derecha.

En esta situación asociamos al j -ésimo individuo una edad aleatoria X_j que indica el instante de entrada en el estudio, y un tiempo Y_j de fallo o censura. Como en el caso de datos censurados a la derecha, se define $t_1 < t_2 < \dots < t_k$, como los diferentes tiempos de fallo, siendo d_j el número de fallos en el instante t_j . Las demás cantidades necesarias para calcular los estadísticos en el caso de censura a la derecha (Kaplan-Meier o Nelson-Aalen) consisten en el número de individuos en riesgo inmediatamente antes del instante t_j , llamémosle r_j . En el caso de censura a la derecha r_j coincide con el número de individuos que están en estudio en el instante 0 y que tienen un tiempo de participación de al menos t_j .

Para los datos truncados a la izquierda, redefinimos estas cantidades r_j como el número de individuos que entraron en estudio antes del tiempo t_j y que tienen un tiempo de participación de al menos t_j , esto es, r_j es el número de individuos i , que verifican que su tiempo de fallo t_i es tal que $x_j < t_i \leq t_j$.

Usando esta nueva definición de r_j , podemos obtener los estimadores de la función de fiabilidad (por ejemplo) para el caso de datos truncados. Sin embargo, hay que tener cuidado con las interpretaciones de esos estimadores. Por ejemplo, el estimador producto límite de la función de supervivencia

(Turnbull) en el tiempo t es ahora un estimador de la supervivencia por encima de t , condicionado a la supervivencia al menor de los tiempos de entrada X . Es decir, estimamos $P(T > t / T \geq X)$

De modo similar, el estadístico de Nelson– Aalen estima la integral de la función de riesgo en el intervalo (X, t) .

El estimador de Turnbull

Turnbull (1976) y Tsai, Jewell y Wang (1987) definen el estimador de la función de fiabilidad o de supervivencia $S(t) = R(t)$ para el modelo de truncamiento a la izquierda y censura a la derecha mediante la siguiente expresión

$$\hat{S}(t) = \prod_{i=1}^n \left(1 - \frac{\mathbf{1}_{\{Y_i \leq t, \delta_i = 1\}}}{\sum_{j=1}^n \mathbf{1}_{\{X_j \leq Y_i \leq Y_j\}}} \right) = \prod_{i=1}^n \left(1 - \frac{d_i}{r_i} \right)$$

Donde d_i representa el número de fallos en el instante Y_i y r_i es el número de individuos en riesgo inmediatamente antes de ese mismo instante. Estas cantidades se calculan como se indica en la expresión anterior.

El anterior es el estimador no paramétrico de máxima verosimilitud de $S(t) = R(t)$ bajo este esquema muestral.

Es importante observar que el estimador de Turnbull en ausencia de truncamiento coincide con el estimador de Kaplan-Meier y en ausencia de censura se reduce al estimador de Linden-Bell (Linden-Bell (1971)) y Woodroffe (1985)). Cuando no hay censura ni truncamiento el estimador coincide con la función empírica de distribución (FES)

El estimador de Turnbull, puede presentar serios problemas en el caso de muestras pequeñas o grandes muestras con pocos valores truncados iniciales. Puede ocurrir que $d_i = r_i$ en algún y_i aunque éste no sea el estadístico de

mayor orden de la serie y_1, y_2, \dots, y_n . En ese caso $\hat{R}(t) = \hat{S}(t) = 0$, para $t \geq y_i$ independientemente de las observaciones que haya después, es decir aunque estemos observando supervivencias y fallos después de este punto. Claramente esta situación no es satisfactoria y es un problema exclusivamente introducido por el caso de truncamiento, que puede inducir pocos casos de individuos en riesgo en la cola de la izquierda.

Estimador de Nelson-Aalen extendido (NAE)

Pan y Chappell (1998) introducen el estimador de *Nelson-Aalen extendido* para la función de supervivencia en el caso de truncamiento a la izquierda y censura a la derecha con el fin de corregir el importante sesgo producido por la subestimación de la supervivencia que acabamos de mencionar.

En presencia de censura a la derecha, Nelson (1969) estima la función de riesgo acumulada $H(t)$ mediante la expresión

$$\tilde{H}(t) = \sum_{i=1}^n \frac{1_{\{y_i \leq t, \delta_i = 1\}}}{\sum_{j=1}^n 1_{\{y_i \leq y_j\}}}$$

que resulta en el estimador de Nelson para la función de supervivencia en la forma

$$\tilde{R}(t) = \tilde{S}(t) = e^{-\tilde{H}(t)}$$

Este estimador es conocido como estimador de Nelson-Aalen o estimador de Breslow, por los trabajos de Aalen (1976, 1978) y Breslow (1972). Fleming y Harrington (1984) recomiendan este estimador como alternativa al estimador no paramétrico de máxima verosimilitud (NPMLE) dado que el anterior tiene el menor error cuadrático medio con datos censurados cuando la verdadera probabilidad de supervivencia sea al menos 0.2.

Pan y Chappell (1998) extienden este estimador para muestras truncadas a la izquierda redefiniendo el número de individuos en riesgo que ahora se obtiene como

$$r_i = \sum_{j=1}^n 1_{\{x_j \leq y_i \leq y_j\}}$$

es decir, introducimos en el cálculo de esta cantidad la condición dada por la variable truncamiento.

Este estimador es denominado por Pan y Chappell Estimador de Nelson–Aalen Extendido, y es recomendado por estos autores frente al estimador NPMLE, porque soluciona el problema de subestimación dado por este último cuando hay truncamiento. Un caso extremo señalan los autores, sería cuando $d_1=r_1$, lo que llevaría a $\hat{R}(t) = \hat{S}(t) = 0$ para todo $t > y_1$. En cambio, $\tilde{R}(y_1^+) = \tilde{S}(y_1^+) = e^{-1} > 0$

. En general, como $\tilde{R}(t) = \tilde{S}(t) \geq \hat{R}(t) = \hat{S}(t)$, el problema de subestimación queda bastante “aliviado”, aunque no completamente solucionado. Además, este estimador tiene propiedades asintóticas similares al estimador NPMLE, siempre que T tenga distribución continua.

En resumen, el estimador de Nelson-Aalen extendido adopta la siguiente forma, para estimar el riesgo acumulado

$$\tilde{H}_e(t) = \sum_{i=1}^n \frac{1_{\{y_i \leq t, \delta_i = 1\}}}{\sum_{j=1}^n 1_{\{x_j \leq y_i \leq y_j\}}} = \sum_{y_i \leq t} \frac{d_i}{r_i}$$

Y esta misma expresión es la que se utiliza para estimar la función de supervivencia o de fiabilidad a través del estimador

$$\tilde{R}_e(t) = \tilde{S}_e(t) = e^{-\tilde{H}_e(t)}$$

4.5.2.6 Comparación de funciones de fiabilidad

Cuando se desea comparar la supervivencia de dos o más grupos de individuos puede utilizarse un test estadístico global, que responda a la pregunta: ¿Todos los grupos presentan la misma supervivencia?

Si el p-valor asociado a ese test es pequeño, lo que nos permite suponer que no todos los grupos son iguales, hay que plantearse una nueva pregunta: ¿Qué grupos son distintos?

Para responder a esa pregunta debemos realizar todas las comparaciones dos a dos, con un test estadístico apropiado, de modo que se pueda inferir, si las diferencias observadas por cada dos curvas de supervivencia pueden ser explicadas o no por el azar.

Una amplia variedad de contrastes no paramétricos, para comparar la igualdad de dos o más funciones de fiabilidad con datos censurados, han sido registrados en la literatura. Los más utilizados son:

El test de *Log Rank*, Mantel-Haenszel (1966), (también llamado test de riesgos proporcionales) es muy potente para detectar diferencias cuando los logaritmos de las curvas de supervivencia son proporcionales (lo que es equivalente a decir que los riesgos son proporcionales). Sin embargo, si las curvas de supervivencia se cortan o se cruzan, el test *Log Rank* tiene problemas para detectar diferencias. En esos casos es más útil el test de *Breslow*, (Breslow, 1970) (también llamado test de Gehan o test de Wilcoxon generalizado), que hace énfasis, de forma especial, en detectar diferencias cuando las curvas se cruzan al principio; por lo cual, este test no resulta adecuado para detectar diferencias a largo plazo. Un test intermedio entre los dos anteriores es el test de *Tarone-Ware*. (Tarone y Ware, 1977).

La mayor parte de estos contrastes, son aproximaciones con grandes muestras a la distribución chi-cuadrado, con el supuesto de que la censura y el tiempo de fallo son variables aleatorias independientes.

En relación al estudio de la potencia de estos contrastes, hay mucho material en la literatura, en distintas situaciones (con pequeñas muestras, cuando el fallo ocurre pocas veces, etc.), Peace y Flora (1978), Latta (1981), Kellerer y Chmelevsky (1983).

Cuando se introduce otra variable que define estratos, las comparaciones que se estudian son las que se basen en los factores o grupos definidos por los estratos existentes. Es el caso que se presenta, por ejemplo, cuando se compara el efecto de dos métodos de manufactura sobre tres líneas de producción.

Los anteriores contrastes sirven para comparar las hipótesis de la forma,

$$H_0 : R_1(t) = R_2(t) = \dots = R_r(t)$$

Frente a la alternativa

$$H_1 : R_l(t) \neq R_{l'}(t)$$

para todo $t > 0$ y para al menos un par de poblaciones l, l' diferentes.

Si $t_1 < t_2 < \dots < t_k$ son k tiempos de la muestra formada por la unión de los individuos de las r muestras. Si además suponemos que, en la muestra j ($j = 1, \dots, r$), ocurren d_{ij} fallos en t_i ($i = 1, \dots, k$) y que n_{ij} están en riesgo en el momento anterior a t_i

Entonces, se definen $d_i = \sum_{j=1}^r d_{ij}$ y $n_i = \sum_{j=1}^r n_{ij}$ como el total de fallos y el total de unidades en riesgo respectivamente en las r poblaciones en el instante t_i ($i = 1, \dots, k$). De modo que, para la j -ésima población el estadístico $Z_j(t)$ queda bien definido por la suma ponderada,

$$z_j(t) = \sum_{i=1}^k w(t_i) \left\{ d_{ij} - n_{ij} \left(\frac{d_i}{n_i} \right) \right\} \quad j = 1, \dots, r$$

Que corresponde a las diferencias entre el número observado de sucesos y su número esperado, bajo la hipótesis nula en la muestra j -ésima; donde la

función peso $w(t_i)$ que aparece en $z_j(t)$ es compartida por todos los r grupos y es la que regula la potencia de los distintos contrastes anteriormente mencionados (*Log Rank*, *Breslow*, etc.)

Las $z_j(t)$, así definidas son la base del estadístico de contraste de las hipótesis anteriormente referenciadas y que viene dado por la forma cuadrática,

$$\chi^2 = [z_1(t), \dots, z_{r-1}(t)] \Sigma^{-1} [z_1(t), \dots, z_{r-1}(t)]'$$

Donde Σ^{-1} es la matriz inversa de varianzas-covarianzas de las componentes seleccionadas. Si la hipótesis nula es cierta, este estadístico debe seguir una distribución ji-cuadrado con $r - 1$ grados de libertad.

4.5.3 Modelos Semiparamétricos

Varias aproximaciones han sido propuestas para modelizar los efectos de las variables explicativas o covariables sobre el tiempo de vida T . Una de estas modalidades corresponde a los modelos de riesgo proporcional, basados en el hecho de que, de las diferentes funciones que pueden describir la distribución de T , la más sencilla de modelizar es la función de riesgo. En consecuencia, una buena aproximación para analizar los efectos de las covariables sobre el tiempo de vida de una unidad, puede ser considerar la función de riesgo condicional de T dado un vector de covariables Z

4.5.3.1 Modelo de riesgo proporcional de Cox

El modelo de riesgos proporcionales introducido inicialmente por Cox (1972) es el modelo de regresión más utilizado en análisis de supervivencia, sin embargo no fue sino a partir del desarrollo del enfoque basado en los procesos de conteo, que este modelo logra su completa madurez, este enfoque ha permitido la verificación de los supuestos de riesgos proporcionales y el estudio de los residuos.

Algunos desarrollos y generalizaciones del modelo de Cox pueden verse en: Cox (1972,1975), Miller (1976), Buckley-James (1979), Tsiatis (1978a), Andersen y Gill (1982), Johansen (1983), Andersen et al. (1993), Therneau y Grambsch (2000).

El modelo de riesgos proporcionales, ofrece una representación más unificada de la relación entre el entorno y el proceso de fallos, dado que nos permite modelar no sólo la relación entre la tasa de fallo y el tiempo, sino también la posible relación con diferentes variables registradas para cada sujeto. Se trata por tanto de calcular la tasa de fallo o de mortalidad como una función del tiempo y de las variables pronóstico o de predicción.

Aunque la idea fundamental es la misma que en cualquier modelo de regresión, aquí la matemática necesaria para la estimación de los coeficientes del modelo se complica sensiblemente, y a pesar de que existen diferentes alternativas, el sistema más utilizado es el de riesgos proporcionales o modelo

de Cox. Éstos modelos se denominan *semiparamétricos*, debido a que asumen la forma paramétrica únicamente para el efecto de las variables pronóstico e incluyen una función de riesgo arbitraria básica (o de referencia) $h_0(t) = \lambda_0(t)$ con forma sin especificar.

Teniendo en cuenta lo anterior, el modelo de riesgo proporcional de Cox como modelo de predicción de fallos toma la forma general:

$$h(t; Z) = h_0(t)e^{b^T Z} = h_0(t)e^{b_1 z_1 + b_2 z_2 + \dots + b_p z_p}$$

Donde t es la edad alcanzada por la unidad (tubería); $h(t, Z)$ es la función de riesgo de la misma, que corresponde a la tasa instantánea de fallo (probabilidad de fallo en el instante $t + \Delta t$ dada la supervivencia hasta el instante t); $h_0(t)$ es una función arbitraria conocida como función de riesgo básica (o de referencia); $Z^T = (z_1, z_2, \dots, z_p)$ es la traspuesta de un vector de covariables (o covariatas) que actúa multiplicativamente sobre la función de riesgo básica; b es un vector (columna de p componentes) de coeficientes que se estiman por regresión o maximizando la verosimilitud parcialmente, esto es sin incluir la función $h_0(t)$.

La función de verosimilitud se maximiza aplicando técnicas numéricas de Newton Raphson, a partir de los datos disponibles. El concepto de riesgo proporcional se aprecia en la Figura 4.7, que muestra el efecto de cada covariable sobre la función de riesgo:

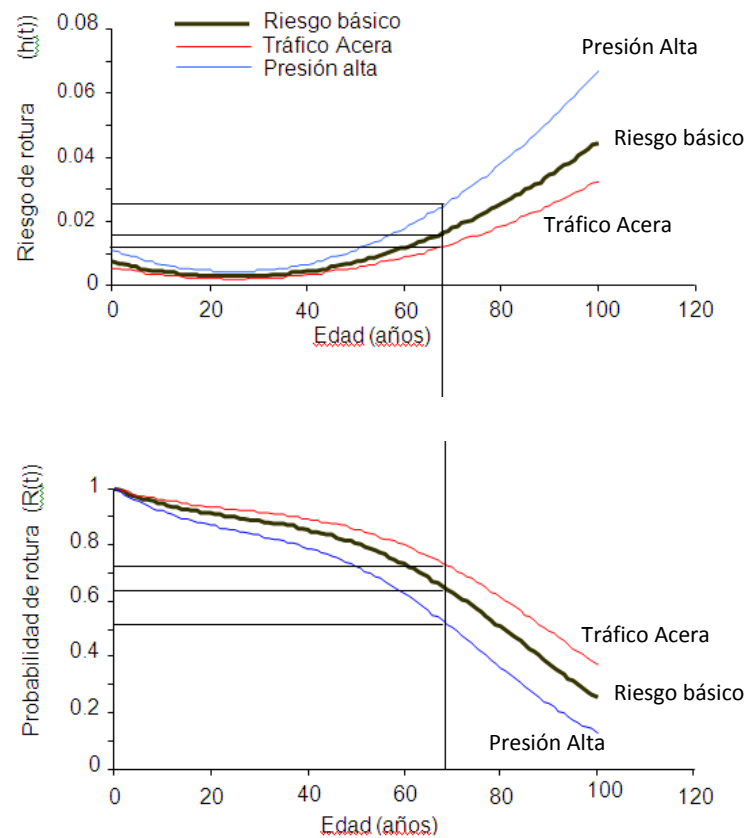


Figura 4.7 Modelo de riesgo proporcional para unidades (conductos) de determinada longitud, que fueron instalados en un determinado año. El riesgo básico (riesgo de referencia o de partida) describe una tubería funcionando a una presión normal y en una zona de tráfico normal. Si la tubería llega a la edad digamos de 70 años, sin roturas su probabilidad instantánea de rotura (digamos de romperse en el próximo año) es de 1,5% aproximadamente. La probabilidad que esta tubería alcance la edad de los 70 años sin reparaciones (sin roturas) es aproximadamente 63%. La línea roja describe la misma tubería en una zona de tráfico leve. El riesgo y la supervivencia cambian por consiguiente; un cambio similar se presenta en la línea azul, que describe a la misma tubería, pero en operación a un alto nivel de presión. Como se puede ver, los factores presión y tráfico, tienen efectos diferentes sobre la operación normal de los conductos.

La función de riesgo básica o de referencia $h_0(t)$ puede ser interpretada como una componente de la edad que depende del tiempo (t) de envejecimiento de la tubería, mientras que las covariables representan factores ambientales y de funcionamiento bajo estrés (Z), que actúan sobre la red para aumentar o reducir su riesgo de fallo. Además los datos pueden ser divididos (o

estratificados) con respecto a una de las covariables si se considera que actúan sobre la función de riesgo de una manera no multiplicativa y otra función de riesgo básica puede proponerse en caso necesario.

El conocimiento a priori de la función de riesgo básica (de referencia) no es necesario para estimar los elementos del vector b , ni para evaluar el efecto de las covariables.

En síntesis cuando el objetivo es evaluar la influencia o el efecto de covariables sobre la función de riesgo, se debe utilizar el modelo de Cox. Pero cuando el objetivo es predecir futuros fallos y averías dentro de un cierto horizonte de tiempo a largo plazo, es más conveniente hacer una hipótesis paramétrica sobre la forma de $h_0(t)$, Kumar y Klefsjö, (1994). Un ejemplo, es el *modelo de riesgo proporcional de Weibull*, donde el $h_0(t)$ de este modelo, es la función de riesgo de Weibull.

Si se asume que $h_0(t)$ es la función de riesgo de una unidad con vector de covariables $Z = 0$ y si admite el modelo de riesgo proporcional de Cox, la función de fiabilidad condicional para T, dado z, será:

$$R(t; z) = e^{-\int_0^t h(t; z) dt} = e^{-\int_0^t h_0(t) e^{b^T Z} dt} = \left[e^{-\int_0^t h_0(t) dt} \right]^{e^{b^T Z}} = [R_0(t)]^{e^{b^T Z}}$$

De forma más general, si se asume que $h_0(t)$ es la función de riesgo de una unidad o individuo con vector de covariables $Z = \bar{Z}$, la función de riesgo condicional para T, dado Z, será:

$$h(t; Z) = h_0(t) e^{b^T (Z - \bar{Z})}$$

Y la función de fiabilidad condicional para T, dado Z, será :

$$R(t; z) = [R_0(t)]^{e^{b^T (z - \bar{z})}}$$

Donde \bar{Z} es un vector de medias, constituido por los valores medios de las covariables componentes.

En el modelo de Cox las variables Z se considera que no dependen del tiempo, y aunque existe una versión del modelo donde las variables son dependientes del tiempo también, nosotros aquí hemos considerado la versión más sencilla.

En general la proporción de riesgo para dos individuos con vector de covariatas Z_i y Z_j

$$\frac{h(t;Z_i)}{h(t;Z_j)} = \frac{h_0(t)e^{b^T Z_i}}{h_0(t)e^{b^T Z_j}} = \frac{e^{b^T Z_i}}{e^{b^T Z_j}} = cte$$

es constante en el tiempo, tal como se expuso anteriormente, por esta razón el modelo de Cox también es conocido como el modelo de *riesgo proporcional*

Modelo de riesgo proporcional estratificado.

Cuando se configuran grupos o estratos (de acuerdo a algún criterio), en un conjunto de datos, la formulación más general del modelo de Cox estratificado para la función de riesgo de un sistema en el j -ésimo estrato ($j = 1, \dots, p$), suele ser expresada por:

$$h_j(t; Z) = h_{0j}(t)e^{b^T Z}$$

Donde las funciones de riesgo básicas $h_{0j}(t)$ en cada uno de los j estratos son arbitrarias y distintas, y el vector de coeficientes b^T es el mismo en todos los estratos. Si la hipótesis de proporcionalidad no se verifica en los p estratos de un factor o covariata, se suele considerar este modelo. En este modelo se asume que las funciones de riesgo son proporcionales dentro del mismo estrato, pero no necesariamente a través de los p estratos. Kalbfleisch y Prentice (1980), es la principal referencia para este modelo.

El concepto de estratificación es muy útil en la modelización de sistemas reparables / no reparables con modelos de riesgo proporcional.

Validación del modelo de riesgo proporcional.

Entre los métodos gráficos para validar el modelo de Cox, el más utilizado está basado en el *paralelismo de las graficas log-log*.

Este método se basa en lo siguiente:

Si se admite el modelo de riesgo proporcional se verifica que,

$$R(t; z) = [R_0(t)]^{e^{b^T z}}$$

$$\ln R(t; z) = e^{b^T z} \ln R_0(t)$$

Y por lo tanto,

$$\ln [-\ln R(t; z)] = b^T z + \ln [-\ln R_0(t)]$$

Consecuentemente, si el modelo de riesgo proporcional es adecuado, las representaciones graficas sobre t de las funciones $\ln [-\ln R(t; z)]$ correspondientes a dos individuos con vectores de covariatas diferentes, Z_1 y Z_2 serán paralelas, ya que su diferencia no dependerá de t ,

$$\ln [-\ln R(t; Z_2)] - \ln [-\ln R(t; Z_1)] = b^T (Z_2 - Z_1)$$

No obstante dada la indecisión que a veces la apreciación visual del paralelismo de estas curvas, han surgido otros métodos gráficos alternativos entre los que se encuentran los gráficos de Andersen (1982), basados en representar el logaritmo de la estimación de la función de riesgo acumulado $H(t)$ para distintos valores de una covariata y observar si existe una relación lineal o no en los gráficos de Arjas (1988).

4.5.3.2 Modelos de vida acelerada.

Ensayos Acelerados

En algunos productos, el tiempo necesario para comprobar la fiabilidad de los mismos puede ser extremadamente largo, dispendioso y por tanto muy costoso.

Los tests de fiabilidad se pueden acelerar incrementando el tamaño muestral, pero teniendo cuidado de que la distribución de los tiempos de vida

de las unidades en estudio, no presente un período de desgaste durante la vida acelerada.

Un ensayo acelerado pone a prueba un producto o unidad bajo condiciones de distribución mucho más desfavorables de las habituales (*condiciones de de sobreesfuerzo o estrés*), por lo que el tiempo de fallo se produce antes.

Por ejemplo, si queremos estudiar los fallos debidos a la corrosión de un elemento a temperaturas y humedades de uso nominales, entonces podemos utilizar el hecho de que el mismo tipo de corrosión ocurrirá mucho antes en condiciones de humedad extrema y temperatura elevada (ensayo en un laboratorio).

La planificación del ensayo acelerado es compleja y debe ser diseñada por los propios ingenieros de diseño del producto.

Es necesario tener un conocimiento previo de los factores que influyen en la aceleración del proceso y de la medida en que estos actúan sobre el tiempo de vida T o sobre alguna transformación de esta variable.

Por ejemplo, si se supone nominalmente que el 99,5% de los motores de propulsión después de 10 minutos operan correctamente bajo altas condiciones de carga y temperatura, (particularmente bajo una tensión de $2 \times 10^8 \text{ N m}^{-2}$ y una temperatura de 450° C). Con base en los datos de algunos ensayos acelerados (que incluyen estas y otras covariables “*aceleradoras*”), se podría pensar, en la estimación de un modelo de regresión de vida acelerada, que explique el tiempo de vida del motor, en función de la carga y la temperatura entre otros factores, para identificar cuáles de ellos, son los de mayor riesgo o influencia sobre la vida de estas unidades y además poder predecir mediante extrapolación muestral del modelo, las condiciones nominales de operación o fiabilidad de estos motores a los 10 minutos de funcionamiento. De esta manera, se podrá ver si el efecto de estas variables explicativas es de *acelere* o *desacelere* del tiempo de vida de estas unidades, al actuar sobre sus condiciones normales (nominales) de operación.

Los modelos de vida acelerada suponen que $\ln T$ se relaciona con el vector Z , a través de una estructura lineal de la forma

$$\ln T = b^T Z + \sigma W$$

o su equivalente

$$T = e^{b^T Z + \sigma W} = e^{b^T Z} e^{\sigma W}$$

Donde T es la variable tiempo de vida (que puede ser Weibull, log-normal o log-logística) de una unidad; Z^T es la transpuesta de un vector de covariatas o variables explicativas de la forma , $Z^T = (1, z_1, z_2, \dots, z_p)$; b^T es la transpuesta de un vector de parámetros de regresión desconocidos , $b^T = (b_0, b_1, \dots, b_p)$; σ es un parámetro de escala y W es una variable de distribución conocida (de valores extremos, normal estándar o logística) que indica el error aleatorio, Kalbfleisch y Prentice (1980).

Estos modelos se denominan también modelos de ensayos de vida acelerada, debido a que el efecto de las variables explicativas usualmente es acelerar o desacelerar el tiempo de vida T , multiplicando su valor su valor de referencia T_0 por un factor constante A_f , denominado factor de aceleración ($A_f = e^{b_1 z_1 + b_2 z_2 + \dots + b_p z_p}$), ya que se verifica que

$$T = e^{b^T Z} e^{\sigma W} = e^{b_1 z_1 + b_2 z_2 + \dots + b_p z_p} \cdot e^{b_0 + \sigma W} = e^{b_1 z_1 + b_2 z_2 + \dots + b_p z_p} \cdot T_0$$

Donde $T_0 = e^{b_0 + \sigma W}$, indica el tiempo de vida de referencia para un vector de covariables $Z^T = (1, 0, 0, \dots, 0)$, es decir, es el tiempo de referencia de una unidad o dispositivo en condiciones *normales* o *usuales* de operatividad, (sin la influencia de las p covariables). Al final, la modelización del efecto del vector Z sobre la aceleración o desaceleración del tiempo de vida T , se mide valora a través de la relación

$$T = A_f T_0$$

Los modelos de vida acelerada se conocen también con el nombre de modelos *log-lineales*, por que las variables explicativas actúan aditivamente sobre el valor medio de la distribución $\ln T$ o multiplicativamente sobre T , a través de la relación anterior.

El método de estimación del vector b^T y el parámetro σ , es el método de máxima verosimilitud, que maximiza con base en una muestra de tamaño n , $((t_i, \delta_i, z_i) \ i=1, \dots, n)$, el logaritmo de la función de verosimilitud dada por

$$L(b^T, \sigma) = \prod_{i=1}^n f_T(t_i; b^T, \sigma, z_i)^{\delta_i} \cdot R_T(t_i; b^T, \sigma, z_i)^{1-\delta_i}$$

Donde los datos con censura por la derecha, contribuyen a esta función con el valor de la función de supervivencia de T en el instante que aparezca la censura, es decir, cuando $\delta_i = 0$.

Bondades del modelo. Si W se escoge adecuadamente, estos modelos de vida acelerada proporcionan:

- Un buen ajuste a los datos de la muestra.
- Sus estimaciones son más precisas, por basarse en pocos parámetros.
- Permiten la extrapolación por fuera del rango de la muestra observada.

4.5.3.3 Residuos de Cox-Snell

Si los parámetros de un modelo seleccionado (entre varias familias), han sido estimados por el método de máxima verosimilitud con datos censurados, la bondad de ajuste del modelo de Cox a los datos observados puede basarse en criterios gráficos utilizando residuales, Smith (1991), en particular los denominados *residuos generalizados de Cox-Snell* (Cox-Snell, 1968), que se definen para el caso de datos completos (sin censura) por,

$$e_i = u_i = -\ln R(t_i; Z, \hat{b}) \quad (4,2)$$

Donde e_i es el residual i-ésimo para la unidad o individuo i de una muestra; $R(t_i; Z, \hat{b})$ es la fiabilidad estimada evaluada en t_i con vector de covariables Z y estimaciones \hat{b} .

De otra parte, ante el problema de los residuos para observaciones con censura, Cox y Snell propusieron una corrección de primer orden de los anteriores residuos (aunque en muchos casos no se tenga en cuenta esta

corrección), Smith(1991), los llamados residuales *ajustados* de Cox-Snell y que autores como Lawless (1982) recomiendan utilizar, cuya forma es,

$$e_i = u_i = -\ln R(t_i; Z, \hat{b}) + 1$$

De otra parte también, cuando la vida T es una variable aleatoria positiva continua con distribución *exponencial*, se verifica que su $H(t)$ es una función lineal de t , de la forma

$$H(t) = \lambda t \quad \text{para } t \geq 0$$

El hecho de que la función de impacto o de riesgo acumulado $H(t)$ venga definida por $H(t) = -\ln R(t; Z, b)$, y que además siga una distribución exponencial de parámetros unidad, de acuerdo a (4,2),

$$e_i = u_i \approx H(u_i)$$

Por ello, los residuos $e_i = u_i$ anteriormente definidos se consideran variables aleatorias, que deberían ser independientes e idénticamente distribuidas con distribución exponencial de parámetro unidad ($\lambda=1$), siempre que la estimación de la función de supervivencia $R(t; Z, b)$, sea correcta.

Por lo tanto, una forma gráfica de verificar la adecuación (la bondad de ajuste) de un modelo de Cox a las observaciones disponibles, será representar los puntos $(u_i, H(u_i))$ en un sistema de coordenadas donde,

$$u_i = -\ln R(t_i; Z, \hat{b})$$

y verificar que son *próximos a la primera bisectriz*, es decir, una gráfica del logaritmo de los residuales u_i contra su valor u_i , debería ser aproximadamente una línea recta con pendiente uno. Cualquier desviación de este tipo, debería ser motivo de preocupación.

4.6. Modelos de fiabilidad en redes de suministro de agua potable.

El análisis de supervivencia en la ingeniería se conoce como análisis de Fiabilidad¹ y está orientado al análisis de los fallos en el tiempo de unidades, dispositivos y componentes de diferente naturaleza (componentes electrónicos, tramos de tubería, electrodomésticos, ordenadores, etc.), con el fin de evaluar la probabilidad de ocurrencia de los fallos y la predicción del instante en que ocurrirán. Con ello se pretende determinar, entre otras cosas, políticas de mantenimiento, plazos de garantías en productos, predicción de costes en mantenimiento de redes de suministro de servicios públicos, etc.

Algunos investigadores del tema han adaptado las técnicas de análisis de supervivencia, utilizadas más frecuentemente en el campo de la biomedicina, a los problemas de fallos de redes de distribución de agua (Clark et al. 1982, 1988; Andreou et al. 1987a). El análisis de supervivencia incorpora el hecho de que, mientras algunas instalaciones fallan (por envejecimiento natural o por una causa provocada), otras no, y todo este acopio de información se considera de vital importancia y de gran impacto a la hora de analizar los fallos en los sistemas de abastecimiento de agua potable. Además las tuberías pueden fallar muchas veces durante su vida útil, y cada vez que un fallo se observa se necesita intervenir inmediatamente la red. Algunos investigadores (Andreou et al. 1987; Eisenbeis 1994; Guftafson y Clancy 1999) han demostrado que el patrón de rotura depende fuertemente del *número de roturas previas que las tuberías han tenido*. De hecho, *el número de roturas previas* es a menudo reportado como el factor más importante al predecir futuros fallos.

El análisis de supervivencia es particularmente útil en este campo cuando los registros de fallos se han conservado en buena parte del historial de averías y roturas de la red.

Sin embargo, un análisis de supervivencia “*de libro*” no siempre se puede aplicar en un estudio de esta naturaleza porque los registros históricos no son lo suficientemente amplios y completos en el tiempo para proporcionar la

¹ La definición utilizada en la ingeniería para la Fiabilidad es la de “probabilidad de que un dispositivo realice su función bajo condiciones establecidas, durante un período de tiempo dado”

valiosa información requerida, especialmente en la modelización estadística del comportamiento de los fallos. Tal es nuestro caso, en el que, como se verá, la información disponible es un factor crítico en la posibilidad de realizar modelizaciones.

Se ha realizado una revisión de algunos de los más importantes artículos e investigaciones publicados en el campo de la evaluación de la fiabilidad de redes de distribución de agua. En base a esa revisión se ha preparado un resumen de los modelos que se emplean y de las variables que en dichos modelos se han manejado, atendiendo también a la importancia o grado de significación que han tenido esas variables al ser utilizadas como explicativas de la duración o supervivencia de los elementos de la red.

Una recopilación reciente de los modelos que fueron creados por universidades y centros de investigación europeos puede encontrarse en Eisenbeis et al.. (2001) y Sægrov et al. (2001). Los modelos se pueden clasificar en dos grandes tipos, de acuerdo con su objetivo:

- Modelos que valoran el estado de las tuberías basándose en métodos estadísticos.
- Modelos que valoran y comparan las estrategias de rehabilitación, basados en hipótesis técnicas y económicas.

Los primeros comienzan por un análisis de los factores que influyen para identificar las variables explicativas del comportamiento de las tuberías (material, diámetro,...). En una segunda fase, y mediante el uso de modelos específicos (Poisson, Weibull,...) se realizan predicciones de supervivencia y fiabilidad. Los modelos incluyen la clasificación de las tuberías.

El segundo grupo de modelos requieren de una descripción de la red (edad, longitud, material, ratio de fallo, etc.) e integran datos económicos en el análisis.

Los modelos desarrollados han sido aplicados a redes de gran longitud, tanto urbanas como rurales, aunque su utilización es aún experimental. El AssetMap (INSA, Lyon, Francia) propuesto por, Malandain (1999) y Malandain

et al. (1999) es uno de los modelos importantes. Estos autores estuvieron analizando la infraestructura de las redes de distribución de agua potable de la comunidad de Lyon y basaron su trabajo en un análisis estadístico de las tasas de fallo con regresión Poisson.

Un segundo modelo es presentado por Eisenbeis (1994), Eisenbeis et al., (1999), Le Gat y Eisenbeis (2000), el cual lleva por nombre Failnet_Stat: Análisis y predicción de fallos en redes de agua, (Cemagref, Francia). Este método ha sido aplicado a redes de agua principales, urbanas y rurales, con registros de mantenimientos cortos (10 años) y largos (50 años) en Burdeos, suburbios de Paris, Alsacia y Carente-Maritime. Su fundamento radica en un análisis estadístico de las frecuencias de fallo con análisis de supervivencia y modelo Weibull.

En el Reino Unido los aportes de Poulton (2001) son bien conocidos con su método GIS y la técnica de análisis Cluster. Por otro lado en países como Noruega y Alemania se han hecho avances valiosos con los aportes de Sægrov, König et al. (2001) y Herz (1996, 1998) respectivamente. Por una parte en Noruega se ha propuesto el método Aquarel: vulnerabilidad en las redes de agua (SINTEF, Noruega), el cual se fundamenta en un análisis estadístico de frecuencias de fallos con procesos de Poisson no homogéneos, y en Alemania se propone el método KANEW: Exploración de las estrategias de rehabilitación (Dresden University of Technology, Germany), el cual muestra un enfoque novedoso a través de la utilización de modelos de supervivencia cohorte y un análisis de la infraestructura de stock existente. Hay que resaltar que a nivel de la Unión Europea se han realizado grandes esfuerzos a través de la subvención de proyectos en el Quinto Programa Marco que contribuye a la implantación de la Acción Clave "Gestión Sostenible y Calidad del Agua" dentro del área temática de Energía, Medio Ambiente y Desarrollo Sostenible. Como resultado de estos esfuerzos encontramos CARE-W, que es un proyecto que pretende desarrollar una plataforma de software dentro del campo de la rehabilitación de redes de agua para apoyar la toma de decisiones gerenciales en los abastecimientos europeos (Sægrov et al., 2003).

La bondad de los modelos anteriores depende de diferentes factores a tomar en consideración, especialmente al momento de intentar extrapolar su utilización a la realidad española. Por ejemplo, resulta fundamental establecer si la previsión se debe centrar más en el corto que en el medio-largo plazo. También dependerá de si se quiere, en la gestión de la red, actuar de manera reactiva (mantenimiento correctivo) o bien proactiva (mantenimiento preventivo). Así mismo, es importante definir con precisión cuáles son los datos mínimos que requiere un correcto análisis estadístico. De hecho, tiene notable interés conocer la sensibilidad de los modelos con respecto a los datos disponibles (número de roturas, material, tipo de terreno, edad, condiciones de tráfico, localización, profundidad, tipos de juntas, etcétera) que de algún modo pueden condicionar su comportamiento futuro. Es pues clara la necesidad de crear una base de datos ligada a la red de distribución que permita alimentar un modelo adecuado.

Parece indudable que un análisis de sensibilidad de los diferentes datos y un buen ajuste de los parámetros a partir de la información que las empresas de agua participantes aporten, resultará esencial a la hora de integrar conceptos y herramientas para adecuar e implementar un modelo estadístico a las necesidades nacionales.

Entre los distintos modelos estudiados en la bibliografía, sólo unos pocos están encaminados particularmente al estudio y análisis de fallos de tuberías con la restricción de *breves historiales de fallos en el tiempo y fuerte censura*. A título informativo, se van a comentar dos de estos estudios, que corresponden a los modelos de Eisenbeis (1994) y Mailhot et al (2000).

- *Eisenbeis (1994) aplicó el modelo de riesgo proporcional de Cox para 4 redes suministro en Francia. El autor observó que las tuberías se comportaban de manera diferente en tres situaciones distintas, en el primer fallo, en los fallos del segundo al cuarto y para fallos de orden quinto en adelante. Por lo tanto, los datos se dividieron en tres partes (o estratos) y se modelaron cada una separadamente. Este investigador*

usó una distribución de cinco parámetros y una distribución Weibull de dos parámetros para cada uno de las dos primeras partes (o estratos), y una distribución exponencial de un parámetro para la última parte. Esto fue posible por el hecho de que el investigador tuvo acceso a datos de 40 a 54 años de fallos en la gran red de tuberías urbana. La disponibilidad de datos permitió el uso del modelo de riesgo proporcional de Cox, el cual requiere que los factores de riesgo de más impacto en los fallos, sean identificados primero mediante análisis de regresión. Esto aumenta significativamente el número de parámetros que deben ser estimados. La función de riesgo obtenida para el modelo de Cox fue evaluada en cada etapa a partir de los datos. Algunos factores de riesgo identificados como significativos para una o más redes fueron: la longitud de la tubería, su diámetro, la corrosividad del suelo, la intensidad del tráfico por encima de la tubería y las instalaciones después de 1966. Eisenbeis (1994) consideró diferentes métodos para la aplicación de su modelo en municipios que tenían breves historiales en sus registros. El ejemplo que dio, fue el de una determinada red con una historia de registros de 5 años (1988-1993), para el cual las primeras tuberías fueron tendidas en 1945. El desarrollo de un modelo de deterioro de las tuberías para esta red, se basó en tomar como referencia un modelo paramétrico de deterioro general desarrollado con redes más grandes, que tenían historiales de fallos registrados más largos. Para ampliar el historial de fallos de la red tomada como ejemplo, utilizó un método de muestreo sobre tuberías instaladas antes de 1988 (en la red de referencia) y agregó estos fallos al historial real de registros de la red ejemplo. Una vez incorporada esta muestra a la historia de la red ejemplo, entonces la propuesta de modelo anteriormente descrita podía ser utilizada.

- Mailhot et al. (2000) presentó una metodología formal para la estimación de parámetros de modelos estadísticos, especialmente en municipios con cortos historiales de fallos. Esta metodología fue aplicada a dos municipios (Quebec y Chicoutimi) que tienen redes de tuberías

muy antiguas (de un siglo) y un historial de fallos registrado de al menos 21 años. Se consideraron 4 modelos de fallos, usando las distribuciones Weibull y exponencial para diferentes tipos de fallos. La estimación de los parámetros, se obtuvo en diferentes periodos de instalación de las tuberías, lo que permitió la identificación de los años 60 como un periodo especialmente crítico para este municipio. Se presentó además una metodología para evaluar, en base a estadísticas, el rendimiento relativo de una distribución de Weibull comparado con una distribución exponencial para un tipo de fallo dado. Esta metodología fue usada en este estudio para predecir los parámetros del modelo de fallos.

La tabla 4.1 de la siguiente página, recoge otros modelos utilizados en el análisis de supervivencia de sistemas de distribución de Agua Potable.

Modelo	Descripción	Artículo
KANEW (Alemania)	Análisis de la infraestructura de stock existente, modelos de supervivencia cohorte	Herz (1996, 1998)
UtilNets (North West Water, Gran Bretaña; Computer Technology Institute, Grecia; SINTEF, Noruega; Tecnic, Italia; UBIS, Alemania)	Valoración física de ratios de fallo considerando el ratio de corrosión y carga externa.	Sægrov y König (2001)
AssetMap (INSA, Lyon, Francia)	Análisis estadístico de las tasas de fallo con regresión Poisson.	Malandain (1999) y Malandain et al. (1999)
Failnet_Stat (Cemagref, Francia)	Análisis estadístico de las frecuencias de fallo con análisis de supervivencia	Eisenbeis (1994), Eisenbeis et al. (1999), Le Gat y Eisenbeis (2000)
Failnet_Reliab (Cemagref, Francia)	Modelización hidráulica considerando fallos de tuberías	Piller et al. (2001), Brémond y Berthin (2001)
Análisis clúster (WRc y Imperial College, Gran Bretaña)	Análisis espacial y clúster de fallos con GIS	Poulton (2001)
Winroc (NTNU, Noruega)	Análisis estadístico de frecuencias de fallos con procesos Poisson no homogéneos	Rostum y Schilling (1999)
Weibull y exponencial	Estimación de parámetros de modelos estadísticos, especialmente con cortos historiales de fallos.	Mailhot et al. (2000)

Tabla 4.1 Algunos modelos utilizados en Análisis de supervivencia.

Otro trabajo que tiene interés, aunque está fuera del campo de la fiabilidad de redes de agua potable, es el de *Iglesias Pérez y González Manteiga (1999)*. Utilizando técnicas de estimación no paramétrica de curvas, definen y estudian el comportamiento asintótico de un estimador de la función

de distribución del tiempo de fallo condicionado al valor de las covariables (covariatas), que denominan estimador límite producto generalizado (LPG). En este trabajo se ilustra el comportamiento de dicho estimador LPG en un problema con datos reales, relativo a la mortalidad en pacientes diabéticos, mostrando su gran potencialidad en el terreno aplicado. El análisis estadístico de los datos pone de manifiesto como el estimador generalizado proporciona una información importante sobre el problema que se aborda, información que en el análisis incondicional permanece oculta. En este trabajo también se aporta una idea nueva para el cálculo del parámetro de suavizado, presente en el estimador LPG, en el complejo contexto de presencia de censura, truncamiento y covariables.

Se presenta a continuación un examen crítico de algunos modelos estadísticos (*Riesgo proporcional, Poisson dependiente del tiempo, y Vida Acelerada, entre otros*) que se han propuesto en la literatura científica para explicar, predecir y cuantificar la rotura de las tuberías o los fallos en la infraestructura de las redes de abastecimiento. El deterioro de la superficie interna de las tuberías sólo se menciona como un factor que contribuye al deterioro estructural de la misma. El enfoque que se sigue es el de Kleiner Y. y Rajani B (2000) que recoge básicamente cuatro aspectos: una breve descripción del modelo, la formulación de su ecuación, algunas limitaciones del modelo y la caracterización de la información que se necesita para su implementación, como se indica posteriormente. La intención de este formato de valoración es facilitar la identificación de las principales características de los modelos estudiados con sus correspondientes limitaciones y requerimientos de información.

4.6.1 Modelos para el análisis del estado de deterioro de una red de agua.

Este apartado presenta una síntesis, de los modelos más conocidos en la literatura, para la cuantificación del estado de deterioro y degradación de las redes de distribución de agua potable (Kleiner, Y., y Rajani, B. (2000)), de significativa importancia en la recopilación de información, para el diseño de bases de datos cercanas al “ideal”, para la gestión y *toma de decisiones* en las

entidades gestoras prestadoras del servicio de abastecimiento del agua potable en los ayuntamientos.

4.6.1.1. Modelos exponenciales y lineales en el tiempo.

1. Modelo propuesto por *Shamir y Howard (1979)*.

$$B(t) = B(t_0) e^{A(t-t_0)}$$

Este modelo predice el N^0 (o tasa) de roturas al año por unidad de longitud, pero se usa para estimar el instante óptimo t de reemplazo de las tuberías de agua en la red. Posteriormente estos mismos autores le combinaron algunos datos económicos a este modelo para la misma predicción del tiempo óptimo necesario, para la renovación de la tubería y obtuvieron otras variantes del modelo, aplicables a la gestión de las redes con criterios económicos.

Notación

$B(t)$ = N^a acumulado de rupturas en el instante t por unidad de longitud por año (en $\text{km}^{-1} \text{año}^{-1}$)

t = tiempo transcurrido (en años) desde el instante t_0

$B(t_0)$ = es el N^0 (o tasa) inicial de rotura en el año t_0 de instalación de la tubería

t_0 = edad de la tubería en el año inicial t_0

A = es una constante, coeficiente del crecimiento de la tasa de roturas o averías (año^{-1})

Datos o covariables que se requieren

Longitud de la tubería, la fecha de instalación t_0 y el historial de roturas (N^0 de roturas previas); Formación esencial de grupos homogéneos de acuerdo a criterios como el tipo de tubería, diámetro, el tipo de suelo, el

tipo de rotura o avería, características de la sobrecarga en el suelo, etc.

2. Modelo propuesto por *Walski y Pelliccia (1982)*.

$$B(t) = C_1 \cdot C_2 \cdot B(t_0) e^{A(t-t_0)}$$

Notación

C_1 = es la relación entre la frecuencia de rupturas para tuberías de hierro fundido (fundición), con alguna (o ninguna) rotura previa y la frecuencia general de rupturas para tuberías del mismo material.

C_2 = es la relación entre la frecuencia de rupturas para tuberías de 500 mm de diámetro y la frecuencia general de rupturas para tuberías de distintos diámetros en general.

Datos que se requieren

Los mismos datos que para el modelo de Shamir y Howard (1979), además de información sobre el método de distribución (reparto) de la tubería de fundición y el diámetro de la tubería.

3. Otros Modelos propuestos por *Walski y Pelliccia (1982)*.

$$3.i) REP = y_1 \cdot e^{y_2 t} \cdot e^{y_3 T} \cdot e^{y_4 PRD} \cdot e^{y_5 DEV} \cdot SL^{y_6} \cdot SH^{y_7} \text{ (Mod exponencial)}$$

$$3.ii) NA = x_1 + x_2 D + x_3 P + x_4 I + x_5 RES + x_6 LH + x_7 T \text{ (Modelo lineal)}$$

Las variables respuesta en estos dos modelos son REP y NA , cuyo significado se explica a continuación

Notación

i) **REP = N° de reparaciones** (que ha tenido la tubería a la edad t)

PRD = diferencial de presión

t = edad de la tubería desde la primera ruptura o avería,

DEV = % de la longitud de la tubería en baja y moderada corrosividad del suelo,

SL = superficie de la tubería en baja corrosividad del suelo,

SH = superficie de la tubería en los suelos altamente corrosivos

ii) **NA = número de años desde la instalación hasta la primera reparación,**

x_i, y_i = parámetros de regresión

D = diámetro de la tubería,

P = presión absoluta dentro de la tubería,

l = % de recubrimiento de la tubería para uso industrial,

RES = % de recubrimiento de la tubería para uso residencial

LH = longitud de la tubería en los suelos altamente corrosivos,

T = tipo de tubería (1 = metálica, 0 = de hormigón armado o concreto reforzado),

Datos o covariables que se requieren (para los dos modelos).

Registro de la fecha de la instalación, historial de roturas y averías, el tipo y diámetro de la tubería, así como información acerca de la presión de funcionamiento, la corrosividad y la composición de la zona y del área que cubre la tubería. Datos adicionales como el tipo de arroturas de la tubería y la capacidad del tubo que se necesita para mejorar el modelo.

4. Modelo propuesto por *Kettler y Goulter (1985)*.

$$N = k_0 Ed$$

Notación

N = número de rupturas o averías por año

K_0 = parámetro de regresión

Datos o covariables que se requieren

Los mismos datos que para el modelo de Shamir y Howard (1979).

5. Modelo propuesto por McMullen (1982).

$$Ed = 65.78 + 0.028 SR - 6.338 pH - 0.049 r_d$$

Notación

Ed = edad de la tubería en la primera ruptura o avería (en años)

SR = resistividad saturada del suelo (ohmios- cm)

pH = pH del suelo

r_d = potencial redox (millivolts)⁶

Datos que se requieren

Los datos necesarios, normalmente no están disponibles; sólo datos esporádicos poco costosos, sin embargo, la continua y extensa recopilación de datos del programa es costosa; es importante la supervisión continua de las propiedades del suelo, donde las condiciones de las aguas subterráneas no han alcanzado el estado de equilibrio o estacionalmente son dependientes

6. Modelo propuesto por Jacobs y Karney (1994).

$$P = a_0 + a_1 Long + a_2 Ed$$

Notación

P = recíproco de la probabilidad de un día sin rupturas o averías
 a_0, a_1, a_2 = coeficientes de regresión

Datos o covariables que se requieren.

Longitud de la tubería (Long), la edad (Ed) y el historial de rupturas; y otros datos que faciliten la formación de grupos homogéneos.

4.6.1.2. Modelos de riesgo proporcional y de vida acelerada.

1. Modelo propuesto por *Marks y otros (1985)*. Riesgo proporcional.

$$h(t, Z) = h_0(t)e^{b^T Z}$$

$$\text{con } h_0(t) = 2 \cdot 10^{-4} - 10^{-5} t + 2 \cdot 10^{-7} t^2$$

Notación

T = tiempo transcurrido hasta la siguiente ruptura o avería

$h(t, Z)$ = función de riesgo

$h_0(t)$ = es la función de riesgo base

Z = vector de covariables

b = vector de coeficientes a ser estimados por máxima verosimilitud

Covariables explicativas

- *El registro natural de la longitud de la tubería*
- *Presión en operación (funcionamiento)*
- *Porcentaje de las tierras bajo desarrollo (en explotación)*
- *Período de la instalación*
- *Edad de la tubería en la segunda (o posteriores) tasa de ruptura o – avería.*
- *El número de rupturas previas en la tubería*
- *Corrosividad del suelo.*

2. Modelo propuesto por *Andreou's et al. (1987, 1987b); Marks et al. (1987)*. Riesgo proporcional.

$$h(t, Z) = h_0(t)e^{b^T Z}$$

Al comienzo en la Etapa inicial, el modelo de Andreou coincide con el modelo de Marks et al. (1985) descrito anteriormente, sólo que en la etapa posterior (posteriormente), Andreou establece que $h=\lambda=e^{Zb^T}$

Notación

h = el riesgo (constante en la última etapa)

Covariables explicativas (o datos) que se requieren

Se requieren los mismos datos de las variables explicativas del modelo de Riesgo proporcional de Marks et al. (1985).

3. Modelo propuesto por *Brémond (1997)*. Riesgo proporcional.

$$h(t, Z) = h_0(t)e^{b^T Z}$$

$$h_0(t) = \lambda\beta(\lambda t)^{\beta-1}$$

Notación

t = tiempo hasta el (próximo) fallo

$h(t)$ = función de riesgo

λ, β = parámetros de forma y escala (respectivamente) de la Distribución de Weibull

Variables explicativas

El número de rupturas/averías (fallos) anteriores

Diámetro de la tubería

Condiciones del suelo

El tráfico de carga

4. Modelo propuesto por Constantino y Darroch (1993), Miller (1993) y Constantino et al. (1996). Modelo de Poisson dependiente del tiempo.

$$H'(t) = \left(\frac{t}{\theta}\right)^\beta$$

Con $\theta = \theta_0 e^{aZ}$

Notación

t = edad de la tubería

$H(t)$ = número promedio de fallos por unidad de longitud a la edad t (en el instante t)

θ, β = parámetros de forma y escala respectivamente

θ_0 = valor del punto de partida (valor basal)

a = vector de coeficientes a ser estimados por regresión;

Z = un vector de covariables que afectan la tasa de rupturas o fallos.

Covariables explicativas

Presión estática promedio.

Condiciones generales del tráfico

Diámetro de la tubería.

Tipo de suelo

5. Modelo propuesto por Lei (1997) y Eisenbeis et al. (1999). Vida acelerada.

$$\ln(T) = \mu + X^T \beta + \sigma Z$$

$$\rightarrow T = f(\mu, \sigma, Z) e^{X^T \beta}$$

Notación

T =tiempo hasta (el próximo) fallo

X =vector de variables explicativas

Z = variable aleatoria con distribución Weibull o Gumbel (distribución de valores extremos, para mínimos)

σ = parámetro a estimar por máxima verosimilitud

β = vector de parámetros estimado por máxima verosimilitud

Nota: Aunque la estructura de los dos modelos es la misma, sin embargo, en el modelo de Eisenbeis (1999), la variable Z sigue una distribución Gumbel, mientras que para el modelo de Lei (1997), Z se distribuye como una Weibull, esencialmente esta es la diferencia (de fondo) entre los dos modelos.

VARIABLES EXPLICATIVAS

Para Lei (1997)

Edad de la tubería por grupos

Tamaño de la tubería

Longitud de la tubería

Material de la tubería (para estratificar los tubos por tipo de material).

Para Eisenbeis et al. (1999)

Registro de la longitud de la tubería

Diámetro de la tubería

Material de la tubería

El tráfico de carga

La acidez del suelo

Humedad del suelo

El número de rupturas anteriores fue considerada como una covariable y también como una variable de estratificación.

4.6.2 Tablas resumen.

MODELOS: Exponenciales y Lineales en el tiempo		
INDICADORES A PREDECIR	INFORMACION REQUERIDA(Covariables)	ESTRUCTURA DEL MODELO
1. No. Roturas/unidad de longitud. B(t)	Longitud, fecha de instalación, historial de roturas (Nº de roturas previas en el instante t), material, diámetro, tipo de suelo, tipo de rotura, tráfico de carga.	$B(t) = B(t_0) \cdot e^{A(t+g)}$ $B(t) = C_1 \cdot C_2 \cdot B(t_0) \cdot e^{A(t+g)}$
2. No. De reparaciones (después de la primera rotura)=REP	Datos, fecha de instalación, historial de roturas, diámetro, material, presión, corrosividad del suelo, tipo de rotura.	$REP = y_1 \cdot e^{y_2 t} \cdot e^{y_3 T} \cdot e^{y_4 PRD} \cdot e^{y_5 DEV} \cdot SL^{y_6} \cdot SH^{y_7}$
3. No. Años transcurridos desde la instalación hasta la primera rotura. NY	Datos, fecha de instalación, historial de roturas, diámetro, material, presión, corrosividad del suelo, tipo de rotura.	$NA = x_1 + x_2 D + x_3 P + x_4 I + x_5 RES + x_6 LH + x_7 T$
4. No. Roturas por año N	Los mismos datos que modelo 1.	$N = k_0 Ed$
5. Edad de la tubería a la primera rotura.	Datos no disponibles normalmente por su alto costo. PH del suelo, Resistividad saturada del suelo y potencial redox (millivolts).	$Ed = 65.78 + 0.028 SR - 6.338 pH - 0.049 r_d$
6. Promedio diario P de tramos de tubería sin roturas (o averías)	Longitud, edad e historial de roturas, formación de grupos homogéneos.	$P = a_0 + a_1 Long + a_2 Ed$

Fuente: Adaptado de Kleiner, Y., and Rajani, B. (2000). *Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models*

Tabla 4.2 Modelos Exponenciales y Lineales empleados en Fiabilidad para la Gestión de redes de agua.

MODELOS MULTIVARIADOS: de Riesgo Proporcional y de Vida Acelerada		
INDICADORES A PREDECIR	INFORMACION REQUERIDA(Covariables)	ESTRUCTURA DEL MODELO
1. Riesgo de fallo de la tubería (en el instante t)	Longitud, presión, fecha de instalación, edad tubería (a partir de segunda tasa de roturas), número de roturas previas, corrosividad del suelo, tiempo hasta la rotura.	$h(t, Z) = h_0(t)e^{b^T Z}$ <p>Donde $h_0(t)$ puede ser:</p> $h_0(t) = 2 \cdot 10^{-4} - 10^{-5} t + 2 \cdot 10^{-7} t^2$ <p style="text-align: center;">o también</p> $h_0(t) = \lambda \beta (\lambda t)^{\beta-1}$
2. Velocidad con que ocurren los fallos	La misma información que en 1.	La de la función de impacto $H(t)$ que corresponda.
3. No. Promedio de fallos por unidad de longitud a la edad t .	Presión estática promedio, diámetro, tipo de suelo, condiciones del tráfico.	La de la función de impacto $H(t)$ que corresponda.
4. Valoración del efecto de las covariables en la aceleración del tiempo de vida.	Edad de la tubería por grupos, tamaño tubería, longitud tubería, estratificar respecto al material. Además: diámetro, tráfico, acidez y humedad del suelo, No. Roturas anteriores.	$H'(t) = \left(\frac{t}{\theta}\right)^\beta$
5. Identificación de factores de riesgo de mayor impacto en el fallo de las tuberías	La misma información que en 1. y en 3.	$\ln(T) = \mu + x^T \beta + \sigma Z$ $\rightarrow T = f(\mu, \sigma, Z) e^{x^T \beta}$

Fuente: Adaptado de Kleiner, Y., and Rajani, B. (2000). *Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models*

Tabla 4.3 Modelos de Riesgo Proporcional y de Vida acelerada empleados en Fiabilidad para la Gestión de redes de agua.

5. Análisis de fiabilidad en una red de suministro de agua potable.

5.1 Caracterización de la red

Los datos considerados en este estudio fueron suministrados por una de las empresas gestoras de abastecimiento de agua potable, localizada en una importante zona del litoral mediterráneo español. La información hace referencia a características básicas de la red, características de cada uno de los tramos² que integran la red (factores de uso y ambientales) y características de los fallos y averías detectados durante el periodo comprendido entre los años 2000 y 2005.

La base de datos original de la red con 32187 registros inicialmente fue sometida a un riguroso proceso de depuración, que consistió en:

- Identificar y eliminar inconsistencias:
 - tramos con edades de fallo negativas
 - tramos de tubería instalados con 100 o más años de antigüedad
 - inconsistencias en los diámetros y en los años de instalación,
- Categorizar algunas variables de las tuberías
- Discriminar entre *fallos naturales*³ y *fallos provocados*⁴

Respecto de este último punto, debe destacarse que los fallos provocados fueron considerados como otra fuente de censura, y que se enfocó el estudio

² En este documento, el término "tramo" se refiere a un segmento de la red, que es homogéneo en el material, diámetro, edad y en los factores medioambientales.

³ Los fallos naturales son los que se producen por envejecimiento, deterioro y degradación natural de las tuberías como consecuencia del paso del tiempo y el uso.

⁴ Los fallos provocados en este estudio se convierten en fuente permanente de censura, son consecuencia de la influencia de factores externos no controlables que actúan sobre la red.

con énfasis en la estimación de la fiabilidad de las tuberías de la red por roturas y averías naturales, como consecuencia del paso del tiempo y el uso.

Esta redefinición de los tipos de censura en la red de datos (censura por fallos provocados en los tubos y censura de tipo I a partir de enero de 2006), provocó que la tasa de censura subiera hasta cerca del 98%, y que número de fallos disponibles para el análisis se redujera a poco más de 500.

Finalmente la base de datos depurada se redujo a 26513 registros, de los cuales, el 98 % corresponden a datos con censura y tan sólo un 2% corresponden a fallos o rupturas por deterioro natural de la tubería.

5.1.1 Estructura original de la base de datos disponible para el estudio

La estructura original de los datos utilizados en este trabajo, conserva el esquema presentado en la sección 4.3 (Figura 4.1), en la que se destacan tres ventanas de observación. La ventana central que corresponde al período de estudio 2000 – 2005 y en la que se recogen los fallos disponibles para el estudio, esto es, los instantes de las roturas y averías de los distintos tramos de tubería observados en este período (*ventana de fallos conocidos*). A su izquierda, está la ventana donde se registran desde un comienzo (1900) los distintos períodos aleatorios de instalación de los conductos y además sin información, sobre posibles fallos que pudieron haber ocurrido en las tuberías, antes del inicio del estudio (2000) y que se desconocen por no haber sido registrados. Finalmente la tercera ventana (a la derecha de la ventana central), representa todas las tuberías que al finalizar el estudio (enero de 2006), no habían registrado ningún fallo, es la ventana con datos *censurados por la derecha*

Las particulares características del método de *Nelson-Aalen extendido* (NAE) para estimar la fiabilidad de las tuberías en condiciones alta censura por la derecha y truncamiento por la izquierda, obliga a establecer las siguientes consideraciones:

El inicio del estudio en el año 2000 induce en cada sujeto una variable de truncamiento X es decir los acontecimientos en cada tramo son desconocidos durante el periodo que va desde su instalación hasta el año 2000.

Además para este esquema particular, se consideran las siguientes variables (medidas en años):

$X = 2000 - \text{año de instalación de la tubería}$ (Edad en años de la tubería al inicio del estudio o *tiempo de truncamiento*), si este valor es negativo, se sustituye por 0, indicando que no hay truncamiento en ese caso.

$T = \text{año de fallo natural registrado} - \text{año de instalación}$ (Tiempo o instante de fallo).

$z_1 = \text{año de fallo provocado} - \text{año de instalación}$ (Tiempo de censura).

$z_2 = 2006 - \text{año de instalación}$ (Tiempo de censura).

Aunque existen dos tipos de censura, en el modelo las trataremos igual, de modo que, hablaremos de Z como la variable aleatoria *censura* en general.

Está claro que los fallos se registran únicamente si tienen lugar después del año 2000, si tienen lugar *antes*, *no hay observaciones*.

Con lo anterior, la manipulación de los datos originales, nos permite a manera de ejemplo, obtener la siguiente información:

ITEM	Año_repara	Año_Inst	OTRAS COVARIABLES	X	T	Z	Y	Estado δ
1		2004		0		2	2	0
2	2006	2004		0	2		2	1
3	2004	2003		0	1	0	1	1
4		2002		0		4	4	0
5		1994		6		12	12	0
6		1994						0
7		1994						0

8		1994						0
9		1994						0
10		1994						0
11		1994						0
12		1994						0
13		1994						0
....
89	2005	1900		100	105	105	1
....

Tabla 5.1 Registro de datos con censura por la derecha y truncamiento por la izquierda

La información original disponible para el estudio de la fiabilidad en la red, con el método de *Nelson-Aalen extendido*, bajo el modelo de truncamiento por la izquierda y censura a la derecha, tendría la siguiente estructura:

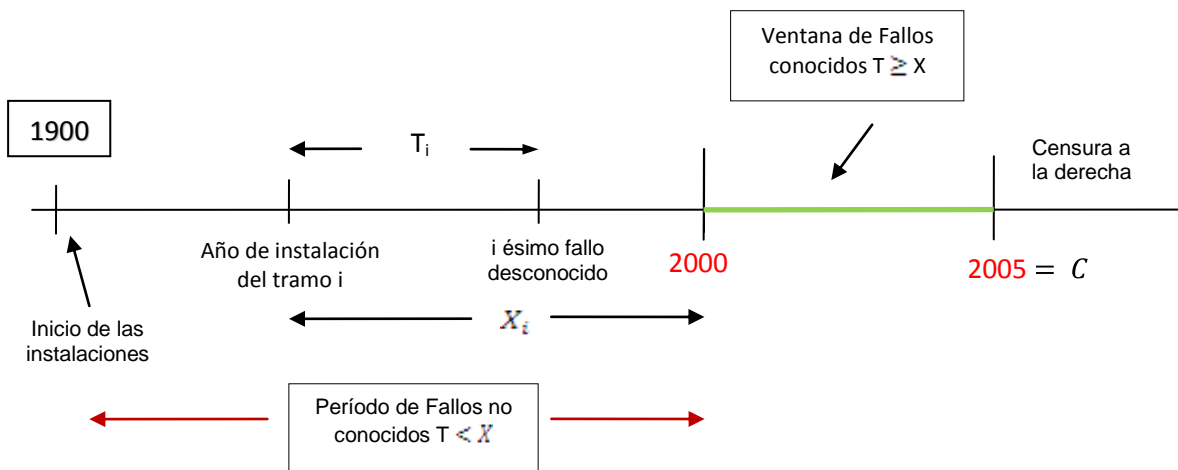


Figura 5.1 Estructura de los datos originales con el esquema de truncamiento por la izquierda y censura por la derecha

5.2 Características de la base de datos del estudio.

En resumen, la base de datos en referencia, se caracteriza por:

- La unidad experimental de observación es el tramo de tubería
- Un corto historial de rupturas de seis años: 2000-2005, que corresponde al período de estudio y/o registro de los fallos.
- Una fuerte censura múltiple del 98%, con truncamiento por la izquierda y un 2% de fallos naturales.
- El período de instalación de las tuberías de la red, llega hasta 2005. Para el estudio se han considerado las tuberías instaladas después de 1948, al observarse fuertes inconsistencias en los datos más antiguos.
- Las principales características físicas y medio ambientales asociadas a cada tramo de la red, así como sus correspondientes códigos son:
 - *Año de instalación*
 - *Año de reparación*
 - *Variable de estado*: Se denota con 0 si el tramo falla por causa provocada o si no ha fallado al finalizar el año 2005. Se denota con 1, si el tramo ha fallado por causa natural
 - *Longitud (en metros)*
 - *Diámetro (en milímetros)*
 - *Presión, en metros de columna de agua (MCA)*. En función de la red y sus condiciones particulares, de topografía, tipo de abonado al que sirve, etc., puede estar entre los 15 y los 30 mca.

- *Tráfico* que soporta la tubería. Tiene los tres niveles siguientes:

TRAFICO	CODIGO
Bajo acera	0
Bajo tráfico normal	1
Bajo tráfico pesado	2

Tabla 5.2 Categorías de la variable *TRAFICO*

- *Tipo de Material* del tramo de tubería. Se consideraron los cuatro tipos de materiales recogidos en la tabla 3.2.

MATERIAL	CODIGO
Fundición Dúctil	FD
Fundición Gris,	FG
Polietileno	P
Uralita.	U

Tabla 5.3 Categorías de la variable *Material*

Además de las anteriores, se generaron para el estudio las siguientes variables:

- *Edad del fallo* en años (Tiempo de fallo). Para los tramos que han fallado es la diferencia entre el año de reparación y el año de instalación. Es la variable respuesta o característica de interés.

- Año de instalación de la tubería categorizado. (*Cat_Año_Ins*):

CATEGORIAS AÑOS DE INSTALACION	CODIGO
1948 – 1959	1
1960 – 1969	2
1970 – 1979	3
1980 – 1989	4
1990 – 1999	5
> 2000	6

Tabla 5.4 Categorías de la variable *Cat_Año_Ins*

- Longitud categorizada (*LENGTH_cat*) en 4 niveles :

LONGITUD (en metros)	CODIGO
< 2,0	1
2,9 – 9,9	2
10 – 49,9	3
50 +	4

Tabla 5.5 Categorías de la variable *Length_cat*

- Diámetro categorizado (*Cat_diam*):

CATEGORIAS DE DIAMETRO (en milímetros)	CODIGO
< 90 mm	1
91 – 175 mm	2
176 – 300 mm	3
> 300 mm	4

Tabla 5.6 Categorías de la variable *Cat_diam*

5.2.1 Presentación y descripción general de la red estudiada

En este apartado se presenta una descripción general de la red, considerando los registros y las características de las distintas variables ya definidas, que hacen parte de la base de datos original, y que son de interés en el desarrollo de este estudio.

A continuación, y a modo de sumario, la tabla 5.6 muestra una síntesis de las características generales más importantes de la red de suministro estudiada en esta investigación.

CARACTERÍSTICAS DE LA BASE DE DATOS

Ubicación Geográfica	<i>Zona del Litoral mediterráneo Español</i>
Longitud total	<i>1028.8 km</i>
Unidad experimental	<i>Tramos de tubería.</i>
Num. De registros	<i>32387</i>
Período Instalación de las tuberías	<i>1890-2005</i>
Historial de roturas/fallos	<i>2000- 2005</i>
Número de fallos registrados	<i>1487</i>

Tabla 5.7 Tabla Resumen de la Red de Suministro objeto de estudio

Los datos utilizados para el análisis empírico exploratorio de los datos han sido suministrados (como ya se ha mencionado) por una compañía de distribución de agua potable del mediterráneo español. Dicha información hace referencia a características básica de la red, características de cada uno de los tramos que integran la red (factores de uso y ambientales) y características de

los fallos y averías detectados, durante el periodo comprendido entre los años 2000 y 2005. A continuación se presenta la descripción de la red estudiada.

- Longitud. Según la información suministrada la longitud total de la red a diciembre de 2005, es de 1028,8 Km repartidos en 32.387 tramos. La longitud que muestra mayor frecuencia es de 0,30 m con un 20.5%, luego siguen longitudes muy variables hasta encontrar una tubería con más de 1800 m de longitud
- Material. Los materiales que se encontraron con una frecuencia menor de 0.2% fueron eliminados del análisis. La tabla 5.7, muestra los 4 materiales resultantes y se observa que la Fundición dúctil (FD) y la Uralita (UR) tienen una frecuencia acumulada del 96% mientras que la Fundición gris (FG) y Polietileno (PE) representan conjuntamente sólo el 4% de la muestra analizada.

Material	Frecuencia	Porcentaje	Porcentaje acumulado
FD	20534	66,1	66,1
UR	9251	29,8	95,9
FG	823	2,6	98,5
PE	452	1,5	100

Tabla 5.8. Frecuencias de Tramos por tipo de material

- Diámetro. Existen varios tramos con diámetro igual a cero, lo que se consideró en las estadísticas como datos perdidos. El 30.4% de los tramos tiene diámetros inferiores a 100 mm; el diámetro con mayor frecuencia es el de 100 mm con un 35.2%.

La distribución de los conductos por diámetro se encuentra en el Tabla 5.2. Tal como se dijo anteriormente el diámetro de 100 mm, es el que se encuentra en mayor proporción en la red, seguido por los diámetros menores o iguales a 70 mm.

Diámetro							
mm	(0; 70]	80	100	150	200	[225; 500]	[600; 1000]
Longitud							
mt	154.350	151.721	298.251	132.834	89.242	137.998	38.034

Tabla 5.9. Longitud total de las tuberías según los diámetros (Enero de 2006)

- Año de Instalación. La edad media de los diferentes conductos no puede ser determinada de manera precisa para toda la red, pero puede decirse que es de aproximadamente 29 años. Los datos que conciernen al año de instalación se encuentran disponibles a partir de 1940; la información anterior a esta fecha ha sido agrupada en el año 1900, lo cual, produce la impresión de ser un valor convencionalmente asignado en aquellos casos en que no se conocía la fecha exacta (por antigua) de la instalación, razón por la cual fueron eliminadas del análisis.

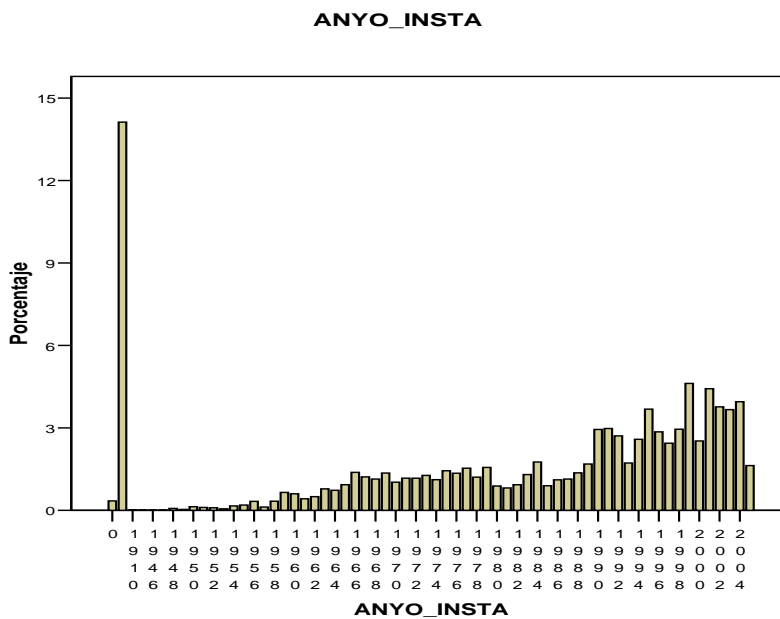


Figura 5.2 Frecuencia de tramos por año de instalación

- Tráfico. Esta variable ha sido clasificada en 3 niveles diferentes conforme a la información facilitada por la concejalía de Tráfico: tramos con tráfico acera (implica que la tubería no cruza la calzada), tramos con tráfico normal y tramos con tráfico pesado y/o intenso.

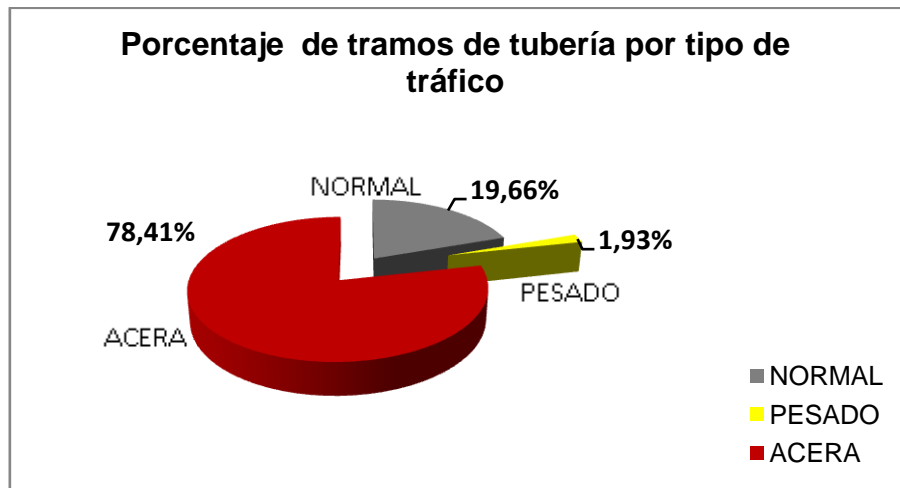


Figura 5.3. Frecuencia de tramos por año tipo de tráfico

- Presión: Las unidades de la presión están dadas en Metros de Columna de agua (m.c.a).

La presión se obtuvo tomando como punto de partida medidas de presión, y extrapolada considerando:

- estructura de la red (sectorización)
- cota sobre el nivel del mar de cada tramo.

Según la gráfica se observa que los valores de la presión van desde 10 hasta 93 m.c.a, las frecuencias más altas, las presiones comprendidas entre 30 y 50 m.c.a

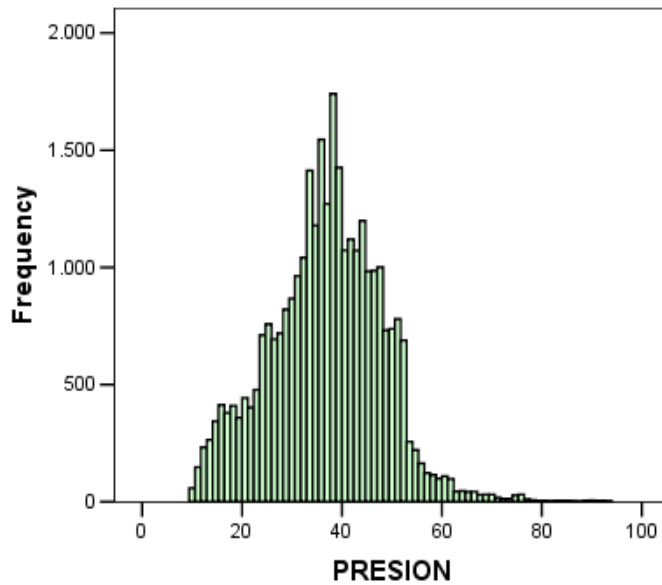


Figura 5.4. Frecuencia de tramos según los valores de la presión

Los fallos registrados están comprendidos entre los años 2000 y 2005, en total en ese periodo se produjeron 1487, distribuidos tal como se muestra en la tabla 5.9 y el gráfico de la Figura 5.4.

Año	Frecuencia	Porcentaje
2000	264	17,8
2001	312	21
2002	302	20,3
2003	143	9,6
2004	215	14,5
2005	250	16,8

Tabla 5.10 Frecuencia de fallos durante el periodo 2000-2005.

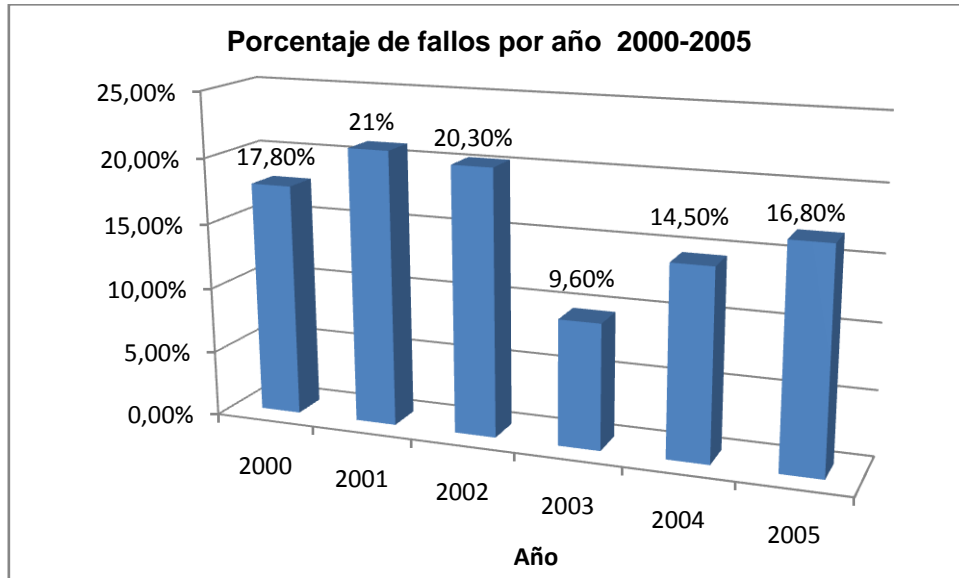


Figura 5.5 Distribución de los tramos según el año de fallo.

Los histogramas a y b muestran como se distribuye el año de instalación con respecto a los tramos fallados y no fallados respectivamente. La mayoría de los fallos ocurren entre los años 1960 y 1980; sin embargo en el año 2002 hay un pico importante de fallos, teniendo en cuenta que han sido instalados recientemente.

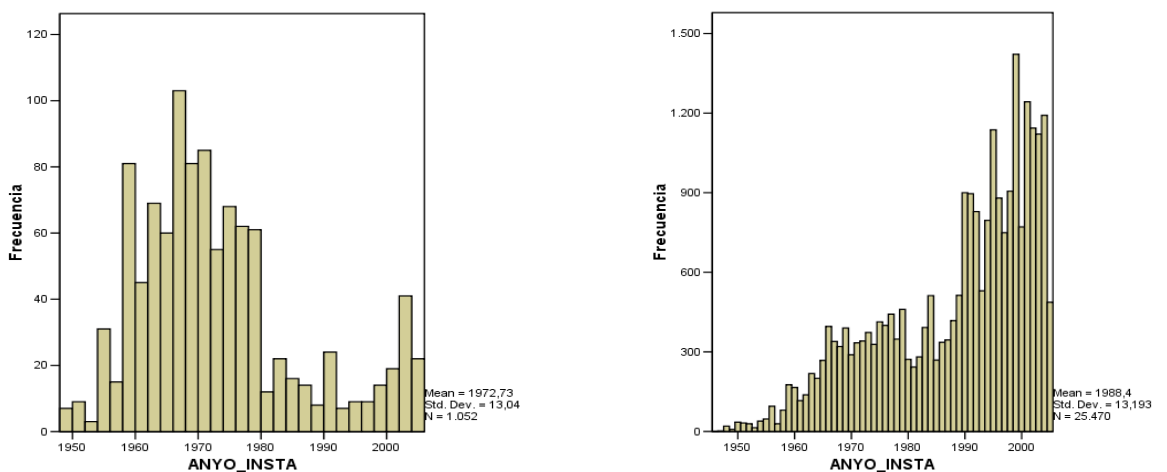


Figura 5.6 Año instalación de los tramos que fallan (izquierda) y de los tramos que no fallan (derecha)

A continuación se realizará el análisis de los fallos teniendo en cuenta el material y el año de instalación de las tuberías.

Calculando el ratio de Fugas/Km. De Red distribuido en los materiales estudiados, se tiene:

Material	Fallos	Km.	Ratio
Fundición Dúctil	317	19809,28	0,016
Fundición Gris	107	4423,8	0,0242
Polietileno	44	3346,22	0,013
Uralita	995	68639,27	0,0145

Tabla 5.11 Ratio de fugas en tuberías según el material.

Se observa que el ratio de fugas en las conducciones de Fundición gris supera el ratio de fugas en el resto de redes consideradas (0,0242).

Siguiendo con el análisis de los materiales, los siguientes gráficos muestran como se relacionan con el año de instalación y con la edad de fallo. Se observa que los materiales instalados más recientemente tienen, lógicamente, unas edades de fallo menores que los materiales más antiguos, como por ejemplo, la FG y el PE. Ello debe ser tenido en cuenta en la interpretación de los resultados que se obtendrán con posterioridad.

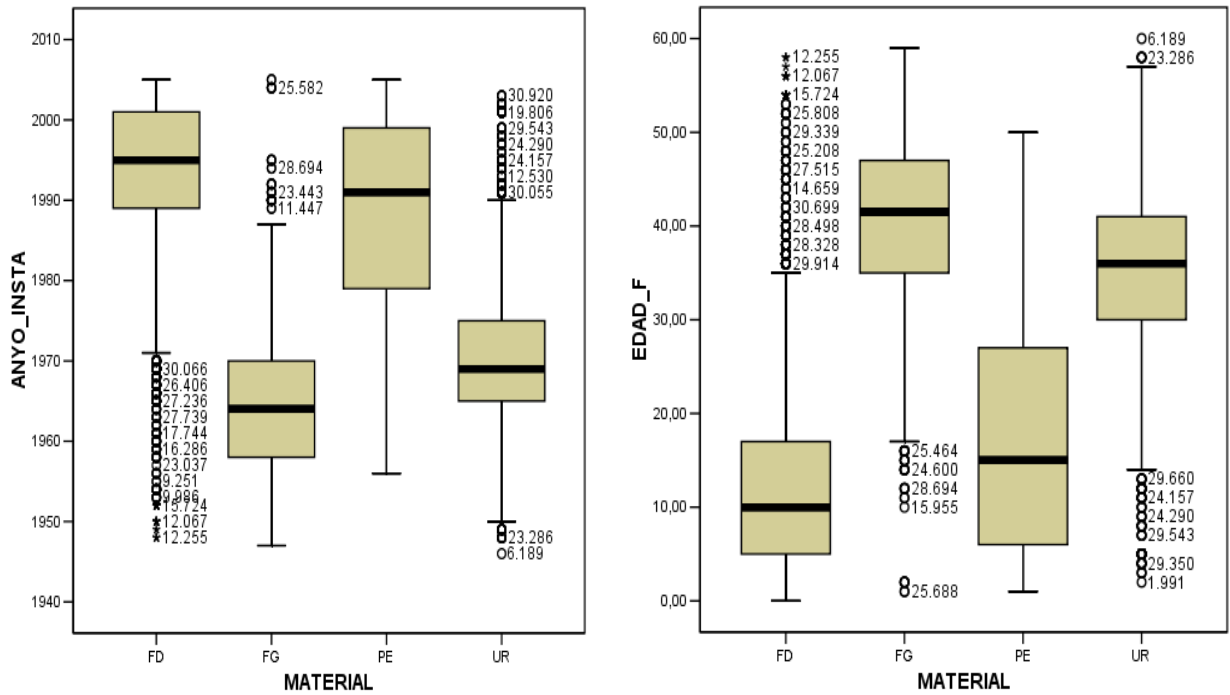


Figura 5.7 Box-Whisker para material vs año instalación⁵ (izquierda) y para material vs edad de fallo (derecha)

Así como se realizó el análisis del material según la fecha de instalación, es interesante realizar un análisis de la distribución de los fallos de acuerdo a la edad de las tuberías; para esto se identificaron las tuberías donde se han producido los fallos, según diferentes rangos de antigüedad. Los rango de edades con mayor frecuencia de fallos son las mayores de 50 años (instaladas antes de 1955) y las tuberías con edades entre 30 y 40 años (instaladas entre 1965- 1975)

⁵ Gráficos elaborados sin considerar los tramos instalados en el año 1900

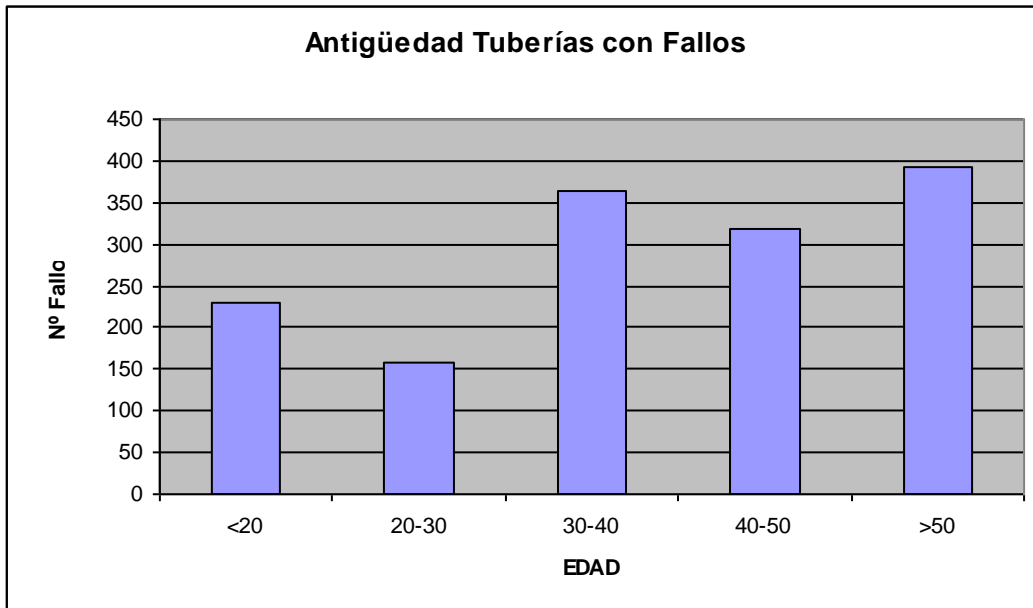


Figura 5.8 Número de fallos según rango de edades de los tramos.

Para valorar adecuadamente la influencia de la edad de las tuberías como causa de los fallos, es preciso contrastar las cifras anteriores de cantidad de fugas por rango de edad, con las longitudes de las tuberías en los diferentes rangos. Según la gráfica las edades con mayor Km de red son los tramos con edades superiores o iguales a 50 años y las tuberías con edades entre 30 y 40 años.

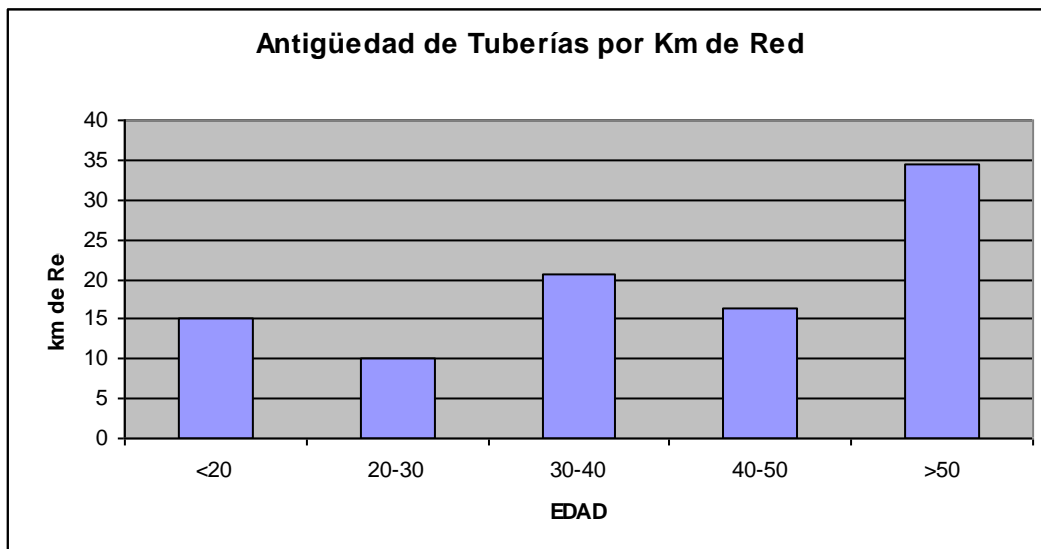


Figura 5.9. Número de Km de red según rango de edades de los tramos.

Observamos en la gráfica siguiente que los rangos de edad que concentran mayor número de fugas, también corresponden con los que agrupan mayor longitud de tuberías.

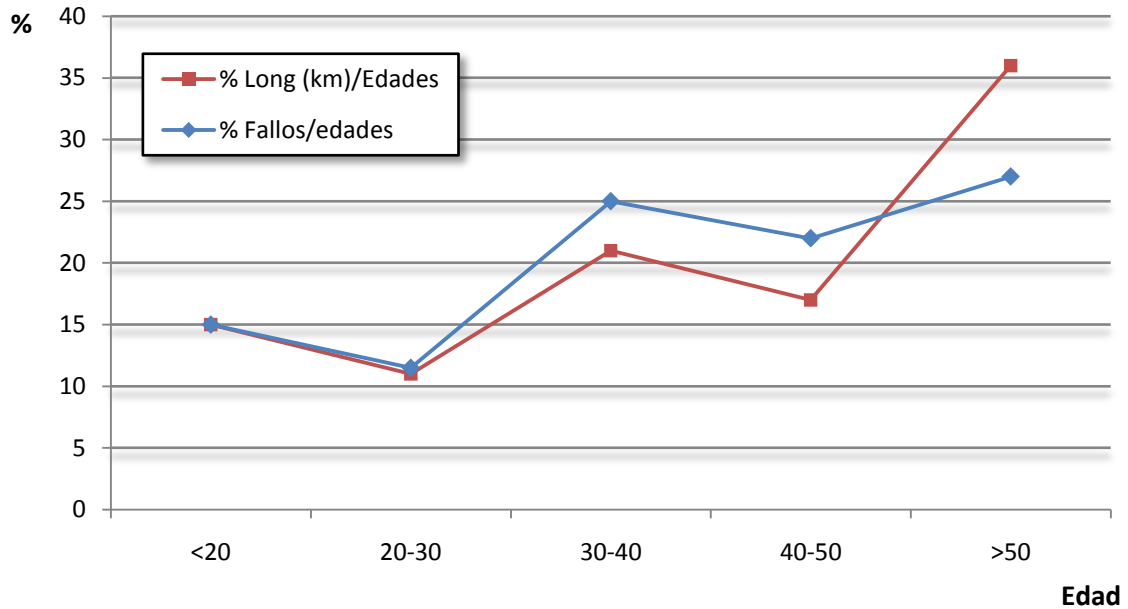


Figura 5.10. Comparación de fallos y longitud en km de red según rango de edades de las tuberías.

Por tanto, con la excepción de las instalaciones más antiguas (>50 años), la incidencia de fugas es más proporcional a la longitud de las tuberías contempladas que a la edad de las mismas.

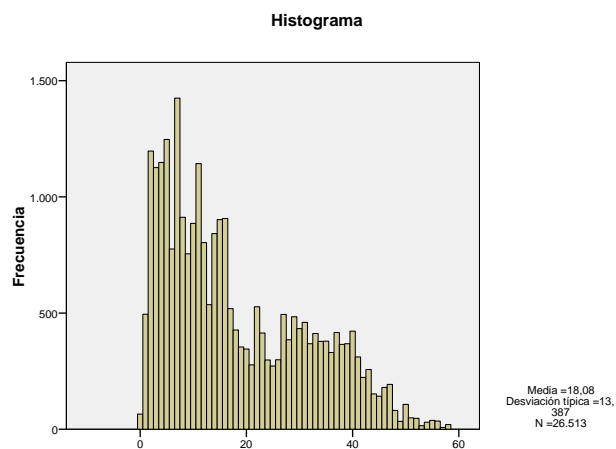


Figura 5.11. Histograma de las edades de las tuberías.

La Figura 5.11 describe el comportamiento general de la edad de las tuberías instaladas desde 1948, en la que se destacan dos “jorobas” del histograma, que se explican por los distintos tipos de material empleados en las tuberías tendidas a lo largo de 58 años de Historia y que se recogen en la Figura 5.12, que muestra la antigüedad de las tuberías por tipo de material.

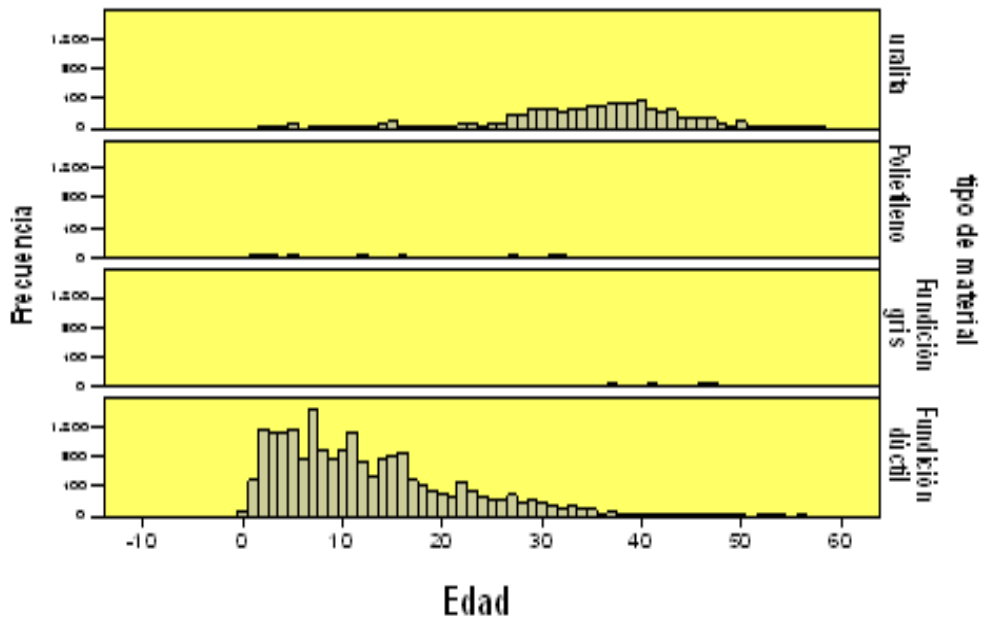


Figura 5.12 Antigüedad de las tuberías por tipo de material (1948-2005)

En conclusión, las tuberías más “jóvenes”, las tendidas en los últimos 25-30 años corresponden principalmente a aquellas cuyo material es predominantemente la fundición dúctil y el polietileno, mientras que las más antiguas (las de mayor edad en funcionamiento, las tendidas hace más de 30 años) están asociados predominantemente con tuberías de Uralita y en menor medida con las de fundición gris (estas últimas con muy poca frecuencia de uso).

El anterior comentario se ratifica con el siguiente Histograma:

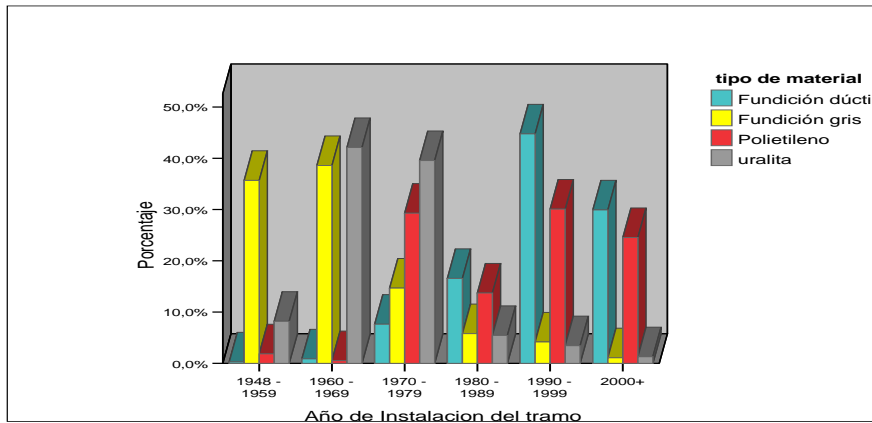


Figura 5.13 Histogramas de los años de Instalación de cuatro tipos de tubería.

La Figura 5.13 muestra la distribución de los materiales según la instalación de los distintos tramos de tubería de la red estudiada, en diferentes décadas desde 1948 hasta 2005. Se observa por ejemplo, que entre el año 48 y finales de la década de los 50's, las tuberías de las redes eran casi en un 40% de fundición gris; en los años 60's seguía predominando la fundición gris (42%) aunque ligeramente superado por la Uralita (45%). En los 70's se observa con fuerza el repunte del polietileno (30%), aunque seguían predominando las redes con tuberías de Uralita. En la década de los 80's, disminuye significativamente el levantamiento de nuevas redes y la fundición dúctil y el polietileno, comienzan a ser los más utilizados del decenio con un 18-20%. A partir de los noventa y hasta el final del estudio (2005), son la fundición dúctil y el polietileno los materiales más comúnmente utilizados en la tendida de nuevas redes de suministro.

La distribución de la edad de la tubería (en años) de la Uralita y la Fundición dúctil se recoge en la Figura 5.14

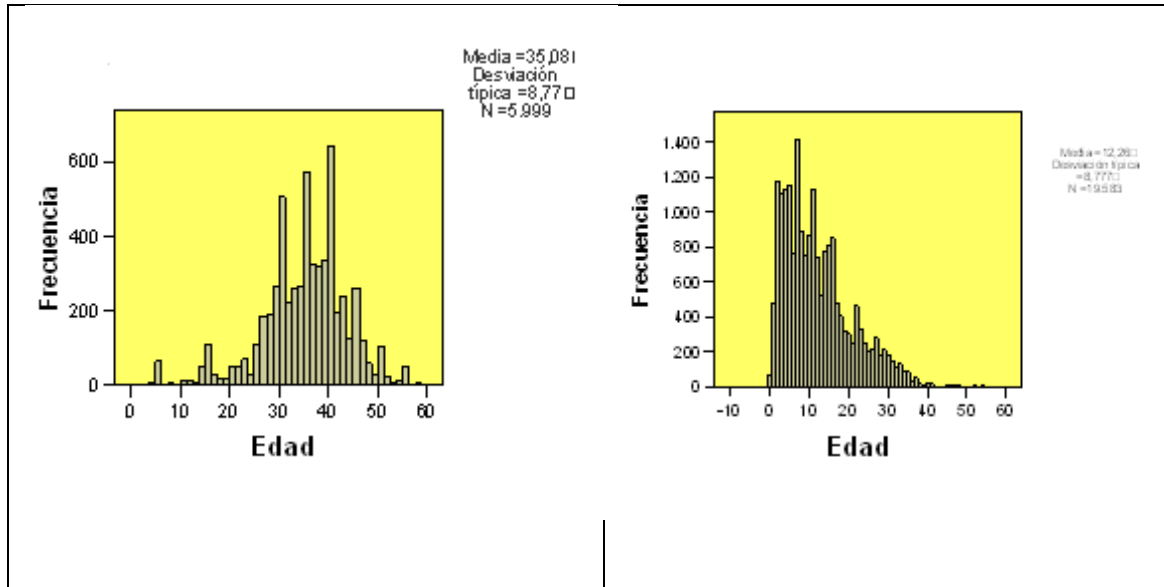


Figura 5.14 Distribución de edad de las tuberías de Uralita (izquierda) y fundición dúctil (derecha).

Que muestra el historial de la antigüedad de las tuberías en forma de histograma de dos de los cuatro tipos de materiales: uno de las más antiguas (la Uralita) y uno de las más jóvenes (la fundición dúctil). La Uralita tiene una edad media de fallos de 35 años y la fundición dúctil de 12,2 años. La Uralita alcanzó su máximo de fallos hace unos 40-42 años; mientras que la fundición dúctil que hace 40 años casi no se utilizaba, lo ha conseguido hace apenas unos 7 años (alcanzando su máxima frecuencia de aparición en este punto). El histograma de la antigüedad de las tuberías para éste último material tiene una clara asimetría, mientras que el de la Uralita tiene un comportamiento más parecido a la normal, con una ligera asimetría negativa.

Si ahora analizamos el recuento de fallos y censuras según los periodos de instalación:

Presencia de fallos		Año de instalación del tramo (Categorizada)						TOTAL % fila
		1948-1959	1960 – 1969	1970 – 1979	1980 – 1989	1990 – 1999	2000+	
tramos con censura (*)	Recuento	648	2727	3895	3627	9070	6007	25974
	% fila	2,5	10,5	15,0	14,0	34,9	23,1	100,0
	% columna	86,4	93,7	96,0	99,3	99,6	99,5	98,0
tramos fallados (**)	Recuento	102	182	164	26	37	28	539
	% fila	18,9	33,8	30,4	4,8	6,9	5,2	100,0
	% columna	13,6	6,3	4,0	,7	,4	,5	2,0
Recuento Total		750	2909	4059	3653	9107	6035	26513
% columna		2,8	11,0	15,3	13,8	34,3	22,8	100,0
TOTAL % columna		100,0	100,0	100,0	100,0	100,0	100,0	100,0

(*) censura múltiple: por fallos provocados o por censura tipo I

(**) Son tramos de tubería que han fallado por el deterioro natural del tiempo y el uso y que se han registrado entre el 2000 y 2005.

Tabla. 5.12 Fallos y Censura según período de instalación

Si se supone que los fallos y la censura se distribuyen de manera uniforme, desde las primeras instalaciones a partir del año 48; la Tabla 5.5 anterior, identifica la(s) década(s) con el mayor porcentaje de tramos de tubería con censura múltiple (incluidos los fallos provocados y la censura tipo I) este es el caso de los años 90-99, donde casi el 35% del historial de las instalaciones desde el año 48, se localizan en este período. Análogamente el período de instalación con el mayor recuento de fallos naturales, corresponde a la década del 60-69, con un porcentaje del 33,8%. En general, la mayoría de los tramos de tubería que han fallado por causas naturales (64,2%) fueron instalados entre los años 60 y 80 y la mayor parte de los tramos (el 58%) que han registrado alguno de los dos tipos de censura (por *fallo provocado* o por *censura a la derecha*) fueron instalados a partir de 1990. Se observa también en la tabla el total de tuberías con censura múltiple, (25974 de las 26513 consideradas en el

estudio, esto es el 98% de los datos) y el total de fallos naturales (2% de 26513, esto es 539 tramos registrados entre 2000 y 2005).

Por último se debe mencionar que ciertos tramos fueron eliminados de la base de datos debido a diversas circunstancias tales como:

- Año de instalación con valores errados (entre 1 y 300 y mayores a 2006),
- Años de instalación igual a 1900, dado que no se puede contrastar la autenticidad de la fecha.
- Tramos cuyo material tiene una proporción menor al 0.2% en la muestra, dado que no resulta de interés el análisis de materiales que tienen valores casi nulos.
- Edades de fallo negativas. Este hecho se presenta cuando la fecha de instalación sobrepasa la fecha de reparación.
- Diámetros iguales a cero.

5.3 Evaluación de la fiabilidad en la red

De acuerdo a las características de la red en estudio, se evaluará y analizará la fiabilidad y el estado técnico de deterioro de la misma, desde 3 ópticas diferentes de estimación. En primer lugar, con técnicas de libre distribución (Kaplan-Meier, método actuarial, Nelson-Aalen extendido.), esto es con métodos no paramétricos. El último método de estas técnicas, se aplica bajo el esquema de truncamiento por la izquierda y alta censura por la derecha

Posteriormente se evalúa la fiabilidad en la red (las tasas de fallo y la predisposición a las roturas) con modelos de riesgo proporcional (modelos semiparamétricos), particularmente adaptando y aplicando el modelo de Cox, que además se utiliza para identificar dentro del listado de características (físico ambientales) asociadas a la red, los factores de mayor influencia de riesgo en la fiabilidad de los conductos, tanto a nivel general como por estratos, según tipo de material y nivel de tráfico de las tuberías; no sin antes verificar los supuestos de proporcionalidad del modelo, estimar sus parámetros con métodos robustos como el de máxima verosimilitud y revisar el ajuste de los datos con los residuos de Cox-Snell.

Finalmente, se identifica el modelo paramétrico que mejor ajuste proporciona para el *Tiempo que tarda una tubería en romperse*. Se evaluarán para ello más de diez distribuciones de probabilidad, usando criterios gráficos (Probability Plot) y numéricos. Se estimarán también los parámetros del modelo seleccionado, utilizando dos métodos, el de máxima verosimilitud y el de mínimos cuadrados, y se valora finalmente, el ajuste del modelo a los datos con los indicadores de Anderson-Darling y de correlación de Pearson.

Los resultados de estos análisis se presentarán en el capítulo 6 (Resultados y análisis).

6. Resultados y análisis

6.1 Evaluación de la fiabilidad en la red con modelos no paramétricos

En ocasiones puede resultar ventajoso, o incluso necesario, comenzar el análisis con métodos no paramétricos, pues éstos no requieren de grandes supuestos previos sobre el modelo de las observaciones. Los modelos no paramétricos son métodos analíticos y gráficos que permiten interpretar los datos obtenidos sin la distorsión que podría causar la elección de un modelo subyacente no demasiado acertado. Sin embargo muchas veces son un paso intermedio hacia la modelación más estructurada (paramétrica) que permite profundizar más en el análisis de las observaciones.

En este estudio se consideran tres de estos métodos. Las tablas de vida y el estimador *Kaplan-Meier*, que se aplican bajo los supuestos de sólo censura por la derecha y suponiendo también que antes del inicio del estudio (año 2000) no ocurrieron fallos ni censuras y el método con el estimador de *Nelson-Aalen* extendido (NAE), que se aplica bajo condiciones de truncamiento por la izquierda y censura por la derecha.

6.1.1 Tablas de vida

Uno de los métodos no paramétricos más clásicos y directos para describir la fiabilidad de una muestra es la *tabla de supervivencia* o *actuarial*, la cual (como anteriormente se describió ampliamente) no es más que una tabla de frecuencias mejorada y ampliada. A partir de ella es posible hacer una primera estimación sobre el comportamiento de la función de supervivencia $R(t)$, de la función de distribución $F(t)$, de la función de densidad $f(t)$, y de la tasa de fallo $h(t)$.

La distribución de los tiempos de fallo como ya se dijo, se divide en un determinado número de intervalos que denotamos $(t_i, t_{i+1}]$. Para cada intervalo se registra el número de observaciones o dispositivos que han entrado en buen estado n_i (número que entra en el intervalo), el número de los que han fallado

d_i (número de eventos terminales), y el número de observaciones perdidas o censuradas en r_i (número que sale en el intervalo). Se calcula a partir de ellos el número de expuestos al riesgo, asumiendo que las pérdidas se producen homogéneamente a lo largo del mismo, su número promedio es $n'_i = n_i - 0.5r_i$. La probabilidad de fallo es la proporción $p_i = d_i/n'_i$ y la de supervivencia es $q_i=1-p_i$.

El análisis de nuestros datos bajo esta perspectiva puede consultarse en la siguiente tabla, que tiene en cuenta sólo los fallos naturales, considerando los provocados como observaciones censuradas, de la que hemos seleccionado sólo algunas columnas de la salida obtenida con SPSS.

Momento t de inicio del intervalo	Número que entra en el intervalo	Número que sale en el intervalo	Número expuesto a riesgo	Número de eventos terminales	% que sobrevive	R(t) estimado	H(t) estimado
0,0	26513	4001	24512,500	30	1,00	1,00	,00
5,0	22482	5100	19932,000	15	1,00	1,00	,00
10,0	17367	4188	15273,000	22	1,00	1,00	,00
15,0	13157	3095	11609,500	14	1,00	1,00	,00
20,0	10048	1831	9132,500	30	1,00	,99	,00
25,0	8187	1848	7263,000	86	,99	,98	,00
30,0	6253	1968	5269,000	83	,98	,96	,00
35,0	4202	1761	3321,500	97	,97	,94	,01
40,0	2344	1278	1705,000	87	,95	,89	,01
45,0	979	579	689,500	51	,93	,82	,02
50,0	349	229	234,500	19	,92	,76	,02
55,0	101	96	53,000	5	,91	,69	,02

a La mediana del tiempo de supervivencia es 55,00

Tabla 6.1 Tabla de vida general de los tramos de tubería de la red con fallos naturales

La Tabla 6.1 adjunta determina la proporción de tuberías que no han fallado (por deterioro normal de desgaste) en el momento t de inicio de un intervalo, Por ejemplo, al inicio del intervalo 40-45, se observa que un 89% de los tubos no había fallado. Se observa también, que la tasa de fallo acumulada al inicio de los 40 años es del 1%. Cómo la distribución del tiempo de fallo es habitualmente asimétrica y sesgada positivamente; por este motivo la mediana

resulta más adecuada que la media. En nuestro caso el 50% de las tuberías no falla antes de los 55 años.

La columna “ $R(t)$ estimado”, nos permite afirmar que una tubería al llegar a los 40 años tiene más posibilidades de alcanzar los 45 años (89 %), que una que alcanza los 45 y pueda llegar a los 50 (82 %); esto mismo, en forma equivalente se puede apreciar al interpretar la columna $H(t)$, donde el riesgo de fallo acumulado de un tramo de tubería, después de alcanzar los 45 años es mayor (0,02), que el riesgo de fallo que presentaría el mismo tramo después de los 40 años (0,01). Lo que se traduce en una mayor predisposición a los fallos por envejecimiento natural después de los 45 años de uso, que después de los 40 años.

Función de supervivencia de tramos de tubería con deterioro natural

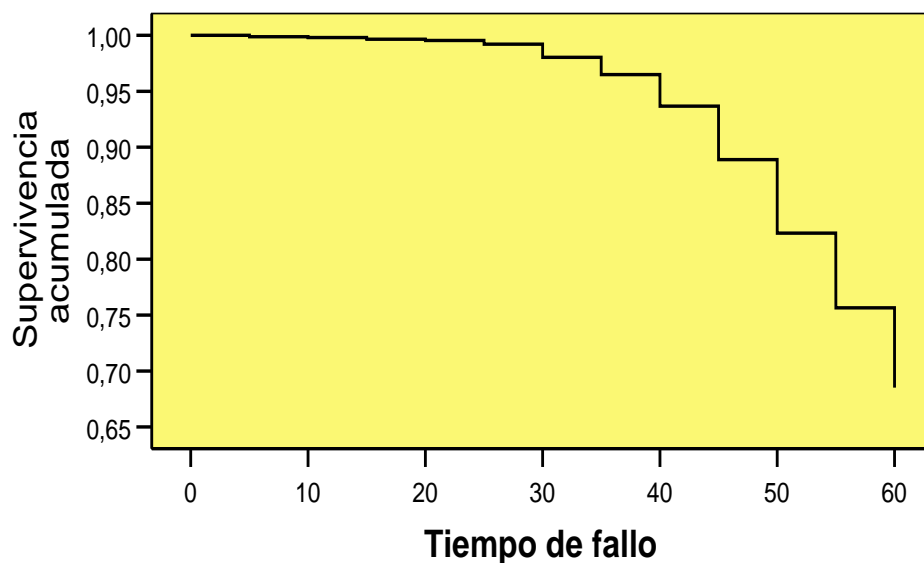


Figura 6.1 Supervivencia estimada por el método de tablas de vida

La Figura 6.1 es la “fotografía” del comentario anterior y en general de lo que ocurre (aunque con menos detalle) al interior de la Tabla 6.1

Función de impacto de la mortalidad de tramos de tubería con fallos naturales

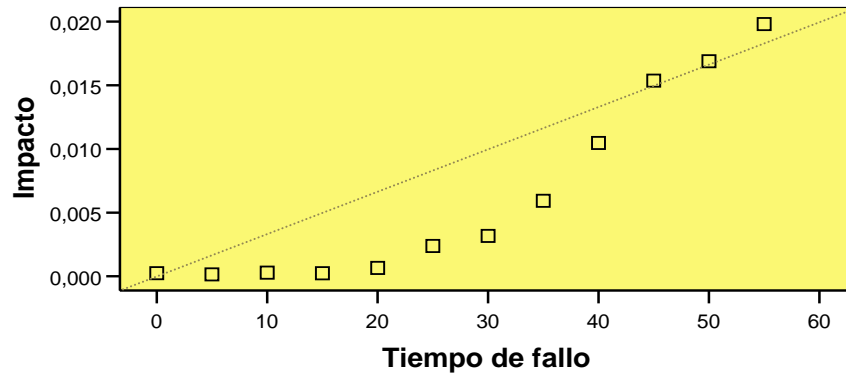


Figura 6.2 Función de Impacto de los fallos de las tuberías

En la Figura 6.2. Se observa que en los primeros 20 años, los fallos en las tuberías no son muy frecuentes, esto es, las tuberías de la red prácticamente no presentan signos de fatiga por envejecimiento u otra causa natural, pues su tasa de fallo acumulada es casi nula, sólo después de los 30 años se insinúa una predisposición creciente al fallo, hasta los 45 años donde se estabiliza alrededor del 0.02.

A continuación se muestra la misma tabla con ambos tipos de fallo: naturales y provocados.

Momento de inicio del intervalo	n_i	C_i	n_i	d_i	p_i	q_i	$R(t)$	$H(t)$
,0	26513	3940	24543,0	91	,00	1,00	1,00	,00
5,0	22482	5089	19937,5	26	,00	1,00	,99	,00
10,0	17367	4172	15281,0	38	,00	1,00	,99	,00
15,0	13157	3072	11621,0	37	,00	1,00	,99	,00
20,0	10048	1790	9153,0	71	,01	,99	,98	,00
25,0	8187	1765	7304,5	169	,02	,98	,96	,00
30,0	6253	1855	5325,5	196	,04	,96	,92	,01
35,0	4202	1672	3366,0	186	,06	,94	,87	,01
40,0	2344	1220	1734,0	145	,08	,92	,80	,02
45,0	979	567	695,5	63	,09	,91	,73	,02
50,0	349	224	237,0	24	,10	,90	,65	,02
55,0	101	95	53,5	6	,11	,89	,58	,02

Tabla 6.2. Tabla de vida de los tramos de tubería con ambos fallos

En la Tabla 6.2. se observa que la predisposición a cualquiera de los dos tipos de fallos, es de (0,02) a partir de los 40 años de instaladas las tuberías, en contraste con los resultados de Tabla 6.1 que a la misma edad es de (0,01). Además la supervivencia de los tubos después de los 40 años con los dos tipos de fallos es menor (80%) en la Tabla 6.2 (lógicamente), que la supervivencia en la Tabla 5.6, a la misma edad (89%).

Función de supervivencia de tramos de tubería con ambos fallos

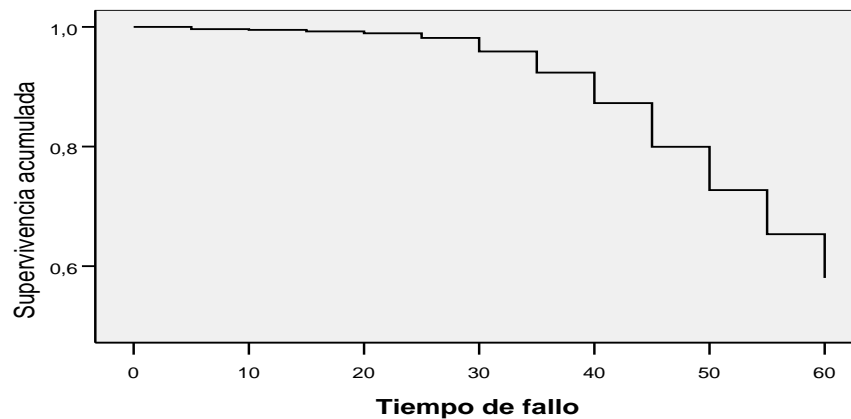


Figura 6.3. Fiabilidad con ambos fallos, por tablas de vida.

La Figura 6.3 corresponde a la representación de la Tabla 6.2, donde se observa en términos generales, que la supervivencia de los tubos expuestos a fallos naturales o provocados es mucho menor que la supervivencia de los tubos que sólo sufren fallos por envejecimiento u otra causa natural.

Función de impacto

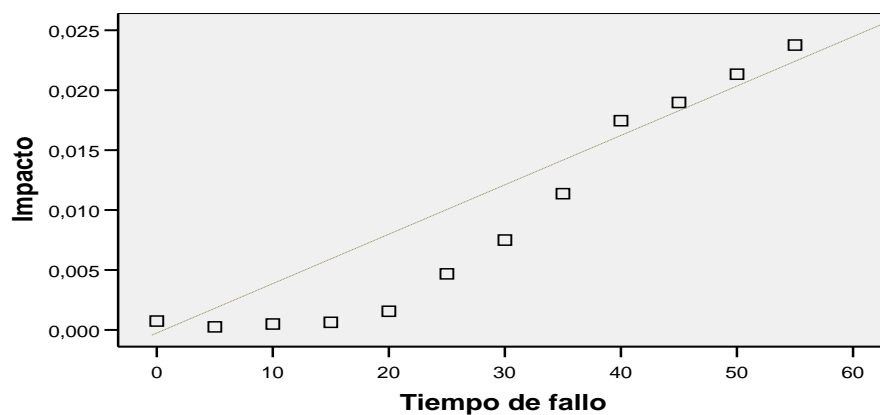


Figura 6.4. Función de impacto con ambos fallos por tablas de vida

En la Figura 6.4 la función de impacto crece más rápidamente, cuando ocurren ambos tipos de fallos, (a partir de los 25 años), que la correspondiente tasa de fallos acumulada (Figura 6.2.) cuando sólo se consideran los fallos naturales en la red, es decir, que la rapidez con que ocurren las averías en tramos expuestos a todo tipo de fallo (tanto el natural, como el provocado por diferentes factores: tráfico, invierno severo, inundaciones, etc), lógicamente se presenta más tempranamente en esta situación y con mayor rapidez (exactamente el doble, 2%) que en la aparición de los fallos por el paso del tiempo y el natural envejecimiento de los conductos. Además, a partir de los 40 años la predisposición a ambos tipos de fallo, parece estabilizarse en una tasa acumulada del 2%.

Frecuentemente, se acostumbra a estudiar la fiabilidad de las covariables por categorías (previa estratificación de la covariable correspondiente) para valorar más exhaustivamente el riesgo y la predisposición a los fallos y averías de las unidades en estudio en cada uno de estos niveles. En lo que sigue, se analizan las tablas de vida (por estratificación) de las covariables *tipo de material, longitud, diámetro y tráfico* en cada uno de sus niveles. Iniciamos el análisis de fiabilidad de la red por estratos, considerando el tipo de *material* y sus categorías.

MATERIAL	INICIO DE INTERVALO	Nº que entra en el intervalo	Nº que sale en el intervalo	Nº expuesto a riesgo	Nº de Fallos en el intervalo	% de Fallos en el intervalo	Estimación de R(t)
Fundición dúctil	,0	19583	3927	17619,5	28	,00	1,00
	5,0	15628	4965	13145,5	13	,00	1,00
	10,0	10650	4035	8632,5	10	,00	1,00
	15,0	6605	2873	17619,5	7	,00	,99
	20,0	3725	1591	2929,5	14	,00	,99
	25,0	2120	1089	1575,5	18	,01	,98
	30,0	1013	667	679,5	8	,01	,97
	35,0	338	213	231,5	1	,00	,96
	40,0	124	51	98,5	0	,00	,96
	45,0	73	36	55,0	3	,05	,91
	50,0	34	25	21,5	2	,09	,83
	55,0	7	7	3,5	0	,00	,83
Fundición gris	,0	552	6	549,0	0	,00	1,00
	5,0	546	0	546,0	0	,00	1,00
	10,0	546	9	541,5	4	,01	,99
	15,0	533	15	525,5	1	,00	,99
	20,0	517	21	506,5	0	,00	,99
	25,0	496	12	490,0	2	,00	,99
	30,0	482	59	452,5	5	,01	,98
	35,0	418	80	378,0	10	,03	,95
	40,0	328	91	282,5	14	,05	,90
	45,0	223	110	168,0	10	,06	,85
	50,0	103	61	72,5	6	,08	,78
	55,0	36	36	18,0	0	,00	,78
Polietileno	,0	379	57	350,5	1	,00	1,00
	5,0	321	55	293,5	2	,01	,99
	10,0	264	69	229,5	1	,00	,99
	15,0	194	34	177,0	1	,01	,98
	20,0	159	23	147,5	7	,05	,93
	25,0	129	46	106,0	12	,11	,83
	30,0	71	54	44,0	8	,18	,68
	35,0	9	1	8,5	0	,00	,68
	40,0	8	3	6,5	1	,15	,57
	45,0	4	1	3,5	0	,00	,57
	50,0	3	3	1,5	0	,00	,57
Uralita	,0	5999	11	5993,5	1	,00	1,00
	5,0	5987	80	5947,0	0	,00	1,00
	10,0	5907	75	5869,5	7	,00	1,00
	15,0	5825	173	5738,5	5	,00	1,00
	20,0	5647	196	5549,0	9	,00	1,00
	25,0	5442	701	5091,5	54	,01	,99
	30,0	4687	1188	4093,0	62	,02	,97
	35,0	3437	1467	2703,5	86	,03	,94
	40,0	1884	1133	1317,5	72	,05	,89
	45,0	679	432	463,0	38	,08	,82
	50,0	209	140	139,0	11	,08	,75
	55,0	58	53	31,5	5	,16	,63

Tabla 6.3 Tabla de vida por tipo de material

La tabla de vida 6.3., sugiere que el 50% de los tubos de fundición dúctil y de fundición gris, aproximadamente fallan antes de los 55 años, por causas naturales de deterioro y envejecimiento, mientras que los tubos de polietileno se deterioran antes de los 50 años por el uso y la degradación natural del paso del tiempo en las tuberías.

Material	Mediana
Fundición dúctil (FD)	55,00
Fundición gris (FG)	55,00
Polietileno (P)	50,00
Uralita (U)	55,00

Tabla 6.4. Mediana estimada del tiempo de supervivencia por material

La supervivencia estimada por tipo de material para el desgaste natural, se aprecia en la siguiente figura:

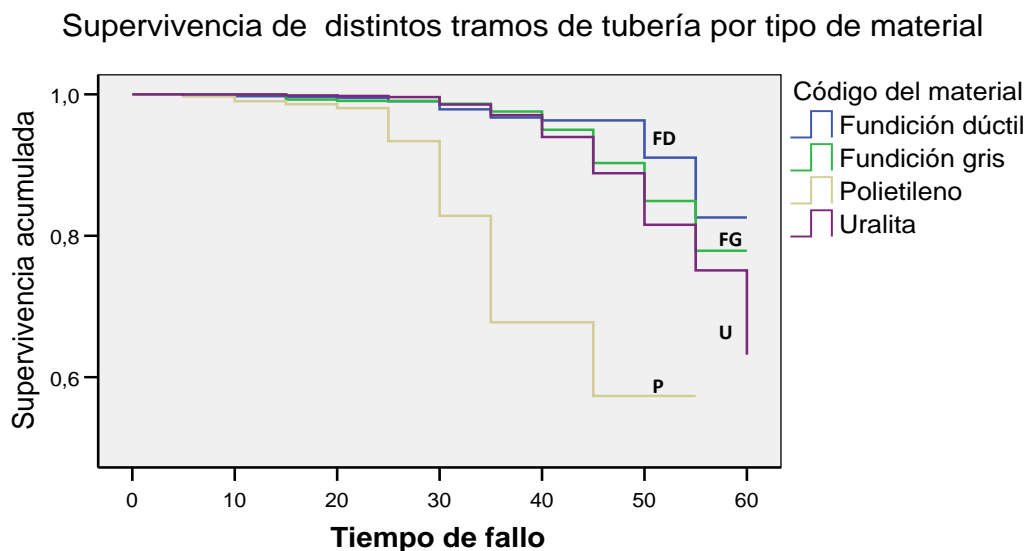


Figura 6.5 Supervivencia estimada según tipo de material

En ésta Figura 6.5 se ve claramente, que el material más “robusto” a los fallos naturales corresponde a los tramos de tubería de Fundición dúctil y el más sensible a los fallos es el Polietileno, esto se aprecia mejor a partir de los 40 años, en donde el polietileno registra un 57% contra un 96% de la F. Dúctil, se observa también una tendencia similar con los restantes materiales. Los tramos de tubería de polietileno, evidentemente están más expuestos a los fallos por degradación y envejecimiento natural que el resto de los materiales.

Por tanto los tramos que más fallan por tipo de material son los fabricados en Polietileno y los que menos lo hacen (por degradación natural) son los de fundición dúctil. La Fundición gris y la Uralita, presentan una supervivencia al desgaste natural, muy similar, lo mismo que su predisposición a los fallos.

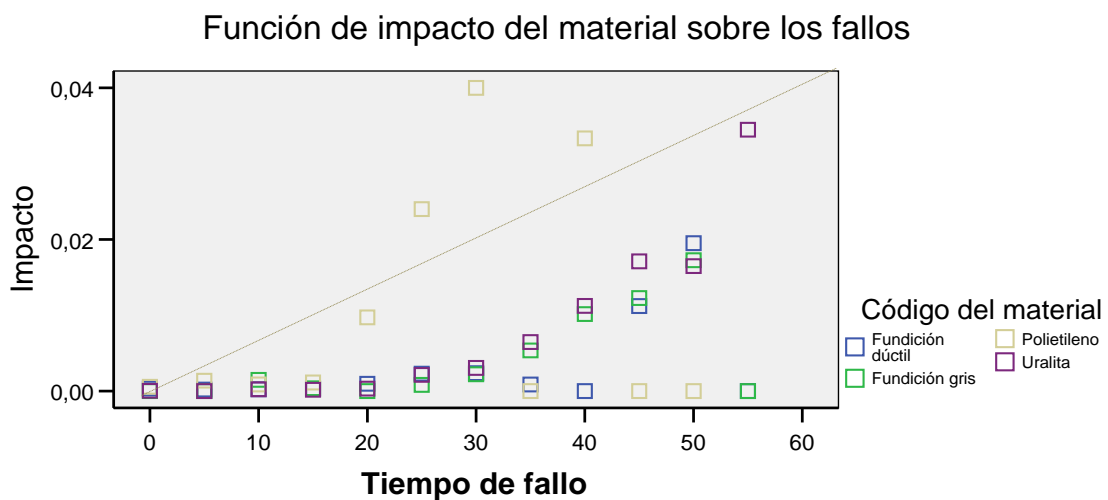


Figura 6.6 Tasa acumulada de fallos según tipo de material

La función de impacto de los materiales, destaca el aumento considerable de la tasa de fallos del Polietileno entre los 20 y los 30 años, con algunos altibajos entre los 35 y los 50 años de deterioro natural de la tubería. En general es el único material de la lista con estas características. Los restantes materiales (F. gris, F. dúctil y Uralita) a partir de los 30 años manifiestan su tendencia creciente de predisposición a los fallos por envejecimiento natural de las tuberías.

La tabla de vida por longitud de tramo es la siguiente:

LONGITUD	Momento de inicio del intervalo	Número que entra en el intervalo	Número que sale en el intervalo	Número expuesto a riesgo	d_i	Proporción que sobrevive	$R(t)$	$H(t)$
<2,0	,0	10100	1792	9204,0	0	1,00	1,00	,00
	5,0	8308	2205	7205,5	1	1,00	1,00	,00
	10,0	6102	1859	5172,5	0	1,00	1,00	,00
	15,0	4243	1250	3618,0	0	1,00	1,00	,00
	20,0	2993	672	2657,0	1	1,00	1,00	,00
	25,0	2320	620	2010,0	0	1,00	1,00	,00
	30,0	1700	560	1420,0	2	1,00	1,00	,00
	35,0	1138	515	880,5	0	1,00	1,00	,00
	40,0	623	379	433,5	1	1,00	1,00	,00
	45,0	243	170	158,0	0	1,00	1,00	,00
	50,0	73	53	46,5	0	1,00	1,00	,00
55,0	20	20	10,0	0	1,00	1,00	,00	
2,0 – 9,9	,000	4317	596	4019	6	1,00	1,00	0,0
	5,0	3715	781	3324,5	2	1,00	1,00	,00
	10,0	2932	717	2573,5	0	1,00	1,00	,00
	15,0	2215	573	1928,5	0	1,00	1,00	,00
	20,0	1642	228	1528,0	2	1,00	1,00	,00
	25,0	1412	324	1250,0	4	1,00	,99	,00
	30,0	1084	369	899,5	9	,99	,98	,00
	35,0	706	297	557,5	7	,99	,97	,00
	40,0	402	196	304,0	11	,96	,94	,01
	45,0	195	125	132,5	8	,94	,88	,01
	50,0	62	44	40,0	2	,95	,84	,01
55,0	16	15	8,5	1	,88	,74	,03	
10,0 – 49,9	,000	7374	931	6908,5	15	1,00	1,00	,00
	5,0	6428	1286	5785,0	9	1,00	1,00	,00
	10,0	5133	1020	4623,0	12	1,00	,99	,00
	15,0	4101	715	3743,5	2	1,00	,99	,00
	20,0	3384	519	3124,5	12	1,00	,99	,00
	25,0	2853	581	2562,5	40	,98	,97	,00
	30,0	2232	658	1903,0	37	,98	,95	,00
	35,0	1537	610	1232,0	44	,96	,92	,01
	40,0	883	471	647,5	45	,93	,86	,01
	45,0	367	214	260,0	20	,92	,79	,02
	50,0	133	85	90,5	10	,89	,70	,02
55,0	38	37	19,5	1	,95	,67	,01	
50,0+	,00	4722	682	4381,0	9	1,00	1,00	,00
	5,0	4031	828	3617,0	3	1,00	1,00	,00
	10,0	3200	592	2904,0	10	1,00	,99	,00

Tabla 6.5 Tabla de vida por longitud del tramo

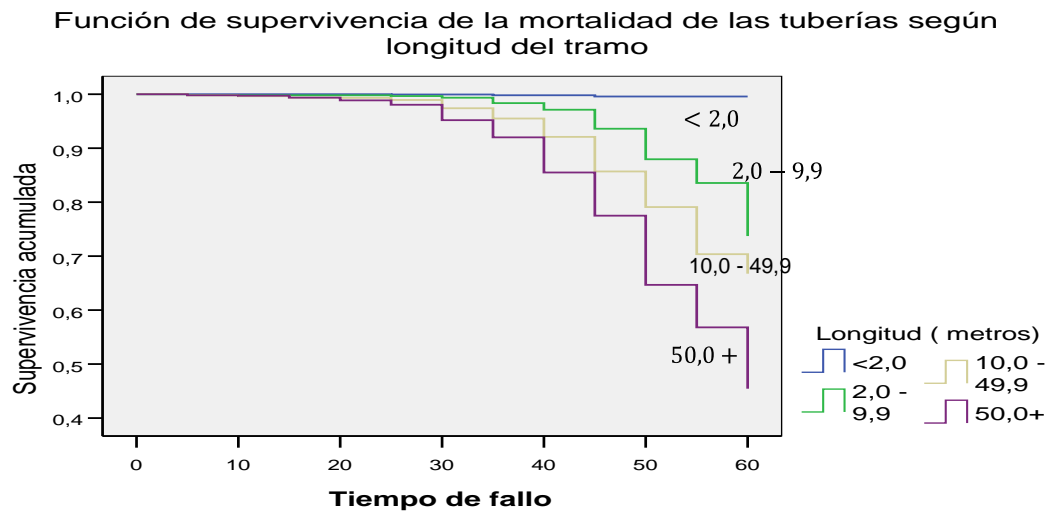


Figura 6.7 Supervivencia de las tuberías según longitud de tramo

La Figura 6.7 es la función de supervivencia de las tuberías según la longitud de los tramos, sugiere que la longitud tiene un fuerte impacto en el análisis de los fallos y en particular, podríamos decir, que entre más cortos son los tramos de tubería de la red, mayor es la "robustez" al deterioro natural y viceversa, entre más largos los tramos, más vulnerables al mismo se insinúan.

De modo que los tramos más expuestos al deterioro natural (a partir de los 30 años) son los de más de 50 metros de longitud.

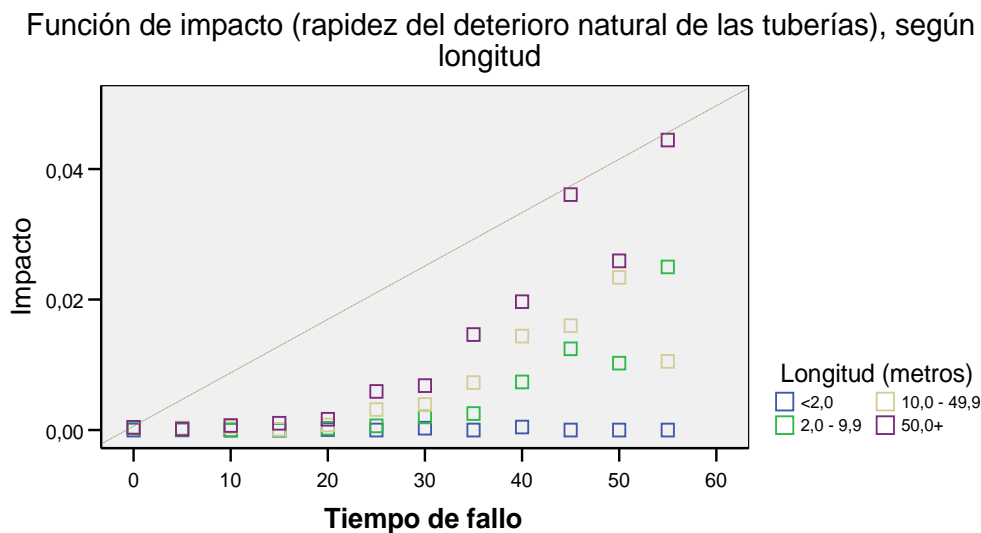


Figura 6.8 Función de Impacto según Longitud de tramo.

En cuanto a la predisposición al fallo o rapidez del deterioro por degradación natural, podríamos decir que casi, los únicos tramos de tubería que muestran una tasa de fallo creciente (a partir de los 40 años) son las de 50 o más metros de longitud, o sea que a partir de los 40 años de envejecimiento natural de la tubería, los tramos más largos, presentan mayor riesgo de fallo (y sólo se estabilizan a partir de los 45 años de funcionamiento) en contraste con los de menos de 2,0 m, cuya tasa de fallo se mantiene prácticamente nula durante toda su vida útil.

A continuación, se muestra la tabla de vida de los conductos según su diámetro (Tabla 6.6)

DIAMETRO DE LA TUBERIA (mm)	Momento de inicio del intervalo	Número que entra en el intervalo	Número expuesto a riesgo	Número de eventos terminales	Proporción que sobrevive	$R(t)$	$H(t)$
<= 90	,000	7403	7334,500	2	1,00	1,00	,00
	5,000	7264	7129,500	2	1,00	1,00	,00
	10,000	6993	6827,500	6	1,00	1,00	,00
	15,000	6656	6418,000	10	1,00	1,00	,00
	20,000	6170	5795,500	25	1,00	,99	,00
	25,000	5396	4780,000	73	,98	,98	,00
	30,000	4091	3472,000	62	,98	,96	,00
	35,000	2791	2229,500	67	,97	,93	,01
	40,000	1601	1164,000	73	,94	,87	,01
	45,000	654	451,000	33	,93	,81	,02
	50,000	215	140,000	10	,93	,75	,01
	55,000	55	29,500	4	,86	,65	,03
91 – 175	,000	14532	13016,500	23	1,00	1,00	,00
	5,000	11478	9632,500	12	1,00	1,00	,00
	10,000	7775	6222,500	13	1,00	,99	,00
	15,000	4657	3648,000	3	1,00	,99	,00
	20,000	2636	2221,000	4	1,00	,99	,00
	25,000	1802	1563,000	10	,99	,99	,00
	30,000	1314	1052,500	18	,98	,97	,00
	35,000	773	581,500	11	,98	,95	,00
	40,000	379	276,000	8	,97	,92	,01
	45,000	165	114,000	7	,94	,87	,01
	50,000	56	47,000	2	,96	,83	,01
	55,000	36	18,500	1	,95	,78	,01
176 – 300	,000	3691	3382,000	4	1,00	1,00	,00
	5,000	3069	2582,500	1	1,00	1,00	,00
	10,000	2095	1814,000	2	1,00	1,00	,00
	15,000	1531	1316,500	1	1,00	1,00	,00
	20,000	1101	981,500	1	1,00	1,00	,00
	25,000	861	800,000	3	1,00	,99	,00
	30,000	736	641,000	3	1,00	,99	,00
	35,000	543	454,500	18	,96	,95	,01
	40,000	348	251,500	6	,98	,93	,00
	45,000	149	118,500	11	,91	,84	,02
	50,000	77	47,000	7	,85	,71	,03
	55,000	10	5,000	0	1,00	,71	,00
301+	,00	887	779,500	1	1,00	1,00	,00
	5,000	671	587,500	0	1,00	1,00	,00
	10,000	504	409,000	1	1,00	1,00	,00
	15,000	313	227,000	0	1,00	1,00	,00
	20,000	141	134,500	0	1,00	1,00	,00
	25,000	128	120,000	0	1,00	1,00	,00
	30,000	112	103,500	0	1,00	1,00	,00

Tabla 6.6 Tabla de Vida de los fallos de las tuberías según su diámetro (mm.)

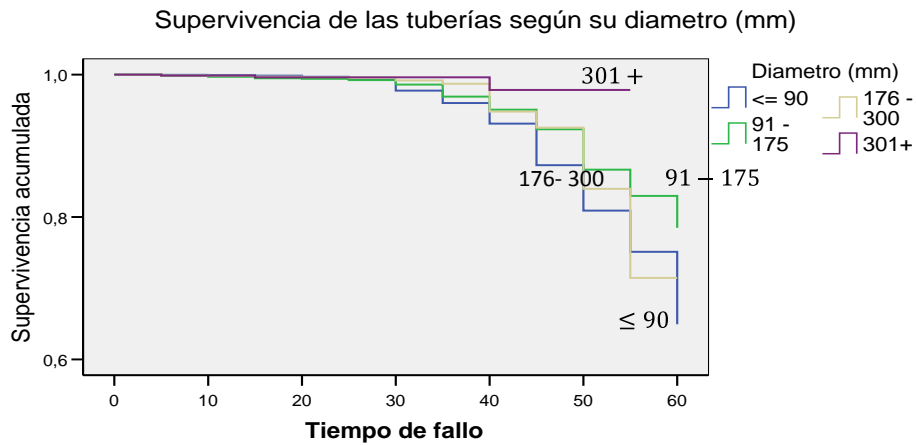


Figura 6.9 Función de Supervivencia según Diámetro

La función de supervivencia de la tabla de vida de las tuberías (según diámetro del tramo), sugiere que la diámetro tiene algún impacto en el análisis de los fallos y en particular, podríamos decir, que entre más ancho es el diámetro de los conductos de la red, más "robusto" es su comportamiento al deterioro natural y viceversa, entre más angosto el tramo, más vulnerable a los fallos es el mismo. Se aprecia por ejemplo, que los tramos con diámetro superior a los 300 mm, a partir de los 50 años mantienen una supervivencia superior al 95 %, en contraste con los tramos de menos de 90 mm, cuya supervivencia está cercana al 80%, para los mismos 50 años de funcionamiento.

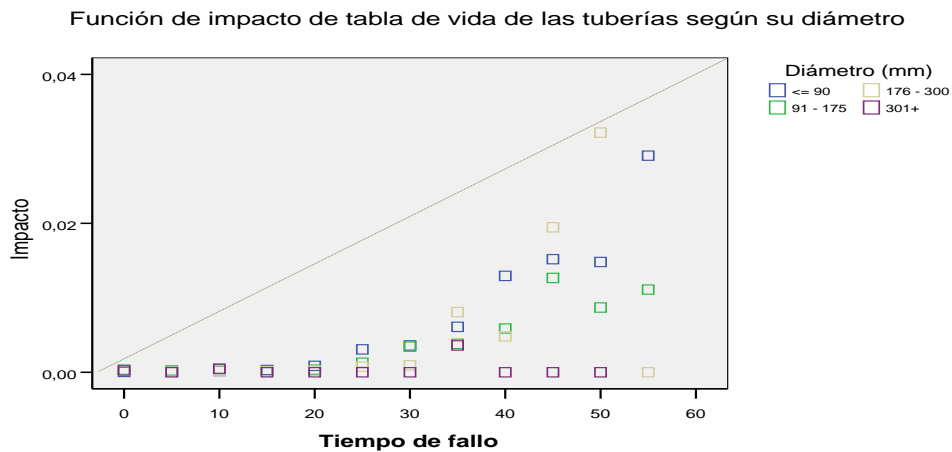


Figura 6.10 Función de impacto según Diámetro.

En cuanto a la rapidez del deterioro natural, se puede decir, que los tramos de tubería que presentan una mayor tasa de fallo creciente (a partir de los 45 años) son los de diámetro comprendido entre 176- 300 mm, o sea que a partir de los 45 años de envejecimiento natural de la tubería, los tramos con dicho diámetro presentan fallos y averías cada con cada vez más frecuencia, en contraste con los tramos de más de 300 mm, donde la velocidad de aparición de los fallos propios del envejecimiento (o predisposición al fallo natural) se mantienen muy estables, con niveles casi nulos durante su vida útil.

Finalmente mostramos la función de supervivencia y la tabla de vida para cada una de las categorías de tráfico:

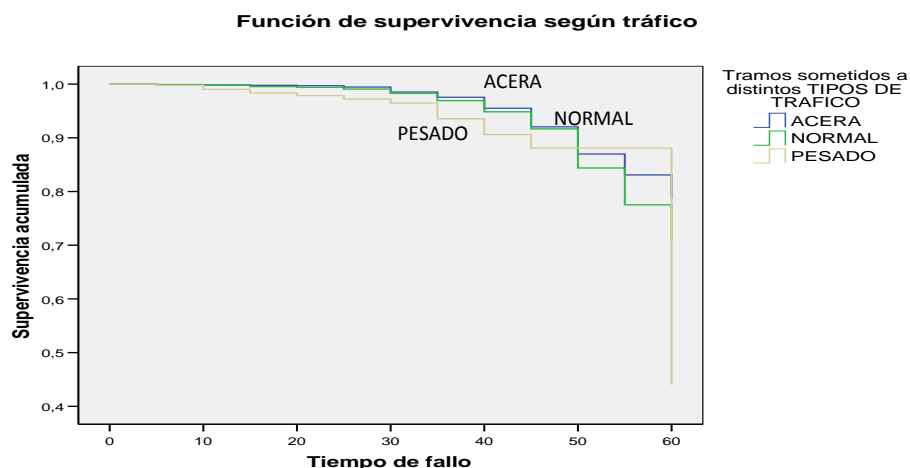


Figura 6.11 Supervivencia de la tuberías según tráfico

La Figura 6.11 sugiere, que para cualquier instante anterior a los 50 años, las tuberías instaladas en zonas de intenso tráfico pesado, sobreviven en menor porcentaje que los restantes tramos de tuberías instaladas en zonas de tráfico normal o acera, es decir, que la mayor parte de los tubos sometidos al tráfico pesado perdura menos tiempo que las tuberías instaladas en zonas de menor intensidad de tráfico, por ejemplo, transcurridos 35 años, sólo sobrevivirán el 91% de los tramos de tubería sometidos a tráfico pesado,

mientras que este porcentaje (para el mismo instante) llega al 95 % para tubos sometidos a

Tramos sometidos a distintos TIPOS DE TRAFICO	Momento de inicio del intervalo	Número que entra en el intervalo	Número que sale en el intervalo	Número expuesto a riesgo	Número de eventos terminales	$R(t)$	$H(t)$
ACERA	,0	20693	3297	19044,500	18	1,00	,00
	5,0	17378	4037	15359,500	10	1,00	,00
	10,0	13331	3265	11698,500	8	1,00	,00
	15,0	10058	2394	8861,000	6	1,00	,00
	20,0	7658	1450	6933,000	18	,99	,00
	25,0	6190	1507	5436,500	50	,99	,00
	30,0	4633	1471	3897,500	39	,98	,00
	35,0	3123	1371	2437,500	51	,96	,00
	40,0	1701	970	1216,000	44	,92	,01
	45,0	687	429	472,500	26	,87	,01
	50,0	232	152	156,000	7	,83	,01
	55,0	73	71	37,500	2	,79	,01
NORMAL	,0	5021	627	4707,500	8	1,00	,00
	5,0	4386	953	3909,500	1	1,00	,00
	10,0	3432	836	3014,000	8	1,00	,00
	15,0	2588	609	2283,500	3	,99	,00
	20,0	1976	348	1802,000	6	,99	,00
	25,0	1622	300	1472,000	12	,98	,00
	30,0	1310	427	1096,500	15	,97	,00
	35,0	868	338	699,000	15	,95	,00
	40,0	515	258	386,000	13	,92	,01
	45,0	244	135	176,500	14	,84	,02
	50,0	95	67	61,500	5	,78	,02
	55,0	23	22	12,000	1	,71	,02
PESADO	,0	483	39	463,500	1	1,00	,00
	5,0	443	108	389,000	3	,99	,00
	10,0	332	82	291,000	2	,98	,00
	15,0	248	81	207,500	1	,98	,00
	20,0	166	26	153,000	1	,97	,00
	25,0	139	20	129,000	1	,96	,00
	30,0	118	39	98,500	3	,94	,01
	35,0	76	24	64,000	2	,91	,01
	40,0	50	28	36,000	1	,88	,01
	45,0	21	11	15,500	0	,88	,00
	50,0	10	7	6,500	0	,88	,00
	55,0	3	2	2,000	1	,44	,13

Tabla 6.7 Tabla de vida para el Tiempo de fallo de tuberías según Tráfico

tráfico normal y a un 96 % para las tuberías que han sido instaladas en zonas peatonales y/o de acera (también después de los 35 años de servicio). Note que el 88% de los tubos instalados por debajo de calles y zonas de mayor intensidad de tráfico habrán fallado entre los 40 y los 50 años. Recordemos que al hablar de fallos en los tubos, estamos haciendo alusión a la aparición de filtraciones de agua en la superficie, resquebrajamiento de cimientos en las calles, etc.

Función de impacto para tuberías según tráfico

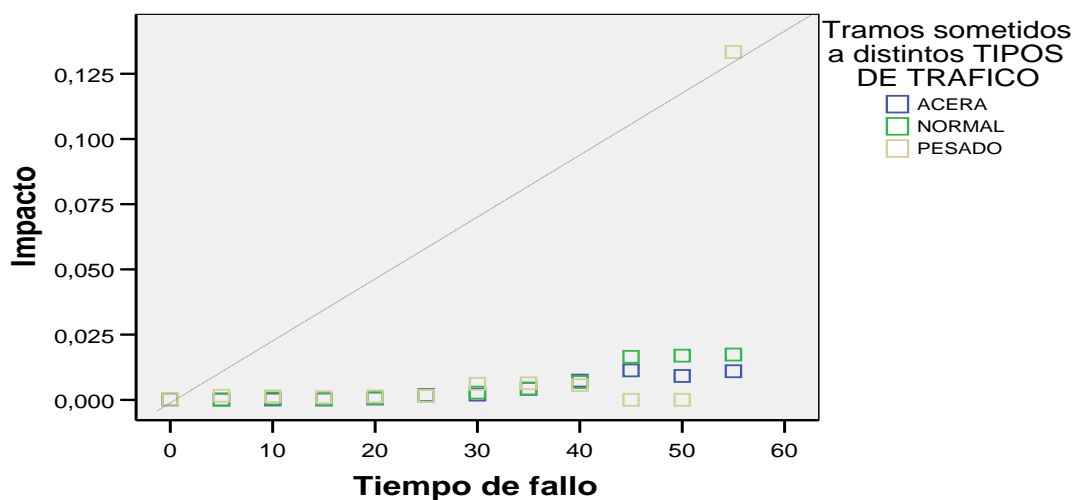


Figura 6.12 Función de riesgo acumulada para las tuberías según tráfico

El riesgo de fallo (tasa instantánea de fallo) de cualquier tramo de tubería para un instante particular t , teóricamente es la pendiente de $H(t)$ en esa edad t , en realidad es la probabilidad condicional de fallo de las tuberías a partir de t , suponiendo que un poco antes de t las tuberías instaladas en una determinada zona (de intensidad de tráfico) no habían fallado. En la práctica nos interesa observar cómo cambia la pendiente de $H(t)$ a través del tiempo, como indicativo de la forma como cambia el riesgo de fallo (rapidez con que ocurren los fallos) de los distintos tipos de tubería.

En particular la Figura 6.12 sugiere que un poco antes de los 55 años, el riesgo de fallo en los tres tipos de tuberías era más o menos el mismo (casi ninguno había fallado), sólo después de los 55 años, el riesgo de fallo en los tramos de tráfico pesado, se “disparó” abruptamente en más de 6 veces, (con relación a las otras dos instalaciones), al pasar de un 1% o 2% a un 13 % (ver tabla de vida).

6.1.2 Estimaciones con el enfoque de Kaplan- Meier

Un estimador para la función de supervivencia fue propuesto por Kaplan y Meier (1958), se conoce como el estimador producto-límite o el estimador de Kaplan-Meier

Si t_i es el tiempo hasta el fallo ordenado, n_i el número de unidades que quedaban en riesgo justo antes del fallo i -ésimo, el número de los que han fallado d_i , el estimador de Kaplan-Meier viene dado por la expresión,

$$\hat{R}(t) = \prod_{j: t_j < t} \frac{n_j - d_j}{n_j}$$

La tabla 6.8, que se presenta a continuación, corresponde a la estimación (puntual y por intervalos del 95% de confianza) de la Fiabilidad por envejecimiento natural de diferentes tramos de tubería, en presencia de una fuerte censura (98%), por el método de Kaplan- Meier, (1958), estimaciones que se consideran globales, sin precisar, si es por la influencia de alguna de las principales características físicas y/o ambientales de los tubos.

Number Time	Number at Risk	Failed	Survival Probability	Standard Error	Lower	Upper
0	26513	19	0,999283	0,0001643	0,998961	0,999605
1	26448	3	0,99917	0,0001769	0,998823	0,999517
2	25953	5	0,998978	0,0001967	0,998592	0,999363
3	24756	1	0,998937	0,0002008	0,998544	0,999331
4	23630	2	0,998853	0,0002095	0,998442	0,999263
5	22482	3	0,998719	0,0002231	0,998282	0,999157
6	21235	4	0,998531	0,0002421	0,998057	0,999006
7	20459	3	0,998385	0,0002564	0,997882	0,998887
8	19034	3	0,998227	0,000272	0,997694	0,998761
9	18122	2	0,998117	0,0002829	0,997563	0,998672
10	17367	3	0,997945	0,0002998	0,997357	0,998533
11	16481	5	0,997642	0,0003289	0,996997	0,998287
12	15338	5	0,997317	0,0003595	0,996612	0,998022
13	14535	3	0,997111	0,0003786	0,996369	0,997853
14	13999	6	0,996684	0,0004167	0,995867	0,9975
15	13157	2	0,996532	0,0004302	0,995689	0,997375
17	11348	2	0,996357	0,0004477	0,995479	0,997234
18	10829	4	0,995988	0,0004838	0,99504	0,996937
19	10402	6	0,995414	0,0005374	0,994361	0,996467
20	10048	4	0,995018	0,0005726	0,993896	0,99614
21	9703	6	0,994402	0,0006249	0,993178	0,995627
22	9426	7	0,993664	0,0006839	0,992324	0,995004
23	8899	9	0,992659	0,0007608	0,991168	0,99415
24	8485	4	0,992191	0,0007957	0,990632	0,993751
25	8187	20	0,989767	0,0009607	0,987884	0,99165
26	7915	11	0,988392	0,0010451	0,986343	0,99044
27	7616	18	0,986056	0,0011788	0,983745	0,988366
28	7122	15	0,983979	0,0012925	0,981446	0,986512
29	6737	22	0,980766	0,0014586	0,977907	0,983624
30	6253	18	0,977942	0,001599	0,974808	0,981076
31	5820	19	0,97475	0,0017535	0,971313	0,978187
32	5360	15	0,972022	0,0018848	0,968328	0,975716
33	4992	13	0,969491	0,0020064	0,965558	0,973423
34	4580	18	0,96568	0,0021903	0,961388	0,969973
35	4202	18	0,961544	0,0023881	0,956863	0,966224
36	3823	20	0,956513	0,0026271	0,951364	0,961663
37	3493	14	0,95268	0,0028093	0,947174	0,958186
38	3077	30	0,943391	0,0032538	0,937014	0,949769
39	2712	15	0,938174	0,0035036	0,931307	0,94504
40	2344	16	0,93177	0,003828	0,924267	0,939272
41	1922	29	0,917711	0,0045747	0,908744	0,926677
42	1611	11	0,911444	0,0049181	0,901805	0,921084

43	1388	25	0,895028	0,0058233	0,883615	0,906441
44	1131	6	0,89028	0,0061065	0,878311	0,902248
45	979	21	0,871183	0,0072595	0,856955	0,885411
46	837	14	0,856611	0,0081157	0,840705	0,872518
47	657	11	0,842269	0,009059	0,824514	0,860024
48	464	1	0,840454	0,0092195	0,822384	0,858524
49	383	4	0,831676	0,010114	0,811853	0,851499
50	349	12	0,80308	0,0126958	0,778197	0,827963
51	242	2	0,796443	0,0134303	0,77012	0,822766
52	193	3	0,784063	0,0150034	0,754657	0,813469
53	146	1	0,778693	0,0158326	0,747661	0,809724

Tabla 6.8 Estimación de la Fiabilidad por KM puntualmente y por intervalos

Se observa por ejemplo, que la probabilidad de que un tubo de cualquier tipo de material o longitud, falle por deterioro natural después de los 53 años es de 77,87 %; también la tabla permite estimar por intervalos la proporción de tubos que por ejemplo, a comienzo de los 53 años no había fallado por causas naturales, esta proporción se estima entre 0,75 y 0,81.con una confianza del 95%.

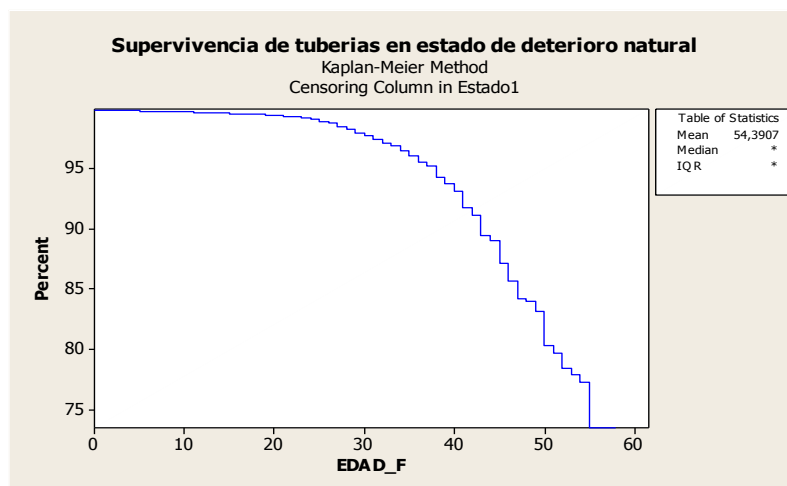


Figura 6.13 Supervivencia de las tuberías por fallos naturales.

El tiempo medio hasta el fallo (MTTF) de los distintos tramos de tubería, se estima en un poco más de los 54 años. Además el % de tramos que a los 55

años no había fallado es un poco más del 78 %, como se observa en la Figura 6.13

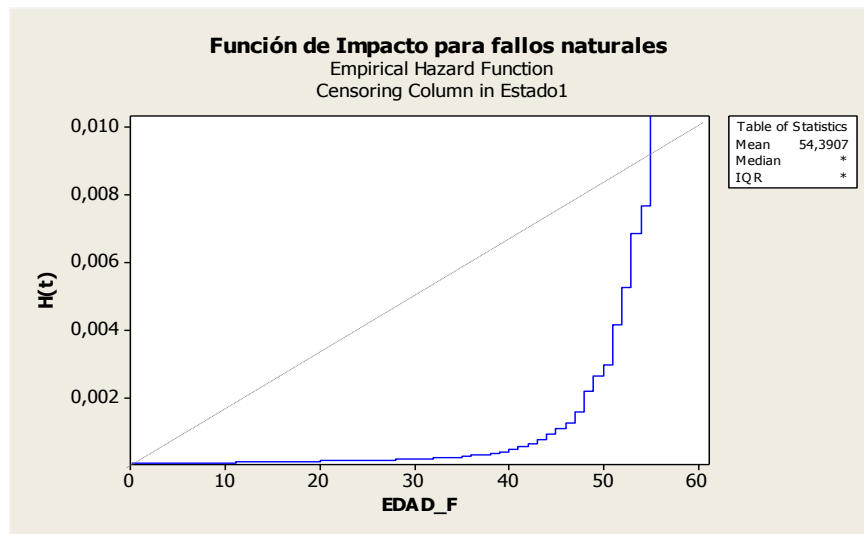


Figura 6.14 Función de Impacto para fallos naturales.

En la Figura 6.14 la tasa de fallos a partir de los 40 años, empieza a crecer, para los tramos de tubería que no están sometidos a factores perturbadores del deterioro natural de envejecimiento de la red (inundaciones, invierno severo, irregularidades en el tráfico, etc.). A partir de los 55 años se “dispara” con gran rapidez el proceso natural de deterioro y envejecimiento de la tubería, lo que se traduce en un mayor riesgo de fallo y por lo tanto, en una mayor predisposición a las averías y filtraciones y por tanto en una mayor necesidad de mantenimiento correctivo (reparaciones, sustituciones, etc.)

Podemos obtener también el estimador Kaplan-Meier para diferentes estratos.

Si consideramos el tipo de material, la tabla 6.9 presenta los siguientes resultados,

Material	Nº de tramos de tubería	Nº de fallos	Edad Media estimada	Censurados	
				Nº	%
Fundición dúctil	19583	104 (0,53%)	55,835	19479	99,5%
Fundición gris	552	52 (9,4%)	54,925	500	90,6%
Polietileno	379	33 (8,7%)	41,967	346	91,3%
Uralita	5999	350 (5,8%)	54,368	5649	94,2%
Global	26513	539 (2,0%)	54,391	25974	98,0%

Tabla 6.9 Número de fallos y censuras por tipo de material

La tabla 6.9 recoge la estratificación de los fallos naturales de las tuberías de la red, por tipo de material, incluyendo el nivel porcentual de censura para cada uno. El material menos utilizado es el Polietileno y el de más frecuente uso en las tuberías es la Fundición dúctil, que también presenta un mayor % de censura. Se destaca también que los materiales con mayor % de fallos son la Fundición gris y el Polietileno 9,4% y 8,7% respectivamente y el de menos % de fallos es la fundición dúctil con un 0,53%:

Funciones de supervivencia de las tuberías según material

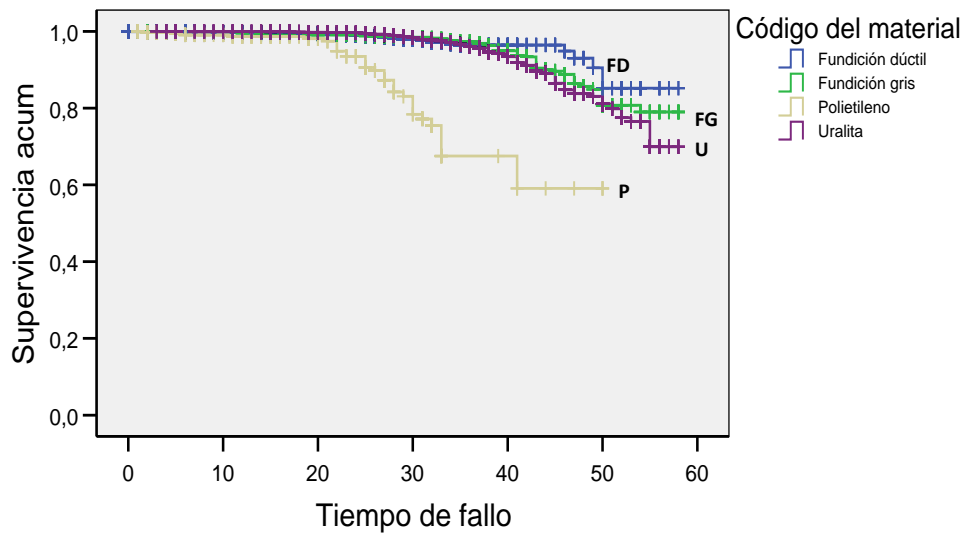


Figura 6.15. Función de fiabilidad por tipo de material.

En la Figura 6.15, se aprecia claramente que la mayor tendencia a los fallos corre por cuenta del Polietileno donde con más frecuencia aparecen los fallos y en contraste, la fundición dúctil es la que presenta mayor resistencia al fallo. La diferencia entre el Polietileno y las restantes queda bien marcada a partir de los 20 años, de modo que, en el Polietileno, la probabilidad de fallo (o de supervivencia) de un tramo de tubería de este material después de 40 años está muy próxima al 58-60%. En general el material con menos posibilidades de duración, a pesar de ser el más “joven” después de los 20 años es el polietileno.

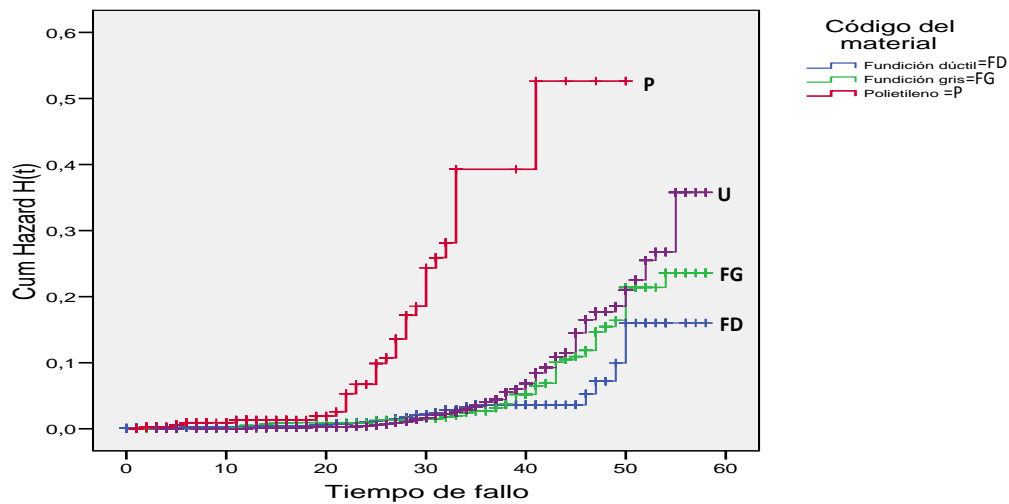
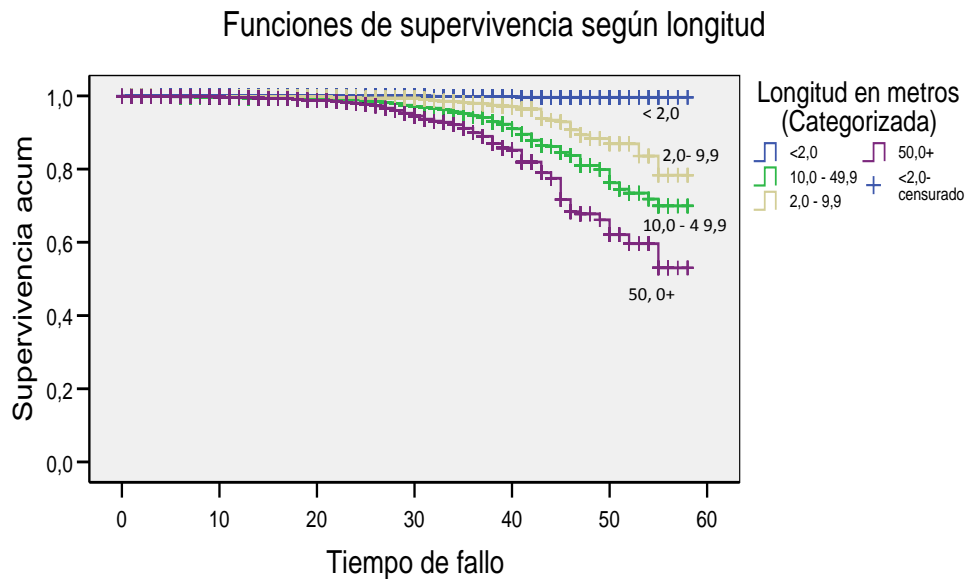


Figura 6.16 Tasa de fallo acumulada para tuberías según tipo de material.

En la Figura 6.16 se aprecia la tendencia marcadamente creciente de la tasa de fallos acumulada del Polietileno, a partir de los 20 años de modo que, después de los 30 años este tipo de material comienza a presentar fallos con mucha frecuencia, obligando a un mantenimiento correctivo/preventivo más frecuente. Los restantes materiales manifiestan una marcada tendencia a sufrir averías, rupturas, escapes, filtraciones, etc, es decir, mayor predisposición a los fallos sólo a partir de los 40 años de uso.

Los resultados de la fiabilidad por longitud del tramo de tubería, según KM se muestran a continuación en la siguiente página, en la Figura 6.17.



En la figura adjunta, se aprecia claramente que la mayor tendencia a los fallos corre por cuenta de las tuberías de más de 50 metros de longitud, son los tramos donde los fallos aparecen más tempranamente y en contraste, los tramos de menos de 2,0 metros, son los que presentan mayor resistencia al fallo. La diferencia entre estas dos longitudes se empieza a notar a partir de los 20 años, de modo que, la probabilidad de fallo en el tramo de mayor longitud a partir de los 50 años está cercana al 50%, según la Figura 6.17. Más exactamente (sin suponer que los datos puedan ajustarse por una determinada distribución teórica), el gráfico lo que nos dice es que hay una sustancial diferencia en la supervivencia de las tuberías de distintas longitudes, por ejemplo, a partir de los 55 años la mayor parte de las tuberías de menos de 2,0 metros (prácticamente todas), tienen una supervivencia del 100%, frente a una supervivencia cercana al 50% para las tuberías de más de 50 metros a la misma edad de los 55 años.

Funcion de impacto para tramos de diferente longitud

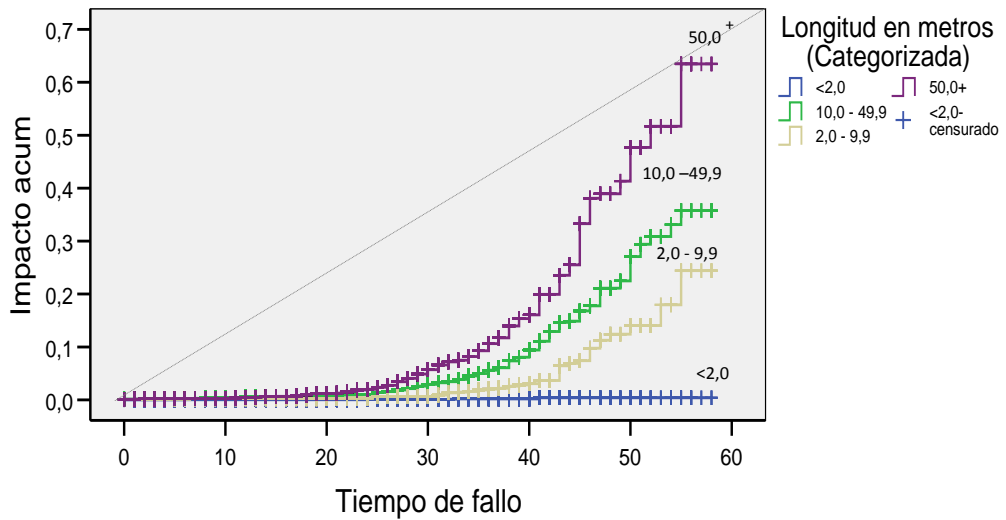


Figura 6.18 Función de Impacto según Longitud

En la Figura adjunta, solo las longitudes de más de 50 metros presentan realmente un preocupante riesgo de deterioro como resultado de su envejecimiento, lo cual se manifiesta marcadamente a partir de los 55 años.

Finalmente en la tabla A de medianas, se observa:

Longitud en metros	Medias
Menos de 2,0	57,908
2,0 – 9,9	55,640
10,0 – 49,9	53,595
Más de 50,0	50,847
Global	54,391

Tabla 6.10 Duración media de los conductos por categoría de Longitud

Se observa en la Tabla 6.10 que los tramos de menor longitud tienen una mayor duración media, que en este caso corresponde a los tubos de menos de

dos metros (57,9 años); siguen en orden de duración decreciente, los tramos de 2 a 10 metros de longitud, cuya duración media es de 55,6 años y finalmente los tramos de más de 50 metros son los que tienen un vida media de 50,8 años.

La supervivencia de las tuberías según el tráfico puede observarse en la siguiente Figura:

Funciones de supervivencia de tuberías según Tráfico

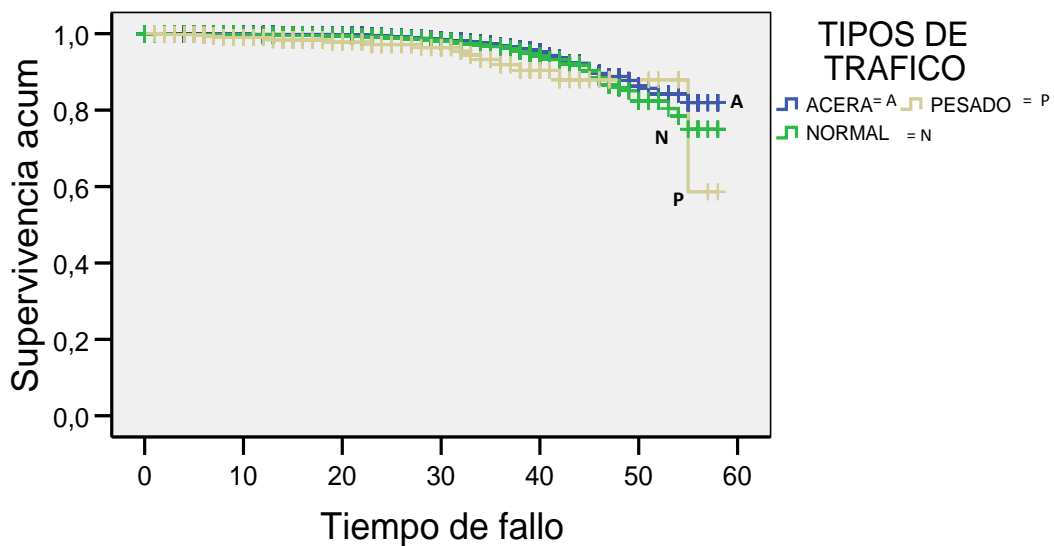


Figura 6.19 Supervivencia de tramos de tubería según su Tráfico

La Figura 6.19 sugiere que la influencia del tipo de Tráfico en los fallos no es significativa antes de los 55 años de servicio de las tuberías, excepto a partir de los 55-56 años, en donde se aprecia claramente el efecto del tráfico pesado sobre la alteración del envejecimiento natural de las tuberías y en donde las posibilidades de supervivencia o probabilidad de fallo para este tipo de tubería (la de tráfico pesado) a partir de los 58-60 años son muy inferiores (60%) frente a las posibilidades de supervivencia (77-80 %) del resto de tubos instalados en zonas de menor intensidad de tráfico (tráfico normal y acera), donde el nivel de “stress”, es muy inferior.

Funcion de impacto según Tráfico

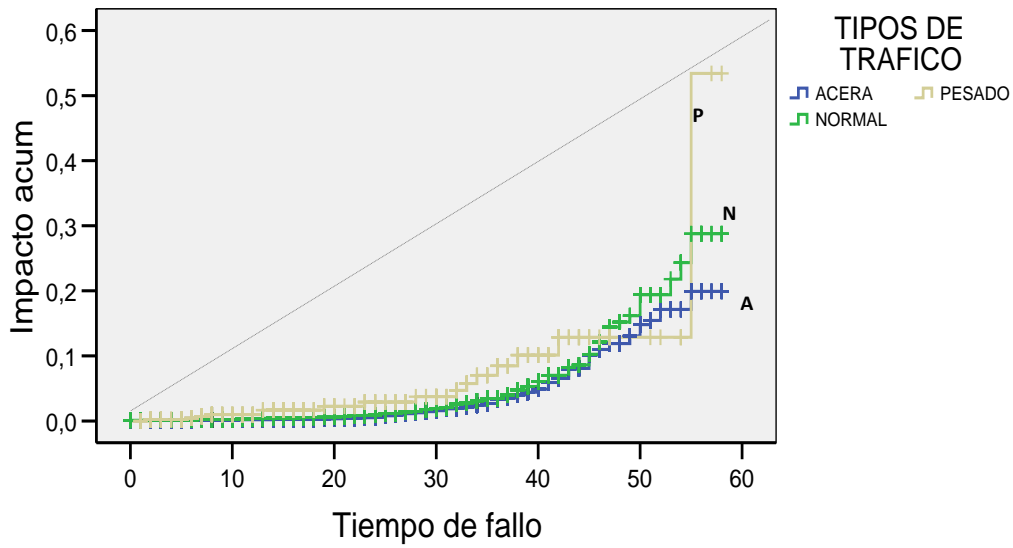


Figura 6.20 Función de Impacto de las tuberías según su tráfico

En la Figura 6.20 se observa que solo los tramos de tubería sometidos a tráfico pesado, manifiestan una tendencia claramente “agresiva” en su tasa de fallo, al pasar abruptamente de una tendencia moderadamente creciente en su predisposición al fallo, entre los 20 y los 45 años, a una situación de crecimiento abruptamente *desmesurado* a partir de los 55 años de servicio, acelerándose de esta manera su proceso de envejecimiento y deterioro natural a partir de este instante.

6.1.3 Estimación de la fiabilidad de la red, con datos truncados por la izquierda y con censura por la derecha, utilizando el método de NAE

En esta primera fase del estudio, debido al corto historial de fallos disponible, sobre el comportamiento del fenómeno que nos ocupa (la estimación de la fiabilidad de los conductos en la red en condiciones de alta censura), se adoptaron básicamente dos enfoques en la aplicación de los métodos de libre distribución utilizados en el desarrollo de este estudio.

En una primera aproximación para el análisis de los datos, se aplicaron las tablas de vida y el método de KM, bajo el esquema de la fuerte censura durante el período del estudio (bien sea por fallos provocados o por censura tipo I u otra similar), con el supuesto de que *no ocurrieron fallos en los conductos de la red antes del año 2000*.

Una segunda estrategia, no descarta del todo la hipótesis de una posible ocurrencia de roturas y censura antes del inicio del estudio, (después de la instalación de varios de los conductos de la red), sólo que, no se tiene ninguna certeza sobre ello, (ni siquiera la más leve sospecha) y, es esta la información que el investigador desconoce y desde luego descarta para su estudio. En este caso, sólo se registran los instantes de fallo T de las tuberías que son observadas durante el período de la referencia y que al año 2000, tenían una edad X desde el año de su instalación, y que además cumplen la condición $T \geq X$, es decir, sólo se consideran tramos de tubería con instantes de fallo T posteriores a X .

La información muestral proporcionada por un tubo que cumple la anterior condición, (en el período 2000-2005), se considera truncada a la izquierda. Por tanto, el estimador que mejor describe esta situación, es el estimador de *Nelson-Aalen extendido* (NAE), que fue ya definido en el capítulo 4.

La tabla 6.11 que se observa a continuación, describe la estimación puntual de la fiabilidad, la tasa de fallos y la función de riesgo acumulado, de diferentes tramos de tubería en forma global y generalizada. Los resultados se obtuvieron con el estimador de *Nelson-Aalen extendido* (NAE).

6.1.3.1 Estimación global de la fiabilidad

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
0,5	19	46	6035	0,003148	0,003148	0,996857
1	3	492	7399	0,000405	0,003554	0,996453
2	5	1192	7816	0,000640	0,004193	0,995815
3	1	1125	7372	0,000136	0,004329	0,995680
4	2	1146	7131	0,000280	0,004610	0,995401
5	3	1244	7124	0,000421	0,005031	0,994982
6	4	772	6678	0,000599	0,005630	0,994386
7	3	1422	6433	0,000466	0,006096	0,993923
8	3	909	5843	0,000513	0,006609	0,993412
9	2	753	5846	0,000342	0,006952	0,993073
10	3	883	5996	0,000500	0,007452	0,992576
11	5	1138	5628	0,000888	0,008340	0,991694
12	5	798	4906	0,001019	0,009359	0,990684
13	3	533	4457	0,000673	0,010033	0,990018
14	6	836	4262	0,001408	0,011440	0,988625
15	2	900	3695	0,000541	0,011982	0,988090
16	0	907	3315	0,000000	0,011982	0,988090
17	2	517	2813	0,000711	0,012693	0,987388
18	4	423	2584	0,001548	0,014241	0,985860
19	6	348	2410	0,002490	0,016730	0,983409
20	4	341	2330	0,001717	0,018447	0,981722
21	6	271	2476	0,002423	0,020870	0,979346
22	7	520	2577	0,002716	0,023587	0,976689
23	9	405	2533	0,003553	0,027140	0,973225
24	4	294	2540	0,001575	0,028714	0,971694
25	20	252	2701	0,007405	0,036119	0,964525
26	11	288	2779	0,003958	0,040077	0,960715
27	18	476	2876	0,006259	0,046336	0,954721
28	15	370	2755	0,005445	0,051781	0,949537
29	22	462	2748	0,008006	0,059787	0,941966
30	18	415	2594	0,006939	0,066726	0,935452
31	19	441	2588	0,007342	0,074067	0,928609
32	15	353	2492	0,006019	0,080086	0,923037
33	13	399	2517	0,005165	0,085251	0,918281
34	18	360	2550	0,007059	0,092310	0,911822
35	18	361	2468	0,007293	0,099604	0,905196
36	20	310	2321	0,008617	0,108221	0,897430
37	14	402	2251	0,006219	0,114440	0,891865
38	30	335	2000	0,015000	0,129440	0,878587

39	15	353	1767	0,008489	0,137929	0,871161
40	16	406	1594	0,010038	0,147967	0,862460
41	29	282	1394	0,020803	0,168770	0,844703
42	11	212	1198	0,009182	0,177952	0,836983
43	25	232	1012	0,024704	0,202656	0,816559
44	6	146	857	0,007001	0,209657	0,810863
45	21	121	769	0,027308	0,236965	0,789019
46	14	166	680	0,020588	0,257553	0,772941
47	11	182	517	0,021277	0,278830	0,756669
48	1	80	353	0,002833	0,281663	0,754528
49	4	30	305	0,013115	0,294777	0,744697
50	12	95	314	0,038217	0,332994	0,716775
51	2	47	221	0,009050	0,342044	0,710317
52	3	44	193	0,015544	0,357588	0,699361
53	1	14	146	0,006849	0,364437	0,694588
54	1	29	131	0,007634	0,372071	0,689306
55	5	33	101	0,049505	0,421576	0,656012
56	0	35	63	0,000000	0,421576	0,656012
57	0	8	28	0,000000	0,421576	0,656012
58	0	20	20	0,000000	0,421576	0,656012

Tabla 6.11 estimación de la fiabilidad de los conductos de la red con el método de NAE

Se observa por ejemplo, que la probabilidad de que un tramo de tubería (independiente de sus características físico ambientales, esto es, tipo de material, longitud, etc.) falle por causas naturales del uso y el paso del tiempo, antes de los 40 años es del 13,75 %, es decir, la fiabilidad del tubo, después de los 40 años es del 86,25% y la fiabilidad de un tubo a los 50 años se estima según la tabla en 0,7168 (71,68%), valor mucho menor que a los 40 años.

También, la tabla permite estimar, la rapidez con que se presentan las roturas, fallos y averías a distintas edades de un tramo de tubería (la rapidez instantánea con que se deteriora o degrada la tubería); en este caso a los 40 años de edad, la tasa de fallos de una tubería se estima en 0,010038 (1,0% aproximadamente), lo que quiere decir, que la probabilidad de que un tubo que no fallado hasta los 40 años, se rompa o falle a los 41 años es del 1,0 %.

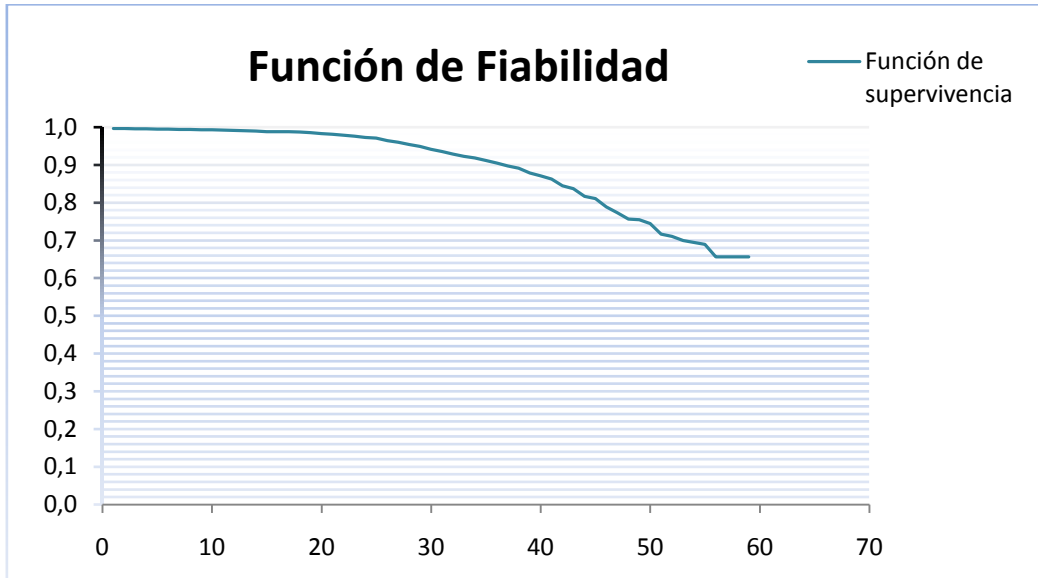


Figura 6.21 Estimación general de la Fiabilidad en la red, utilizando el método de NAE

La Figura 6.21 corresponde a la representación de la Tabla 6.11, donde se observa en términos generales, que la supervivencia de los tubos que se deterioran y fallan por causas naturales del paso del tiempo, tienen un rango de variación que oscila entre el 0,66 y 1,0. Se observa también que sólo a partir de los 28 – 30 años, empieza a decaer la fiabilidad de los conductos de la red, situándose por debajo del 0,95, es decir, a partir de estos años se empieza a notar el deterioro de la red, que hasta ese instante presenta una tasa de fallo acumulada del 6,0% (0,0597)aproximadamente. Se observa también que la fiabilidad de las tuberías en general por encima de los 58 años de operación de la red, es del 65,6 % ; valor que está por debajo de las correspondientes estimaciones en este mismo instante de tiempo, con los dos primeros métodos.

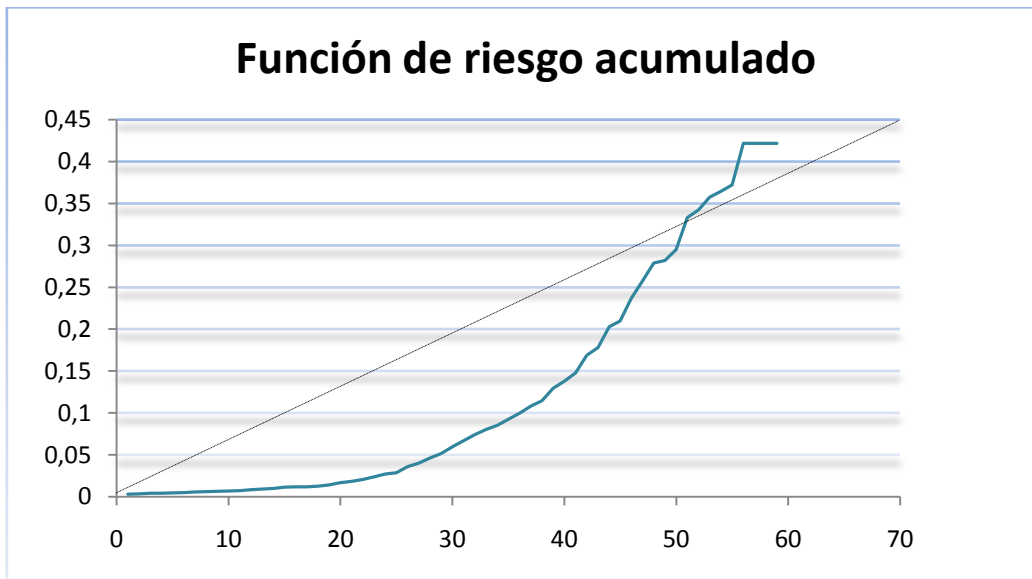


Figura 6.22 Estimación global del riesgo acumulado en la red, utilizando el método de NAE

En la Figura 6.22. se observa que la predisposición acumulada al fallo en los primeros 20 años es casi nula, no llega al 2,0%, es decir, las roturas y averías en la red en estos primeros años no son muy frecuentes, sus tuberías prácticamente no presentan signos de fatiga por envejecimiento u otra causa natural, sólo después de los 30 años, se insinúa una marcada predisposición al fallo en forma sostenida y continúa, de manera creciente hasta los 50 años de operación, con tendencia a decrecer o al menos a estabilizarse después de los 55 años.

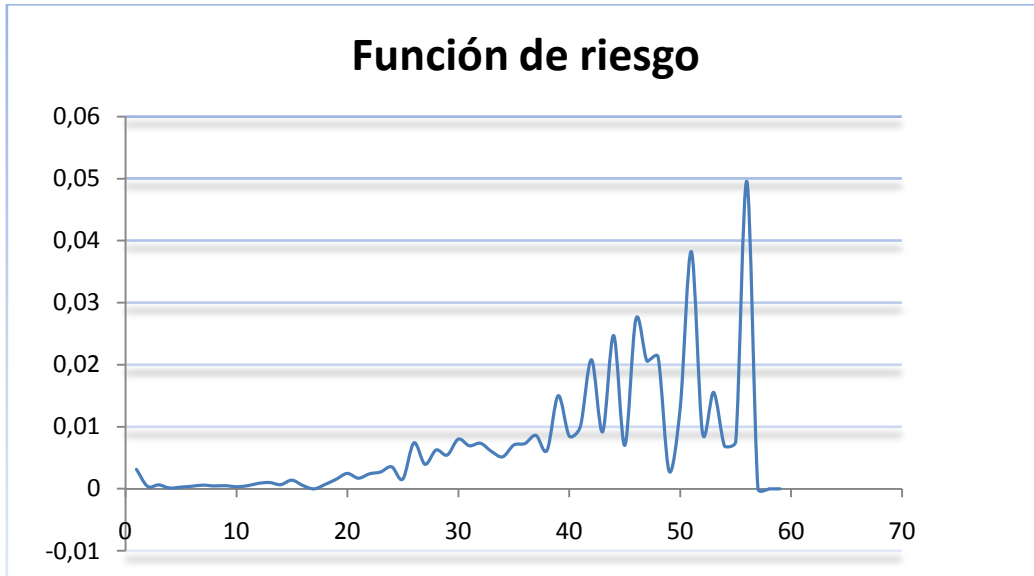


Figura 6.23 Estimación global del riesgo de fallo en la red, utilizando el método de NAE

Se observa en la Figura 6.23. que en los primeros 25 años la tasa instantánea de fallo es casi nula; solo a partir de los 28 años, se empiezan a dar cambios significativos, que aunque intermitentes muestran una tendencia creciente hasta los 48 – 50 años, con máximos locales en los 45, 50 y 55 años de operación. Se ve por ejemplo, que la probabilidad de que un tubo que ha sobrevivido sin fallos por 40 años, sobreviva más de 41 años en las mismas condiciones, es casi 4 veces mayor que la de uno que ha llegado a los 50 años sin averías y lo pueda hacer en las mismas condiciones por más de 51 años. Además en todo estos 58 años de seguimiento a la red, la fuerza de mortalidad más alta (FOM) de sus tuberías o tasa instantánea de fallo más alta, se localiza (en la anterior figura) a los 55 años de operación de la red, con un valor que no supera el 5,0% (0,049505).

En lo que sigue, se analiza la fiabilidad de los tramos de tubería de la red, con este método, considerando distintos criterios de estratificación por tipo de material, tráfico, etc.

6.1.3.2 Estimación estratificada de la fiabilidad por tipo de material.

MATERIA L	Tiempo	Tasa de fallo	Tasa de fallo acumulada	Fiabilidad
Fundición Dúctil	0,5	0,003243	0,003243	0,996762
	5	0,000288	0,004766	0,995245
	10	0,000351	0,006939	0,993085
	15	0,000000	0,008601	0,991436
	20	0,001494	0,013160	0,986926
	25	0,003489	0,022663	0,977591
	30	0,001266	0,035569	0,965056
	35	0,003984	0,054346	0,947105
	40	0,000000	0,054346	0,947105
	45	0,000000	0,054346	0,947105
	50	0,064516	0,204973	0,814669
	55	0,000000	0,204973	0,814669
Fundición Gris	0,	0,000000	0,000000	1,000000
	5	0,000000	0,000000	1,000000
	10	0,043478	0,043478	0,957453
	15	0,038462	0,204099	0,815382
	20	0,000000	0,204099	0,815382
	25	0,054054	0,258153	0,772477
	30	0,013889	0,272042	0,761822
	35	0,007463	0,312954	0,731284
	40	0,000000	0,388493	0,678078
	45	0,006536	0,488741	0,613398
	50	0,057471	0,620962	0,537427
	55	0,000000	0,642701	0,525870
Polietileno	0	0,000000	0,000000	1,000000
	5	0,013514	0,023718	0,976561
	10	0,000000	0,034829	0,965771
	15	0,000000	0,046733	0,954342
	20	0,000000	0,073760	0,928894
	25	0,051948	0,227961	0,796156
	30	0,064516	0,392272	0,675520
	35	0,000000	0,619725	0,538092
	40	0,000000	0,619725	0,538092
	45	0,000000	0,819725	0,440553
	50	0,000000	0,819725	0,440553
	55	0,000000	0,819725	0,440553

Uralita	0	0,000000	0,000000	1,000000
	5	0,000000	0,011364	0,988701
	10	0,000000	0,011364	0,988701
	15	0,004630	0,045856	0,955180
	20	0,003891	0,067664	0,934575
	25	0,007799	0,086369	0,917256
	30	0,007186	0,123115	0,884162
	35	0,007685	0,157172	0,854557
	40	0,011494	0,206755	0,813219
	45	0,034965	0,300565	0,740400
	50	0,025907	0,374614	0,687555
	55	0,086207	0,518734	0,595273

Tabla 6.12 Estimación de la fiabilidad por tipo de material utilizando el método de NAE

El análisis de la fiabilidad de los conductos de la red por tipo de material, con el método de NAE, muestra las siguientes características:

El material menos predispuesto a los fallos por el uso y el paso del tiempo es la fundición dúctil (FD) que a los 55 o más años tiene una supervivencia por encima del 80 % (0,814669); este material presenta unas diferencias muy significativas con el resto de los materiales, a partir de los primeros 20 años de operación de la red y se acentúan estas con el paso del tiempo a partir de los 25 años. La figura 6.24 también sugiere que, el material con tendencia a fallar más tempranamente en los primeros 23 años es la fundición gris (FG) y después de los 30 años el más vulnerable (aparentemente) de los materiales a las roturas y averías naturales es el polietileno.

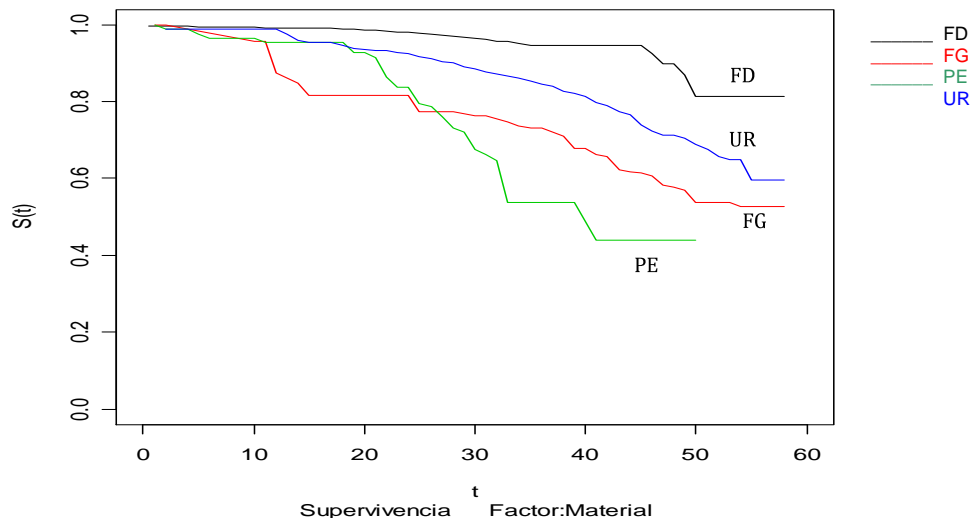


Figura 6.24 Fiabilidad de las tuberías por tipo de material utilizando el método de NAE

El segundo material más resistente a los fallos pero con marcadas diferencias con la FD, es la uralita que después de los 55 años, tiene una fiabilidad cercana al 60 % (0,595273). Después de los 28-30 años, la anterior figura sugiere que los tramos de tubería aparentemente más frágiles son de polietileno.

Es importante destacar, que esta situación no se veía tan clara con los dos métodos anteriores, donde por ejemplo el método de Kaplan-Meier, no

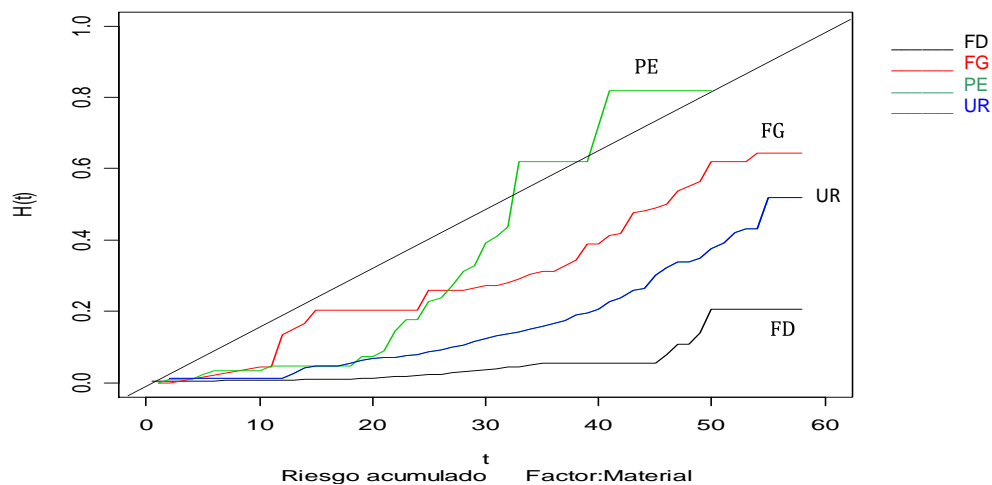


Figura 6.25 Riesgos acumulados por tipo de material utilizando el método de NAE

mostraba diferencias tan importantes en la fiabilidad antes de los 40 años entre la FD, la FG y la Uralita, que si se aprecian con claridad con el método de NAE, evidenciándose con ello su alto *poder de discriminación* en este sentido.

La Figura 6.25, sugiere que durante los primeros 12 años ninguno de los materiales de las tuberías de la red, tiene alguna tendencia a la predisposición de los fallos; sólo entre los 12 y los 27 años, la FG muestra alguna tendencia creciente al fallo respecto a los restantes materiales. Después de los 30 de funcionamiento de la red la situación es más clara, el polietileno, presenta (aparentemente) una tendencia “más agresiva” a las roturas y averías naturales, que el resto de los materiales. Lo que sí es cierto es que, con este método se confirma una vez más de que el material más resistente y el de menos predisposición a los fallos es la FD, le sigue la uralita y posteriormente la FG (Figuras 6.24 y 6.25).

6.1.3.3 Estimación estratificada de la fiabilidad por tráfico

TIPO DE TRAFICO	Tiempo	Tasa de fallo	Tasa de fallo acumulada	Fiabilidad
ACERA	0,5	0,002229	0,002229	0,997773
	5	0,000536	0,003925	0,996082
	10	0,000215	0,005604	0,994411
	15	0,000349	0,007816	0,992214
	20	0,001606	0,012002	0,988070
	25	0,007006	0,026447	0,973899
	30	0,003586	0,046125	0,954923
	35	0,004722	0,067051	0,935147
	40	0,006601	0,099554	0,905241
	45	0,024955	0,166418	0,846693
	50	0,018692	0,220516	0,802105
	55	0,027397	0,271812	0,761997
NORMAL	0	0,005160	0,005160	0,994853
	5	0,000000	0,007461	0,992567
	10	0,001669	0,009907	0,990142
	15	0,000000	0,016201	0,983929
	20	0,000000	0,022303	0,977944
	25	0,006110	0,041494	0,959355
	30	0,005386	0,063166	0,938787
	35	0,006186	0,093584	0,910661
	40	0,012539	0,137003	0,871967
	45	0,022472	0,197521	0,820763
	50	0,035294	0,308684	0,734413
	55	0,043478	0,400972	0,669669
PESADO	0	0,000000	0,000000	1,000000
	5	0,000000	0,007519	0,992509
	10	0,000000	0,033269	0,967278
	15	0,000000	0,049663	0,951550
	20	0,000000	0,079074	0,923971
	25	0,000000	0,106852	0,898659
	30	0,000000	0,131852	0,876471
	35	0,000000	0,206718	0,813249
	40	0,000000	0,260137	0,770946
	45	0,000000	0,297174	0,742915
	50	0,000000	0,297174	0,742915
	55	0,333333	0,630508	0,532322

Tabla 6.13 Fiabilidad estimada por tipo de tráfico utilizando el método de NAE

El análisis de la fiabilidad de los conductos de la red atendiendo a las zonas por donde circulan las tuberías de la red y que se denomina tipo de tráfico (acera (A), normal (N) y pesado (P)) muestra las siguientes características

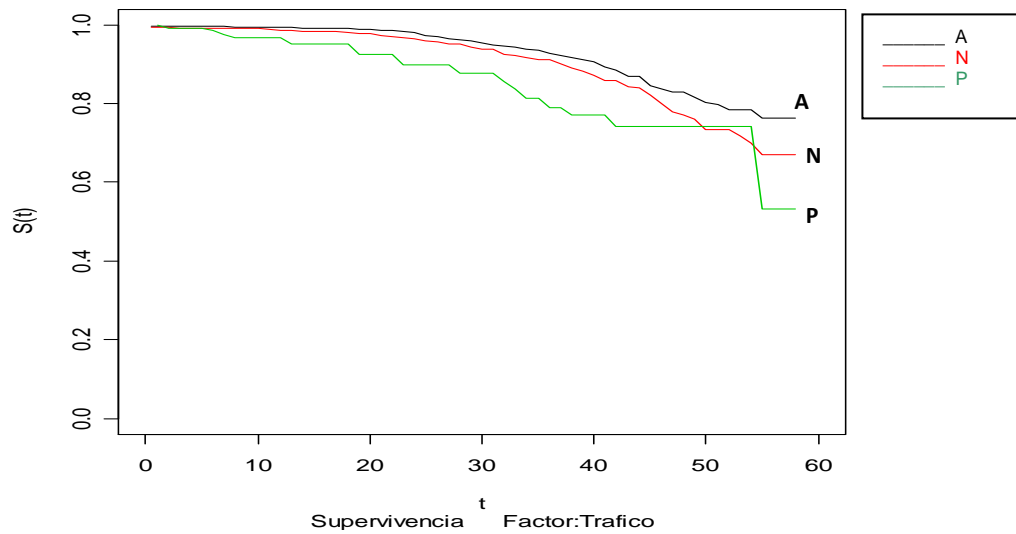


Figura 6.27 Fiabilidad estimada de las tuberías por tipo de tráfico

El tráfico donde las tuberías tienen menos predisposición a los fallos por el uso y el paso del tiempo es el tipo acera (A), que durante los primeros 25 años mantienen la fiabilidad de los conductos casi intacta, por encima del 97%. Este tipo de tráfico no difiere mucho del tráfico normal (N), hasta los 48-50 años, que es cuando se observan diferencias significativas, especialmente a los 55 años.

Un poco antes de los 10 primeros años, se empiezan a acentuar las diferencias entre el tráfico pesado y los dos restantes, los tráficos normal y acera, de tal modo que los conductos más vulnerables a roturas y averías por el uso y el paso del tiempo en casi todo el historial de fallos de la red, son los que están instalados en zonas de tráfico pesado.

Como en el caso de los materiales, el análisis de la fiabilidad con este enfoque de NAE, queda lo suficientemente aclarado y explicado para el caso del tipo de tráfico por donde circula la red, lo que no sucedía con los métodos de KM y de las tablas de supervivencia, como se aprecia en las Figuras 6.11 y 6.19.

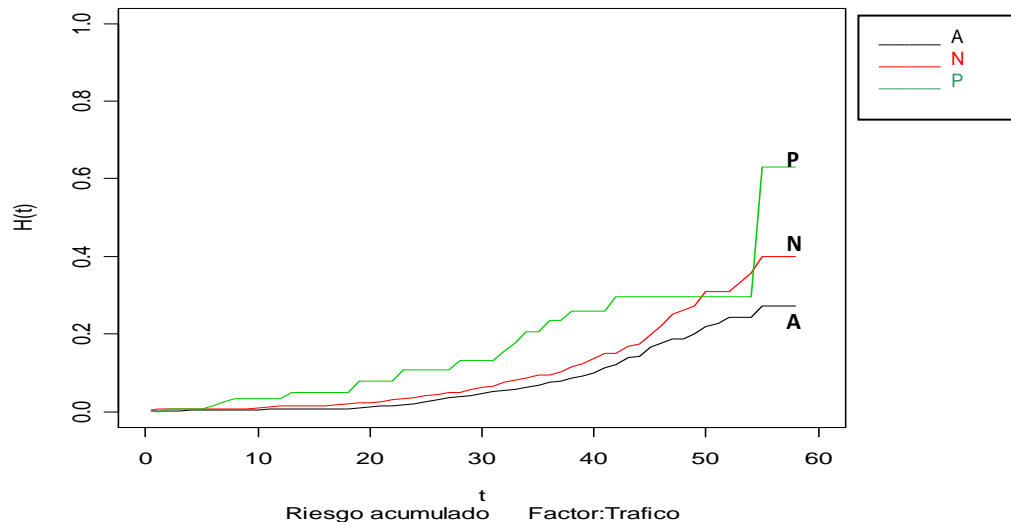


Figura 6.28 Riesgos acumulados por tipo de tráfico

La Figura 6.28 resalta desde los primeros 8-10 años de funcionamiento de la red, la mayor predisposición al fallo en los conductos instalados en zonas de alta intensidad de tráfico (tráfico pesado), con una tendencia que crece más rápidamente en este tipo de tráfico, que el crecimiento que se observa en los tráfico normal y acera hasta los 40- 42 años, acompañado de una relativa estabilidad hasta los 54 años; con un salto abrupto a partir de los 55 años que arroja una predisposición creciente acumulada al fallo del 63%.

Los tubos instalados en las dos restantes zonas de tráfico, la normal y la acera, igualmente y en forma muy similar muestran una tendencia creciente a las roturas y averías hasta los 43-45 años, a partir del cual se infiere, que la zona donde se presenta la mayor predisposición a las roturas y averías naturales de las tuberías es la zona del tráfico acera, seguida del tráfico normal.

6.2 Evaluación de la fiabilidad en la red con modelos semiparamétricos

En este apartado el objetivo es identificar los factores pronóstico, es decir las variables que pueden explicar o influir en las roturas o averías naturales de una tubería.

El modelo de riesgos proporcionales introducido inicialmente por Cox (1972) es el modelo de regresión más utilizado en análisis de supervivencia, sin embargo no fue sino a partir del desarrollo del enfoque basado en los procesos de conteo, que este modelo logra su completa madurez, este enfoque ha permitido la verificación de los supuestos de riesgos proporcionales y el estudio de los residuos generalizados (Ansell J.I y Phillips M.J 1994). Algunos desarrollos y generalizaciones del modelo de Cox pueden verse en: Therneau. Y Grambsch (2000).

Una vez ajustado el modelo de Cox mediante el software SPSS, podemos analizar la estimación de sus coeficientes y su significancia en la siguiente tabla. A partir de la significancia de sus coeficientes podemos identificar los factores o variables explicativas que tienen mayor impacto o influencia en el comportamiento de la tasa de roturas y o averías de los conductos

COVARIATAS	B	Sig.	Exp(B)
Longitud	,004	,000	1,004
Diámetro	-,003	,003	,997
Presión	,023	,000	1,023
TRAFICO		,015	
Acera	-,723	,006	,485
Normal	-,548	,044	,578
MATERIAL		,000	
Fundición Dúctil	,331	,023	1,393
Fundición Gris	-,132	,473	,876
Polietileno	1,728	,000	5,629

Tabla 6.14 Estimación del modelo de Cox

En la tabla 6.14 se observa por ejemplo, que todas las variables, salvo el material *Fundición Gris* son altamente significativas en el modelo. Por lo que el material fundición gris no se diferencia del material de comparación que en nuestro caso es la Uralita, que por ello no figura en la tabla. En consecuencia, el riesgo que tiene una tubería de Uralita o de fundición gris es el mismo si el resto de variables permanecen constantes.

Codificaciones de variables categóricas(b,c)

	Frecuencia	(1)	(2)	(3)
Mat_cod(a)	1=Fundición dúctil	18858	1	0
	2=Fundición gris	528	0	1
	3=Polietileno	325	0	0
	4=Uralita	5576	0	0
Traf_cod(a)	0=ACERA	19884	1	0
	1=NORMAL	4920	0	1
	2=PESADO	483	0	0

a Codificación de parámetros de indicador

b Variable de categoría: Mat_cod (Código del material)

c Variable de categoría: Traf_cod (Código del tráfico de carga)

Tabla 6.14 a. Variables Dummy en el modelo de riesgos proporcionales de Cox

Hemos también probado a introducir interacciones del material con el resto de covariables y dichas interacciones han resultado no significativas.

La *Longitud*, por tener un coeficiente con signo positivo y altamente significativo de 0,004, está aumentando el riesgo normal de fallo en una tubería, en un 0,4 % por cada metro adicional de tramo de tubería que se extienda, es decir, entre más larga es la tubería, mayor es el riesgo de fallo de la misma.

El *Diámetro*, que es otra variable que aporta información al modelo, con un coeficiente negativo altamente significativo de -0,003, tiene un efecto contrario que reduce el riesgo de fallo (la tasa de roturas y averías) en los conductos en un 0,3 %, por cada (mm) adicional de la boca del tubo, o sea que entre más ancho es el diámetro del conducto menor es el riesgo de rotura.

El *tráfico* Acera y Normal, tienen también un efecto similar al anterior, sobre la tasa normal de fallos de reducir esta en un 51,5 % y en un 42,2% respectivamente; todo lo anterior desde la óptica del tráfico pesado de la red, es decir, en *condiciones normales de operación*.

Un análisis similar, se aplica al efecto de la *presión* sobre la tasa de fallos de los conductos de la red en condiciones normales de operatividad. En este caso por cada metro de columna de agua (mca) de presión adicional, que circule por cada tramo de tubería, el riesgo normal de fallo se incrementa en un 2,3%, es decir, entre mayor sea la presión que circule por los conductos, mayor es el riesgo de fallo de los mismos.

En cuanto a la curva de supervivencia obtenida del modelo de Cox en condiciones normales de operación de la red, en media de las covariatas (es decir, por tramo con características igual a las medias de la variables), los tramos de la red con 55 años o más de operación tienen una supervivencia cercana al 79 % y después de los 40 años, la supervivencia de los conductos a las roturas y averías es del 95%, como se aprecia en la figura siguiente.

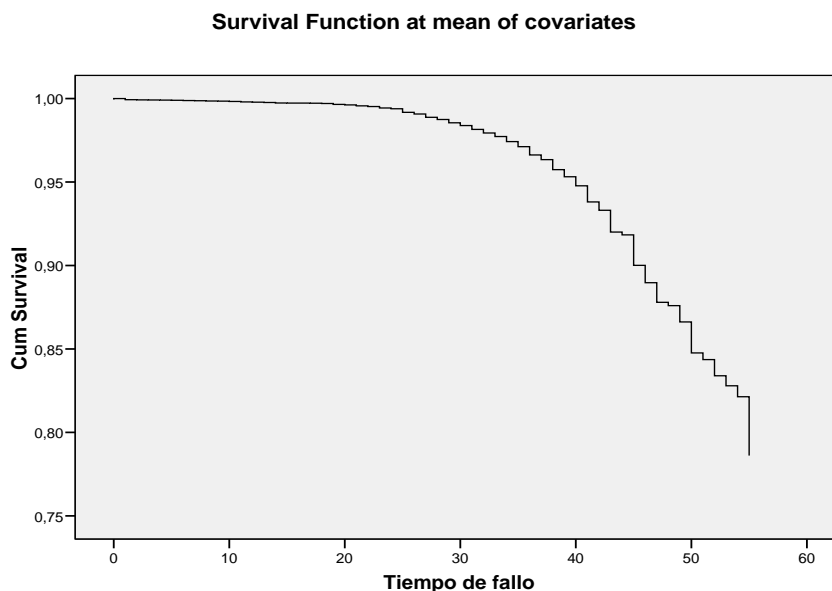


Figura 6.30 Función de supervivencia para el valor medio de las covariables.

En relación con la predisposición al fallo en el tiempo o rapidez con que se degrada la tubería en general, esta inicia su ciclo de deterioro a partir de los 25-28 años, al comienzo muy lentamente (con pequeños cambios, insignificantes en forma creciente), pero luego a partir de los 40 años hasta los 55 años se acentúan estos y, el desgaste y deterioro de los conductos pasa de un 5% a casi un 20%, es decir se cuadruplica en 15 años de operación o funcionamiento normal de la red. La siguiente Figura 6.31, ilustra las anteriores observaciones.

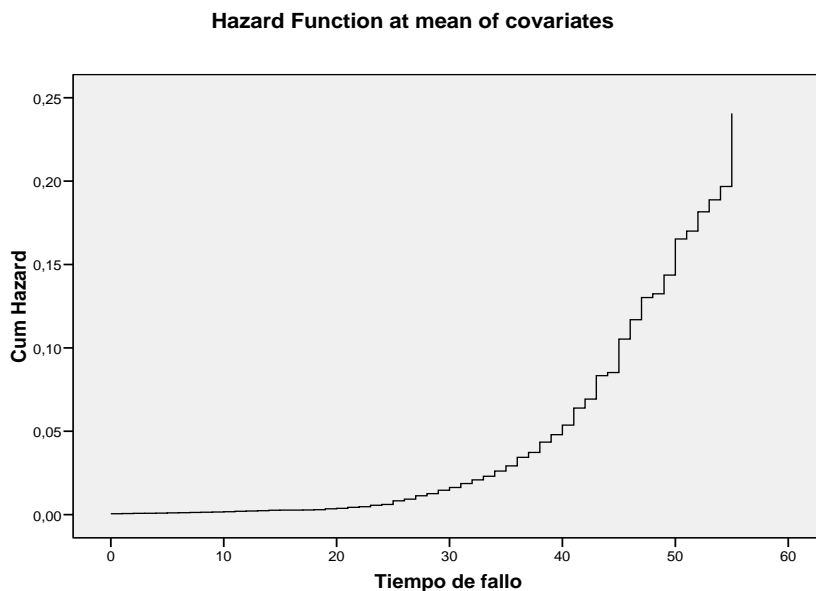


Figura 6.31 Función de impacto para el valor medio de las covariables.

6.2.1 Validación de los supuestos de Proporcionalidad de Cox

En lo que sigue de esta sección, se verifica la hipótesis general de proporcionalidad de Cox, para las distintas covariables de cada tramo de tubería, esto es, tipo de material, tráfico, longitud y diámetro.

El método gráfico que se utiliza, es el del *paralelismo de las gráficas log-log*, que ya fue presentado en la sección 4.5.3.1.

En el caso de los *materiales*, el resultado de la validación de los supuestos del modelo se recoge en el siguiente gráfico.

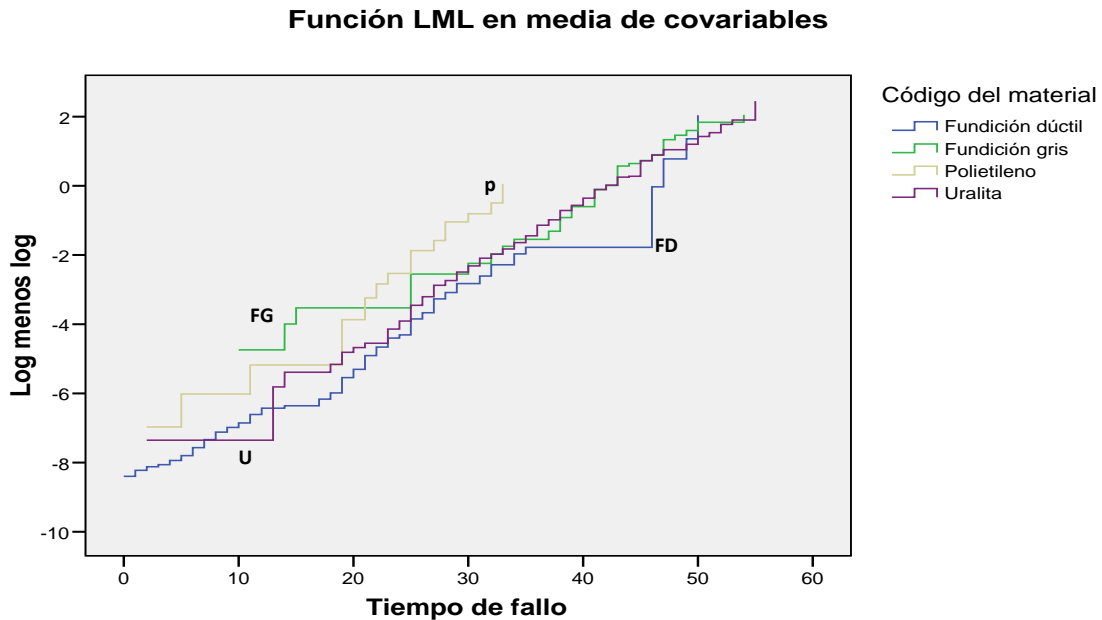


Figura 6.32 Paralelismo de las *gráficas log-log* para verificar los supuestos de Cox con el tipo de material

El paralelismo que se insinúa en la figura 6.32, con los distintos tipos de material, (aunque este criterio es un poco subjetivo) nos sugiere que puede admitirse la hipótesis global de riesgos proporcionales, para el caso de esta covariable. Esta configuración es un indicador del cumplimiento de los supuestos del método de Cox de riesgos proporcionales para el caso ya referenciado, es decir, por tipo de material.

Para el factor tipo de *Tráfico*, el resultado es similar como se aprecia en Figura 6.33, incluso podríamos decir que el resultado es menos subjetivo en su interpretación que en el caso de la covariable anterior (el *material*), apreciándose una tendencia lineal mucho más definida, comprobándose de esta manera, la admisión de la hipótesis global de riesgos proporcionales de Cox para esta covariable *Tráfico*.

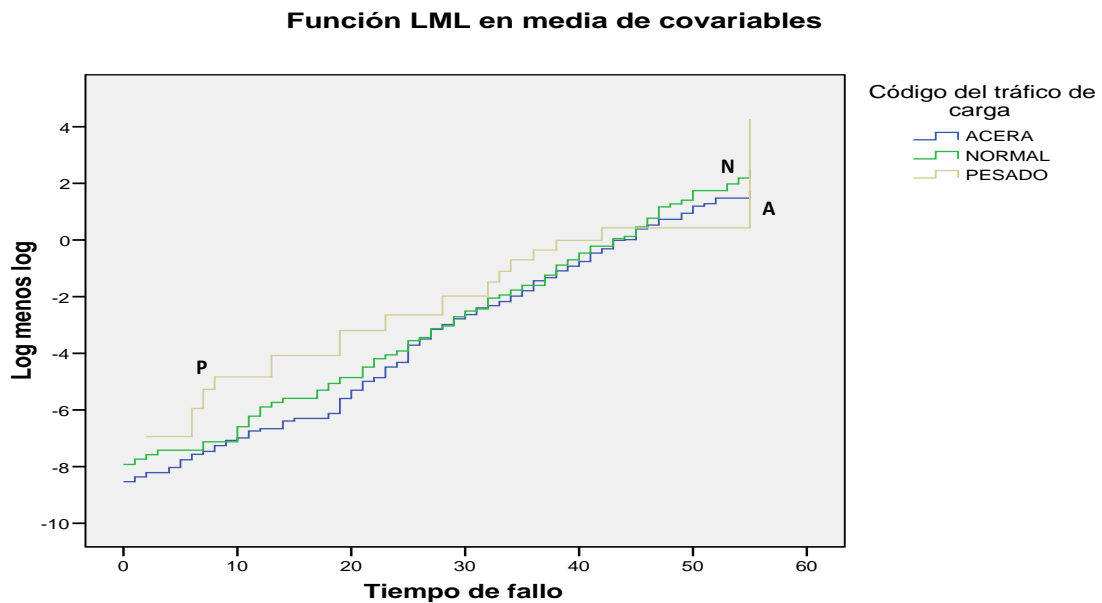


Figura 6.33 Paralelismo de las *gráficas log-log* para el tipo de tráfico

Para los factores *longitud en metros* y *diámetro en milímetros* de cada tramo de tubería, los resultados de las gráficas, también sugieren cada una cierta tendencia al paralelismo por parte de las distintas categorías en cada covariable (como se aprecia en las Figuras 6.34 y 6.35 respectivamente), confirmándose de esta manera, la admisión de los supuestos de proporcionalidad de Cox en cada una de ellas.

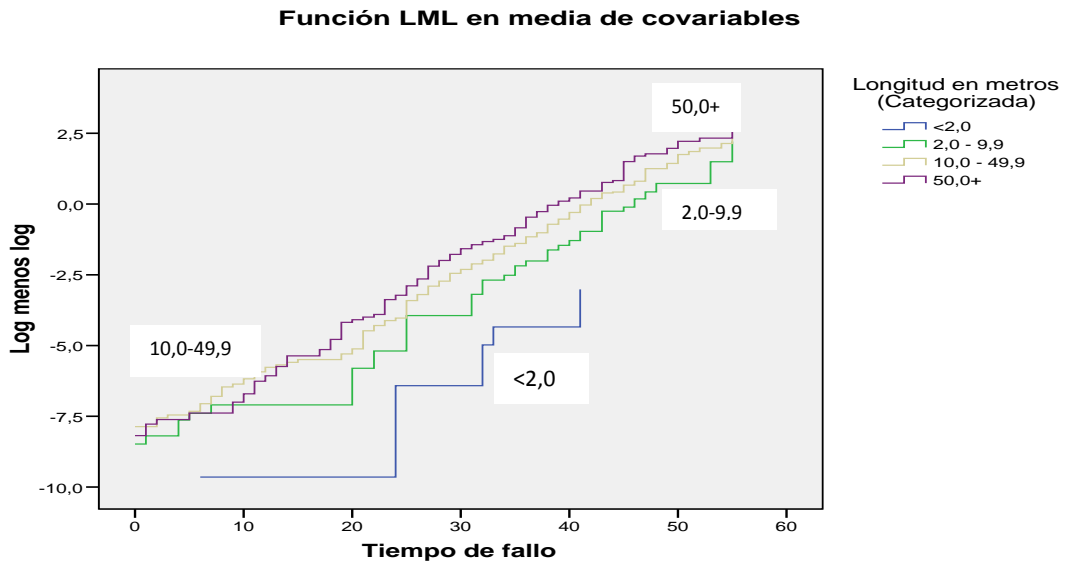


Figura 6.34 Paralelismo de las gráficas log-log para la *longitud en metros*, de cada tramo de tubería

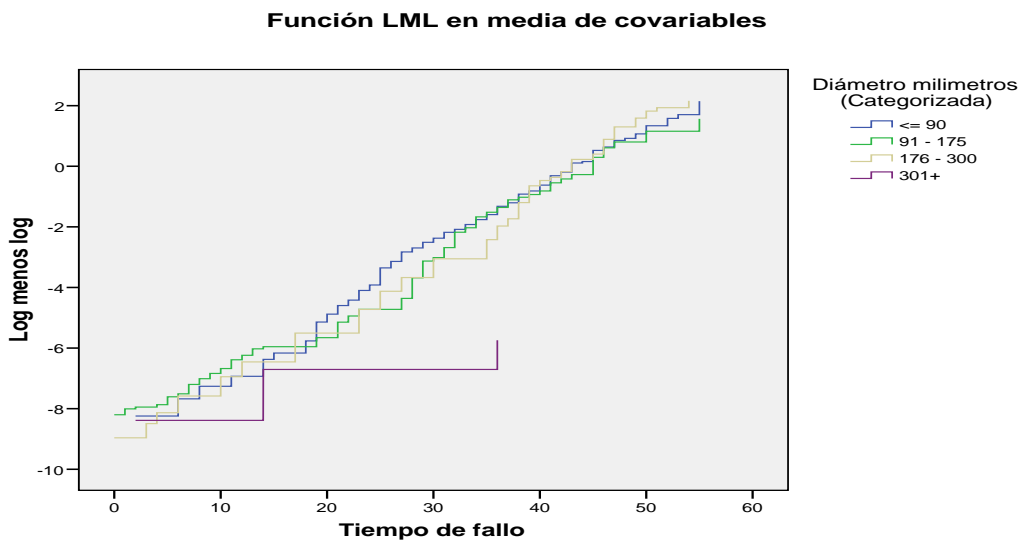


Figura 6.35 Paralelismo de las gráficas log-log para el Diámetro

El modelo de Cox estratificado también constituye una de las maneras de corregir el modelo de Cox general, cuando no se cumple el supuesto de riesgos proporcionales para alguna de las covariables. En este caso suele correrse el modelo estratificando por la covariable que no cumple con el supuesto de riesgo proporcional. Este procedimiento permite corregir el sesgo en la estimación del parámetro que puede presentarse cuando se viola el supuesto de riesgo proporcional. Sin embargo, presenta una desventaja y es que no existe ningún parámetro que permita estimar el efecto de la covariable de estratificación.

Pero además podemos utilizarlo para comparar las diferentes curvas de supervivencia según los estratos. A modo de ejemplo mostramos las gráficas del modelo de Cox estratificado por materiales y por tráfico, pues ilustran bien cual es el efecto de ambos factores.

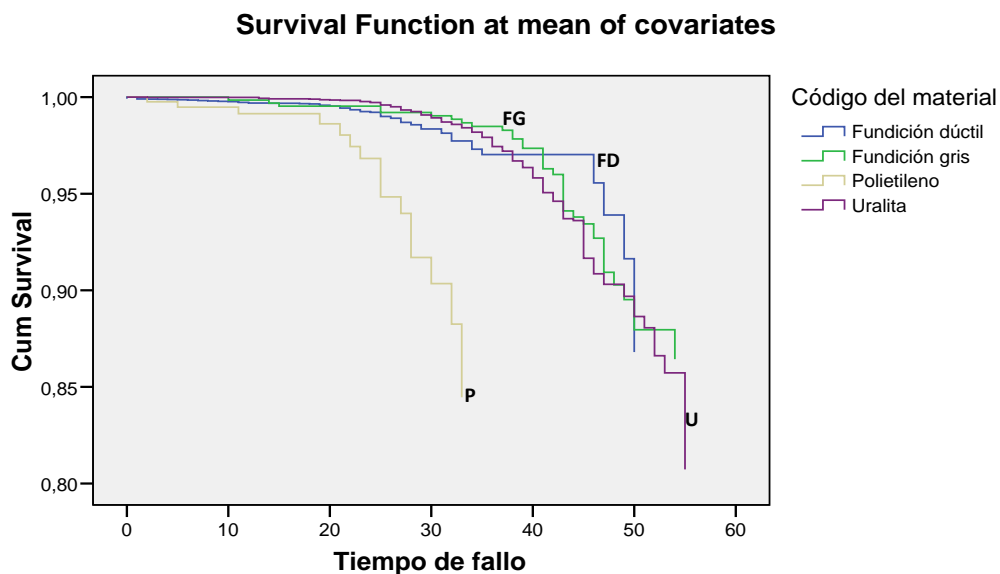


Figura 6.36 Función de Supervivencia según tipo de material

La supervivencia de los materiales según nuestro modelo de riesgo proporcional, sugiere claramente que el material menos robusto a las roturas a partir de los 20 años, corresponde al Polietileno, donde por ejemplo, la supervivencia a los 32 años de este material, está por debajo del 85% en contraste con la fundición dúctil y los otros que están por encima del 95%.

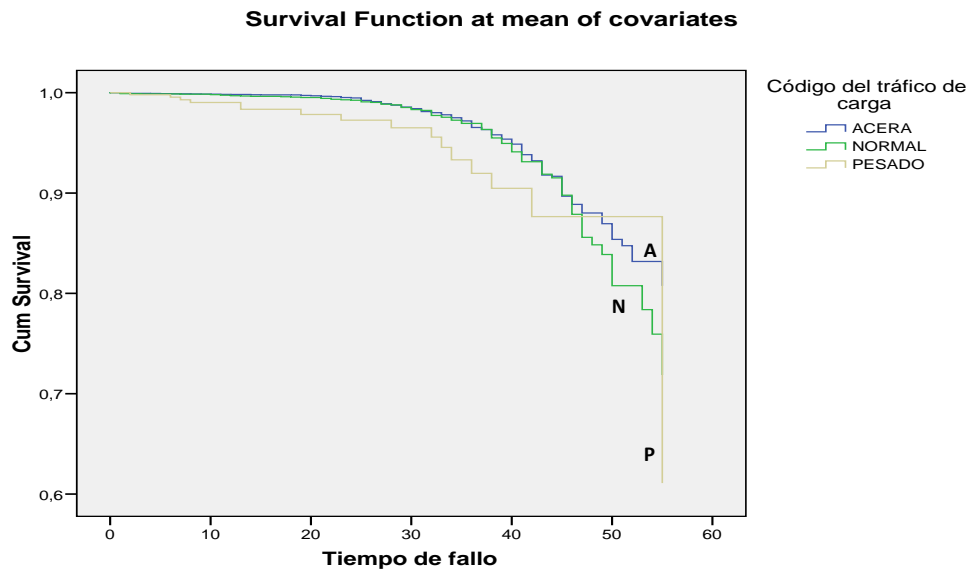


Figura 6.37 Función de Supervivencia según tipo de tráfico

En las diferentes modalidades de tráfico (según los Riesgos de Cox), las curvas de supervivencia de las tuberías, muestran con claridad que en el tráfico pesado es donde se presenta el mayor riesgo a las roturas entre los 10 y los 46-47 años de operación normal de la red; la supervivencia de los conductos en el tráfico acera y en el normal es más o menos la misma hasta los 48 años de antigüedad de la red y sólo a partir de este instante, se evidencian diferencias entre estas dos modalidades de instalación de las tuberías.

En lo que sigue, se analiza el modelo de Cox por tipo de rotura, natural o provocada, con el fin de comparar separadamente el comportamiento de cada una de estas categorías. Al respecto, cuando se mezclan los dos tipos de averías (naturales y provocadas), el aumento de heterogeneidad en los datos, hace que sólo mantengan su poder de discriminación global las variables material y tráfico (Tabla 6.15), en contraste con lo que acontece, cuando sólo se considera el modelo de Cox con solo fallos naturales (Tabla 6.14) en donde se aprecia claramente el alto poder de discriminación de las variables Presión, longitud, tráfico, diámetro y tipo de material.

COVARIATAS	B	Sig.	Exp(B)
LENGTH	,001	,232	1,001
DIAMETRO	,000	,884	1,000
PRESION	-,005	,212	,995
MATERIAL		,000	
Fundición Dúctil	1,484	,000	4,410
Fundición Gris	-,219	,183	,803
Polietileno	1,189	,000	3,285
TRAFICO		,000	
Acera	,317	,169	1,373
Normal	-,095	,688	,909

Tabla 6.15. Modelo de Cox con los dos tipos de fallos mezclados (naturales y provocados).

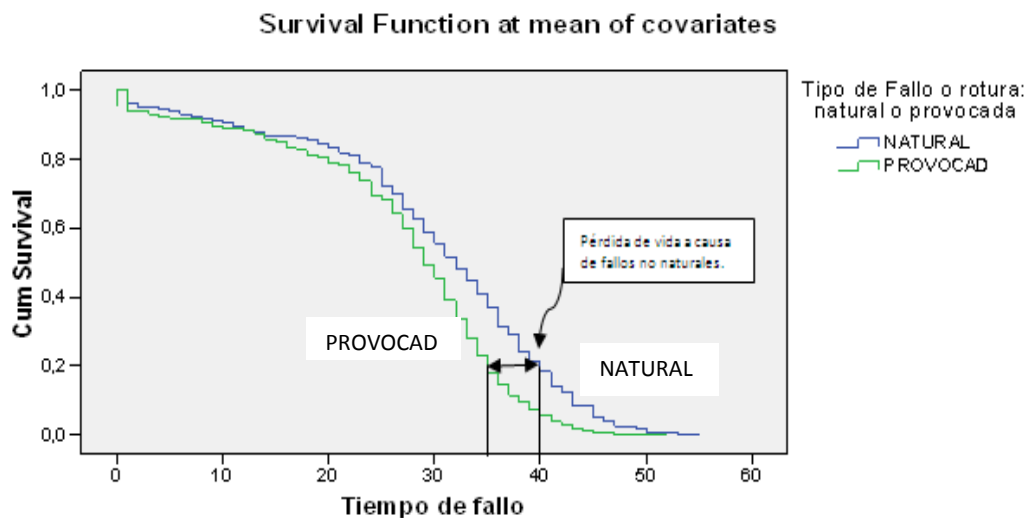


Figura 6.38 Función de supervivencia según tipo de fallo.

En la Figura 6.38, se aprecia la pérdida de vida en los conductos de la red, causada por factores no naturales, es decir por ejemplo, si el 20 % de las tuberías que funcionan y se deterioran normalmente con el uso y el paso del tiempo fallan antes de los 40 años, las mismas tuberías bajo la influencia de otros factores diferentes al envejecimiento y desgaste natural, fallarán con la misma probabilidad por lo menos 4- 5 años antes. Estas diferencias se empiezan a manifestar a partir de los 15-16 años de operación de la red.

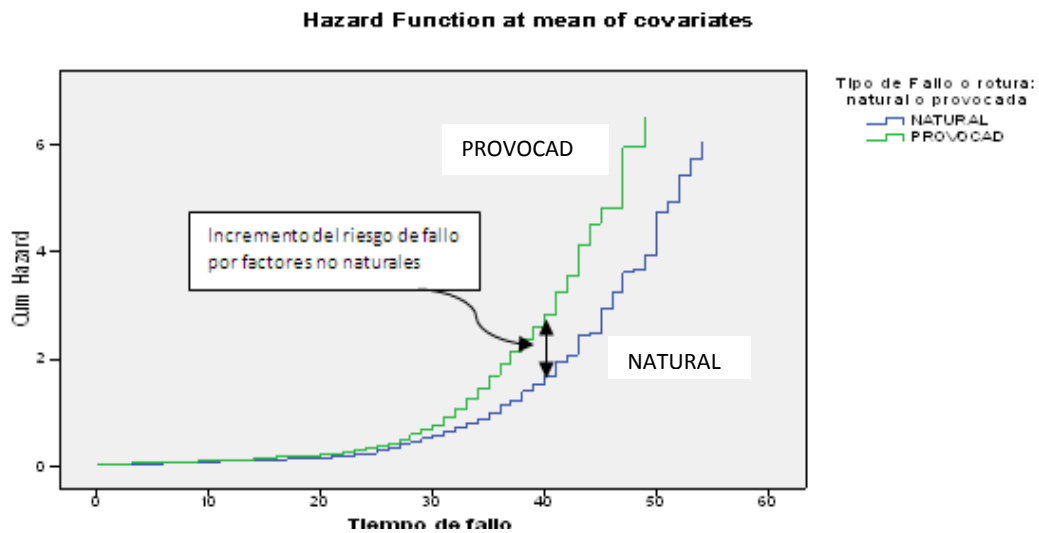


Figura 6.39. Función de impacto según tipo de fallo

De otra parte, la Figura 6.39 permite apreciar la forma cómo se incrementa la predisposición al fallo en ambos tipos de rotura; a partir de los 27- 28 años el riesgo de fallo es significativamente mayor en los conductos con roturas ocasionadas por causas diferentes al envejecimiento y desgaste natural (provocadas), por ejemplo, a los 40 años el incremento $\Delta H(t)$ en el riesgo de fallo en los dos tipos de averías es apreciable (1,156) y este incremento en el riesgo, se acentúa en forma creciente, a partir de los 40 años en adelante de operación continua de la red.

t	$R_n(t)$	$R_p(t)$	$\Delta R(t)$	$H_p(t)$	$H_n(t)$	$\Delta H(t)$
0	0,966	0,953	0,013	0,048	0,034	0,014
1	0,960	0,942	0,018	0,06	0,041	0,019
2	0,951	0,939	0,012	0,062	0,05	0,012
3	0,949	0,93	0,019	0,072	0,052	0,02
4	0,944	0,923	0,021	0,08	0,057	0,023
5	0,938	0,918	0,02	0,085	0,064	0,021
6	0,931	0,916	0,015	0,088	0,072	0,016
7	0,924	0,906	0,018	0,098	0,079	0,019
8	0,917	0,897	0,020	0,109	0,087	0,022
9	0,912	0,889	0,023	0,118	0,092	0,026
10	0,905	0,886	0,019	0,121	0,100	0,021
11	0,892	0,881	0,011	0,126	0,114	0,012
12	0,885	0,874	0,011	0,135	0,122	0,013
13	0,878	0,858	0,020	0,153	0,131	0,022
14	0,868	0,85	0,018	0,163	0,142	0,021
15	0,865	0,831	0,034	0,185	0,145	0,04
17	0,863	0,825	0,038	0,192	0,148	0,044
18	0,858	0,811	0,047	0,209	0,154	0,055
19	0,842	0,803	0,039	0,22	0,172	0,048
20	0,835	0,788	0,047	0,239	0,181	0,058
21	0,819	0,782	0,037	0,246	0,200	0,046
22	0,808	0,761	0,047	0,273	0,213	0,06
23	0,786	0,737	0,049	0,306	0,241	0,065
24	0,775	0,696	0,079	0,363	0,255	0,108
25	0,724	0,683	0,041	0,382	0,323	0,059
26	0,701	0,641	0,06	0,445	0,356	0,089
27	0,656	0,598	0,058	0,514	0,422	0,092
28	0,628	0,542	0,086	0,613	0,465	0,148
29	0,588	0,492	0,096	0,709	0,531	0,178
30	0,556	0,456	0,100	0,786	0,586	0,200
31	0,516	0,391	0,125	0,940	0,662	0,278
32	0,481	0,339	0,142	1,082	0,732	0,35
33	0,449	0,281	0,168	1,269	0,8	0,469
34	0,408	0,232	0,176	1,462	0,896	0,566
35	0,371	0,182	0,189	1,702	0,992	0,71
36	0,316	0,145	0,171	1,932	1,151	0,781
37	0,290	0,115	0,175	2,164	1,237	0,927
38	0,243	0,094	0,149	2,367	1,415	0,952
39	0,214	0,074	0,14	2,603	1,542	1,061
40	0,184	0,058	0,126	2,851	1,695	1,156
41	0,141	0,039	0,102	3,248	1,955	1,293
42	0,124	0,029	0,095	3,556	2,088	1,468
43	0,088	0,016	0,072	4,127	2,435	1,692
44	0,084	0,011	0,073	4,53	2,479	2,051
45	0,053	0,008	0,045	4,804	2,942	1,862
47	0,026	0,003	0,023	5,952	3,632	2,32
49	0,019	0,001	0,018	6,541	3,948	2,593

Tabla 6.16 Comparación modelos de Cox

En la columna $\Delta R(t)$, de la tabla anterior, se registran también los instantes de máximo y mínimo incremento en la supervivencia de los de los conductos de la red, por factores no naturales, por ejemplo, a los 11 -12 años de operación de la red, se detecta el menor incremento y a los 35 se registra el mayor.

La columna $\Delta H(t)$, también registra el máximo y el mínimo incremento del riesgo de fallo de las tuberías igualmente por factores no naturales. Esto es, a los 11 años se registra el menor incremento y a los 49 años de operación de la red se presenta el mayor incremento de riesgo de fallo, 2,593.

Los resultados de la bondad del ajuste del modelo de Cox, se recogen en la siguiente tabla.

Model Summary and Parameter Estimates

Dependent Variable: Hazard function $H(U_i)$

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,997	189566,703	1	532	,000	0,000203	0,9874

The independent variable is Residuos de Cox Snell.

Tabla 6.16 a. Bondad de ajuste modelo de Cox

En general si existiese un buen ajuste (por parte de los datos) al modelo de Cox, se debería cumplir que los residuos U_i de Cox-Snell, siguen aproximadamente un modelo Exponencial con parámetro $\lambda = 1$, lo que significa que la función de impacto o de Riesgo Acumulado debería ser de la forma $H(t) \cong t$, o lo que es lo mismo $H(U_i) \cong U_i$, donde U_i representa, el riesgo de Cox-Snell i-ésimo, cuya gráfica se aprecia en la Figura 6.40, es decir, una recta que pasa por el origen y que tiene pendiente $\lambda \cong 1$, cuyo valor ha sido estimado con nuestros datos en $0,9874 \cong 1$

La bondad de ajuste de los datos (tiempos de fallo natural de los conductos) observados a la recta en mención es de 99,7% ($R^2= 0,997$), como se observa en la tabla anterior.

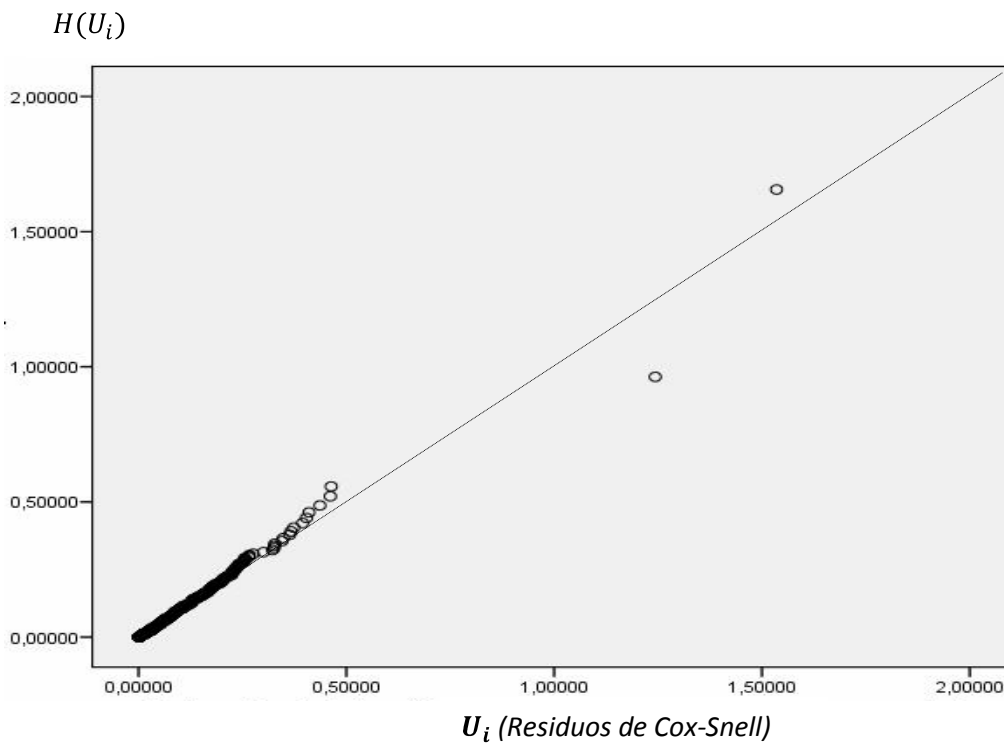


Figura 6.40 Ajuste de los datos al modelo de Cox

Se concluye que los supuestos del Modelo de Cox son válidos en el caso de nuestros datos y además este modelo se ajusta perfectamente a estos mismos datos con un 99,7% de variabilidad explicada por el modelo (ver Tabla 6.16)

6.3 Evaluación de la fiabilidad en la red con modelos paramétricos

Como alternativa de solución a las restricciones con las estimaciones de libre distribución (no permitir estimaciones más allá del último tiempo de fallo muestral, ni permitir la estimación del MTTF con datos censurados) se presentan los modelos paramétricos, que se consideran en este estudio para analizar el tiempo de vida y la modelización del riesgo de fallo de los conductos de nuestra red de suministro de agua potable. Entre estos modelos, se destacan las distribuciones, exponencial, Weibull, Logística, Lognormal y otras más que son mencionadas en este trabajo. La utilización de estos métodos, consiste en asumir ciertas funciones de probabilidad para la variable tiempo de espera. Si esta asunción es razonable, (y para averiguarlo habrá que realizar pruebas de bondad del ajuste a los datos), los métodos paramétricos se consideran más *potentes y poderosos* que los no paramétricos.

El método consiste en estimar, por máxima verosimilitud, los parámetros característicos de la distribución, y usar su normalidad asintótica para realizar la estimación por intervalos y los contrastes de hipótesis que correspondan.

Las pruebas de bondad de ajuste se realizan con la habitual prueba basada en la ji-cuadrado, con la del logaritmo del cociente de verosimilitudes o con métodos gráficos. Una útil referencia para profundizar en estos temas, es el trabajo realizado por Johnson y Kotz (1970), quienes tratan extensamente catálogos matemáticos y las propiedades estadísticas de las distribuciones tratadas en este documento, a su vez también provee referencias concernientes a sus áreas de aplicación.

Los resultados de este apartado con algunos modelos paramétricos han sido los siguientes:

6.3.1 Identificación del Mejor Modelo para los Datos

Lo primero que convendría es tratar de identificar la distribución que mejor se aproxima a las observaciones mediante el uso de gráficos de probabilidad. Este tipo de gráficos muestran la función de distribución linealizada de una distribución teórica junto con una nube de puntos que representa estimaciones (no paramétricas) puntuales de la función de distribución de T. Evidentemente, cuanto más se aproxime la nube de puntos a la recta que aparece en el gráfico, tanto mejor será el ajuste.

El programa estadístico MINITAB permite automatizar este tipo de cálculos para identificar y describir gráficamente la distribución que mejor se ajusta a un conjunto de observaciones: La Figura 6.41, muestra los resultados del ajuste de los datos con cuatro diferentes modelos, como se aprecia en la siguiente página.

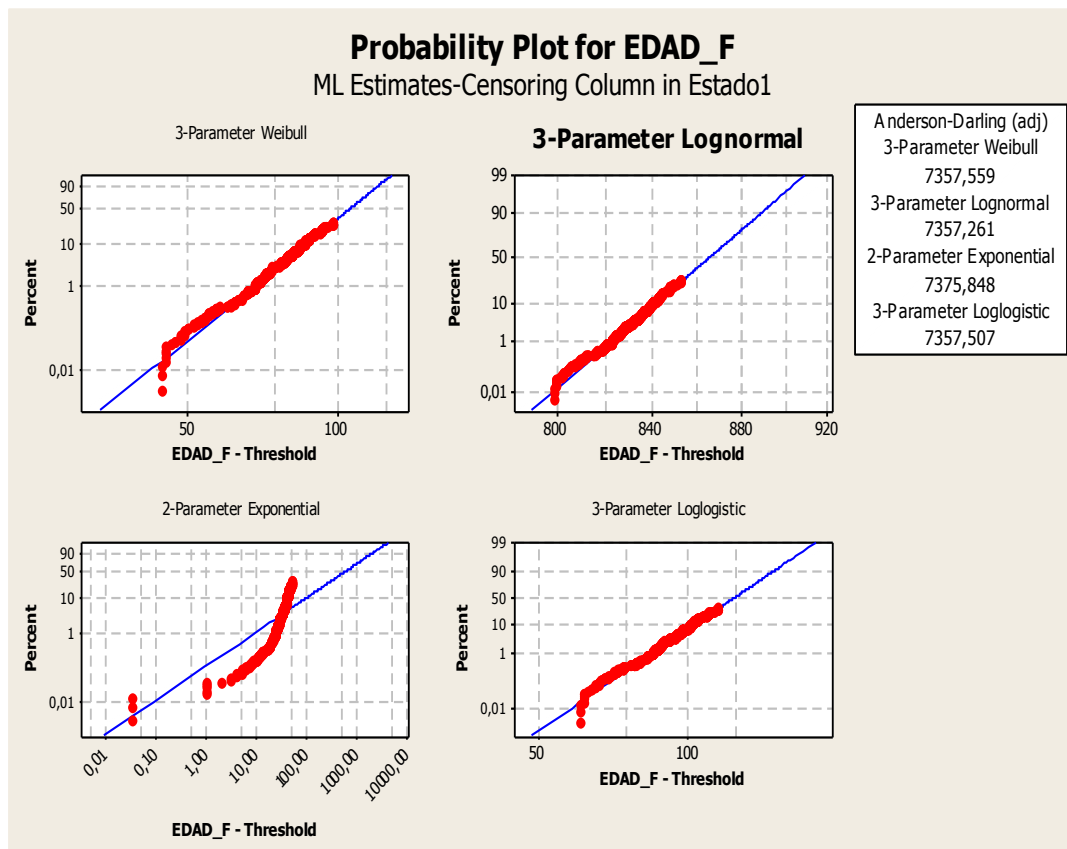


Figura 6.41 Función de distribución linealizada para cada modelo.

En la Figura 6.41 se aprecia claramente, que los modelos que mejor ajuste presentan a la diagonal del gráfico de probabilidad corresponden a las distribuciones: Weibull, Lognormal y Loglogística con 3 parámetros cada una, como se mencionó anteriormente.

Por su parte el estadístico Anderson-Darling nos da una medida de lo alejadas que se encuentran las observaciones de la recta que representa la función de distribución. La Distribución de mejor ajuste debe estar en correspondencia con el valor más pequeño del estadístico de Anderson-Darling, y/o con el Coeficiente de Correlación de Pearson, de mayor valor en la lista; en este caso es la *Lognormal con 3 parámetros* la escogida, aunque otros modelos, como la Loglogística y la Weibull, ambas de 3 parámetros (con un excelente ajuste, $R^2 = 0,95$), son opciones igualmente atractivas, que no difieren en casi nada con la distribución elegida.

Distribution	Anderson-Darling	Correlation Coefficient
Weibull	7361,488	0,938
Lognormal	7362,171	0,889
Exponential	7361,777	*
Loglogistic	7361,573	0,935
<i>3-Parameter Weibull</i>	7357,690	0,995
<i>3-Parameter Lognormal</i>	7357,259	0,995
2-Parameter Exponential	7361,698	*
<i>3-Parameter Loglogistic</i>	7357,482	0,995
Smallest Extreme Value	7361,054	0,991
Normal	7357,269	0,995
Logistic	7359,263	0,992

Tabla 6.17 Bondad de ajuste de los modelos paramétricos

El método utilizado fue el método de los Mínimos Cuadrados, aunque el método de Máxima Verosimilitud, entrega resultados también consistentes muy parecidos. Se concluye por tanto, que el modelo que mejor se ajusta a la edad

(en años) de los conductos de la red, en el instante de su rotura o avería natural (por deterioro, envejecimiento u otra causa resultado del paso del tiempo) es precisamente un Modelo Lognormal con tres parámetros. El siguiente paso es estimar los parámetros de este modelo, para proceder posteriormente a las estimaciones de la supervivencia y la predisposición a las roturas y otros eventos de interés.

6.3.2 Estimación de los Parámetros del Modelo

Los resultados de las estimaciones de los parámetros básicos del modelo seleccionado, esto es, la distribución Lognormal con sus tres parámetros (escala, umbral y localización) así como la estimación de algunas de las características físicas del modelo, se presentan a continuación.

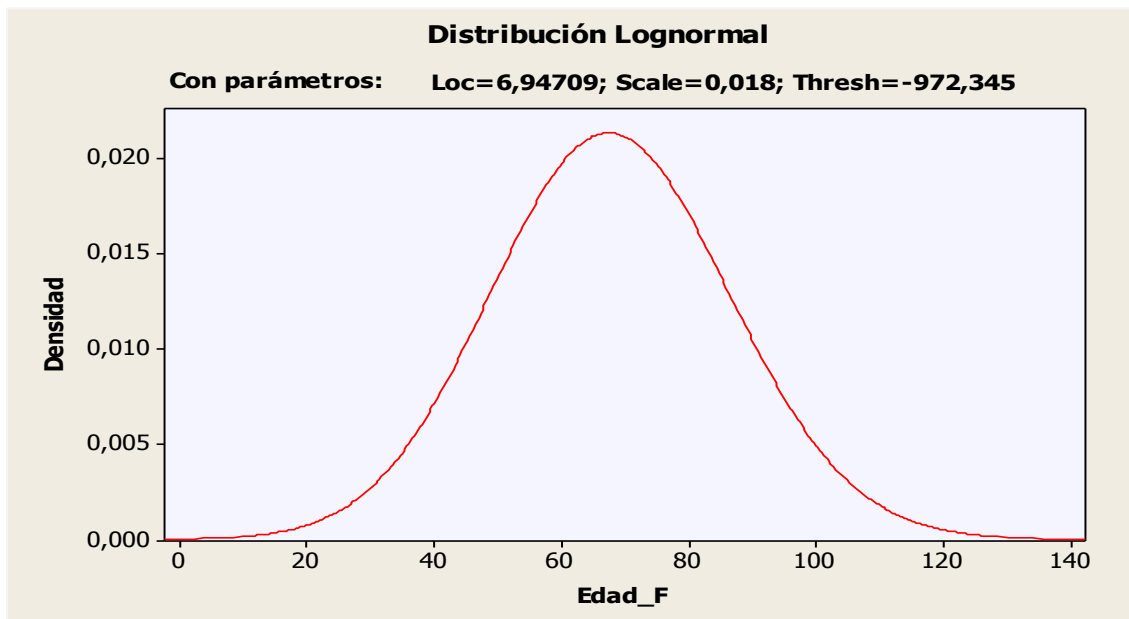


Figura 6.42 Función de densidad para el modelo lognormal estimado con 3 parámetros

Distribution Analysis: EDAD_F				
Variable: EDAD_F				
Censoring Information Count				
Uncensored value 520				
Right censored value 25928				
Censoring value: Estado1 = 0				
Estimation Method: Least Squares (failure time(X) on rank(Y))				
Distribution: 3-Parameter Lognormal				
Parameter Estimates				
Parameter	Estimate	Standard Error	95,0% Normal CI	
			Lower	Upper
Location	6,94709	0,0181651	6,91148	6,98269
Scale	0,01809	0,0006189	0,01692	0,01935
Threshold	-972,345	18,901500	-1009,39	-935,299

Cuadro 6.1 Estimación de las características físicas del modelo.

Además hay que realizar un análisis descriptivo de los tiempos de fallo (media, mediana, percentiles, etc.) usando la distribución teórica ajustada, en este caso la log-normal.

De la salida de MINITAB anterior (Cuadro 6.1) se desprende que los valores encontrados para cada uno de los parámetros básicos de la distribución en referencia son:

Characteristics of Distribution				
	Estimate	Standard Error	95,0% Normal CI	
			Lower	Upper
Mean (MTTF)	67,9413	1,04702	65,8892	69,9934
Standard Deviation	18,8238	0,54941	17,7772	19,9320
Median	67,7710	1,03788	65,7368	69,8052
First Quartile (Q1)	55,1548	0,71307	53,7572	56,5524
Third Quartile (Q3)	80,5422	1,39194	77,8140	83,2703
Interquartile Range (IQR)	25,3873	0,74065	23,9764	26,8813

Cuadro 6.2 Descriptivo de las características de la distribución

el de localización 6,947, el de escala 0,181 y el de umbral o desplazamiento - 972,345. En cuanto a algunas características físicas de esta distribución, se

observa por ejemplo, que el Tiempo medio hasta la rotura o avería natural (MTTF) de un conducto de la red, (por el uso y paso del tiempo u otra causa natural), está entre los 67 y los 68 años. Además después de los 67.77 años sobrevivirán a las roturas, averías y en general al deterioro o degradación natural de la red el 50 % de los conductos y el 25% sobrevivirán a los 80,5 años.

Percent	Percentile	Standard Error	95,0% Normal CI	
			Lower	Upper
0,001	-9,47215	1,2712	-11,9636	-6,98065
0,01	0,0849761	1,02801	-1,92988	2,09983
0,1	11,2113	0,755401	9,73075	12,6919
1	24,8996	0,474422	23,9698	25,8295
2	29,8304	0,413578	29,0198	30,641
3	32,9715	0,395634	32,1961	33,7469
4	35,3409	0,394877	34,5669	36,1148
5	37,2723	0,402589	36,4832	38,0614
6	38,9192	0,414801	38,1062	39,7322
7	40,3654	0,429463	39,5236	41,2071
8	41,662	0,445447	40,789	42,5351
9	42,8427	0,462104	41,937	43,7484
10	43,9307	0,479049	42,9918	44,8697
20	52,0524	0,640547	50,7969	53,3078
30	57,9489	0,78156	56,4171	59,4807
40	63,0142	0,911304	61,228	64,8003
50	67,771	1,03788	65,7368	69,8052
60	72,5498	1,16835	70,2598	74,8397
70	77,6868	1,31141	75,1165	80,2571
80	83,7309	1,48261	80,825	86,6367
90	92,1706	1,72578	88,7881	95,5531
91	93,3115	1,75896	89,864	96,759
92	94,5523	1,79513	91,0339	98,0707
93	95,9184	1,83504	92,3217	99,515
94	97,4461	1,87978	93,7618	101,13
95	99,1911	1,93102	95,4064	102,976
96	101,245	1,9915	97,3416	105,148
97	103,775	2,06627	99,7254	107,825
98	107,148	2,16637	102,902	111,394
99	112,485	2,32571	107,927	117,044

Cuadro 6.3 Descriptivo de los percentiles.

La Interpretación de los resultados de la salida anterior, referente a los percentiles, no es otra cosa que la estimación de la fiabilidad de los distintos conductos de la red para diferentes porcentajes. En la primera columna, se ve por ejemplo, que al valor 10 (de la columna “Percent”) le corresponde el Percentil **43,9307**, que indica que, el 10% de las tuberías presentará averías y/o roturas un poco antes de los 43,9 años, bajo el supuesto de que el tiempo hasta el fallo (o rotura) de las tuberías, siguen un modelo Lognormal de 3 parámetros cuyas estimaciones fueron proporcionadas por el método de Mínimos Cuadrados. Por tanto la fiabilidad con el modelo Lognormal a los 43,9 años es del 90%.

La función de supervivencia de la edad de fallo, suponiendo que su distribución es Lognormal con 3 parámetros, en presencia de alta censura.

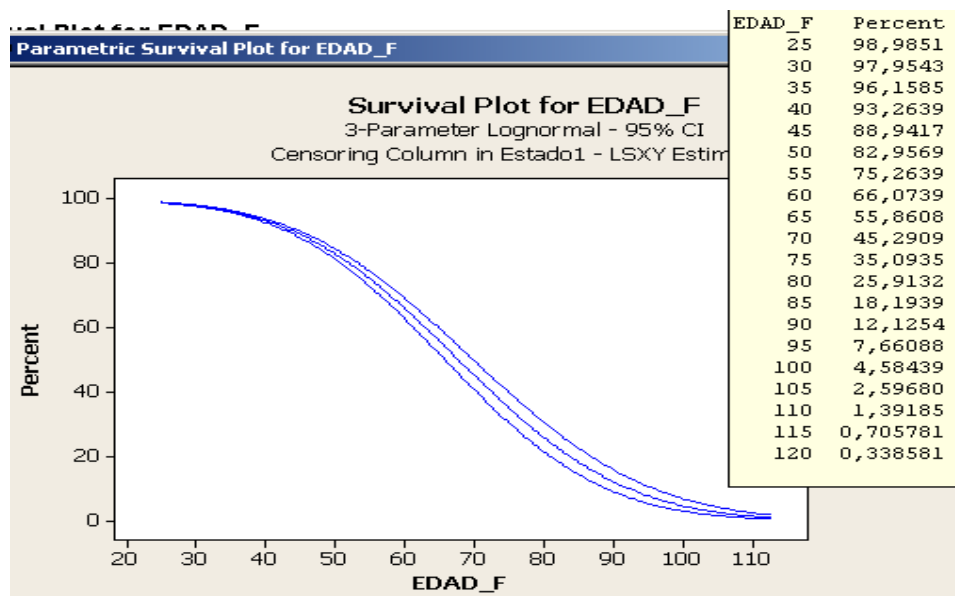


Figura 6.43 Función de supervivencia para el modelo lognormal.

La triple trayectoria de la Figura 6.43 se explica, por tratarse de una estimación de la curva por intervalos, con un 95 % de confianza. La tabla adjunta, está constituida por puntos sacados de la curva en referencia, y corresponden a la estimación de la fiabilidad de tramos de tubería de la red en operación hasta los 120 años.

Respecto a la predisposición al fallo en el tiempo o la rapidez con que se degrada la tubería en general, esta inicia su ciclo de deterioro a partir de los 30

años, como lo indica la curva, al comienzo muy lentamente (con pequeños cambios, insignificantes en forma creciente), pero luego a partir de los 50 años hasta los 80 años se acentúan estos y, el desgaste y/o deterioro de los conductos pasa de un 2% a un 7%, es decir se cuasi cuadruplica en 30 años de operación de la tubería

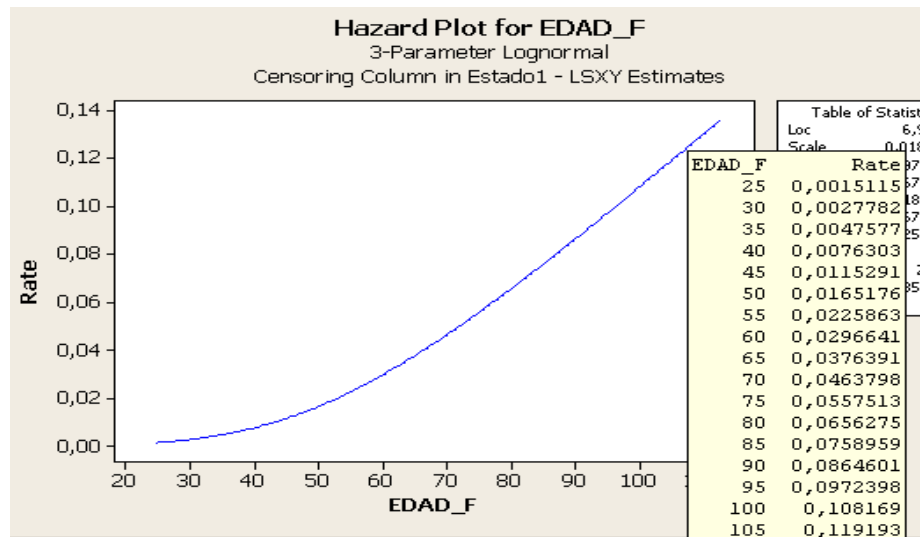


Figura 6.44 Función de impacto para el modelo lognormal.

Se aprecia en la tabla de la Figura 6.44, por ejemplo, que los conductos que un poco antes de los 50 años no han fallado, tienen un riesgo (o predisposición) de fallo del 1,65 % al llegar a esa edad. Es de anotar que después de los 80 años, la tasa instantánea de roturas, más o menos se estabiliza más o menos en 1,0% , entre un año y el siguiente. El modelo seleccionado, esto es la distribución *Lognormal* con tres parámetros:

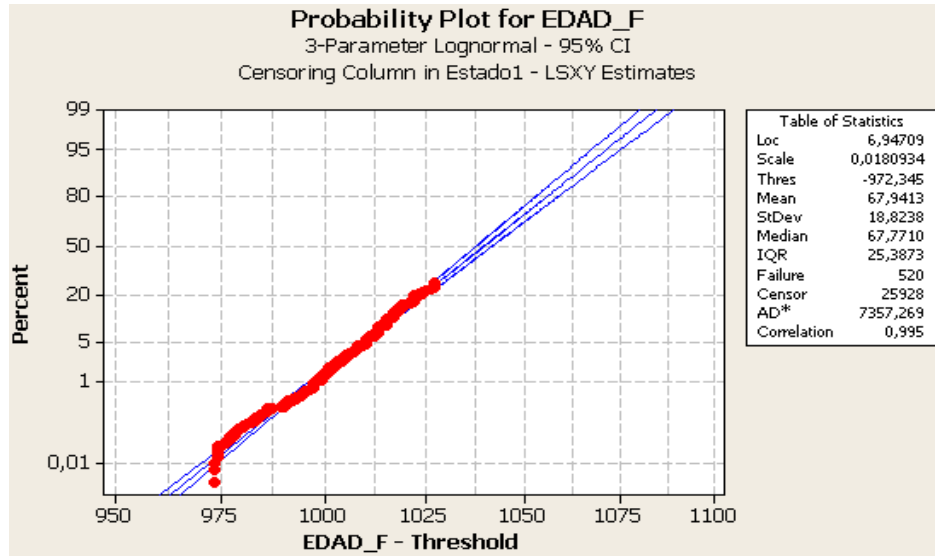


Figura 6.45 Función de distribución linealizada para el modelo lognormal.

La Figura 6.45 muestra el ajuste amplificado de los datos de la variable *Edad de fallo* (EDAD_F), de las tuberías al modelo Lognormal, con una banda de confianza del 95%, en presencia de la fuerte censura; además en el recuadro de la derecha se recogen algunas características físicas del modelo en referencia.

6.4 Comparación de las estimaciones de $R(t)$ con los cuatro métodos

La tabla 6.18, recoge estimaciones en general de la fiabilidad de las tuberías de la red, haciendo uso de métodos no paramétricos (Kaplan- Meier y NAE), semi-paramétricos (Riesgos Proporcionales de Cox) y Paramétricos (modelo Lognormal), en diferentes instantes de tiempo t . Las estimaciones aquí consideradas hacen referencia a tiempos de fallos múltiples de cinco con alta censura en la información.

t	<i>K – Meier</i>	<i>M. Cox</i>	<i>Lognormal</i>	<i>NAE</i>
0,5		0,999000	0,999902	0,996856649
5	0,998719	0,999000	0,999710	0,994981932
10	0,997945	0,998000	0,999207	0,992575786
15	0,996532	0,997000	0,997997	0,988089848
20	0,995018	0,996000	0,995319	0,981722121
25	0,989767	0,992000	0,989851	0,964525363
30	0,977942	0,984000	0,979543	0,935451798
35	0,961544	0,971000	0,961585	0,905196207
40	0,931770	0,948000	0,932639	0,862459925
45	0,871183	0,900000	0,889417	0,789018963
50	0,803080	0,848000	0,829569	0,716774544
55	0,734000	0,786000	0,752639	0,656012392
60	0,710000	0,725000	0,660739	0,656012392

Tabla 6.18 Estimaciones de la fiabilidad de la red con cuatro métodos diferentes

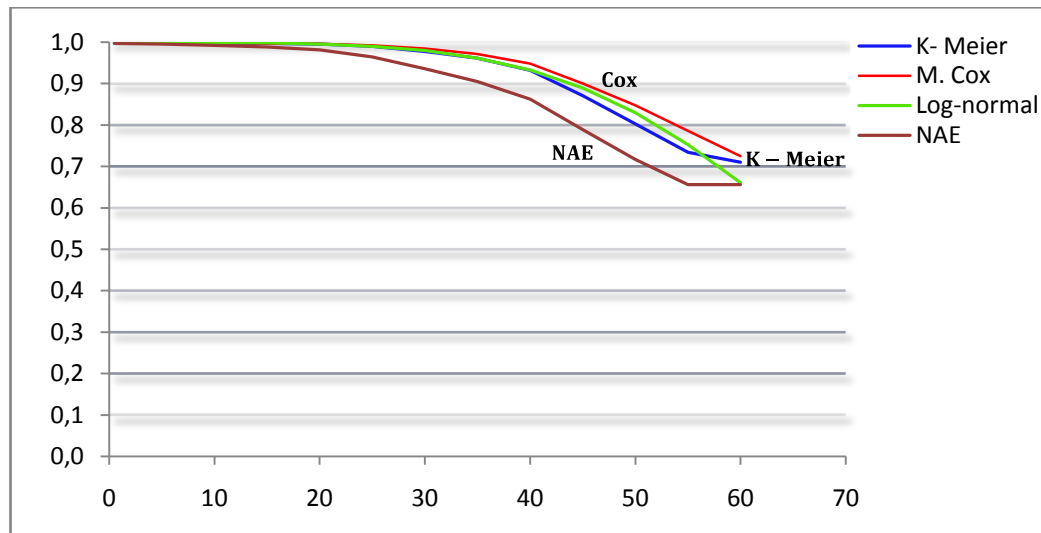


Figura 6.46 Estimación de la fiabilidad de los conductos de la red con 4 Diferentes modelos

Con base en los datos de la Tabla 6.18 y en las curvas de la Figura 6.46, las estimaciones con el Método de Kaplan-Meier (línea azul) y las estimaciones con el modelo Lognormal (línea verde), no presentan mayores diferencias hasta los 40 años de operación de la red.

Aunque las estimaciones de Cox muestran una ligera tendencia a las sobrestimación de la supervivencia de las tuberías, tampoco es que difieran significativamente estas con las estimaciones de los dos anteriores modelos (esto se observa claramente en la figura hasta los 35-40 años), razón por la cual, se puede afirmar que, los métodos de KM, Lognormal y Cox son muy semejantes en sus estimaciones.

Donde sí se presentan diferencias bastante evidentes es, entre el método de Nelson Aalen extendido (NAE) y los tres métodos anteriores.

Esto se aprecia particularmente en la Figura 6.47, donde se muestran las diferentes estimaciones de KM (línea azul) y las de NAE (línea roja). Se observa por ejemplo, que a partir de los 25 años de funcionamiento de la red, las dos curvas (líneas) se empiezan a separar alcanzando diferencias importantes con el paso del tiempo.

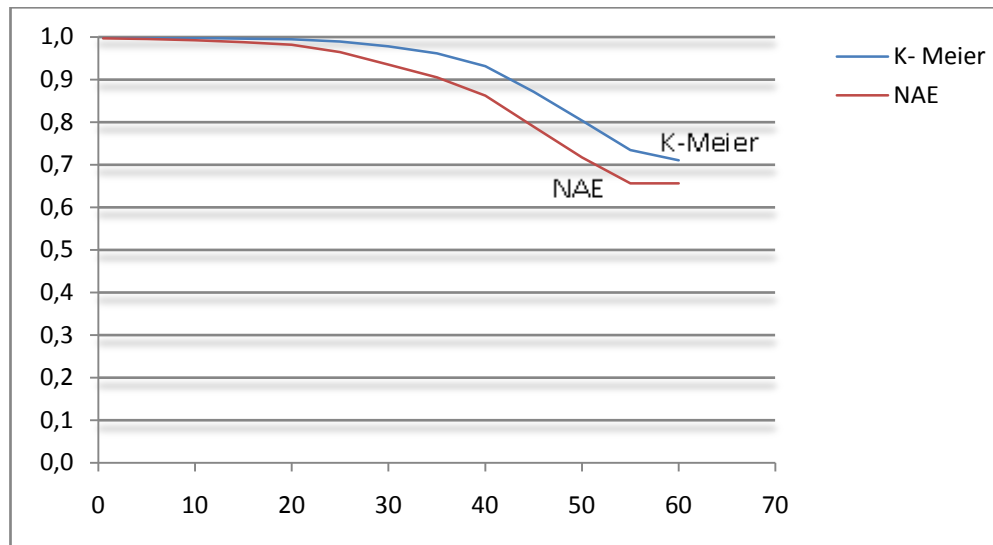


Figura 6.47 Funciones de fiabilidad estimadas con KM y el método de NAE

El hecho de que el método de KM no considere la existencia y/o posibilidad de fallo a la izquierda del inicio del estudio (año 200), y sólo considere información parcial (censura) o/y completa (fallos naturales) recogida durante el período 2000-2005 para construir la curva (azul) de KM; es lo que explica en gran medida, que las estimaciones de KM superen a las estimaciones de NAE, en buena parte del historial de roturas de la red.

Lo anterior se complementa por el hecho de que, considerar los datos con truncamiento (en el estimador de NAE), en realidad equivale a suponer (implícitamente) la posible existencia de fallos y censuras antes del 2000, así no hayan sido registrados ni considerados explícitamente en los cálculos de las estimaciones; lo cierto es, que de alguna manera su influencia es evidente en el cálculo de la magnitud del estimador.

En cualquier caso, el estimador de Kaplan-Meier es un estimador de la función de supervivencia de la variable T (tiempo de fallo); mientras que el estimador de Nelson-Aalen extendido estima la función de supervivencia de la variable aleatoria T condicionada a la restricción $T \geq X$. Es decir, no se refieren a la misma variable.

Finalmente, pareciese ser que el método de NAE en cierta medida, lo que está es corrigiendo el sesgo producido por las sobrestimaciones de KM, al considerar no sólo la censura por la derecha, sino también el truncamiento a la izquierda del inicio del estudio (año 2000)

7. Necesidades de información para la modelización estadística

7.1 Generalidades

Las bases de datos con ciertas características que podríamos llamar “ideales”, (en cuanto a claridad, precisión, suficiencia, oportunidad, calidad y confiabilidad de sus registros) son la referencia básica obligada para el buen diseño de indicadores de gestión. Estos indicadores permiten evaluar y mejorar, de forma cuantitativa y continua, el desempeño de todo tipo de organizaciones y en particular de las entidades prestadoras del servicio de abastecimiento de agua potable y saneamiento (mediante la comparación consigo mismo y con los mejores gestores del abastecimiento de agua potable, a través del *benchmarking*⁶). Algunos aspectos susceptibles de control son el servicio al cliente, el subsistema técnico-operativo, la gestión financiera, el recurso humano, etc.

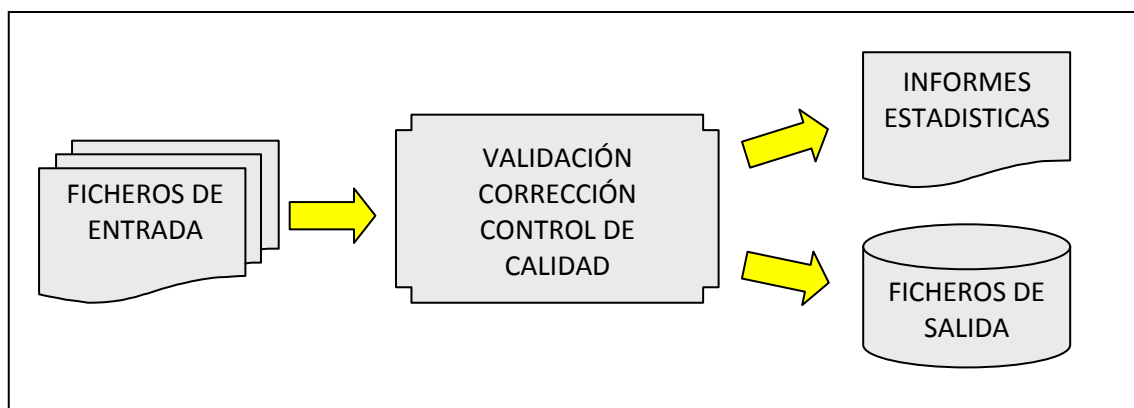


Figura 7.1. Proceso de Gestión de la Información.

Las entidades gestoras del abastecimiento en España se enfrentan a múltiples dificultades en *la toma de decisiones*, por no contar con sistemas suficientemente desarrollados de captura, almacenamiento y gestión de la

⁶ Para los propósitos de este documento, el *benchmarking* debe entenderse como “El proceso de búsqueda continua de las mejores prácticas en la gestión administrativa, técnica y financiera de los organismos operadores de agua potable y saneamiento, mediante indicadores de gestión”. También se define como “la búsqueda de las mejores prácticas que conducirán a un excelente desempeño”

información (es decir, de contar con bases de datos confiables). Ello afecta a las áreas de mejoramiento de sus prácticas de calidad, optimización de sus recursos para la prestación de un buen servicio, etc. Esos sistemas serían la base de la *retroalimentación* (a través de los indicadores de gestión) en la industria del agua.

Cabrera R.E et al. (1999)⁷, lo precisan claramente cuando afirman: “...*el desconocimiento del rendimiento de la red, del estado de las conducciones o de la condición en la que se encuentra el parque de contadores (si es que éste existe) limita la posibilidad de una decisión. Las empresa gestoras de abastecimientos por su parte realizan estudios, basados generalmente en la experiencia, que les permiten dilucidar si el abastecimiento puede resultar rentable o no....*”.

Nuestro interés en lo que resta del documento, se centra fundamentalmente en responder a la pregunta: *¿qué tipo de información debería recogerse para modelizar el comportamiento de los sistemas de distribución de agua frente al fallo?*

Para ello se van a utilizar dos aproximaciones. Por un lado, se aprovechará la revisión bibliográfica realizada y buena parte de la descripción resumida realizada en el apartado 4.3 (Síntesis de algunos modelos estadísticos utilizados en la predicción de fallos en redes de abastecimiento de agua potable). Por otro lado, se realizó una valoración propia de la base de datos utilizada en este trabajo, para configurar a partir de este análisis, algunos lineamientos (sugerencias y recomendaciones) sobre el tipo de información requerida, para garantizar la calidad y confiabilidad del dato.

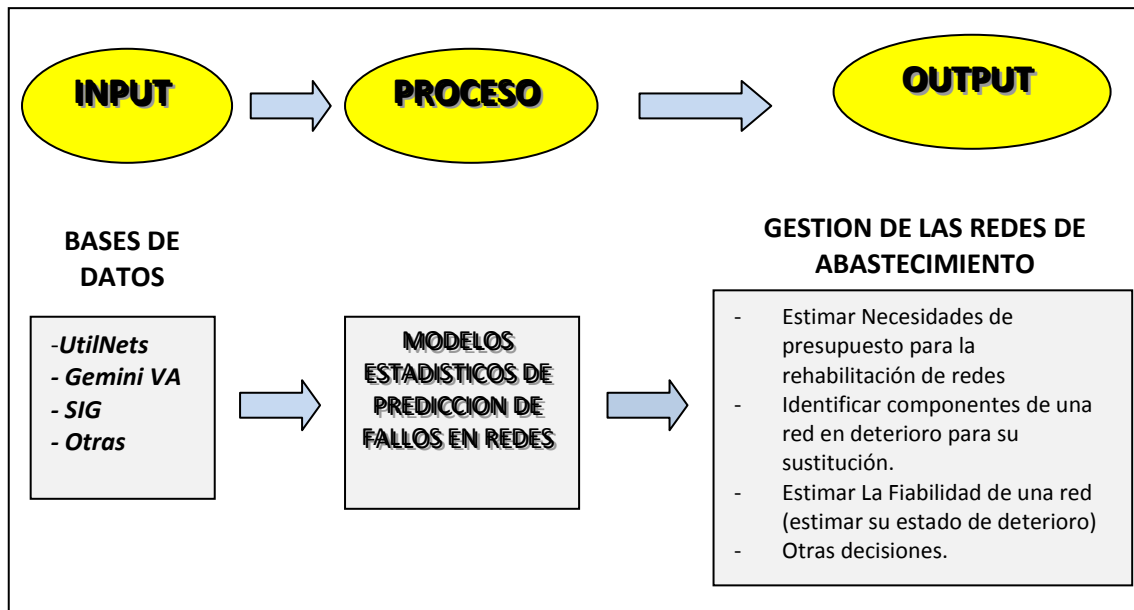
Debe tenerse en cuenta que esos datos son la “materia prima” que alimenta los modelos estadísticos de predicción de fallos, base del esfuerzo por una gestión *proactiva* eficiente en las estrategias de conservación y rehabilitación de las redes de abastecimiento y sus componentes. En consecuencia, sin datos de calidad difícilmente pueden ser tomadas decisiones acertadas.

⁷ “Indicadores de Gestión. Un mecanismo para el control de abastecimientos” Grupo Mecánica de Fluidos, UPV. (1999)

7.2 Bases de datos en fiabilidad

La mayor parte de los servicios públicos de abastecimiento de agua más grandes de Europa, han venido utilizando en su sistema de gestión bases de datos que, a pesar de haber sido diseñadas y construidas con las últimas exigencias de la tecnología actual, no dejan de presentar algunas limitaciones en su implementación. Por ejemplo, adolecen de exceso de módulos y variables registradas en el sistema, insuficiencia de información (registro únicamente del primer fallo de cada tubería, no se discriminan los fallos por tipo de rotura, consideración de un solo tipo de material en las tuberías, etc.). Tal es el caso, por ejemplo, de los siguientes sistemas de información: “*UtilNets (1997)*”, creado con fondos de la Unión Europea para las empresas de servicios públicos de abastecimiento de agua. El sistema Noruego “*Gemini VA*” con funciones de mapeo temático automatizado (que también incluye registros de fallos y reparaciones de tuberías, con el mismo propósito de gestión que el anterior, pero en Noruega). Los *Sistemas de Información Geográfica SIG*⁸ (1966 USA), que además de sus funciones de mapeo temático, permite mejorar la tradicional presentación de los resultados de los modelos de predicción estadística, de forma tal que se facilite la interpretación de estos resultados, para dar apoyo eficiente y eficaz a la Gestión de las redes de abastecimiento de Agua potable.

⁸ Un Sistema de Información Geográfica (SIG o GIS, en su acrónimo inglés) es una integración organizada de hardware, software, datos geográficos y personal, diseñado para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión.



(Fuente: Elaboración propia).

Figura 7.2 Esquema general del proceso de Modelización estadística para la Gestión de Redes de Abastecimiento de Agua Potable.

La idea es avanzar hacia la *gestión proactiva* en la Industria del Agua, donde las redes y sus componentes (tuberías, codos, bombas, válvulas, etc.) puedan ser rehabilitadas en el momento oportuno, antes de que ocurran los fallos. En la actualidad, se está trabajando en la integración de diversos modelos de predicción en estas bases de datos, de modo que estas (con los modelos integrados) se conviertan en una poderosa herramienta de planificación, en particular para la rehabilitación de las redes.

Al respecto Ansell, J.I. y Phillips, M.J. (1994), destacan algunos errores que usualmente cometen los encargados de la recolección de datos históricos en análisis de fiabilidad, extensivos a las redes de abastecimiento de agua Potable.

- **Recolección de datos genéricos.** Algunas bases de datos recogen información en escenarios y entornos muy generales que sólo pueden ser procesados para sistemas y/o componentes igualmente muy generales, ignorando totalmente el contexto original de los datos, por ejemplo, no es suficiente con registrar el fallo de una tubería (que es un

elemento reparable), sino que además es necesario precisar en que situación ocurrió, si fue por una u otra causa, natural, provocada, etc.

- **El uso de datos cualitativos.** Es muy frecuente también, en los análisis de la fiabilidad de las tuberías, incurrir sistemáticamente en la no utilización de datos cualitativos (a pesar de estar estos registrados en la base de datos), con el pretendido fin de alcanzar “*un mayor impacto en el análisis y comprensión de los datos*”.
- **Clarificación de definiciones.** Otro problema es la ausencia absoluta de claridad en la definición de algunas covariables o la ambigüedad con que se las trata, tal es el caso de las covariables *modos de fallo, síntoma de la rotura y Causa de la rotura*. Incluso, en el mismo sistema de distribución del líquido y sus componentes (tubos, válvulas, bombas, etc.), con frecuencia no se definen con claridad o no se especifica con precisión, cuáles son las unidades de medida que se van a utilizar en el registro de los datos.
- **Validación de los datos.** Otro error frecuente en el que se incurre al utilizar o construir bases de datos en Fiabilidad, es la no comprobación o validación de los datos registrados en ella, (comprobación que debería hacerse incluso antes de incluirlos en la referida base de datos) dando lugar a reiteradas inconsistencias.
- **Uso de datos secundarios.** El hecho de que la mayor parte de los datos sean registros sobre mantenimiento, fallos (averías, roturas, degradación, rehabilitaciones) e inventarios de tuberías y otros componentes de la red y que la principal preocupación del operario que recopila los datos no sea la fiabilidad del dato, sino el interés por no incurrir en un reporte incompleto de la información registrada y un deficiente análisis de los fallos , hace que se pierda de vista la importancia del procesamiento de información primaria y que se incurra por lo tanto en las siguientes prácticas *de pérdida de información primaria*:

- a) No anotar la frecuencia con que ocurren las incidencias (fallos, roturas), de las tuberías durante su vida útil, con la consecuente imposibilidad de poder analizar el *Nº de fallos (roturas) previos* en un instante dado en una tubería, que es un factor significativo importantísimo en la predicción de futuros fallos con los modelos de Walski y Pellicia, (1982); Marks et al. (1985); Andreou's et al.(1987)
 - b) Cuando se registra un fallo, el mismo no se anota inmediatamente en el instante en que ocurre (fecha, lugar y hora), en lugar de esto sólo se registran las fechas inicial y final del trabajo de reparación.
 - c) Otro ejemplo de pérdida de información primaria, se aprecia cuando en la tabla, en una donde debería figurar determinada clasificación, un operario de mantenimiento ha registrado “Desconocido” o “Ver Nota” en el momento de la anotación. Este tipo de registro termina interpretándose como un valor “*faltante*” o valor ““missing”. En nuestro estudio, este es el caso que se presenta con las variables TRAFICO y PRESION.
- **Aplicabilidad de los datos.** La relevancia de los datos en su aplicabilidad, es también un elemento importante en una base de datos histórica que se utilice para hacer estudios de fiabilidad. “¿Son aplicables los datos a estudios futuros?”; “¿Son los datos aplicables al estudio que vamos a realizar?”; “están actualizados los datos con las últimas mejoras tecnológicas o de procedimiento realizadas al sistema y/o alguna de sus componentes?”

Resumiendo, la no verificación (validación) de los datos, la falta de claridad en la definición de algunas variables, la inaplicabilidad de la información, el recoger datos demasiado genéricos y secundarios y la incapacidad de utilizar información cualitativa, invalidan el uso de algunas bases de datos históricas para estudios de Fiabilidad.

7.3 Variables de mayor influencia en el fallo

Aunque la tarea de hacer una recopilación exhaustiva y meticulosa en la literatura sobre este tema (por cierto muy numerosa), no es precisamente el objetivo de este trabajo, si creemos oportuno presentar al menos un resumen (de forma general y un tanto esquemática), de los resultados que muchos estudiosos del tema, han logrado identificar conjuntamente desde mediados de los años 60 (Mosevoll, G. (1994).)

Variables estructurales	Variables externas o ambientales	Variables internas	Variables de mantenimiento
Localización de la tubería	Tipo de suelo	Velocidad del agua	Fecha del fallo
Diámetro	Carga sobre la tubería	La presión del agua	Fecha de la reparación
Longitud	Aguas subterráneas	La calidad del agua	Ubicación del fallo
Año de construcción	Corriente aislada directa	Golpe de ariete	Tipo de fallo
Material de la tubería	Tasa de fugas	Corrosión interna	Historial de fallos previos (Nº fallos previos)
Método de articulación	Otras redes		
Protección interna	Sal para descongelar carreteras		
Protección externa	Temperatura	Actores	
Tipo de Presión	La corrosión externa		
Grosor de la pared			
Profundidad del tubo			
Condiciones de asentamiento del tubo			

Fuente: Rostum J. (2000). *Statistical Modelling of Pipe Failures in Water Networks* (Tesis doctoral)

Tabla 7.1 Factores que afectan el deterioro estructural de la distribución del agua, según Rostum et al (1997).

Al respecto, autores como Goulter and Kazemi (1988), Shamir and Howard (1979) y otros, coinciden en la identificación de un número mínimo de

factores o covariables que son causas generales del deterioro de las tuberías en las redes de distribución de agua potable, afectando de una forma u otra la normal operatividad de dichas instalaciones. Rostum et al. (1997) van más allá al agrupar en 4 categorías o tipos de factores, las variables de mayor significancia en el deterioro estructural de los sistemas de abastecimiento de agua; estas categorías corresponden a factores: *Estructurales o físicos, externos o ambientales, hidráulicos o internos y Factores de mantenimiento.*

Si bien es cierto, la mayor parte de estas covariables se mantienen constantes en el tiempo, es decir, no dependen de este, otras como la calidad del agua (el PH y algunas características físico químicas del líquido), la velocidad de circulación del agua y la carga que se aplica sobre la tubería (de la tierra, de la intensidad del tráfico), sí que dependen del tiempo. La tabla 7.1 recoge en forma resumida las cuatro categorías mencionadas.

7.4 Problemática de la base de datos analizada

El análisis de la base de datos objeto de estudio ha proporcionado valiosos resultados en lo que respecta a las estimaciones de la supervivencia de las tuberías, la predisposición a los fallos en el tiempo, y la identificación de factores de riesgo en el comportamiento de la tasa de roturas y averías y en otros temas de interés.

Sin embargo, al comienzo del estudio se detectaron algunas *debilidades* en la estructura general de la base de datos original que bien vale la pena mencionar a de cara a su mejora y por el interés general que puede tener la evitación de ese tipo de errores:

- Ausencia total del registro de fallos recurrentes en cuanto a los tiempos de fallo, es decir, sólo se consideró un único fallo en cada una de las tuberías, como si estas fuesen únicamente *elementos no reparables* (el registro de las diferentes reparaciones de la tubería en su vida útil es de vital importancia en la modelación estadística).
- Existen algunas columnas, que no aportan información a los modelos de predicción utilizados en este estudio (algunas por no estar definidas). Como por ejemplo podemos citar *DCONS_ID*, *DROTU_ID*, *NOMBRE* y *PEP_SAP*, entre otras. Se trata de variables que pueden ser de interés en otras áreas de actividad, pero no lo son para el análisis estadístico realizado. No se plantea su eliminación.
- Ausencia de información sobre la calidad del agua (pH, contaminación, turbiedad, etc.), tipo de suelo en que se instala la tubería y otras sobre la corrosión interna y externa de la tubería.
- El sistema actual de codificación de los materiales de los conductos, se debe actualizar en lo que corresponde a los materiales plásticos (algunas nuevas variedades de Polietileno, desarrolladas en los últimos 5 años). Es de anotar que en la actualidad y con carácter general, los poco definidos sistemas de codificación de los materiales son fuente de muchos problemas para los gestores de la rehabilitación de las redes.

- Si el interés del estudio fuese la valoración del estado estructural de deterioro de una buena parte del sistema de abastecimiento (no sólo el deterioro de las tuberías), la información actual no sería suficiente, pues se deberían incluir datos adicionales sobre características de otros componentes como válvulas, hidrantes, bombas, codos, etc., en este caso el volumen de información que se necesitaría para la modelización con fines de predicción sería mucho mayor.
- No se incluyó la definición y recolección de datos (covariables) importantes, que establecieran si la tubería *había sido intervenida o no* anteriormente antes del último fallo detectado en un instante dado y si la intervención (si este fuera el caso), se debía a una operación de mantenimiento, sustitución y/o *rehabilitación*⁹ de la misma.

⁹ La Rehabilitación de un Sistema o dispositivo (tubería), comprende actividades de restauración, mejoramiento del desempeño de la tubería; también incluye mantenimiento y reparación, así como, renovación y reemplazo o sustitución de la misma.

7.5 Contenido y características recomendables en una base de datos de vida de redes de suministro de agua

Es complicado hablar de una base “ideal”, aplicable a todos los estudios de fiabilidad, por cuanto cada estudio puede tener sus propios objetivos de investigación. A su vez el diseño de buenas bases de datos está condicionado en gran medida por esos objetivos. Sin embargo con base en el análisis de las secciones 4.6.2, 7.3 y 7.4, se proponen a continuación algunas pautas y lineamientos (de carácter general), para orientar el trabajo en este sentido.

Los datos que se buscan deben concernir a servicios de suministro de aguas que pretendan conservar registros sobre un período suficientemente largo de todos los trabajos de mantenimiento y en particular de todas las roturas y fugas reparadas sobre los conductos. Sería deseable que los registros de mantenimiento más cortos estén por lo menos entre cinco y diez años.

Se requieren los siguientes tres bloques de información:

- Información básica, referida al conjunto de la red
- Información detallada de las características de cada uno de los tramos que integran la red
- Información relativa a los fallos y averías detectados y reparados.

Cada uno los dos últimos bloques constituye una base de datos, relacionadas entre sí a través del código de tramo, tal cómo se representa en la Figura 7.3.

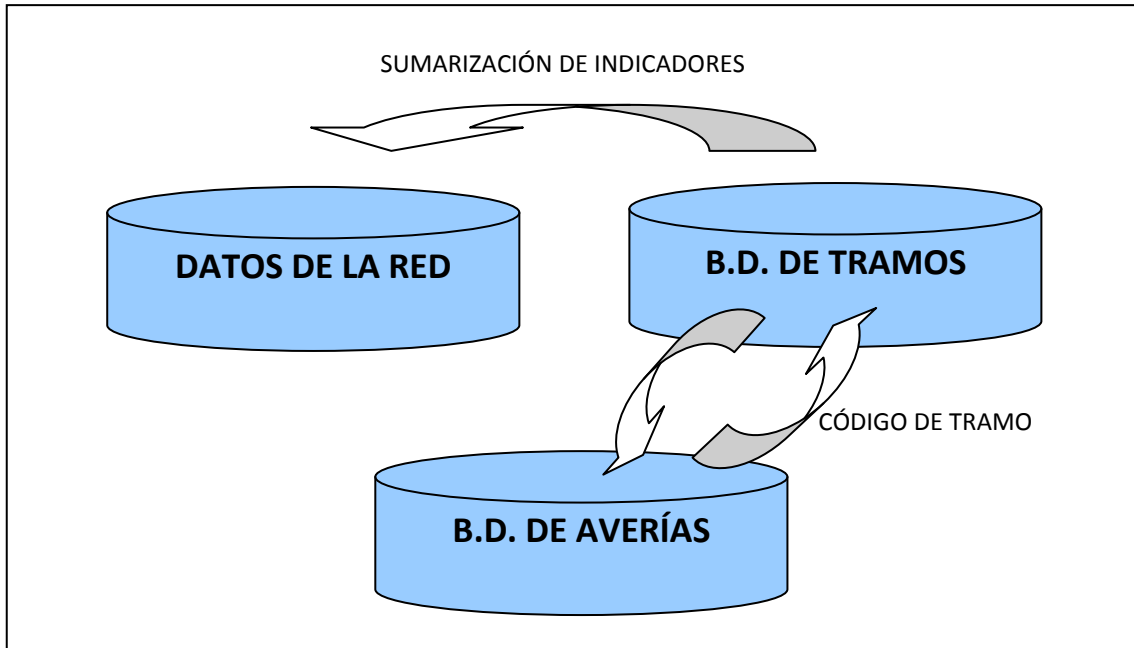


Figura 7.3. Estructura de Bases de Datos

7.5.1. Características de la red

Los datos relativos al conjunto de la red, que son considerados necesarios, son los que figuran en la Tabla 7.2.

Esta base de datos, es en parte una tabla resumen de una serie de indicadores descriptivos de la red, y por otra parte contiene información que si puede adoptar la forma de una base de datos.

Otra información que sería conveniente tener:

- Fechas en las que se han producido cambios en las técnicas de instalación, si se pueden establecer "hitos" o fechas clave.
- Fechas en las que se han producido cambios consistentes en el uso de nuevos materiales (juntas, revestimientos, etc.), si se pueden establecer "hitos" o fechas clave.

Variable	Descripción
KMTOTAL	Longitud de la red
KMSECTOR1	Distribución de los conductos por sectores o distritos
KMSECTOR2	Id.
....
NUM_FUENTES	Origen del agua: Número de zonas de alimentación o captación del agua
ID_FUENTE1	Nombre de la captación 1
ID_FUENTE2	Nombre de la captación 2
....
Q_FUENTE1	Porcentaje de suministro de la fuente 1
Q_FUENTE2	Porcentaje de suministro de la fuente 2
....
pH_FUENTE1	Calidad del agua de la captación 1: Análisis químicos tales como; oxígeno disuelto, pH, conductividad del agua
pH_FUENTE2	Id.

Tabla 7.2. Datos relativos al conjunto de la red.

7.5.2. Características de los tramos

Para la descripción de los tramos de conducto que integran la red, se considera necesario disponer, para cada uno de ellos, de la información recogida en la Tabla 7.3. Esencialmente, las características observadas en los tramos de tuberías son de tipo físico- medioambiental (localización del tramo y su entorno, fecha de instalación, características físicas, etc.)

Variable	Descripción
ID_TRAMO	Código de identificación del tramo
SECTOR	Sector de la ciudad/red
CALLE	Calle
FECHA_INST	Fecha de instalación
DIAMETRO	Diámetro de la conducción (en mm)
MATERIAL	Material de la conducción
LONGITUD	Longitud del tramo (en m)
CORROSIVIDAD	Medición de las características del suelo o sustrato en que se coloca la instalación, a través del pH o su resistividad. Puede ser cuantitativa o cualitativa. Podían ser varias variables que recogieran varios aspectos del suelo: humedad, tipo de suelo,...
TRAFICO	Esta variable busca determinar si el tráfico es débil o elevado. Se puede medir, por ejemplo, de acuerdo al número de vehículos por hora o el tipo de carretera (carretera local o nacional). Puede ser cuantitativa o cualitativa.
PRESION	Presión de funcionamiento del tramo (mca)
USUARIOS	Usuarios/km. Este dato puede ser obtenido contando el número de usuarios por calle a partir de los ficheros de usuarios del Servicio de Agua y midiendo la longitud de cada tramo
FUENTE	Identificación de la fuente de abastecimiento
ACERA_CALZADA	Bajo acera o calzada. Este factor caracteriza el hecho que el conducto se encuentra bajo la calzada o la acera
PROFUNDIDAD	Profundidad de la instalación
TIPO DE SERVICIO	Industrial, doméstico, servicios públicos,...

Tabla 7.3. Datos de los tramos

Puede valorarse la necesidad de incluir datos sobre otras características físicas de la tubería (como la rugosidad interior/externa), o como variable secundaria el número de averías que ha sufrido el tramo, la fecha de la última

reparación o sustitución, características del agua circulante (si es que existe variación en la misma) o las que en cada caso se consideren adecuadas.

7.5.3. Información sobre fallos y averías.

La información considerada necesaria para la adecuada descripción de los fallos las consiguientes reparaciones es la que se recoge en la Tabla 7.4.

Variable	Descripción
ID_TRAMO	Código de identificación del tramo en que produce el fallo
ID_AVERIA	Código de identificación de la avería. Asocia ésta con un parte de avería/reparación.
FECHA_FALLO	Fecha del fallo o del informe de
TIPO_FALLO	Tipo de fallo (ej.: natural, provocado,...)
CAUSA_FALLO	Causa del fallo, si se conoce
FECHA_REP	Fecha de la reparación o del informe de la reparación
TIPO_REP	Tipo de reparación (sustitución/ reparación, elemento afectado, ...)

Tabla 7.4. Datos de los fallos

Puede considerarse alguna otra variable, como el síntoma del fallo, el equipo responsable de la reparación, etc.

Debe definirse también el tratamiento que se vaya a dar a los tramos reemplazados. Un opción recomendable sería incluir en la base de datos de tramos el campo “fecha de la última sustitución”, que en principio coincidiría con la fecha de instalación, pero que con el tiempo serviría mejor que ésta para el cálculo de la antigüedad de la tubería.

La relación de variables recogidas en la estructura anterior está en parte contenida en la tabla de Rostum et al (1997) y esencialmente es una

combinación de tres tipos de factores, todas las variables de *mantenimiento* con una parte importante de las variables *estructurales* y de las *ambientales*.

En cuanto a la implementación del sistema, se debe tener cuidado, de no diseñar voluminosas BD con excesivas e inviables cantidades de información, que sólo contribuirían a complicar su operatividad y análisis, así como a un mayor consumo de recursos de todo orden. En lugar de esto, se sugiere a las empresas de servicios públicos que empiecen por recoger inicialmente los datos estrictamente necesarios, como parte de su funcionamiento normal, teniendo en cuenta de no caer en los errores anteriormente señalados en la sección 7.4.

La alimentación de algunos de los modelos estadísticos descritos anteriormente en la sección 4.6.2, con este tipo de información, produce los conocidos *indicadores de gestión*, que conforme se mencionó anteriormente, son la base para la gestión en las redes de abastecimiento, cuya lista se describe a continuación.

7.6 Indicadores estadísticos para la gestión

Las siguientes son algunas de los indicadores y características asociadas a las redes de tuberías en los sistemas de distribución de agua potable, que se podrían estimar cuantitativamente para apoyar la gestión de estos sistemas, desde el punto de vista técnico-operativo y de calidad en el servicio:

1. Número de roturas anual por unidad de longitud (Km^{-1} , año^{-1}).
2. Numero de reparaciones (después de la primera rotura).
3. Edad de la tubería a la primera rotura.
4. Numero de roturas por año.
5. Número de años transcurridos desde la instalación hasta la primera rotura.
6. Número promedio de tramos de tubería sin roturas (o averías).
7. Riesgo de fallo de la tubería (en el instante t).
8. Rapidez o velocidad con que ocurren los fallos (Tasa de fallos)
9. Valoración del efecto de las covariatas en la aceleración del tiempo de vida.
10. Número promedio de fallos por unidad de longitud a la edad t .
11. Identificación de los factores de mayor impacto en el fallo de las tuberías (Identificación de factores de riesgo).
12. Comparación del riesgo de fallo en poblaciones no homogéneas.

Como ejemplo podemos considerar el indicador 1. (*Número de roturas anual por unidad de longitud (Km^{-1} , año^{-1})*), que se obtuvo con el modelo de *Shamir y Howard (1979)*, a partir de variables como: Longitud, fecha de instalación (del tramo), número de roturas previas, etc. y algunas variables ambientales (tipo de suelo, carga, tipo de rotura, etc.) entre otras. Se utiliza muy frecuentemente para estimar el instante de sustitución de un tramo de tubería y bien pudiera constituirse en la base para definir una nueva política de rehabilitación de los conductos de la red. Por ejemplo, “*reemplazar un tramo con más de x fallos al año por Km de longitud*” o bien considerar otro indicador (estimado con otro modelo) para definir la misma política, como por el ejemplo, “*si la tasa de reparaciones o de fallo es superior 0,2, reemplazar la tubería*”.

7.7 Explotación de la base de datos

En las secciones anteriores, se ha definido ya la estructura y contenidos que se proponen para la Base de Datos de uso en fiabilidad, y se han identificado los principales indicadores de utilidad para la gestión del mantenimiento y renovación de las redes de suministro. Se plantea ahora la definición de las características básicas del sistema de explotación de esa base de datos, como un sistema de ayuda a la toma de decisiones.

La figura 7.4 presenta una idea esquemática de este sistema. Tras disponer de la base de datos propuesta, con la información requerida por los modelos predictivos seleccionados, se dispondría de un software de cálculo y predicción. Este software suministraría esencialmente dos tipos de información:

- Información sobre la fiabilidad y el riesgo de fallo de los tramos integrantes de la red. Puede pensarse inicialmente en dos formas de presentación de esta información:
 - Un sistema más o menos automático que identificara cuáles son los tramos de la red con mayor riesgo de fallo, a partir de la información disponible. Esa información podría ser base para políticas de renovación y mantenimiento preventivo.
 - La posibilidad de que el analista solicite la evaluación de riesgo para un tramo determinado.

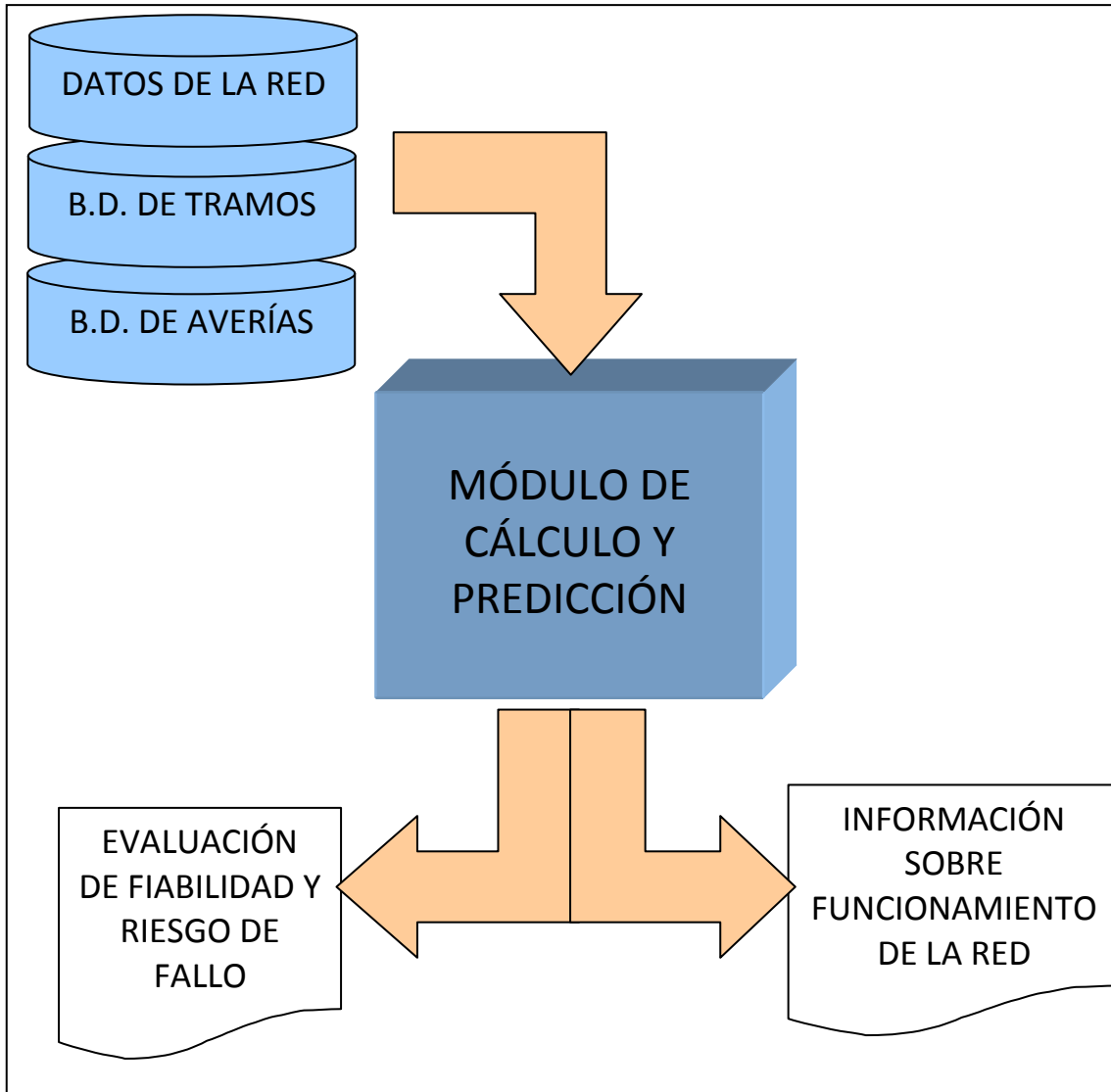


Figura 7.4. Esquema de explotación de la base de datos

- Información de carácter más general sobre el funcionamiento de la red. Esa información se estructuraría a partir de algunos de los indicadores antes descritos, por ejemplo con una estructura de “cuadro de mando”. Esta información permitiría el seguimiento a largo plazo del rendimiento global de la red y de las incidencias que le afectan.

8. Conclusiones y sugerencias

A continuación se presentan las conclusiones que se desprenden del trabajo realizado. Para más claridad esas conclusiones se han estructurado en varios bloques:

8.1 Conclusiones sobre la base de datos utilizada

La base de datos objeto de estudio ha permitido obtener valiosos resultados en lo que respecta a las estimaciones de la supervivencia de las tuberías, la predisposición a los fallos en el tiempo, y la identificación de factores de riesgo en el comportamiento de la tasa de roturas y averías y en otros temas de interés.

No obstante, dicha base de datos adolece de una serie de debilidades que, al menos en parte, podrían ser corregidas sin grandes dificultades. Podemos recordar las ya citadas en el documento:

- Errores repetidos en algunas variables (año de instalación, diámetros, etc.)
- Gran frecuencia de año de instalación igual a 1900, valor convencional usado al no disponer de la fecha auténtica.
- Edades de fallo negativas. Este hecho se presenta cuando la fecha de instalación es mayor que la fecha de reparación, consecuencia de la forma de reflejar la sustitución de tramos en la base de datos.

Ello obligó a una labor de depuración de la base de datos, que redujo, desafortunadamente, el número de fallos disponibles. Dentro de esta depuración, se eliminaron también aquellos tramos cuyo material tiene una proporción menor al 0.2% en la muestra, dado que no resulta de interés el análisis de materiales que tienen frecuencias casi nulas.

Desde 1948 a nuestros días, el material que más se ha empleado en las instalaciones de los diferentes tramos de tubería de la red, corresponde a la

Fundición dúctil con un 73,9%, (además de ser el más “joven”, pues se comienza a emplear con fuerza a partir de 1970); le sigue la Uralita en cuanto a la frecuencia de utilización, con un 22,6 %, sólo que este material es de los más antiguos. El material que menos se ha utilizado es el polietileno con un 1,4%, aunque es uno de los más “jóvenes”.

8.2 Conclusiones sobre los modelos no paramétricos

Los modelos no paramétricos son métodos analíticos y gráficos que permiten interpretar los datos obtenidos sin la distorsión que podría causar la elección de un modelo subyacente no demasiado acertado.

Se han empleado en la estimación de la fiabilidad de la red, en primer lugar las tablas de vida (método actuarial) y las estimaciones por el método de Kaplan-Meier, y posteriormente se han hecho estimaciones por el método de Nelson-Aalen (NAE) considerando el enfoque de la censura y el truncamiento en los datos, como se explicó anteriormente (Figura 5.1). Los resultados sugieren lo siguiente.

- La fiabilidad en la red por tipo de material, es muy similar con los dos primeros métodos (ver Figuras 6.5 y 6.15). Con el método de NAE, se aprecian importantes y marcadas diferencias en la fiabilidad de los materiales, a partir de los 27-30 años de operación de la red (ver Figura 6.24), por ejemplo, el material que parece tener un mejor comportamiento, desde el punto de vista de su fiabilidad, es la fundición dúctil, le sigue la uralita y finalmente la fundición gris. En el caso del polietileno, los resultados no son muy coherentes con la realidad y se sospecha que las razones están en:
- El insuficiente número de fallos observados (Tabla 6.9), la falta de información sobre el tipo de polietileno utilizado (lo hay de alta y baja densidad y de diferente espesor) y los posibles fallos ocasionados por problemas técnicos de instalación en los distintos tramos de la red, pueden ser los factores que expliquen en alguna medida, los *inquietantes* y contradictorios resultados obtenidos en la estimación de la fiabilidad del polietileno. (con Kaplan Meier y con las tablas de vida). Por cierto, este material es considerado de excelentes características técnicas, con muchísimas propiedades en cuanto a su manejo, flexibilidad y resistencia (a la corrosión, a la abrasión, a los agentes químicos, que no se oxida ni

se pudre, etc.), razón por la que, se debe ser cauto y especialmente cuidadoso en las apreciaciones y conclusiones que se extraigan sobre la estimación de su fiabilidad.

- La consideración de los fallos provocados (en parte evitables con una gestión adecuada) implica una reducción en la probabilidad de supervivencia de los tramos: por ejemplo para la edad 40 años, si sólo se consideran los fallos naturales la supervivencia alcanza el 89%, mientras que la inclusión de los fallos provocados reduce ese valor al 80%.

Tal como el sentido común y la práctica lo sugieren, se ha comprobado con estas técnicas de libre distribución que:

- Por *longitud*, los tramos que menos fallan por causas naturales son los de menos de 2,0 metros y los que más fallan son los de más de 50 metros.
- Por *diámetro*, los que menos fallan por el paso natural del tiempo, son los tramos con más de 300 m.m de diámetro y los que más fallan son los de menos de 90 m.m.
- Por *tráfico*, el tráfico pesado es el que tiene mayor efecto sobre la alteración del envejecimiento natural de las tuberías y es en esta modalidad donde se presenta las mayores posibilidades de fallo, en contraste con las tuberías instaladas en el tráfico acera, donde las probabilidades de supervivencia son mayores.
- Las estadísticas anteriores sobre deterioro de las tuberías de la red resultantes de la aplicación de Kaplan-Meier y de las Tablas de vida, si bien es cierto son cifras muy coherentes, no dejan de ser sobrestimaciones (por las razones anteriormente dadas), cuyo sesgo se desconoce.
- A nivel general se observa, que la vida mediana de las tuberías de la red alcanza valores altos, más del 50% de los tramos tiene una vida superior a los 55 años.

8.3 Conclusiones sobre los modelos Semiparamétricos: factores influyentes en la supervivencia

El modelo de riesgos proporcionales, ofrece una representación más unificada de la relación entre el entorno y el proceso de fallos, dado que nos permite modelar no sólo la relación entre la tasa de fallo y el tiempo, sino también la relación simultánea con las diferentes variables registradas para cada sujeto.

- Las variables Longitud, Diámetro, Presión, Tipo de material y Tráfico tienen un efecto significativo sobre el riesgo de fallo de las tuberías:
- El riesgo disminuye al aumentar el diámetro y al disminuir la longitud y la presión de trabajo.
- Existen diferencias significativas en el comportamiento de algunos materiales. Tomando como referencia la uralita, la fundición gris no se diferencia de ella, mientras que la fundición dúctil y el polietileno (muy especialmente éste último) tienen un comportamiento peor.
- Respecto al tráfico soportado por los tramos, y frente a la referencia que corresponde a “tráfico pesado”, tanto el “tráfico normal” como “bajo acera” significan reducciones significativas del riesgo.
- En síntesis, la incorporación de estas variables en el modelo mejora significativamente el pronóstico de la fiabilidad y del riesgo en los conductos de la red.
- Es importante anotar, que se han probado e introducido distintos tipos de interacciones, en particular del material con el resto de covariables y dichas interacciones han resultado nada significativas.
- De los 25 años en adelante, la consideración de los fallos provocados (además de los naturales) genera una aceleración en el aumento del riesgo de fallo. Por ejemplo, el incremento en la función de riesgo acumulada, que a los 38 años es del 0.952, en sólo 7 años, a los 45 de edad de las tuberías, alcanza el valor de 1.862, casi el doble.

8.4 Conclusiones sobre los modelos paramétricos

- De los once modelos paramétricos considerados en el proceso de simulación de la edad hasta el fallo de un tramo de tubería, el de mejor ajuste a los datos disponibles (de acuerdo a los indicadores de ajuste de Anderson-Darling y de coeficiente de correlación de Pearson) es la distribución Lognormal de 3 parámetros (ver tabla 6.15), seguido muy de cerca por una Normal y por una Loglogística también de 3 parámetros. Los parámetros del modelo Lognormal seleccionado, fueron estimados con los métodos de Máxima Verosimilitud y de Mínimos Cuadrados y los resultados son similares lo que refuerza su elección
- Una vez estimado el modelo paramétrico nos permite estimar la probabilidad de fallo a cualquier edad, lo cual es una ventaja frente a los no paramétricos y semiparamétricos que sólo estiman las probabilidades para las edades de los datos existentes.
- A partir de estas probabilidades estimadas con este modelo lognormal, la edad media (MTTF) hasta el fallo por envejecimiento y desgaste natural de un tramo de tubería, se estima en 67,94 años, y se prevé que la mitad de los conductos fallará antes de los 67,77 años. También se prevé que a los 72,5 años sólo sobrevivirán a las roturas/averías (fallos) propias del deterioro y envejecimiento de la red, el 40% de sus conductos.

8.5 Comparación de los modelos.

- Durante los primeros 35 años de operación de la red, la fiabilidad de las instalaciones con el método de Cox, es muy similar a la fiabilidad obtenida con el método de Kaplan- Meier y a la obtenida paramétricamente con el modelo Lognormal de 3 parámetros. Incluso podríamos afirmar que en general los resultados de las estimaciones de la fiabilidad y de la rapidez/velocidad con se presentan los fallos y las averías en los conductos de la red, es muy similar con los tres métodos. Esta consistencia refuerza la confianza en los análisis realizados.
- De los métodos considerados en este estudio, el método de Riesgo Proporcional de Cox presenta las ventajas de su robustez con pequeñas muestras, simplicidad de manejo y de permitir la identificación de los factores de riesgo.
- La consideración del truncamiento por la izquierda (además de la censura por la derecha) en los cálculos del estimador de Nelson Aalen extendido, mejora de forma importante (sobre todo en precisión con datos censurados), las sobreestimaciones producidas con Kaplan - Meier y con las tablas de vida, corrigiendo en gran medida el sesgo ocasionado por dichos métodos, por lo cual, se concluye que, el método de Nelson Aalen extendido, es el de mejores resultados en cuanto a las estimaciones de la fiabilidad de los conductos de la red, dentro de los no-paramétricos considerados en este estudio.
-

8.6 Recomendaciones para una base de datos de supervivencia en redes de suministro de agua potable.

- Desafortunadamente, como ya se observó, la mayor dificultad de desarrollo y aplicación de modelos y técnicas estadísticas para valorar el presente y el futuro de las redes de suministro de agua potable, es la falta de datos en cantidad y calidad. Se echa en falta la disponibilidad de historiales de roturas de más años y el registro de variables que consideren la recurrencia y la reparabilidad de las tuberías (tales como “nº de rupturas/fallos previa(o)s por tramo, el registro de “causas de fallos provocados”, “promedio anual de fallos”...etc.).
- Se sugiere un mayor esfuerzo en el diseño de bases de datos que tengan en cuenta estos factores, con mayor disponibilidad de personal cualificado, de recursos, de nuevas tecnologías, para evitar datos confusos, pérdida de información, información no sistematizada (archivos de papel, hojas sueltas, etc). En consonancia con lo anterior, se propone también avanzar en la búsqueda de un modelo económico que valore la optimización de recursos como herramienta básica en la gestión en las redes de suministro de agua potable.
- A corto y mediano plazo, se sugieren también mayores esfuerzos para que los modelos de predicción (y sus salidas) hagan parte integral de los sistemas de información (bases de datos) sobre fallos y averías en la industria del agua, como un importante paso hacia la gestión proactiva de la rehabilitación y sustitución de las redes y sus conductos.
En particular, las salidas o indicadores de los modelos registrados en las *Tablas resumen 4.6.2* (Tablas 4.2 y 4.3) se pueden integrar a futuras bases de datos, de modo que, puedan ser utilizadas como indicadores en la estimación de futuras necesidades de presupuesto y de evaluación de estrategias para la planificación de la rehabilitación de la red en general.
- En los últimos años, una amplia gama de nuevos materiales de tubería se encuentran disponibles en el mercado, razón por la cual, el sistema actual

de codificación de los existentes (en las bases de datos), se debe ampliar, para incluir la descripción de las nuevas tecnologías, y poder así eliminar los múltiples problemas que generan estos pobres sistemas de codificación a los gestores y administradores de estos sistemas.

Por último se incluyen las nuevas líneas de investigación que este trabajo sugiere.

8.7 Nuevas líneas de investigación

El desarrollo de este estudio, permitió identificar las siguientes sugerencias, sobre posibles trabajos y líneas de investigación para el futuro:

- Adecuación de las bases de datos y los modelos de análisis a la condición de “Equipos reparables” que tomen este tipo de instalación
- Evaluación del comportamiento de los diferentes modelos con distintos niveles de censura, con especial atención a las situaciones de alta censura, por ser los más desfavorables.
- De cara a la utilización de estos análisis por parte de las empresas prestadoras del servicio, sería interesante el desarrollo de un software que incorporara a la gestión de los datos la evaluación de riesgos.
- La principal crítica a los modelos de predicción (no paramétricos) utilizados habitualmente en condiciones de alta censura, es la sistemática tendencia a la *sobrestimación*, de la distribución del tiempo de vida de una unidad y la *falta de indicadores de sensibilidad e incertidumbre*. Pedrosa (2006), entre otros, argumenta que las medidas de incertidumbre en la mencionada predicción, son necesarias y puntualiza que deben llevarse a cabo mediante intervalos de confianza, expresando dicha incertidumbre en forma probabilística. Siguiendo esta sugerencia en el campo de la fiabilidad, se propone como una posible vía de investigación en el campo de la fiabilidad, la utilización de técnicas de *bootstrap*, para la cuantificación (por intervalos de confianza) de la mencionada incertidumbre.
- Ya que el método de KM, sólo es aplicable cuando los sucesos a observar (fallo, muerte, etc.) son dicotómicos, es decir fallo SI/NO y no es adecuado en situaciones donde se presentan fallos que no lo son y por el contrario, tienen diferentes causas (natural, provocado, etc.), se propone reenfocar este estudio con el modelo de *riesgos competitivos*, que plantea, que existe un estado inicial en un individuo (en operación o

vivo) y k posibles estados finales, cada uno con su correspondiente función de riesgo $h_k(t)$, las que deben ser estimadas mediante esta metodología; se dice que esta es una situación de *riesgos competitivos* o de *múltiples estados*.

- Puesto que la aplicación de K-M con ponderaciones (KM_p), (Bahrawar y otros (2005)), eliminó (en nuestro caso) toda la censura de los datos originales y convirtió la información original en datos completos (al tratar de corregir el sesgo de las sobreestimaciones que se generaban por KM) y esto no era lo que se quería, (pues estaríamos desperdiciando valiosa información de los datos censurados), se sugiere como posible tema de investigación, avanzar en la redefinición y búsqueda de un factor alternativo de ponderación w'_i con la siguiente estructura

$$w'_i = \frac{n_i - \alpha * c_i}{n_i} \quad \text{Con } 0 \leq \alpha \leq 1$$

De modo que, para cada nivel de censura de los datos originales, se pueda escoger el α óptimo que garantice las mejores estimaciones de la supervivencia, esto es, sin sobredimensionar los estimados de fiabilidad de los conductos de la red, con el menor margen de sesgo posible y con una adecuada tendencia de la cola de las estimaciones hacia el valor 0. La solución a este problema es motivo de un cuidadoso análisis, por la cantidad y calidad de simulaciones que conlleva su solución.

9. Referencias

- Aalen, O. (1976), Nonparametric inference in connection with multiple decrement models, *Scand. J. Statistics*, vol. 3, pg. 15–27.
- Aalen, O. (1978) “Non-parametric estimation of partial transition probabilities in multiple decrement models”, *Ann. Statist.* vol. 6 pp. 534–545.
- Aalen, O. (1978). Non-parametric inference for a family of counting processes. *Ann. Statist*, 6, 701-726.
- Agterberg, F. P. (1974). *Geomathematics*. (Journal Elsevier)
- Altman, N. y Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plan. Inf.* 46, 195-214.
- Altschuler, B. (1970). Theory for the measurement of competing risks in animal experiments. *Math. Biosci.*, 6, 1-11
- Andersen, P. (1982). Testing Goodness of Fit of Cox-Regression and Life Models. *Biometrics*, 32, 67-77
- Andersen, P. K. y Gill, R. D. (1982). "Cox's regression model for counting processes: a large sample study ", *Ann. Statist.*, 10, 1100-1120.
- Andersen, P.K., Borgan, O., Gill, R.D. y Keiding, N. (1993). *Statistical Models Based on Counting Processes* . Springer-Verlag, New York
- Andreou, S. (1987). Maintenance Decisions For Deteriorating Water Pipelines. *Journal of Pipelines*, 7, pp. 21-31.
- Andreou, S. A., Marks, D. H., and Clark, R. M. (1987a). “A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory.” *Advance in Water Resources*, 10, 2-10.
- Andreou, S. A., Marks, D. H., and Clark, R. M. (1987b). “A new methodology for modelling break failure patterns in deteriorating water distribution systems: Applications.” *Advance in Water Resources*, 10, 11-20.
- Ansell, J.I. and Phillips, M.J. (1994). *Practical Methods for Reliability Data Analysis*. Claredon Press- Oxford.
- Arjas, E. (1988) A Graphical Method for Assesing Goodness of Fit in Cox's Proportional Hazards Models. *JASA*, 83, 204-212
- Bahrawar J., S.Waqar A., Shuhrat S., M.Fazli Q. (2005) Weighted Kaplan Meier Estimation of Survival Function in Heavy Censoring
- Bain, L. J., y Engelhardt, M. (1980). "Probability of correct selection of Weibull v ersus gamma based on likelihood ratio", *Comm. Statist.-Theory Meth.*, 9, 375-381.

- Bartholomew, D. J. (1957) "A problem in life testing", J. Amer. Statist. Assoc., 52, 350- 355
- Bartholomew, D. J. (1963). "The sampling distribution of an estimate arising in life testing", Technometrics, 5, 361-374
- Baur, R. and Herz, R. (2002). Selective inspection planning with aging forecast for sewer types. Water Sci. Technol., 46.
- Beran, R. (1981). Nonparametric regression with randomly censored data. Technical report. Univ. California, Berkeley.
- Berrettoni, J.N. (1964). Practical applications of the Weibull distribution. Ind. Qual. Control, 21, 71-79
- Berslow, N.E. (1991). Introduction to Kaplan and Meier (1958) Nonparametric estimation from incomplete observations. En Breakthroughs in Statistics II, Kotz Sand Johnson N.L. (eds), pp: 311-318. New York: Springer.
- Billingsley, P. (1986), Probability and Measure (Second Edition), New York: John Wiley & Sons.
- Bowman, A., Hall, P. y Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions Biometrika. 85, 799-808.
- Brémond, B., (1997). "Statistical modelling as help in network renewal decision." European Commission Co-operation on Science and Technology (COST), Committee C3 – Diagnostics of Urban Infrastructure, Paris, France.
- Brémond, B. et Berthin, S. 2001. Reliability of a drinking water supply system. IWA specialized conference : system approach to leakage control and water distribution systems management, 167-174, Brno.
- Breslow, N. (1972), "Discussion on Professor Cox's paper", J. Roy. Statist. Soc. A, vol. 34, pg. 2160–2170.
- Breslow, N.E. (1974). Covariate Analysis of Censored Survival Data. Biometrics, v. 30, 89-99.
- Breslow, N. y Crowley J. (1974). "A large sample study of the life table and product limit estimates under random censorship", Ann. Statist., 2, 437- 453.
- Buckley, J. y James, I. (1979). Linear regression with censored data Biometrika. 66, 429-436.
- Camarinopoulos, L., Pampoukis, G. and Preston, N. (1996b). Reliability of a Water Supply Network. In: Proceedings of ESREL'96, Probabilistic Safety Assessment and Management.
- Chernoff, H., and Lieberman, G. J. (1954), Use of normal probability paper. Journal of the American Statistical Association, 49, 778-785.

- Clark, R. M., Stafford, C. L., and Goodrich, J. A. (1982). "Water distribution systems: A spatial and cost evaluation." *J. Water Resources Planning and Management Division, ASCE*, 108(3), 243-256.
- Clark, R.M., Eilers, R. G., and Goodrich, J. A (1988). Distribution system: Cost of repair and replacement. *Proc., Conf. On Pipeline Infrastructure*, B. A. Bennett, ed., ASCE, New York, 428-440
- Cohen, A. C. (1961). "Tables for maximum likelihood estimation: singly truncated and singly censored samples", *Technometrics*, 3, 535-541.
- Cohen, A. C. (1965). "Maximum likelihood estimation in the Weibull distribution based on complete and censored samples", *Technometrics*, 17, 579-588.
- Collett D (2003) *Modelling survival data in medical research* (Chapman & Hall/CRC)
- Constantine, A. G., and Darroch, J. N. (1993). "Pipeline reliability: stochastic models in engineering technology and management." S. Osaki, D.N.P. Murthy, eds., World Scientific Publishing Co.
- Constantine, A. G., Darroch, J. N., and Miller, R. (1996). "Predicting Underground Pipe Failure." Australian Water Works Association.
- Cox D.R. & Oakes D. (1984). *Analysis of survival data* (Chapman & Hall)
- Cox D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34: 187-220.
- Cox, D. R. (1975). Prediction intervals and empirical Bayes confidence, in *Perspectives in Probability and Statistics*, J. Gani. Editor. London: Academic Press.
- Cox, D. R. (1980). *Point processes In series: Monographs on applied probability and statistics*. Chapman and Hall, London.
- Cox, D.R. and Snell, E.J., (1968), A general definition of residuals (with discussion), *Journal of the Royal Statistical society B*, 30, 248-275
- Crowder, M. J., Kimber, A, C., Smith, R. L. And Sweeting, T. J. (1991), *Statistical analysis of Reliability Data*, New York: Chapman & Hall
- Dabrowska, D. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.* 17, 1157-1167.
- Davis, D. J. (1952). "An analysis of some failure data", *J. Amer. Statist. Assoc.*, 47, 113-150
- Deb, A.K. et. al. (1998). Quantifying future rehabilitation and replacement needs of water mains. *AWWA Res. Fdn.*, Denver.
- Dixon, W. J (1960). "Simplified estimation from censored normal samples", *Ann. Math. Statist.*, 31, 385-391.

- Efron, B. (1967). "The two-sample problem with censored data", Proc. Fifth Berkeley Symp. Math. Statist. Probab., Univ. California Press, 4, 831-853.
- Eisenbeis, P., (1994). "Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable, PhD thesis, University Louis Pasteur of Strasbourg, collection Etudes Cemagref no17.
- Eisenbeis, P., Røstum, J. and Le Gat, Y. (1999). Statistical Models for Assessing the Technical State of Water Networks - Some European Experiences. In: Proceedings of annual conference of AWWA, Chicago, Illinois, 20 -24 June 1999.
- Elnaboulsi, J., and Alexandre, O. (1996). Le renouvellement des réseaux d'eau potable: Une revue critique. Can. Water Resour. J., 21(4), 341-354 (in French)
- Epstein, B. (1958). "The exponential distribution and its role in life-testing", Ind. Qual. Control , 15, 2-7.
- Epstein, B. y Sobel, M. (1954). "Some theorems relevant to life testing from an exponential distribution", Ann. Math. Statist., 25, 373-381
- Epstein, B. y Tsao, C. K. (1953). "Some test based on ordered observations from two exponential populations", Ann. Math. Statist. 24 , 456-466.
- Feinleib, M. (1960). "A method of analyzing lognormally distributed Survival data with incomplete follow-up", J. Amer. Statist. Assoc. , 55, 534-545.
- Fleming, T. R. y Harrington, D. P. (1991). Counting Processes and Survival Analysis , John Wiley, New York.
- Fleming, T.R. y Harrington D.P. (1984), Nonparametric estimation of the survival distribution in censored data, Comm. Statist.–Theor. Meth., vol. 20, 2469–2486.
- Gehan, E. A. (1965a). "A generalized two-sample Wilcoxon test for comparing arbitrarily singly censored samples", Biometrika, 52, 203-223.
- Gehan, E. A. (1965b). A generalized two-sample Wilcoxon test for doubly censored data, Biometrika, 52, 650-653.
- González-Manteiga, W. y Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. J. Nonparametric Statist. 4, 65-78.
- Goulter, I. and Kanzemi, A. (1988). Spatial and Temporal Groupings of Water Mains Pipe Breakage in Winnipeg. Canadian Journal of Civil Engineering, 5, pp. 91-97.
- Greenwood, M. (1926). The natural duration of cancer. Reports on Public Health and Medical Subjects. London: Her Majesty's Stationery Office; 33. 1-26.

- Gross, A. J. y Clark, V. A. (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*, John Wiley, New York.
- Gu, M. G. y Zhang C. H. (1993). "Asymptotic properties of self-consistent estimators based on doubly censored data", *Ann. Statist.*, 21, 611-624.
- Gupta, S. S. y Groll, P. A. (1961). "Gamma distribution in acceptance sampling based on life tests", *J. Amer. Statist. Assoc* 65, 942-970.
- Gustafson, J-M., and Clancy, D. V. (1999a). "Modelling the occurrence of breaks in cast iron water mains using methods of survival analysis" *Proc. AWWA Annual Conference, Chicago*
- Hardle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.
- Harter, H. L. y Moore, A. H. (1965). "Maximum likelihood estimation of the parameters of the gamma and Weibull populations from complete and from censored samples", *Technometrics*, 7, 639-643.
- Harter, H.L. (1967). "Maximum-likelihood estimation of the parameters of a four-parameter generalized gamma population for complete and censored samples", *Technometrics*, 9, 159-165.
- Herz R. K. (1998). Exploring rehabilitation needs and strategies for water distribution networks. *J. Water SRT-Aqua*, 47, 1-9.
- Herz, R. K. (1996). Ageing processes and rehabilitation needs of drinking water distribution networks. *Journal of Water Supply Research and Technology-Aqua*, 45, pp. 221-231.
- Herz, R. K. (1998). "Exploring rehabilitation needs and strategies for water distribution networks." *J. Water SRT – Aqua*, 47(6), 275-283.
- Hill, B. M. (1963). "The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic", *J. Amer. Statist. Assoc.*, 58, 72-84.
- Høyland, A. and Rausand, M. (1994). *System Reliability Theory: Models and Statistical Methods*. John Wiley & Sons, Inc, New York.
- Iglesias Pérez, M.C (2003) Estimation of the conditional distribution function with truncated and censored data: an application to the study of the mortality of diabetics. *Spanish Statistics*. 45, 275-301
- Iglesias Pérez, M.C. y González Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *J. Nonparametric Statist.* 10, 213-244.
- Jacobs, P., and Karney, B. (1994). "GIS development with application to cast iron water main breakage rate." 2nd Int. Conf. on Water Pipeline Systems, BHR Group Ltd., Edinburgh, Scotland.

- Johansen, S. (1978). "The product limit estimate as a maximum likelihood estimate", *Scand. J. Statist.*, 5, 195-199.
- Johnson, N.L. y Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions*, Vol. 1. Boston: Ed. Houghton Mifflin.
- Kalbfleisch, J. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York.
- Kao, J. H. K (1959). "A graphical estimation of mixed Weibull parameters in life testing of electron tubes", *Technometrics* 1, 389-407
- Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53: 457-481
- Karaa, F.A., and Marks, D. H. (1990). Performance of water distribution networks: Integrated approach. *J. Perform. Constr. Facil.*, 4(1), 51-67
- Kellerer, A.M. y Chemelevsky, D. (1983). Small-sample Properties of Censored-Data Rank Tests. *Biometrics*, 39, 675-682.
- Ketler, F. A., and Goulter, I:C (1985). "An analysis of pipe breakage in urban water distribution networks." *Can. J. Civ Eng.*, 12, 286-293.
- Kim, J.H. (1992). Optimal rehabilitation/replacement for water distribution systems. PhD thesis, University of Texas, Austin.
- Kimball, A. W. (1960). Estimation of mortality intensities in animal experiments. *Biometrics*, 16, 505-521
- Kirmeyer, G. J., Richards, W., and Smith, C. D. (1994). "An assessment of water distribution systems and associated research needs." AWWA Research Foundation, Denver. Colo.
- Klein, J. and Moeschberger, M. (1997). *Survival analysis: Techniques for Censored and Truncated data*. Springer, New York.
- Kleiner, Y. and Rajani, B. (2000). *Comprehensive Review of Structural Deterioration of Water Mains: Statistical Models*.
- Kleiner, Y., Adams, B.J. and Rogers, J.S. (1998b). Selection and scheduling of rehabilitation alternatives for water distribution systems. *Water Resources Research*, 34, pp. 2053-2061.
- Koch, G., and Link, R. F. (1970). *Statistical Analysis of Geological Data* (J. Wiley)
- Koul, H.L., Susarla, V. y Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* 9, 1276-1288.
- Kumar, D. and Klefsjö, B. (1994). Proportional hazards model - a review. *Reliability Engineering & System Safety*, 44, pp.177-188.

- Latta, R.B. (1981). A Monte Carlo Study of Some Two-Sample Rank Tests with Censored data. *JASA*, 76, 713-719
- Lawless J.F.(1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons.
- Lawless, J.F. (1987). Regression methods for Poisson process data. *Journal of American Statistical Association*, 82, pp. 808-815.
- Le Gat Y, Eisenbeis P (2000). Using maintenance records to forecast failures in water networks. *Urban Water* 2, 173-181.
- Lee E.T. (1992). *Statistical Methods for Survival Data Analysis*. Wiley New York.
- Lei, J. (1997). "Statistical approach for describing lifetimes of water mains – Case Trondheim Municipality." SINTEF Civil and Environmental Engineering, Report No. 22F007.28, Trondheim, Norway.
- Lei, J. and Sægrov, S. (1998). Statistical approach for describing lifetimes of water mains. *Water Science and Technology*, 38, No. 6, 1998, pp. 209-217
- Leiderman, P .H., Babu, D., Kagia, J., Kraemer, H. C. y Leiderman, G. F. (1973). "African infant precocity and some social influences during the first year", *Nature*, 242, 247-249
- Lieblein, J. y Zelen, M. (1956). "Statistical investigation of the fatigue life of deep groove ball bearings", *J. Res. Nat. Bur. Stand.*, 57, 273-316
- Lynden-Bell, D. (1971). "A method for allowing for known observational selection in small samples applied to 3CR quasars," *Mon. Nat. Royal Astr. Soc.* vol. 155 pp. 95–118.
- Madiec, H., Botzung, P., Bremond, B. and Eisenbeis, P. (1997). Implementation of a probability model for renewal of drinking water networks. *Water Supply*, 14, (3/4), pp. 347-351.
- Mailhot, A., Duchesne, S., Musso, E., Villeneuve, J. P. (2000). Modélisation de l'évolution de l'état structural des réseaux d'égout : application à une municipalité du Québec. *Can. J. Civ. Eng.* 27(1): 65–72
- Malandain, J. (1999). Modélisation de l'état de santé des réseaux de distribution d'eau pour l'organisation de la maintenance. Etude du patrimoine de l'agglomération de Lyon. Thèse de Doctorat n° 99 ISAL 0040 de l'Institut National des Sciences Appliquées de Lyon, Laboratoire URGC / Hydrologie Urbaine, 206 p.
- Malandain, J., Le Gauffre, P. and Miramond, M. (1999). Modeling the aging of water infrastructure. In: *Proceedings of the 13th EJSW*, 8 September – 12 September, Dresden University of Technology. ISBN: 3-86005-238-1.

- Manjarrés, L. (2006). Modelización Estadística para la Predicción de Fallos en los Conductos de Agua Potable. Memoria para Obtención del DEA. UPV.
- Mantel, N. (1966). Evaluation of Survival Data and Two New Rank order Statistics Arising in its Consideration. *Cancer Chemotherapy*, 50, 163-170
- Mantel, N. (1967). "Ranking procedures for arbitrarily restricted observations", *Biometrics*, 23, 65-78.
- Marks, H. D., Andreou, S., Jeffrey L., Park, C., and Zaslavski, A. (1987). "Statistical models for water main failures" US Environmental Protection Agency (Co-operative Agreement CR810558) M.I.T Office of Sponsored Projects No. 94211. Boston, Mass.
- Marks, H. D., et al., (1985). "Predicting urban water distribution maintenance strategies: A case study of New Haven Connecticut." US Environmental Protection Agency (Co-operative Agreement R810558-01-0).
- Martínez A N. (1998). Análisis de Regresión con datos Censurados. Aplicación al Estudio de Factores Pronóstico en la Supervivencia al Cáncer de Mama. (Tesis doctoral)
- McMullen, L. D., (1982). "Advanced Concepts in Soil Evaluation for Exterior Pipeline Corrosion", Proceeding AWWA Annual Conference, Miami.
- Meeker W. Q. And Escobar L.A (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons.
- Miller Jr. Rupert G. (1981). *Survival Analysis* Wiley series in probability and mathematical statistics, N.Y.
- Miller, R. B. (1993). Personal communications. CSIRO Division of Mathematics and Statistics, Glen Osmond. Australia.
- Miller, R. G. (1981). *Survival Analysis*, John Wiley , New York.
- Miller, R. G. (1983). "What price Kaplan-Meier?", *Biometrics* 39, 1077-1081.
- Miller, R.G. (1976). Least Squares Regression with Censored Data. *Biometrika* 63, 521-532.
- Morales, D., Pardo, L. and QuesadaV. (1991). "Estimation of a survival function with doubly censored data and Dirichlet process prior knowledge on the observable variable" *Comm. Statist.-Simul.*, 19, 349-361.
- Morris, R.E. (1967). Principal Causes and Remedies of Water Main Breaks. *Journal AWWA*, pp. 782-798.
- Mosevoll, G. (1994). Vedlikehold og fornyelse av VA- ledninger: Modeller for tilstands-prognose / Funksjonskrav til informasjonssystemer. Dr.ing avhandling, Institutt for Vassbygging, Norges Tekniske Høgskole Universitetet i Trondheim.

- Mykland, P. A. y Ren, J.- J. (1996). "Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data", *Ann. Statist.*, 24, 1740-1764.
- Nelson W.B (1970). Graphical Analysis of Accelerated life test data with the inverse power law model. *IEEE Trans. Realiab.* R21, 2-11.
- Nelson, W. (1982). *Applie Life Data Analysis*, John Wiley New York.
- Nelson, W. B. (1969). "Hazard plotting for incomplete failure data," *J. Quality Technology* vol. 1 pp. 27–52,.
- Nelson,W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analyses*, John Wiley New York
- O. Piller ; B. Bremond ; P. Eisenbeis (2001). Modélisation des réseaux d'alimentation en eau potable avec une demande dépendant de la charge et une prise en compte des points hauts. *Orbx Cemagref Bordeaux, IWA specialised conference : system approach to leakage control and water distribution systems management*, 16-18 may 2001, Brno, CZE. 2001, 72-80, Communication à un congrès
- O`Day, D. K., Weiss, R., Chiavari., and Blair, D. (1986). Water main evaluation for rehabilitation/replacement. Rep. American Water Works Association Research Foundation, Denver, and U.S. Environmental Protection Agency Cincinnati.
- Oakes, D. (1986). "Semiparametric inference in a model for association in bivariate survival data", *Biometrika*, 73, 353-361.
- Pan W. and Chappell, R. (1998). "Estimating survival curves with left truncated and interval censored data via the EMS algorithm," To appear in *Comm. in Statist.–Theory and Methods*.
- Pan W. y Chapell R.(1998). A Nonparametric Estimator of Survival Functions for Arbitrarily Truncated and Censored Data. *Lifetime Data Analysis*, vol. 4, pg. 187–202. Kluwer Academic Publishers, Boston. 1998
- Peace, K.E. y Flora, R.E. (1978). Size and Power Assessment of Tests of Hypotheses on Survival parameters. *JASA*, 73, 129-133.
- Pelletier G., Mailhot A., Villeneuve J. (2003). *Modeling Water Pipe Breaks - Three case Studies*
- Peterson, A. V. (1977). "Expressing the Kaplan-Meier estimate as a function of empirical subsurvival funcions", *J. Amer. Statist. Assoc.*, 82 , 854-858.
- Peto R, Lee and Paige W.S (1972). Statistical Analysis of the Bioassay of Continuos carcinogens.*Br. J. Cancer*, 26, 258- 612.
- Peto. R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics*, 22, 86-91.

- Pike, M. C. (1966). A method of analysis of a certain class of experiments in carcinogenesis, *Biometrics*, 22, 142-161
- Poulton M (2001). Elements of an integrated approach for studying leakage and bursts in water distribution of pipelines. Doctoral thesis, Imperial College of Science, Technology and Medicine, London).
- Prentice, R., Williams, B. and Peterson A. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68, (2), pp. 373-379.
- Rajani, B. and S. McDonald "Water Mains Break Data on Different Pipe Materials for 1992 and 1993", Report No. A-7019.1, National Research Council of Canada, Ottawa, 1995.
- Rausand, M. and Reinertsen, R. (1996). Failure mechanisms and life models. *International Journal of Reliability, Quality and Safety Engineering*, 3, No. 3, pp. 137-152.
- Rostum J. (2000). Statistical Modelling of Pipe Failures in water Networks. (PhD thesis)
- Rostum J., Schilling W. predictive service-life models user for water network management. 14th EJSW, Dresden, September 8-12, 1999
- Sægrov, S., König, A., et al. (2001). "Evaluation Of UtilNets-a decision support systems for water mains rehabilitation". 16-5-2001. IWA International Conference in Brno, Czech Republic. Ref Type: Conference Proceeding.
- Saegrov S, et al. (2003). Computer aided rehabilitation of water networks. Pumps, electromechanical devices and systems applied to urban water management, I: 279-287, Balkema
- Sarabia, J.M., et al. (2006). "Estadística Actuarial: teoría y aplicaciones". Pearson Prentice Hall.
- Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plan. Inf.* 35, 65-75.
- Shamir, U. and Howard, D.D. (1979). An Analytic Approach to Scheduling Pipe Replacement, *Journal AWWA*, May, 71, pp. 248-258.
- Stacy, E. W. y Mihram, G. A. (1965). "Parameter estimation for a generalized gamma distribution", *Technometrics*, 7, 349-358
- Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivar. Analysis* 45, 89-103.
- Susan, M. (2001). Using weighted Kaplan Meier statistics in non-parametric comparisons of paired censored survival outcomes. *Biometrics* 57: 361-368.
- Tang, M.-X., Tsai, W.-Y., Marder, K. and Mayeaux, R. (1995). "Linear rank test for doubly censored data", *Statistics in Medicine* 14, 2555-2563.

- Tarone, R. y Ware, J. (1977). On Distribution- Free Tests for equality of Survival Distributions. *Biometrika*, 64, 156-160.
- Therneau, T.M. y Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. N.Y. Springer-Verlag.
- Trujillo, R.A. (1995). *Bedarfsprognose und Strategieentwicklung für die Rehabilitation stätischer Wasserrohrnetze*. Institut für Städtebau und Landesplanung, Universität Karlsruhe.
- Tsai, W.Y., Jewell, N.P. y Wang, M.C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* 74, 883-886.
- Tsiatis, A. A. (1978a) A large sample study of the estimate for the integrated hazard function in Cox's regression model. for survival data. University of Wisconsin, Department of Statistics, Technical Report N° 526.
- Turnbull, B. W. (1974). "Nonparametric estimation of a survivorship function with doubly censored data", *J. Amer. Statist. Assoc.*, 69, 169-173.
- Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc. B*, 38, 290-295.
- Van Keilegom, I. y Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. math.* 49, 467-491.
- Wagner, J.M., Shamir, U. and Marks, D.H. (1988a). Water Distribution Reliability: Analytical Methods. *Journal of Water Resources Planning and Management*, May, 114, pp. 253-275.
- Walski, T. M., and Pelliccia, A. (1982). "Economic analysis of water main breaks." *J. AWWA*, 74(3), 140-147.
- Wand, M.P. y Jones, M.C. (1995). *Kernel smoothing*. Chapman and Hall, London.
- Wang, Z. y Gardiner, J. C. (1996). "A class of estimators of the survival function from interval-censored data", *Ann. Statist.*, 24, 647-658.
- Weibull W (1951). A Statistical Distribution Function of Wide Applicability. *Journal of Applied Mechanics*, 18, 293- 297.
- Wellner, J. A. (1982). "Asymptotic optimality of the product limit estimator", *Ann. Statist.*, 10, 595-602
- Wilk, M. B., Gnanadesikan, R. y Huyett, M. J (1962b). "Estimation of parameters of the gamma distribution using order statistics", *Biometrika*, 49, 525-545.
- Wilk, M. B., Gnanadesikan, R. y Huyett, M. J. (1962a) . "Probability plots for the gamma distribution", *Technometrics* 4 , 1-20

Woodroffe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* 13, 163-177.

Anexos

Estimación de la fiabilidad por tipo de material por el método de NAE

a. Para la fundición dúctil (FD)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
0.5	19	46	5859	0,003243	0,003243	0,996762
1	3	474	7211	0,000416	0,003659	0,996348
2	3	1172	7626	0,000393	0,004052	0,995956
3	1	1105	7199	0,000139	0,004191	0,995818
4	2	1130	6962	0,000287	0,004478	0,995532
5	2	1154	6955	0,000288	0,004766	0,995245
6	3	761	6543	0,000459	0,005225	0,994789
7	3	1412	6301	0,000476	0,005701	0,994316
8	3	890	5668	0,000529	0,006230	0,993789
9	2	748	5592	0,000358	0,006588	0,993434
10	2	867	5690	0,000351	0,006939	0,993085
11	4	1123	5305	0,000754	0,007693	0,992336
12	3	742	4584	0,000654	0,008348	0,991687
13	0	524	4166	0,000000	0,008348	0,991687
14	1	779	3943	0,000254	0,008601	0,991436
15	0	811	3410	0,000000	0,008601	0,991436
16	0	852	3063	0,000000	0,008601	0,991436
17	2	482	2548	0,000785	0,009386	0,990658
18	2	406	2316	0,000864	0,010250	0,989803
19	3	322	2118	0,001416	0,011666	0,988402
20	3	302	2008	0,001494	0,013160	0,986926
21	4	243	1985	0,002015	0,015175	0,984939
22	3	460	1924	0,001559	0,016734	0,983405
23	3	332	1678	0,001788	0,018522	0,981648
24	1	254	1534	0,000652	0,019174	0,981008
25	5	205	1433	0,003489	0,022663	0,977591
26	2	218	1333	0,001500	0,024164	0,976126
27	4	276	1252	0,003195	0,027359	0,973012
28	3	178	1062	0,002825	0,030184	0,970267
29	4	212	971	0,004119	0,034303	0,966279
30	1	185	790	0,001266	0,035569	0,965056
31	2	150	662	0,003021	0,038590	0,962145
32	3	109	534	0,005618	0,044208	0,956755
33	0	133	439	0,000000	0,044208	0,956755
34	2	90	325	0,006154	0,050362	0,950885

35	1	87	251	0,003984	0,054346	0,947105
36	0	33	168	0,000000	0,054346	0,947105
37	0	55	140	0,000000	0,054346	0,947105
38	0	21	89	0,000000	0,054346	0,947105
39	0	17	79	0,000000	0,054346	0,947105
40	0	19	70	0,000000	0,054346	0,947105
41	0	18	58	0,000000	0,054346	0,947105
42	0	5	48	0,000000	0,054346	0,947105
43	0	6	44	0,000000	0,054346	0,947105
44	0	3	41	0,000000	0,054346	0,947105
45	0	12	40	0,000000	0,054346	0,947105
46	1	8	40	0,025000	0,079346	0,923720
47	1	7	36	0,027778	0,107124	0,898415
48	0	7	37	0,000000	0,107124	0,898415
49	1	2	30	0,033333	0,140457	0,868961
50	2	3	31	0,064516	0,204973	0,814669
51	0	2	27	0,000000	0,204973	0,814669
52	0	8	27	0,000000	0,204973	0,814669
53	0	3	19	0,000000	0,204973	0,814669
54	0	9	16	0,000000	0,204973	0,814669
56	0	4	7	0,000000	0,204973	0,814669
57	0	1	3	0,000000	0,204973	0,814669
58	0	2	2	0,000000	0,204973	0,814669

b. Para la fundición gris (FG)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
1	0	2	6	0,000000	0,000000	1,000000
2	0	4	4	0,000000	0,000000	1,000000
10	1	0	23	0,043478	0,043478	0,957453
11	0	2	24	0,000000	0,043478	0,957453
12	2	2	22	0,090909	0,134387	0,874251
14	1	5	32	0,031250	0,165637	0,847353
15	1	4	26	0,038462	0,204099	0,815382
16	0	7	24	0,000000	0,204099	0,815382
17	0	2	20	0,000000	0,204099	0,815382
19	0	2	24	0,000000	0,204099	0,815382
20	0	10	26	0,000000	0,204099	0,815382
22	0	3	23	0,000000	0,204099	0,815382
23	0	3	20	0,000000	0,204099	0,815382
24	0	5	21	0,000000	0,204099	0,815382
25	2	1	37	0,054054	0,258153	0,772477
26	0	4	39	0,000000	0,258153	0,772477
27	0	5	43	0,000000	0,258153	0,772477
28	0	2	60	0,000000	0,258153	0,772477
30	1	4	72	0,013889	0,272042	0,761822
31	0	20	103	0,000000	0,272042	0,761822
32	1	6	110	0,009091	0,281133	0,754928
33	1	8	110	0,009091	0,290224	0,748096
34	2	21	131	0,015267	0,305491	0,736762
35	1	8	134	0,007463	0,312954	0,731284
36	0	6	142	0,000000	0,312954	0,731284
37	2	35	142	0,014085	0,327038	0,721056
38	2	23	125	0,016000	0,343038	0,709611
39	5	8	110	0,045455	0,388493	0,678078
40	0	24	131	0,000000	0,388493	0,678078
41	4	25	165	0,024242	0,412735	0,661838
42	1	19	158	0,006329	0,419064	0,657662
43	8	3	142	0,056338	0,475402	0,621635
44	1	20	147	0,006803	0,482205	0,617421
45	1	8	153	0,006536	0,488741	0,613398
46	2	29	163	0,012270	0,501011	0,605918
47	5	57	134	0,037313	0,538324	0,583726
48	1	13	81	0,012346	0,550670	0,576563

49	1	3	78	0,012821	0,563490	0,569219
50	5	15	87	0,057471	0,620962	0,537427
51	0	20	70	0,000000	0,620962	0,537427
52	0	15	63	0,000000	0,620962	0,537427
53	0	2	48	0,000000	0,620962	0,537427
54	1	9	46	0,021739	0,642701	0,525870
55	0	10	36	0,000000	0,642701	0,525870
56	0	10	26	0,000000	0,642701	0,525870
57	0	3	16	0,000000	0,642701	0,525870
58	0	13	13	0,000000	0,642701	0,525870

c. Para el polietileno (PE)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
1	0	16	101	0,000000	0,000000	1,000000
2	1	16	98	0,010204	0,010204	0,989848
3	0	17	82	0,000000	0,010204	0,989848
4	0	8	77	0,000000	0,010204	0,989848
5	1	25	74	0,013514	0,023718	0,976561
6	1	11	90	0,011111	0,034829	0,965771
7	0	6	84	0,000000	0,034829	0,965771
8	0	12	83	0,000000	0,034829	0,965771
9	0	1	77	0,000000	0,034829	0,965771
10	0	12	92	0,000000	0,034829	0,965771
11	1	5	84	0,011905	0,046733	0,954342
12	0	41	80	0,000000	0,046733	0,954342
13	0	6	45	0,000000	0,046733	0,954342
14	0	5	45	0,000000	0,046733	0,954342
15	0	6	43	0,000000	0,046733	0,954342
16	0	16	41	0,000000	0,046733	0,954342
17	0	4	27	0,000000	0,046733	0,954342
18	0	2	29	0,000000	0,046733	0,954342
19	1	6	37	0,027027	0,073760	0,928894
20	0	6	39	0,000000	0,073760	0,928894
21	1	3	61	0,016393	0,090154	0,913791
22	4	4	72	0,055556	0,145709	0,864409
23	2	3	66	0,030303	0,176013	0,838607
24	0	7	63	0,000000	0,176013	0,838607
25	4	8	77	0,051948	0,227961	0,796156
26	1	9	91	0,010989	0,238950	0,787455
27	3	19	89	0,033708	0,272657	0,761354
28	3	9	76	0,039474	0,312131	0,731886
29	1	1	64	0,015625	0,327756	0,720539
30	4	2	62	0,064516	0,392272	0,675520
31	1	19	56	0,017857	0,410129	0,663564
32	1	25	36	0,027778	0,437907	0,645386
33	2	8	11	0,181818	0,619725	0,538092
39	0	1	2	0,000000	0,619725	0,538092
41	1	2	5	0,200000	0,819725	0,440553
44	0	1	5	0,000000	0,819725	0,440553
47	0	1	4	0,000000	0,819725	0,440553
50	0	3	3	0,000000	0,819725	0,440553

d. Para la Uralita (UR)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
2	1	0	88	0,011364	0,011364	0,988701
3	0	3	91	0,000000	0,011364	0,988701
4	0	8	92	0,000000	0,011364	0,988701
5	0	65	92	0,000000	0,011364	0,988701
7	0	4	43	0,000000	0,011364	0,988701
8	0	7	83	0,000000	0,011364	0,988701
9	0	4	164	0,000000	0,011364	0,988701
10	0	4	191	0,000000	0,011364	0,988701
11	0	8	215	0,000000	0,011364	0,988701
12	0	13	220	0,000000	0,011364	0,988701
13	3	3	225	0,013333	0,024697	0,975606
14	4	47	242	0,016529	0,041226	0,959612
15	1	79	216	0,004630	0,045856	0,955180
16	0	32	187	0,000000	0,045856	0,955180
17	0	29	218	0,000000	0,045856	0,955180
18	2	15	216	0,009259	0,055115	0,946377
19	2	18	231	0,008658	0,063773	0,938218
20	1	23	257	0,003891	0,067664	0,934575
21	1	25	408	0,002451	0,070115	0,932287
22	0	53	558	0,000000	0,070115	0,932287
23	4	67	769	0,005202	0,075316	0,927450
24	3	28	922	0,003254	0,078570	0,924437
25	9	38	1154	0,007799	0,086369	0,917256
26	8	57	1316	0,006079	0,092448	0,911696
27	11	176	1492	0,007373	0,099821	0,905000
28	9	181	1557	0,005780	0,105601	0,899783
29	17	249	1646	0,010328	0,115929	0,890538
30	12	224	1670	0,007186	0,123115	0,884162
31	16	252	1767	0,009055	0,132170	0,876192
32	10	213	1812	0,005519	0,137689	0,871370
33	10	250	1957	0,005110	0,142798	0,866929
34	14	249	2093	0,006689	0,149487	0,861149
35	16	266	2082	0,007685	0,157172	0,854557
36	20	271	2010	0,009950	0,167123	0,846096
37	12	312	1968	0,006098	0,173220	0,840953
38	28	291	1784	0,015695	0,188915	0,827857
39	10	327	1576	0,006345	0,195260	0,822620

40	16	363	1392	0,011494	0,206755	0,813219
41	24	237	1166	0,020583	0,227338	0,796652
42	10	188	990	0,010101	0,237439	0,788645
43	17	223	824	0,020631	0,258070	0,772541
44	5	122	664	0,007530	0,265600	0,766746
45	20	101	572	0,034965	0,300565	0,740400
46	11	129	473	0,023256	0,323821	0,723380
47	5	117	343	0,014577	0,338398	0,712911
48	0	60	232	0,000000	0,338398	0,712911
49	2	25	194	0,010309	0,348707	0,705600
50	5	74	193	0,025907	0,374614	0,687555
51	2	25	124	0,016129	0,390743	0,676554
52	3	21	103	0,029126	0,419869	0,657133
53	1	9	79	0,012658	0,432528	0,648867
54	0	11	69	0,000000	0,432528	0,648867
55	5	23	58	0,086207	0,518734	0,595273
56	0	21	30	0,000000	0,518734	0,595273
57	0	4	9	0,000000	0,518734	0,595273
58	0	5	5	0,000000	0,518734	0,595273

Estimación de la fiabilidad según el Tráfico por el método de NAE

a. Tráfico Acera (A)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
0.5	11	15	4934	0,002229	0,002229	0,997773
1	2	422	6023	0,000332	0,002561	0,997442
2	3	976	6314	0,000475	0,003037	0,996968
3	0	956	5923	0,000000	0,003037	0,996968
4	2	928	5670	0,000353	0,003389	0,996616
5	3	1018	5597	0,000536	0,003925	0,996082
6	2	609	5204	0,000384	0,004310	0,995700
7	1	1110	5037	0,000199	0,004508	0,995502
8	2	713	4563	0,000438	0,004947	0,995066
9	2	587	4520	0,000442	0,005389	0,994625
10	1	702	4643	0,000215	0,005604	0,994411
11	3	855	4341	0,000691	0,006295	0,993724
12	1	627	3826	0,000261	0,006557	0,993465
13	0	446	3468	0,000000	0,006557	0,993465
14	3	635	3294	0,000911	0,007468	0,992560
15	1	668	2866	0,000349	0,007816	0,992214
16	0	713	2636	0,000000	0,007816	0,992214
17	0	402	2235	0,000000	0,007816	0,992214
18	1	346	2047	0,000489	0,008305	0,991729
19	4	265	1913	0,002091	0,010396	0,989658
20	3	272	1868	0,001606	0,012002	0,988070
21	4	208	1968	0,002033	0,014034	0,986064
22	2	435	2059	0,000971	0,015006	0,985106
23	6	314	2042	0,002938	0,017944	0,982216
24	3	221	2004	0,001497	0,019441	0,980747
25	15	210	2141	0,007006	0,026447	0,973899
26	7	231	2184	0,003205	0,029652	0,970783
27	12	367	2240	0,005357	0,035009	0,965596
28	7	300	2146	0,003262	0,038271	0,962452
29	9	399	2109	0,004267	0,042539	0,958353
30	7	276	1952	0,003586	0,046125	0,954923
31	12	353	2007	0,005979	0,052104	0,949230
32	4	271	1926	0,002077	0,054181	0,947261
33	7	291	1944	0,003601	0,057782	0,943856
34	9	280	1979	0,004548	0,062329	0,939573
35	9	257	1906	0,004722	0,067051	0,935147
36	17	242	1819	0,009346	0,076397	0,926448
37	6	327	1748	0,003432	0,079830	0,923274

38	11	273	1531	0,007185	0,087014	0,916664
39	8	272	1347	0,005939	0,092954	0,911236
40	8	312	1212	0,006601	0,099554	0,905241
41	14	213	1041	0,013449	0,113003	0,893148
42	7	163	881	0,007946	0,120948	0,886080
43	14	176	740	0,018919	0,139867	0,869474
44	1	106	627	0,001595	0,141462	0,868088
45	14	89	561	0,024955	0,166418	0,846693
46	5	132	487	0,010267	0,176685	0,838044
47	4	131	364	0,010989	0,187674	0,828885
48	0	54	237	0,000000	0,187674	0,828885
49	3	23	212	0,014151	0,201824	0,817238
50	4	73	214	0,018692	0,220516	0,802105
51	1	35	141	0,007092	0,227608	0,796436
52	2	23	119	0,016807	0,244415	0,783163
53	0	13	94	0,000000	0,244415	0,783163
54	0	8	81	0,000000	0,244415	0,783163
55	2	29	73	0,027397	0,271812	0,761997
56	0	25	42	0,000000	0,271812	0,761997
57	0	4	17	0,000000	0,271812	0,761997
58	0	13	13	0,000000	0,271812	0,761997

b. Tráfico Normal (N)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de fiabilidad
0.5	5	7	969	0,005160	0,005160	0,994853
1	1	62	1248	0,000801	0,005961	0,994056
2	1	195	1356	0,000737	0,006699	0,993324
3	1	151	1312	0,000762	0,007461	0,992567
4	0	212	1328	0,000000	0,007461	0,992567
5	0	201	1383	0,000000	0,007461	0,992567
6	0	137	1341	0,000000	0,007461	0,992567
7	1	290	1288	0,000776	0,008237	0,991797
8	0	171	1157	0,000000	0,008237	0,991797
9	0	154	1182	0,000000	0,008237	0,991797
10	2	167	1198	0,001669	0,009907	0,990142
11	2	266	1129	0,001771	0,011678	0,988390
12	2	157	932	0,002146	0,013824	0,986271
13	1	85	852	0,001174	0,014998	0,985114
14	1	161	831	0,001203	0,016201	0,983929
15	0	191	733	0,000000	0,016201	0,983929
16	0	173	618	0,000000	0,016201	0,983929
17	1	96	528	0,001894	0,018095	0,982068
18	1	70	494	0,002024	0,020119	0,980082
19	1	79	458	0,002183	0,022303	0,977944
20	0	64	422	0,000000	0,022303	0,977944
21	2	62	457	0,004376	0,026679	0,973674
22	2	78	463	0,004320	0,030999	0,969477
23	1	80	436	0,002294	0,033292	0,967256
24	1	64	478	0,002092	0,035385	0,965234
25	3	33	491	0,006110	0,041494	0,959355
26	1	49	528	0,001894	0,043388	0,957539
27	3	98	564	0,005319	0,048708	0,952460
28	1	64	536	0,001866	0,050573	0,950684
29	4	56	555	0,007207	0,057780	0,943857
30	3	122	557	0,005386	0,063166	0,938787
31	1	77	510	0,001961	0,065127	0,936948
32	6	69	493	0,012170	0,077298	0,925614
33	2	86	495	0,004040	0,081338	0,921882
34	3	73	495	0,006061	0,087399	0,916312
35	3	87	485	0,006186	0,093584	0,910661
36	0	57	435	0,000000	0,093584	0,910661
37	4	72	430	0,009302	0,102887	0,902229
38	5	52	383	0,013055	0,115941	0,890527

39	3	70	352	0,008523	0,124464	0,882970
40	4	85	319	0,012539	0,137003	0,871967
41	4	58	275	0,014545	0,151549	0,859376
42	0	38	249	0,000000	0,151549	0,859376
43	4	48	219	0,018265	0,169814	0,843822
44	1	29	191	0,005236	0,175049	0,839416
45	4	26	178	0,022472	0,197521	0,820763
46	4	33	168	0,023810	0,221331	0,801452
47	4	43	134	0,029851	0,251181	0,777881
48	1	26	103	0,009709	0,260890	0,770366
49	1	7	80	0,012500	0,273390	0,760796
50	3	22	85	0,035294	0,308684	0,734413
51	0	11	64	0,000000	0,308684	0,734413
52	0	17	59	0,000000	0,308684	0,734413
53	1	1	42	0,023810	0,332494	0,717133
54	1	16	40	0,025000	0,357494	0,699427
55	1	3	23	0,043478	0,400972	0,669669
56	0	10	19	0,000000	0,400972	0,669669
57	0	3	9	0,000000	0,400972	0,669669
58	0	6	6	0,000000	0,400972	0,669669

c. Tráfico Pesado (P)

Tiempo	Nº de fallos	Nº de censuras	Nº en riesgo	Función de riesgo	Función de riesgo acumulado	Función de supervivencia
1	0	5	112	0,000000	0,000000	1,000000
2	1	18	133	0,007519	0,007519	0,992509
3	0	13	126	0,000000	0,007519	0,992509
4	0	3	125	0,000000	0,007519	0,992509
5	0	24	139	0,000000	0,007519	0,992509
6	1	26	129	0,007752	0,015271	0,984845
7	1	22	105	0,009524	0,024795	0,975510
8	1	25	118	0,008475	0,033269	0,967278
9	0	11	127	0,000000	0,033269	0,967278
10	0	12	137	0,000000	0,033269	0,967278
11	0	17	142	0,000000	0,033269	0,967278
12	0	14	132	0,000000	0,033269	0,967278
13	2	2	122	0,016393	0,049663	0,951550
14	0	37	122	0,000000	0,049663	0,951550
15	0	32	86	0,000000	0,049663	0,951550
16	0	21	59	0,000000	0,049663	0,951550
17	0	17	44	0,000000	0,049663	0,951550
18	0	7	37	0,000000	0,049663	0,951550
19	1	4	34	0,029412	0,079074	0,923971
20	0	5	35	0,000000	0,079074	0,923971
21	0	1	39	0,000000	0,079074	0,923971
22	0	5	39	0,000000	0,079074	0,923971
23	1	6	36	0,027778	0,106852	0,898659
24	0	9	41	0,000000	0,106852	0,898659
25	0	4	35	0,000000	0,106852	0,898659
26	0	5	36	0,000000	0,106852	0,898659
27	0	8	45	0,000000	0,106852	0,898659
28	1	1	40	0,025000	0,131852	0,876471
29	0	2	45	0,000000	0,131852	0,876471
30	0	12	47	0,000000	0,131852	0,876471
31	0	3	39	0,000000	0,131852	0,876471
32	1	6	44	0,022727	0,154579	0,856775
33	1	15	44	0,022727	0,177307	0,837523
34	1	3	34	0,029412	0,206718	0,813249
35	0	7	36	0,000000	0,206718	0,813249
36	1	4	36	0,027778	0,234496	0,790969
37	0	3	33	0,000000	0,234496	0,790969
38	1	4	39	0,025641	0,260137	0,770946
39	0	6	37	0,000000	0,260137	0,770946

40	0	6	32	0,000000	0,260137	0,770946
41	0	7	34	0,000000	0,260137	0,770946
42	1	6	27	0,037037	0,297174	0,742915
43	0	2	20	0,000000	0,297174	0,742915
44	0	7	18	0,000000	0,297174	0,742915
45	0	3	12	0,000000	0,297174	0,742915
46	0	1	10	0,000000	0,297174	0,742915
47	0	7	9	0,000000	0,297174	0,742915
51	0	1	9	0,000000	0,297174	0,742915
52	0	1	9	0,000000	0,297174	0,742915
54	0	5	8	0,000000	0,297174	0,742915
55	1	0	3	0,333333	0,630508	0,532322
57	0	1	2	0,000000	0,630508	0,532322
58	0	1	1	0,000000	0,630508	0,532322