



**ESCUELA SUPERIOR POLITÉCNICA DEL
LITORAL**

Instituto de Ciencias Matemáticas

Ingeniería en Estadística Informática

**“Simulación y evaluación de la confiabilidad de un proceso de sistema de
producción”**

TESIS DE GRADO

Previa a la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentada por:

Jorge Simón Loor Quevedo

GUAYAQUIL – ECUADOR

AÑO

2003

AGRADECIMIENTO

En principio a Dios, por ser mi guía y estar a mi lado en los momentos más difíciles. A mis padres quienes mantuvieron hacia mi un apoyo constante e incondicional, quienes me enseñaron el sendero correcto y me impartieron sus principios, a mi hermana quien fue mi soporte en toda mi época de aprendizaje, a mi hermanito gracias a quien obtuve el carácter para poder ser un buen ejemplo. A mi director de tesis por sus conocimientos.

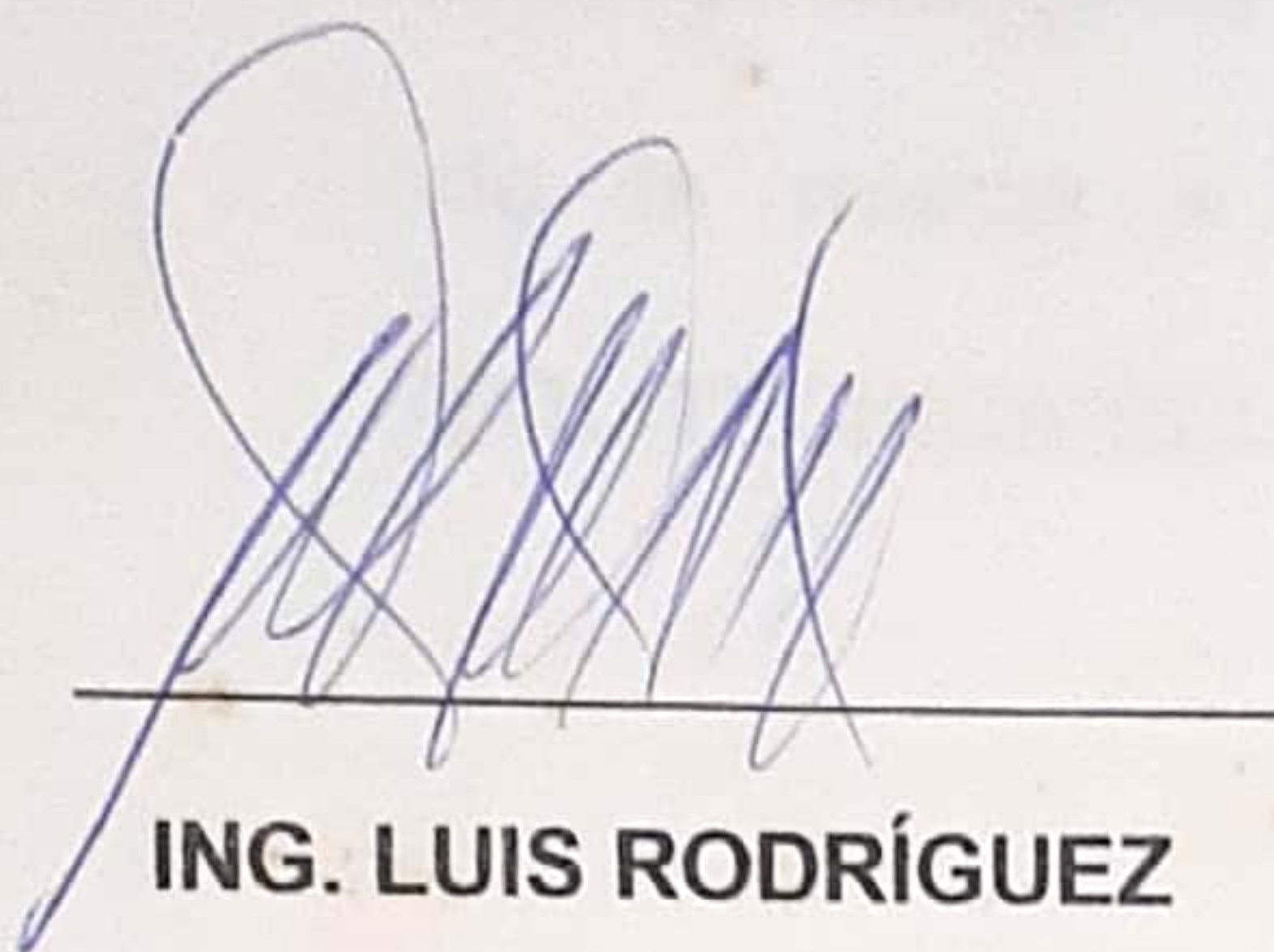


A.F. 136522

DEDICATORIA

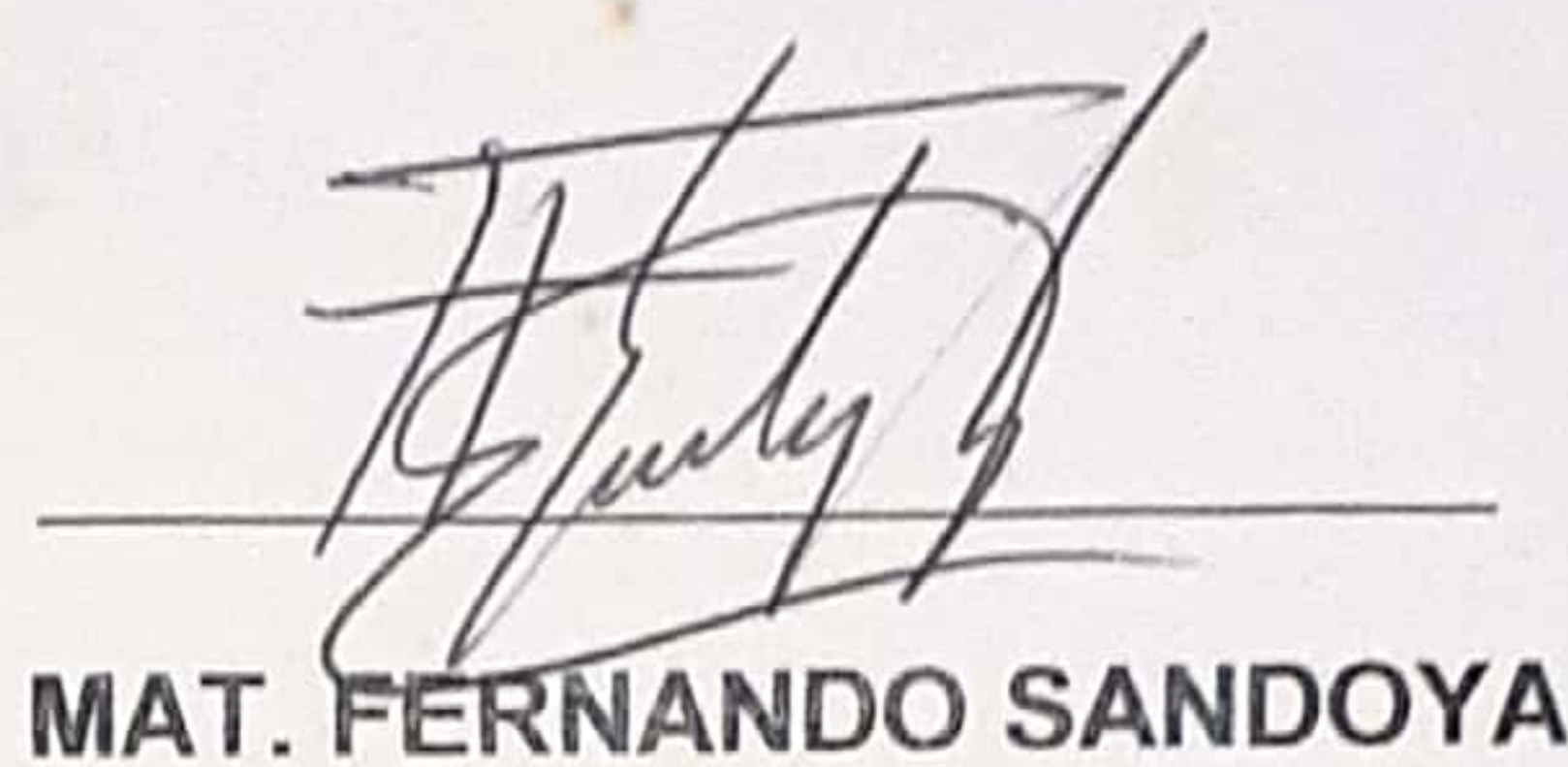
A las tres mujeres de mi vida: Mi
madre, mi hermana y mi amor
Mayra.

TRIBUNAL DE GRADUACIÓN



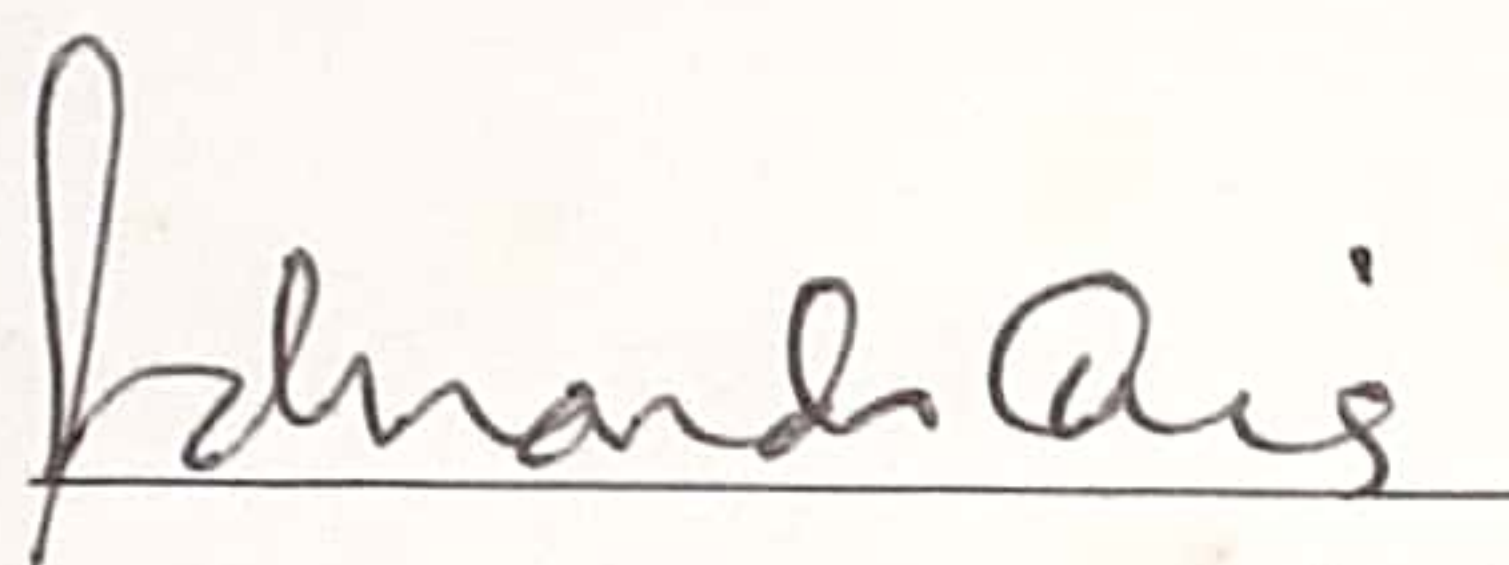
ING. LUIS RODRÍGUEZ

SUBDIRECTOR DEL ICM



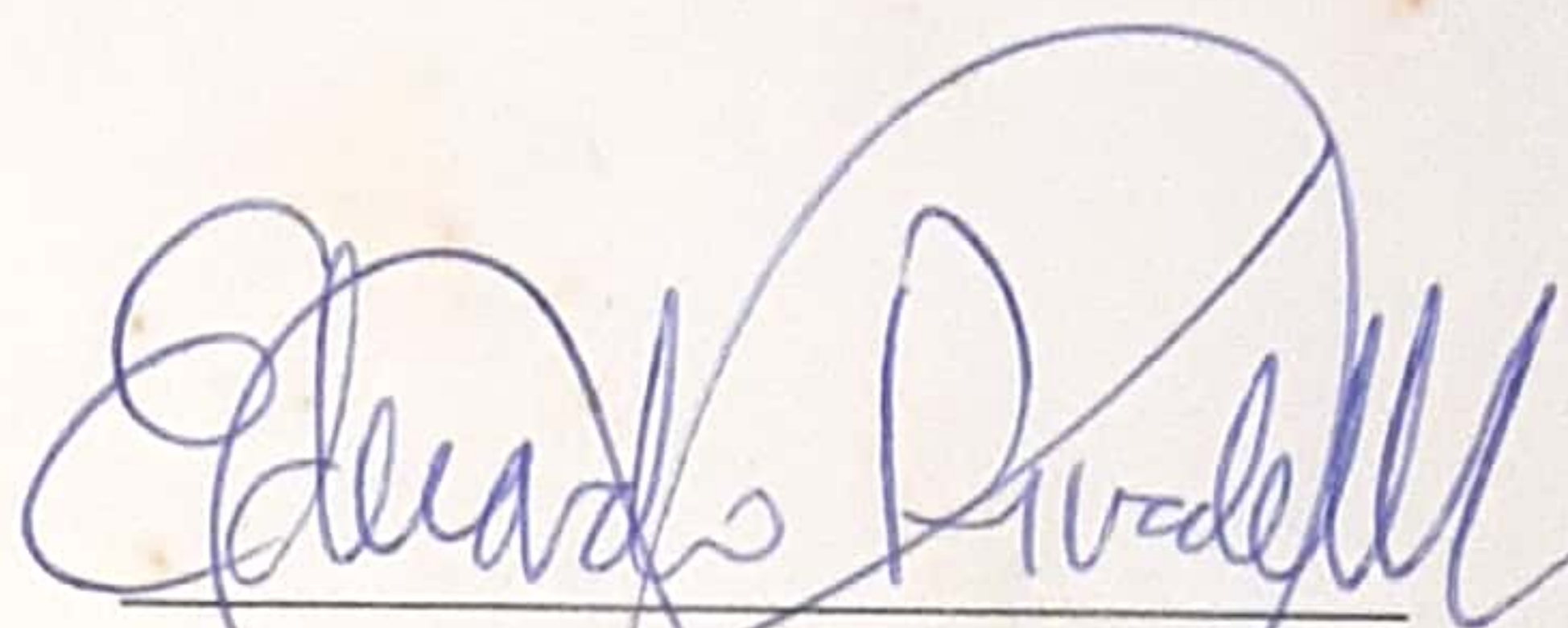
MAT. FERNANDO SANDOYA

DIRECTOR DE TESIS



ING. EDUARDO ORCÉS

VOCAL



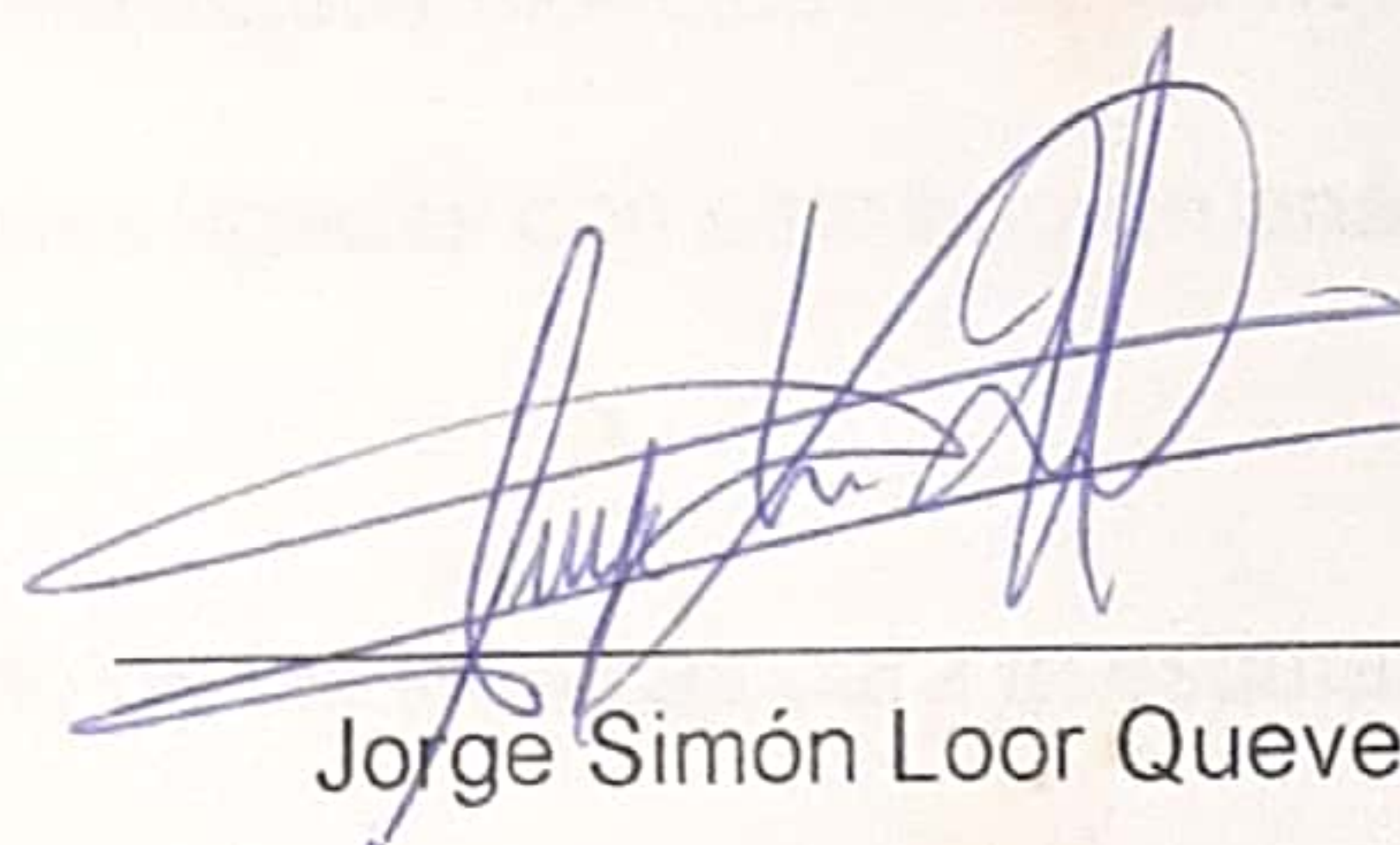
MAT. EDUARDO RIVADENEIRA

VOCAL

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta tesis de grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”

(Reglamento de graduación de la
ESPOL)



Jorge Simón Loor Quevedo

RESUMEN

En el presente trabajo se estudia la confiabilidad y la supervivencia de datos en una empresa que comercializa equipos de tecnología para desarrollar labores administrativas cotidianas. De igual forma se analizarán las fallas de existentes en un proceso de manufactura dentro de una empresa que se dedica a la elaboración de llantas.

En la primera parte se realiza un preámbulo de la historia e importancia de la confiabilidad de aparatos o sistemas ya sea en el campo de la medicina o en el campo industrial; dando a conocer también algunos conceptos básicos de la teoría de la confiabilidad o supervivencia.

En la segunda parte se manifiestan los diversos modelos de supervivencia, y demás técnicas estadísticas y actuariales ligadas con este tipo de análisis.

En la tercera parte se realiza un estudio de las técnicas para la estimación no paramétrica, apropiado para este análisis de los tiempos de falla o también denominados tiempos de supervivencia.

En la cuarta parte se realizarán los respectivos cálculos para estimar de forma paramétrica y no paramétrica la función de supervivencia de los datos obtenidos dentro de la empresa comercializadora de equipos con tecnología.

En esta cuarta parte, se realizará también el análisis estadístico de las fallas existentes en la producción de la empresa fabricante de llantas.

En la quinta y última parte se realizarán las conclusiones obtenidas del desarrollo del estudio, junto con las respectivas recomendaciones para la obtención de mejoras para próximos análisis.

1. INTRODUCCION, MOTIVACION Y EJEMPLOS

1.1. Introducción

1.1.1. Definición de la Tecnología

1.1.2. Importancia de la Tecnología

1.1.3. Análisis de la Tecnología

1.1.4. Función de Supervivencia

1.1.5. Tiempo Esperado de Vida

1.1.6. Supervivencia

1.1.7. Tiempo Esperado de Vida

1.1.8. Funciones de Supervivencia Exponenciales

1.1.9. Con respecto a la Función de Riesgo

INDICE GENERAL

	Pág.
RESUMEN	II
INDICE GENERAL	III
ABREVIATURAS	IV
SIMBOLOGÍA	V
INDICE FIGURAS	VI
INDICE TABLAS	VII
INTRODUCCIÓN	1
I. INTRODUCCIÓN, IMPORTANCIA Y EJEMPLOS	3
1.1. Introducción.....	3
1.1.1. Definición de Confiabilidad y Metodología.....	7
1.2. Importancia de la Confiabilidad de Sistemas.....	18
1.3. Análisis de Supervivencia.....	24
1.3.1. Funciones de Supervivencia.....	30
1.3.2. Utilización de datos para estimar la Función de Supervivencia.....	34
1.3.3. Tiempo Promedio de Fallas.....	41
1.3.4. Funciones de Supervivencia Exponenciales.....	42
1.3.5. Con respecto a la Función de Riesgo.....	44

1.3.6. Esperanza de vida a la edad t	47
--	----

II. MODELOS, CENSURAS Y VEROSIMILITUD PARA DATOS DE TIEMPOS DE FALLA **56**

2.1. Modelos para Procesos Continuos de Tiempos de Falla.....	56
2.1.1. Funciones de Distribución de Tiempos de Falla.....	57
2.1.2. Modelos de Riesgo.....	63
2.1.2.1. Modelo de Riesgo Constante.....	64
2.1.2.2. Modelo de Riesgo Potencia.....	72
2.1.2.3. Modelos de Riesgo Polinomiales.....	77
2.1.2.4. Riesgos Exponenciales.....	78
2.2. Modelos para Datos Discretos de un Proceso Continuo.....	81
2.2.1. Modelo de Tiempo de Falla Multinomial.....	81
2.2.2. Función de Distribución Acumulada Multinomial.....	84
2.3. Censuras.....	86
2.3.1. Mecanismos de Censuras.....	86
2.3.1.1. Censura Tipo I.....	89
2.3.1.2. Censura Tipo II.....	91
2.3.2. Supuestos Importantes en los Mecanismos de Censura.....	94
2.4. Verosimilitud.....	95
2.4.1. Verosimilitud basada en Métodos Estadísticos.....	95

2.4.2. Especificando la Función de Verosimilitud.....	97
2.4.3. Contribución a la Función de Verosimilitud.....	99
III. ESTIMACIÓN NO PARAMÉTRICA	103
3.1. Introducción	103
3.2. Estimación Individual de intervalos de datos censurados.....	104
3.3. Estimación de múltiples datos censurados.....	106
3.4. Estimación de múltiples datos censurados con tiempos exactos de falla.....	109
3.5. Tiempos de censura indefinidos.....	110
3.6. Cálculo de Tablas de Vida.....	113
3.7. Intervalos de confianza para datos con censura simple.....	122
3.7.1. Intervalo de confianza de punto centrado con aproximación normal para $F(t_i)$	122
3.8. Intervalos de Confianza de punto centrado para datos con censura múltiple.....	124
3.8.1. Fórmula de Greenwood.....	124
3.8.2. Intervalo de confianza de punto centrado con aproximación normal para $F(t_i)$	130
3.9. Bandas de Confianza Simultáneas.....	133
3.9.1. Bandas de Confianza Simultáneas para $F(t)$ a partir de tamaños de muestras grandes.....	134

3.9.2. Determinación del rango de tiempo para las bandas de confianza para $F(t)$	137
3.10. Censura Arbitraria.....	138

**IV. ESTIMACIÓN Y PREDICCIÓN PARA LOS DATOS OBTENIDOS EN
UNA EMPRESA QUE COMERCIALIZA EQUIPOS DE IMPRESIÓN Y
FOTOCOPIADO, Y EN UNA INDUSTRIA MANUFACTURERA DE
LLANTAS** **139**

4.1. Análisis de los tiempos de falla en los equipos comercializados por una empresa de tecnología.....	139
4.1.1. Alcance del estudio.....	140
4.1.2. Contrastes de Hipótesis para pruebas de bondad de ajuste.....	148
4.1.3. Función de Supervivencia y de Riesgo.....	152
4.1.4. Modelo Multinomial.....	159
4.1.5. Estimación No Paramétrica.....	164
4.1.5.1. Estimación de Kaplan – Meier.....	169
4.1.6. Diagnóstico para ajuste a un modelo paramétrico tomando en cuenta las censuras.....	175
4.2. Análisis estadístico de las fallas en la producción de una empresa de manufactura.....	187

4.2.1. Selección de los tipos de llantas para la aplicación del análisis estadístico.....	190
4.2.2. Definición de Variables.....	195
4.2.3. Análisis Univariado.....	196
4.2.3.1. Estadística Descriptiva.....	198
4.2.3.2. Pruebas de Bondad de Ajuste.....	212
4.2.3.3. Pruebas de Hipótesis e Intervalos de Confianza para la media de la variable Cantidad de unidades defectuosas para los tres gremios.....	216
4.2.4. Análisis Bivariado.....	220
4.2.4.1. Tablas de Contingencia.....	221
4.2.4.2. Análisis de Correlación.....	229
V. CONCLUSIONES Y RECOMENDACIONES	231
5.1. Conclusiones.....	232
5.1.1. Conclusiones para el estudio de la confiabilidad de una empresa que comercializa equipos con tecnología.....	232
5.1.2. Conclusiones para el estudio de las fallas y confiabilidad dentro del proceso de producción de una empresa que manufactura llantas.....	236

5.2. Recomendaciones..... 239

5.1.1. Recomendaciones para el estudio de la confiabilidad de una
empresa que comercializa equipos con tecnología..... 239

5.1.2. Recomendaciones para el estudio de las fallas y confiabilidad
dentro del proceso de producción de una empresa que
manufactura llantas..... 241

ANEXOS

BIBLIOGRAFÍA

ABREVIATURAS

<i>MTTF</i>	Tiempo Promedio de Falla
<i>MTBF</i>	Tiempo Promedio entre Fallas
<i>MTU</i>	Tiempo Promedio Disponible
<i>MTD</i>	Tiempo Promedio No Disponible
<i>MTR</i>	Tiempo Promedio de Reparación
<i>MACMT</i>	Tiempo Promedio de Mantenimiento Correctivo
<i>TTF ó T</i>	Tiempo de Falla
<i>fda</i>	Función de Distribución Acumulada
<i>fdp</i>	Función de Densidad de Probabilidad
<i>fs</i>	Función de Supervivencia
<i>fr</i>	Función de Riesgo
<i>IFR</i>	Tasa de Fallas Creciente
<i>DFR</i>	Tasa de Fallas Decreciente
<i>MLCG</i>	Generador Congruencial Lineal Multiplicativo
<i>EP</i>	Estadístico de Prueba
<i>RR</i>	Región de Rechazo

SIMBOLOGÍAS

$S(y)$	Función de Supervivencia de Y
$E(y)$	Valor Esperado de Y
$F(y)$	Distribución Acumulada de Y
$f(y)$	Densidad de Probabilidad de Y
$r(t)$	Tiempo de vida residual a la edad t
$h(t)$	Función de Riesgo de T
Δy	Variación en Y
I_A	Función Indicadora del evento A
$S_n(y)$	Función Empírica de Supervivencia de Y
p	Proporción de fallas
n	Tamaño de muestra
μ	Media
σ^2	Varianza
σ	Desviación Estándar
H_0	Hipótesis nula
H_1	Hipótesis alterna
α	Nivel de significancia
Z	Distribución Normal Estándar
X^2	Distribución Ji- cuadrado

df.	Grados de libertad
(α, β)	Parámetros de forma y escala
Y_i	Variable Aleatoria de Supervivencia i
t_i	Tiempo de censura fijado para Y_i
δ_i	Indicador de censura para Y_i
π_i	Probabilidad multinomial que una unidad falle en el intervalo i
p_i	Probabilidad condicional que una unidad falle en el intervalo i
n_i	Número en riesgo para el intervalo i
n_i^*	Número efectivo en riesgo para el intervalo i
d_i	Número de fallas para el intervalo i
r_i	Número de unidades que sobreviven al intervalo i. (censuras al intervalo i)
$z_{(p)}$	Cuantil p de la distribución Normal estándar
$e_{(a,b,1-\alpha/2)}$	Factores de precisión en la aproximación de Bandas Simultáneas de Confianza.
X_{ij}	Variable i para el gremio j.
μ_0	Media fijada
σ_{ij}	Covarianza entre la variable i y la variable j

r_{ij}	Estimador del coeficiente de correlación lineal entre la variable i y la variable j
t_d	Tiempo en el cual el equipo ha estado disponible.
t_u	Tiempo en el cual el equipo ha estado indisponible.
t_r	Tiempo de reparación de un equipo.
t_c	Tiempo de mantenimiento correctivo de un equipo.
k_F	Número total de fallos de los equipos en determinado período de observación.
k_d	Número total de tiempos en los cuales los equipos han estado disponibles.
k_u	Número total de tiempos en los cuales los equipos han estado indisponibles.
k_r	Número total de reparaciones de los equipos en determinado período de observación.
k_c	Número total de mantenimientos correctivos de los equipos en determinado período de observación.

INDICE DE FIGURAS

		Pág.
Figura 1.1	Interpolación de una función de tiempo.....	22
Figura 1.2	Forma General de una función de Supervivencia.....	33
Figura 1.3	Histograma del Ejemplo1.3 (Placebo).....	39
Figura 2.1	Curva de Riesgo en el tiempo de vida para los humanos.....	60
Figura 2.2	Partición del eje de tiempo para el Modelo Multinomial.....	82
Figura 2.3	Gráfico de π_i	83
Figura 4.1	Partición del eje de tiempo para la Línea de Equipo 1.....	160
Figura 4.2	Partición del eje de tiempo para la Línea de Equipo 2.....	161
Figura 4.3	Función de Supervivencia e Intervalos de Confianza para Línea de Equipo 1.....	166
Figura 4.4	Función de Supervivencia e Intervalos de Confianza para Línea de Equipo 2.....	168
Figura 4.5	Estimación de Kaplan - Meier para los datos de la Línea de Equipo 1.....	172
Figura 4.6	Estimación de Kaplan - Meier para los datos de la Línea de Equipo 2.....	174
Figura 4.7	Diagnóstico para el fenómeno de los datos de la Línea de Equipo 1.....	175

Figura 4.8	Diagnóstico para el fenómeno de los datos de la Línea de Equipo 2.....	176
Figura 4.9	Densidad de probabilidad de Weibull para los datos de la Línea de Equipo 1.....	179
Figura 4.10	Supervivencia de Weibull para los datos de la Línea de Equipo 1.....	180
Figura 4.11	Riesgo de Weibull para los datos de la Línea de Equipo 1.....	180
Figura 4.12	Densidad de probabilidad de Weibull para los datos de la Línea de Equipo 2.....	181
Figura 4.13	Supervivencia de Weibull para los datos de la Línea de Equipo 2.....	182
Figura 4.14	Riesgo de Weibull para los datos de la Línea de Equipo 2.....	182
Figura 4.15	Producción Total Mensual para cada gremio.....	193
Figura 4.16	Porcentaje Total de fallas por mes para cada gremio.	194
Figura 4.17	Histograma de Frecuencias de las Unidades Producidas para el gremio <i>Camion</i>	199
Figura 4.18	Histograma de Frecuencias de las Unidades Defectuosas para el gremio <i>Camion</i>	200
Figura 4.19	Histograma de Frecuencias del Porcentaje de fallas para el gremio <i>Camion</i>	202

Figura 4.20	Histograma de Frecuencias de las Unidades Producidas para el gremio <i>Cmt. Bias</i>	203
Figura 4.21	Histograma de Frecuencias de las Unidades Defectuosas para el gremio <i>Cmt. Bias</i>	205
Figura 4.22	Histograma de Frecuencias del Porcentaje de fallas para el gremio <i>Cmt. Bias</i>	206
Figura 4.23	Histograma de Frecuencias de las Unidades Producidas para el gremio <i>Cmt. Radial</i>	208
Figura 4.24	Histograma de Frecuencias de las Unidades Defectuosas para el gremio <i>Cmt. Radial</i>	209
Figura 4.25	Histograma de Frecuencias del Porcentaje de fallas para el gremio <i>Cmt. Radial</i>	211

INDICE DE TABLAS

		Pág.
Tabla I	Cuadro de resultados del experimento del Ejemplo 1.1.....	26
Tabla II	Cuadro de resultados del experimento del Ejemplo 1.2.....	28
Tabla III	Cuadro de resultados del experimento del Ejemplo 1.3.....	38
Tabla IV	Tabla de datos del Ejemplo 1.4.....	51
Tabla V	Promedio Residual de la esperanza de vida a la edad t para los datos del Ejemplo 1.4.....	53
Tabla VI	Datos de ejemplo de Riesgo Constante.....	69
Tabla VII	Contribución de la Censura a la Función de Verosimilitud.	99
Tabla VIII	Datos de fechas de instalación y de primera llamada para los equipos de tecnología.....	143
Tabla IX	Bondad de Ajuste Geométrico para los datos de la Línea de Equipo 1.....	150
Tabla X	Bondad de Ajuste Geométrico para los datos de la Línea de Equipo 2.....	151
Tabla XI	Funciones de Riesgo y Supervivencia a partir de la Distribución Geométrica para ambas líneas de equipo.....	154
Tabla XII	Bondad de Ajuste Exponencial para los datos de la Línea de Equipo 1.....	157
Tabla XIII	Bondad de Ajuste Exponencial para los datos de la Línea de Equipo 2.....	157

Tabla XIV	Funciones de Densidad, Riesgo y Supervivencia a partir de la Distribución Exponencial para ambas líneas.....	158
Tabla XV	Modelo Multinomial a partir de la Distribución Geométrica para la Línea de Equipo 1.....	160
Tabla XVI	Modelo Multinomial a partir de la Distribución Exponencial para la Línea de Equipo 1.....	161
Tabla XVII	Modelo Multinomial a partir de la Distribución Geométrica para la Línea de Equipo 2.....	162
Tabla XVIII	Modelo Multinomial a partir de la Distribución Exponencial para la Línea de Equipo 2.....	162
Tabla XIX	Tabla de Vida para los datos de la Línea de Equipo 1.....	165
Tabla XX	Tabla de Vida para los datos de la Línea de Equipo 2.....	167
Tabla XXI	Estimación de Kaplan - Meier para los datos de la Línea de Equipo 1.....	171
Tabla XXII	Estimación de Kaplan - Meier para los datos de la Línea de Equipo 2.....	173
Tabla XXIII	Valores de <i>MTTF</i> del diagnóstico para los datos de la Línea de Equipo 1.....	176
Tabla XXIV	Valores de <i>MTTF</i> del diagnóstico para los datos de la Línea de Equipo 2.....	177

Tabla XXV	Parámetros del modelo de Weibull y de <i>MTTF</i> para los datos de la Línea de Equipo 1.....	179
Tabla XXVI	Parámetros del modelo de Weibull y de <i>MTTF</i> para los datos de la Línea de Equipo 2.....	181
Tabla XXVII	Modelo Multinomial a partir de la Distribución de Weibull para la Línea de Equipo 1.....	183
Tabla XXVIII	Modelo Multinomial a partir de la Distribución de Weibull para la Línea de Equipo 2.....	184
Tabla XXIX	Comparación entre los modelos de Weibull y Kaplan - Meier para los 20 datos originales de cada línea.....	185
Tabla XXX	Codificación de los tipos de llanta seleccionados.....	191
Tabla XXXI	Tabla de la toma de Datos.....	192
Tabla XXXII	Definición de Variables según el tipo de llanta.....	196
Tabla XXXIII	Estadística Descriptiva de la Cantidad de Unidades Producidas para el gremio <i>Camion</i>	198
Tabla XXXIV	Estadística Descriptiva de la Cantidad de Unidades Defectuosas para el gremio <i>Camion</i>	200
Tabla XXXV	Estadística Descriptiva del Porcentaje de fallas para el gremio <i>Camion</i>	201
Tabla XXXVI	Estadística Descriptiva de la Cantidad de Unidades Producidas para el gremio <i>Cmt.Bias</i>	203

Tabla XXXVII	Estadística Descriptiva de la Cantidad de Unidades Defectuosas para el gremio <i>Cmt.Bias</i>	204
Tabla XXXVIII	Estadística Descriptiva del Porcentaje de fallas para el gremio <i>Cmt.Bias</i>	206
Tabla XXXIX	Estadística Descriptiva de la Cantidad de Unidades Producidas para el gremio <i>Cmt.Radial</i>	207
Tabla XL	Estadística Descriptiva de la Cantidad de Unidades Defectuosas para el gremio <i>Cmt.Radial</i>	209
Tabla XLI	Estadística Descriptiva del Porcentaje de fallas para el gremio <i>Cmt.Radial</i>	210
Tabla XLII	Pruebas de Bondad de Ajuste para la Cantidad de Unidades Defectuosas y para el porcentaje de fallas para cada gremio.....	212
Tabla XLIII	Contraste de Hipótesis para la media poblacional de la Cantidad de Unidades Defectuosas del gremio <i>Camion</i>	217
Tabla XLIV	Contraste de Hipótesis para la media poblacional de la Cantidad de Unidades Defectuosas del gremio <i>Cmt.Bias</i>	218
Tabla XLV	Contraste de Hipótesis para la media poblacional de la Cantidad de Unidades Defectuosas del gremio <i>Cmt.Radial</i>	219

Tabla XLVI	T.C. Tipo de Llanta – Porcentaje de Fallas.....	222
Tabla XLVII	T.C. Cant. Unid. Producidas – Cant. Unid. Defectuosas para el gremio <i>Camion</i>	224
Tabla XLVIII	T.C. Cant. Unid. Producidas – Cant. Unid. Defectuosas para el gremio <i>Cmt.Bias</i>	226
Tabla XLIX	T.C. Cant. Unid. Producidas – Cant. Unid. Defectuosas para el gremio <i>Cmt.Radial</i>	228
Tabla L	Análisis de Correlación entre las variables Cant. Unid. Producidas y Cant. Unid. Defectuosas para cada gremio	230

INTRODUCCIÓN

El estudio de nuevas técnicas que permitan un incremento en los niveles de confiabilidad (fiabilidad, mantenibilidad, disponibilidad) de los productos de alto riesgo tecnológico, ha suscitado un gran interés en los últimos años en las sociedades más avanzadas las cuales demandan cada vez mejores productos.

De esta manera el objetivo es presentar los resultados del desarrollo de un programa de confiabilidad, basado en la evaluación y seguimiento de un conjunto de indicadores, y su aplicación a la industria.

Las actuales estrategias de mantenimiento exigen que se controle las tasas de fallas, los tiempos promedios de reparación y los tiempos de uso de los sistemas compuestos por máquinas y equipos. En otras palabras se pide que se controle la fiabilidad, mantenibilidad y disponibilidad de los sistemas.

¿Qué beneficio brinda el tener equipos con tasas de fallas bajas, tiempos de reparación cortos y tiempos de uso prolongados? La respuesta es obvia, las empresas se vuelven más productivas y competitivas, porque sus costos de mantenimiento se reduciría y las máquinas pasarían menos tiempo paradas lo que se traduciría en un mayor rendimiento y por lo tanto mayores ingresos al menor costo.

Uno de los parámetros que sirve para conocer el estado de las máquinas con respecto a su operación es la confiabilidad.

Se debe de analizar la funcionalidad del sistema, tomando en cuenta los dos tipos básicos de conexiones que son: en serie y en paralelo.

Se dice que los componentes están conectados en serie, si una falla en cualquiera de ellos causa que el sistema falle. Se dice que los componentes están conectados en paralelo, cuando el sistema falla en el instante que todos lo hacen.

Por otro lado, se puede medir la confiabilidad de una línea de equipos mediante un estudio del tiempo de vida. Se toma una muestra de estos equipos y se analiza el tiempo hasta que cada uno de los equipos presente fallas. Esta una forma de medir la confiabilidad de ciertos dispositivos para los cuales no se puede tener acceso a la información de su manufactura.

La confiabilidad de alguna máquina industrial puede ser medida también realizando un estudio de los productos defectuosos fabricados por esta, es decir que se podría aplicar un análisis directamente a la máquina o realizar un estudio de las fallas que esta genere en la producción.

CAPÍTULO 1

1. INTRODUCCIÓN, IMPORTANCIA Y EJEMPLOS

En el presente capítulo se presentará la importancia de la confiabilidad de aparatos, sistemas y demás dispositivos; ya sea en el campo de ingeniería o de medicina. Se darán a conocer algunos conceptos básicos para el análisis de confiabilidad y supervivencia de datos. Se explicará además, con ejemplos, en que consisten estos análisis de supervivencia.

1.1. Introducción.

Para introducirnos en el tema de la Confiabilidad es preciso primero entender su significado y en donde nace o se concibe la idea de realizar estudios de confiabilidad para sistemas o dispositivos en general.

El desarrollo de las concepciones y técnicas para el análisis de confiabilidad de componentes, equipos y sistemas ha estado asociado al desarrollo de tecnologías complejas y de alto riesgo, tales como la aeronáutica, militar y nuclear. Las primeras preocupaciones surgieron en el sector aeronáutico. Antes de la Segunda Guerra Mundial se realizaron cálculos de seguridad para aviones polimotores capaces de volar con uno o más motores parados.

Durante la guerra de Corea el Departamento de Defensa de los Estados Unidos realizó estudios de confiabilidad de equipos electrónicos militares, cuyos fallos estaban ocasionando graves pérdidas económicas y disminución de la efectividad militar. La relación entre confiabilidad, costos y mantenimiento adquirió gran importancia. Las compras de equipos electrónicos por las fuerzas armadas de los Estados Unidos fueron reglamentadas según especificaciones de confiabilidad de los equipos.

En la década de 1950 comenzó el desarrollo de la industria nuclear, y los conceptos relacionados con la confiabilidad fueron usados de forma creciente en el diseño de las plantas nucleares y de sus sistemas de seguridad.

Hasta principios de los años 60 los estudios teóricos y prácticos sobre confiabilidad eran realizados fundamentalmente en los Estados Unidos y la Unión Soviética. En esta década los estudios se extienden hacia otros países y también hacia otras tecnologías.

Tiene lugar además un gran desarrollo de los fundamentos y de los conceptos teóricos relacionados con la confiabilidad, se produce la consolidación de una Teoría de la Confiabilidad. En esta época se expone por primera vez una teoría matemática de la confiabilidad (Barlow and Proschan (1964, 1975) y Gnedenko et al. (1965)).

El estudio de nuevas técnicas que permitan un análisis de la confiabilidad de equipos y sistemas ha suscitado un gran interés en los últimos años en sectores como el del automóvil, transporte, ingeniería civil, naval, etc., donde se ha observado un incremento progresivo en la aplicación de ciertos elementos de los programas de confiabilidad. Tradicionalmente, en dichos sectores, se ha venido manteniendo el riesgo en niveles bajos adoptando unos márgenes de seguridad elevados en la fase de diseño del producto. Sin embargo, en la actualidad es difícil de mantener estos procedimientos en un mercado tan competitivo, donde la competencia, la incorporación de nuevas tecnologías y el aumento

de la complejidad de los sistemas supone un riesgo tanto económico como de seguridad industrial y social.

La evaluación de la confiabilidad de equipos es importante tanto en la primera etapa de vida del equipo, es decir, en la fase de diseño, donde el análisis se realiza mediante el uso de modelos probabilísticos o métodos de simulación, como a lo largo de su vida operativa donde se deben validar los requisitos establecidos en la fase de diseño a partir de la información obtenida de la experiencia operacional del equipo.

Para realizar esta validación, en muchas ocasiones, es posible utilizar una serie de medidas que permiten una evaluación cuantitativa de la confiabilidad. Estas medidas, que pueden ser calculadas a partir del histórico de mantenimiento y fallos del equipo, integran una serie de factores tales como: la fiabilidad esperada a lo largo del tiempo, la mantenibilidad (en el caso de equipos reparables), o los niveles de disponibilidad del equipo durante su vida operativa.

A continuación se presentan una serie de medidas que permiten realizar un seguimiento de la confiabilidad de equipos.

1.1.1. Definición de Confiabilidad y Metodología.

La confiabilidad es un término que engloba factores tales como la *fiabilidad, disponibilidad y mantenibilidad*. En este ámbito el concepto de confiabilidad resulta especialmente importante en equipos o sistemas cuyo fallo lleve asociado un alto riesgo tecnológico entre los cuales puede incluirse equipos y sistemas. El análisis de confiabilidad es necesario desde la primera etapa del ciclo de vida de un equipo, es decir, desde la fase de diseño. Así, durante el diseño del equipo se debe garantizar que el mismo cumple unos requisitos mínimos en términos de fiabilidad, disponibilidad, y mantenibilidad. En esta etapa existen múltiples técnicas en función del tipo y alcance del estudio que pueden ser aplicadas con el fin de conseguir niveles de calidad aceptables en relación a la confiabilidad entre las más utilizadas se encuentran: Análisis de Modos y Efectos del Fallo (AMFE), Diagramas de bloques de fiabilidad, Árboles de fallos, Técnicas de predicción de la fiabilidad, Análisis de mantenibilidad, etc., técnicas que permiten determinar parámetros críticos de confiabilidad así como verificar si se alcanzan los objetivos previstos de confiabilidad durante la fase de diseño.

Una vez concluido el diseño de un producto en el cual se han especificado unos requisitos de confiabilidad en función del producto a diseñar, requisitos que en ocasiones tienen un carácter contractual, se debe proceder a validar si dichos requisitos se cumplen durante la vida operacional del equipo. Por otro lado, el realizar un seguimiento de la confiabilidad del equipo permite realimentar la fase de diseño y corregir problemas que puedan surgir en los niveles de confiabilidad y, en definitiva, llevar a cabo una mejora del producto. Para realizar el seguimiento de los requisitos de confiabilidad es posible la utilización de una serie de parámetros que permiten conocer el estado del equipo respecto a los diferentes factores que engloban la confiabilidad, los cuales se encuentran relacionados entre sí.

De esa forma garantizar una fiabilidad y una disponibilidad adecuada resulta posible si se consigue un buen diseño, se realiza un mantenimiento adecuado y se persigue una mejora continua del diseño y del mantenimiento. Para establecer un seguimiento del cumplimiento de los requisitos de confiabilidad de los equipos y sistemas resulta necesario establecer una metodología que permita de una forma estructurada realizar la evaluación y posterior seguimiento de aquellos indicadores que se consideren

representativos dentro de las diferentes áreas que engloba la confiabilidad. Dicha metodología consta de cuatro etapas:

- a) Obtención de la información,
- b) Diseño de indicadores,
- c) Diseño de equipos, y
- d) Evaluación y seguimiento de indicadores.

Etapas que se describen brevemente en los siguientes apartados.

a) Obtención de información.

La metodología se inicia con la obtención de la información necesaria, la cual tiene que provenir de tres fuentes principales: diseño, mantenimiento y operación. El tipo de seguimiento que pueda establecerse posteriormente va a depender, lógicamente, de la información existente relativa a los equipos y sistemas que se incluyan en el análisis y de cómo se encuentre la misma disponible. No obstante, básicamente, dicha información vendrá referida a históricos de mantenimientos, correctivos y preventivos, en lo relativo al número de actividades desarrolladas, duración de las mismas, tiempos de indisponibilidad, costes, etc.

b) Diseño de indicadores.

Una vez obtenida la información de partida, la siguiente etapa de la metodología consiste en diseñar aquellos indicadores de confiabilidad más adecuados a la aplicación concreta. Dichos indicadores serán función del tipo de datos operacionales que se dispongan. En principio, unos pocos indicadores deben ser suficientes para realizar el seguimiento de la confiabilidad. Sin embargo, puede ser necesario disponer de un conjunto más amplio de indicadores que permitan, en caso de observarse desviaciones en el comportamiento de los primeros, realizar un seguimiento de las causas que han originado dichas desviaciones. Los indicadores a seleccionar para realizar un seguimiento adecuado dependen lógicamente del tipo de equipo que se esté analizando, debiendo considerar dos aspectos fundamentales del mismo:

- si es o no un equipo reparable,
- si se encuentra normalmente en operación o en espera.

En función de los dos aspectos anteriores se deben seleccionar los parámetros más adecuados para realizar el análisis de confiabilidad.

Como se ha indicado anteriormente, el análisis de confiabilidad pretende asegurar ciertos parámetros tales como la fiabilidad, disponibilidad, y mantenibilidad del equipo. Por tanto resulta necesario seleccionar medidas adecuadas que permitan un seguimiento en cada una de las áreas anteriores. A continuación se sugieren dentro de cada una de las áreas de interés posibles medidas que pueden ser utilizadas para realizar el seguimiento.

La **fiabilidad** de un equipo se define como la probabilidad de que un dispositivo realice adecuadamente su función a lo largo del tiempo cuando opera en el entorno para el cual ha sido diseñado. Una de las medidas básicas de la fiabilidad para el caso concreto de equipos no reparables es el tiempo medio hasta el fallo (*MTTF*) el cual puede estimarse, si se disponen de los tiempos hasta el fallo de n equipos, a partir de:

$$MTTF = \frac{\sum_{i=1}^n t_i}{n}$$

siendo t_i el tiempo total de operación del equipo i en el período de observación.

En el caso concreto de equipos reparables se puede estimar la fiabilidad a partir del tiempo medio entre fallos (*MTBF*) como:

$$MTBF = \frac{\sum_{i=1}^n t_i}{k_F}$$

siendo t_i el tiempo de observación de n equipos y k_F el número total de fallos de los equipos durante dicho período de observación.

Otras medidas adicionales relacionadas con la fiabilidad de equipos reparables son el tiempo medio en que el equipo está disponible (MUT) y el tiempo medio que se encuentra indisponible (MDT). Medidas que pueden ser evaluadas como:

$$MUT = \frac{\sum_{i=1}^n (t_d)_i}{k_d}$$

$$MDT = \frac{\sum_{i=1}^n (t_u)_i}{k_u}$$

donde t_d y t_u son los tiempos en que el equipo i se encuentran disponibles e indisponibles, respectivamente, en el intervalo de observación, y k_d y k_u , son respectivamente, el número total de tiempos en que el equipo i ha estado disponible e indisponible, siendo n el total de equipos observados.

La ***mantenibilidad*** de un equipo o sistema es la probabilidad de que pueda ser restaurado a uno de sus estados operacionales en un periodo de tiempo dado cuando la acción de reparación se efectúa de acuerdo a procedimientos previamente establecidos. Si disponemos de los tiempos medios de reparación de n equipos podemos obtener una estimación del tiempo medio de reparación como:

$$MRT = \frac{\sum_{i=1}^n (t_r)_i}{k_r}$$

donde k_r es el número total de reparaciones de los equipos durante el periodo de tiempo considerado y t_r el tiempo de reparación correspondiente al equipo i . El MRT puede ser también estimado a partir de:

$$MRT = MACMT - MDT$$

siendo $MACMT$ el tiempo medio de mantenimiento correctivo y MDT el tiempo medio técnico. Donde $MACMT$ puede ser evaluado, si se conocen los tiempos medios de correctivo de n equipos, como:

$$MACMT = \frac{\sum_{i=1}^n (t_c)_i}{k_c}$$

siendo t_c el tiempo de mantenimiento correctivo del equipo i en el periodo de tiempo analizado y k_c el número total de mantenimientos correctivos de los equipos durante el periodo de tiempo considerado.

La **disponibilidad** se utiliza para medir el efecto combinado de la fiabilidad, la mantenibilidad y el apoyo logístico en la efectividad operacional del equipo. Para realizar un análisis de disponibilidad del equipo se debe tener en cuenta los correctivos y fallos así como las actividades de mantenimiento programado que le aplican. Dicha disponibilidad puede ser aproximada, para equipos no reparables, mediante:

$$A = \frac{MTTF}{MTTF + MTTR}$$

y como:

$$A = \frac{MTTF}{MTTF + MTTR}$$

c) Diseño de equipos.

La etapa de diseño de equipos consiste en establecer los sistemas, equipos y componentes en alcance del estudio. En esta etapa

resulta interesante agrupar los equipos de acuerdo a características comunes. Así, por ejemplo, se puede considerar un sistema completo, para realizar su seguimiento como un único grupo que podemos denominar grupo principal. Dicho grupo principal puede ser dividido a su vez en diferentes subconjuntos, cada uno de los cuales podría corresponder a subsistemas del grupo principal, el cual podría ser de nuevo dividido en los diferentes equipos que lo constituyen, los cuales pueden ser agrupados en, por ejemplo, componentes mecánicos, eléctricos, e instrumentación y control. El establecimiento de esta jerarquía nos permite realizar, posteriormente, un seguimiento del grupo principal y utilizar el resto de subgrupos en el caso de observarse desviaciones, de forma que permitan investigar cual ha sido el subsistema o equipo concreto que ha originado dicha desviación.

d) Evaluación y seguimiento de indicadores.

Los indicadores seleccionados deben ser evaluados para equipos y periodos concretos de tiempo. El análisis de los resultados obtenidos del proceso de evaluación puede ser dirigido a los valores absolutos que toma dicho indicador o a la tendencia del mismo. Si se realiza el análisis sobre el valor absoluto del indicador

en cada periodo de tiempo de la historia del equipo puede compararse dicho valor con una referencia previamente fijada con el fin de determinar si el mismo presenta una desviación significativa o no.

La evaluación y análisis de los indicadores sobre equipos y periodos de tiempo determinados permite servir de apoyo en el proceso de toma de decisiones sobre el comportamiento de los equipos respecto a sus características de confiabilidad.

La Confiabilidad en la práctica.

El cálculo de la confiabilidad en la práctica no es sencillo como lo parece. ¿Por qué?

Primero, porque la obtención o recolección de datos sobre los tiempos de falla no es una tarea sencilla. Cuando hay registros, estos no están completos o presentan anomalías, otra razón es que hay que dar capacitación a los operadores que son los que toman los datos.

Segundo, debe estar definido claramente los niveles a cuales se va calcular la confiabilidad, es decir, a nivel de sistema, subsistema o componente. Es decir definir el alcance del estudio, analizando los datos que se tienen.

Tercero, es necesario definir lo que es una falla o modo de falla y el nivel de criticidad de la misma.

Cuarto, no se sabe cuales son los componentes de un sistema ni como se relacionan entre sí.

¿Cómo se resuelven estos contratiempos?

En primer lugar hay que definir que va a realizar nuestro sistema, definiendo el alcance.

En segundo lugar hay definir que se va considerar como fallas.

En tercer lugar, hay que entrenar a los operadores para que anoten de la manera más precisa los tiempos que van a ser analizados dentro del estudio.

Pero eso no es todo, ya que implementar todo lo anterior no es fácil, requiere paciencia y tiempo. Además conocer la estructura de un sistema no es sencillo y si no se conoce la estructura no se podría. Finalmente hay que tomar en cuenta el período de medición.

1.2. Importancia de la Confiabilidad de Sistemas.

El campo de aplicación de la Teoría de la Confiabilidad se amplía constantemente. Todos los sistemas de ingeniería, simples y complejos, pueden beneficiarse de la aplicación integrada de los conceptos de esta teoría en sus fases de planeación, diseño y operación. En la fase de operación y de mantenimiento de los equipos, los estudios de confiabilidad ayudan en la toma de decisiones sobre la calidad y frecuencia de los mantenimientos.

Un aumento de la confiabilidad conlleva, en general, el aumento a corto plazo de los costos.

Pero este aumento de la confiabilidad puede revertirse en ganancia en un plazo mayor, y puede significar, por otra parte, una disminución de riesgos para la salud y la vida de las personas, y para el medio ambiente.

La salud y riesgo en la vida de las personas es uno de los principales factores para los cuales debería considerarse un Análisis de Confiabilidad de los Sistemas, ya que si no se obtiene un alto grado de confianza en el desempeño de los distintos equipos y componentes que se implementan en hospitales y demás centros de salud, se pondría en riesgo la vida misma de uno o más individuos.

Por esto es preciso que se tenga un nivel elevado de confianza en estos sistemas (especialmente si se trata de equipos para operación y de quirófano), ya que si se tiene un nivel de confianza para el desempeño correcto de los equipos de un 95%, no es muy aceptable, debido a que esto pudiere significar que "de cada 100 enfermos tratados con estos equipos mueren 5 a causa del mal funcionamiento de los mismos"; y esto no resulta muy beneficioso.

Para el caso de Confiabilidad en sistemas industriales, el análisis resulta muy ventajoso debido a que se puede obtener resultados favorables en cuanto a calidad, mayor rentabilidad a largo plazo, bajo costo de mantenimiento, optimización de los procesos, reducción de tiempos de fallos en las máquinas o componentes.

El proceso conocido como *Análisis de Confiabilidad de Sistemas* está compuesto por una serie de etapas que conducen no solo a la evaluación precisa de determinados parámetros sino a un profundo conocimiento del sistema por parte de quien lo realiza.

Como se mencionó anteriormente, antes de comenzar cualquiera de estos análisis será necesario definir clara y explícitamente el alcance y el grado de detalle. En la práctica estos parámetros se suelen fijar de forma iterativa, ya que, con el avance del estudio se llega a una comprensión y definición más exacta de los mismos. La correcta elección es de gran importancia, ya que se trata de incluir todo aquello que tenga importancia para el análisis pretendido y que sea abordable tanto por disponer de información sobre aspectos cualitativos y/o datos técnicos como por disponer de conocimientos sobre su comportamiento y métodos de resolución adecuados. Tanto el alcance como el grado de detalle deben ser

complementados con el horizonte temporal del análisis así como con la definición de los objetivos del mismo y medios disponibles para la realización. Con todo ello se llega a una primera aproximación válida y realista al contexto general del análisis y se puede plantear una estrategia global de resolución.

A la hora de averiguar la evolución de la calidad de un elemento cualquiera, así como de su confiabilidad, aparece la pregunta:

¿Cómo ha evolucionado hasta el momento, y cómo lo hará en el futuro?

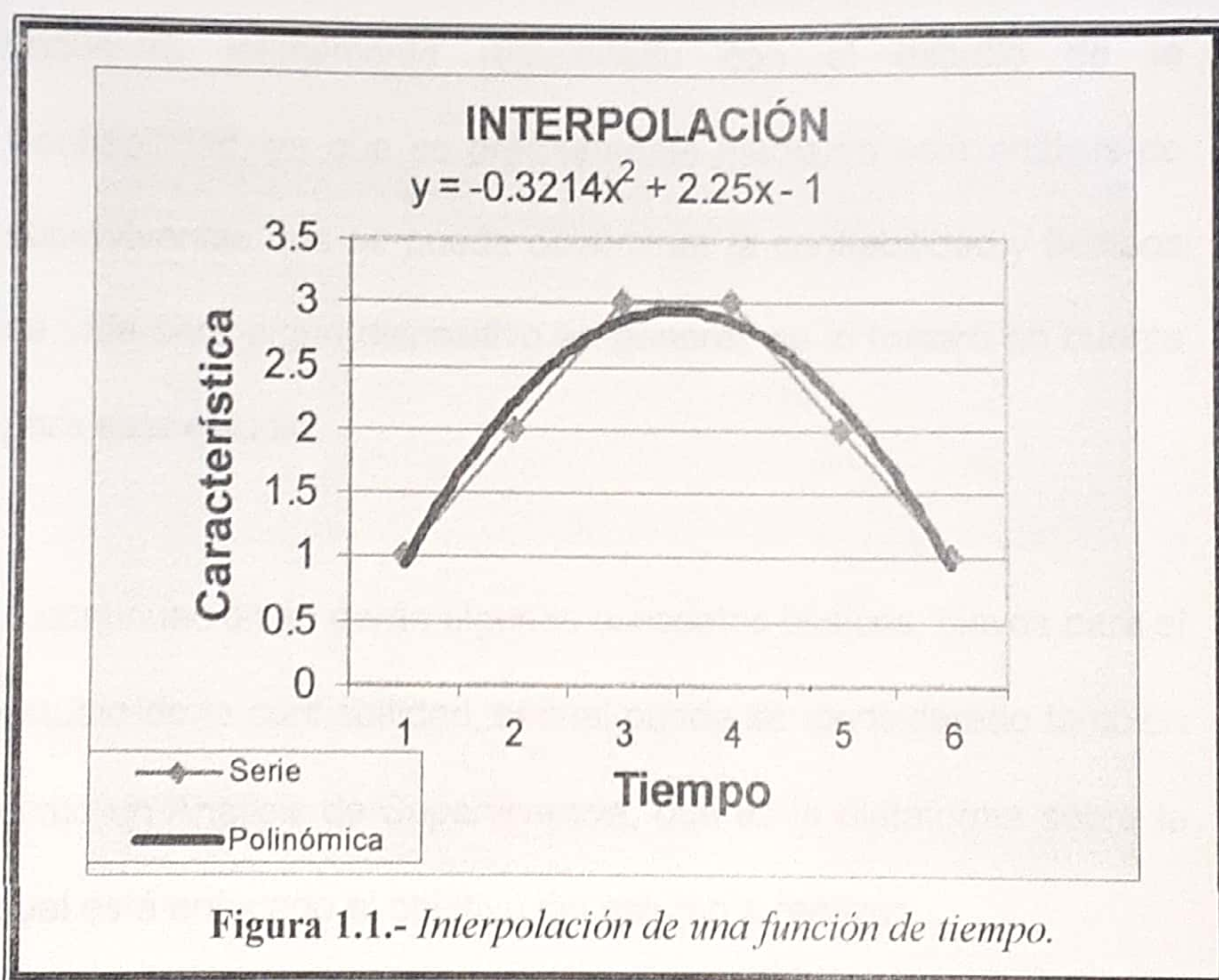
Esta pregunta, tendría respuesta, en el caso de disponer de una función $f(t)$, que dependa del tiempo, la que nos determine la característica que buscamos, en un instante determinado.

El encontrar esta función, no es en absoluto, una tarea fácil; para ello, podemos optar por 2 caminos:

- 1.) Poseer una base de datos, correspondiente a todas las eventualidades de cualquier tipo que le han ocurrido al elemento (fallos, reparaciones, tiempos de reparación, tiempos entre fallos, dinero que cuesta cada reparación, etc...) en función del tiempo.

Con ella, se puede aproximar una función f por interpolación (existen diversos métodos de interpolación, Newton, Lagrange,...), que dependerá del tiempo; de esta forma, podemos conocer el estado de cierta característica a lo largo del tiempo. La principal dificultad que tiene este procedimiento, es el conseguir dicha base de datos temporal (y con ella una o varias series temporales); de ahí la importancia de su existencia.

Aquí se tiene un ejemplo de interpolación; los puntos son la representación de los datos reales:



2.) Aproximar una función f , que se supone representa la evolución a lo largo del tiempo. Este segundo procedimiento, conlleva numerosos peligros, ya que dependiendo de la función que aproximemos, obtendremos resultados diferentes.

Otro camino a seguir resulta de implementar el Análisis de Supervivencia para estas máquinas, equipos o componentes, el cual puede tomar como instante o tiempo cero el momento en el cual se inicia la observación o toma de datos.

Dado que este Análisis de Fallas y Supervivencia de datos se encuentra íntimamente relacionado con el estudio de la Confiabilidad, ya que es precisamente mediante este análisis de supervivencia que se puede determinar la confiabilidad y tiempos de vida para algún dispositivo en general; se lo tomará en cuenta para este estudio.

A continuación se darán algunos conceptos básicos, claves para el estudio de la confiabilidad, el cual puede ser considerado también como un Análisis de Supervivencia, que es la plataforma sobre la cual está enfocado el objetivo del estudio a realizar.

1.3. Análisis de Supervivencia.

El Análisis de Supervivencia involucra el estudio de distribuciones de Tiempos de vida. Estos tiempos de vida pueden ser medidos en personas como por ejemplo pacientes con algún tipo de enfermedad, o para máquinas, componentes o paquetes de software.

Como se mencionó anteriormente este análisis considera como tiempo cero el instante en que se inicia la inspección, si es que no se posee datos de tiempos anteriores, pero cuando el caso proviene de un experimento, entonces resulta más sencillo determinar el instante inicial o tiempo cero.

El análisis se lo realizará enfocado a las aplicaciones en equipos y asuntos de manufactura, pero se tomará en cuenta un ejemplo biológico para realizar una analogía del estudio de la supervivencia.

Los datos y análisis de tiempos de vida son encontrados con mayor frecuencia en el ambiente de la ingeniería, ya que siempre

constituye una tarea de estudios y seguimientos no tan simples para el buen desempeño de las máquinas y de los procesos de manufactura en sí.

Los siguientes ejemplos ilustran en mejor forma en que consiste y el objetivo que persigue el análisis de la duración o tiempo de vida tanto en seres vivos como en un componente o máquina.

Ejemplo 1.1.

Tiempo transcurrido hasta la aparición de un tumor en ratones

Los datos presentados a continuación fueron tomados de un estudio diseñado para chequear si el cambio de las condiciones experimentales tenía algún efecto en el "tiempo hasta que el tumor aparezca" en ratones inyectados con agente cancerígeno. El tiempo de inyección con el agente fue tomado como tiempo cero. Luego los ratones fueron observados: fue de interés el tiempo transcurrido hasta que el tumor se desarrolle en los ratones (1).

(1). Peter J. Smith, 2002, *Analysis of Failure and Survival Data*, Primera Edición, Pag. 2.

Autopsia encontró:	Edades Individuales de muerte (días)									
	Ratones Convencionales									
Tumor	381	477	485	515	539	563	665	582	603	
	616	624	650	651	656	659	672	679	698	
	702	709	723	731	775	779	795	811	838	
No Tumor	45	198	215	217	257	262	266	371	431	
	447	454	459	477	479	484	500	502	503	
	505	508	516	531	541	553	556	570	572	
	575	577	585	588	594	600	601	608	614	
	616	632	632	638	642	642	642	644	644	
	647	647	653	659	660	662	663	667	667	
	673	673	677	689	693	718	720	721	728	
	760	762	773	777	815	886				
	Ratones Libres de Gérmenes									
Tumor	546	609	692	692	710	752	753	781	782	
	789	808	810	814	842	846	851	871	873	
	876	888	888	890	894	896	911	913	914	
	914	921	921	921	926	936	945	1008		
No Tumor	412	524	647	648	695	785	814	817	851	
	880	913	942	986						

Tabla I.- Cuadro de resultados del experimento del Ejemplo 1.1.

Resulta dificultoso detectar cuando un tumor se desarrolla, algunos ratones fueron inspeccionados en el instante de sacrificio o en el instante de muerte. Para este caso de supervivencia, un tiempo de vida se refiere al tiempo transcurrido hasta que el tumor aparece. Esto es difícil de medir con exactitud.

Algunos ratones muertos o en sacrificio no mostraron evidencia de algún tumor presente y esto refleja una complicación en el análisis de estos datos.

Si en el momento del sacrificio no se presenta el tumor, entonces es obvio que el tiempo de aparición sería mayor que el tiempo de sacrificio. Si al momento de la muerte, un tumor está presente, entonces el tiempo de aparición del tumor es menor que el tiempo de muerte (ya que los tumores no son rápidamente letales). Estos efectos son ejemplos de **Censuras** que son características comunes de datos de tiempos de vida.

Más adelante se tratará con profundidad el tema de los datos censurados o también denominado Censura de datos.

Ejemplo 1.2.

Datos de fallas en el proceso de fundición del aluminio

Para este ejemplo se tomará en cuenta un horno para la fundición de aluminio (1).

(1) Peter J. Smith, 2002, Analysis of Failure and Survival Data, Primera Edición, Pag. 1.

La alúmina es fundida en una caja o cubículo forrado de acero el cual está construido para soportar temperaturas extremadamente altas. En el proceso de fundición, el aluminio es producido como un derivado cuando el cubículo de acero funciona semejante a una batería con alúmina derretida como electrolitos. El cubículo necesita ser reemplazado cuando se producen rajaduras, permitiendo que impurezas penetren dentro del proceso. Los datos de tiempos de falla que a continuación se listan fueron parte de un gran conjunto de datos y representan días de servicio hasta que el cubículo fue reemplazado.

Período hasta ocurrencia de Falla (días)					
1540	1415	660	999	1193	1006
869	1035	797	296	775	1424
1169	1500	728	670	841	

Tabla II .- Cuadro de resultados del experimento del Ejemplo 1.2.

Por tanto el objetivo de realizar el análisis de supervivencia de máquinas o componentes es averiguar cuál o cuáles son los factores preponderantes que afectan el tiempo de vida de los mismos.

Estos análisis contribuyen principalmente a establecer el grado de confiabilidad de un determinado proceso

En este ejemplo los datos proporcionados son *tiempos de falla* que permiten visualizar en un determinado período la situación del proceso y a su vez realizar una estimación del tiempo de vida promedio de un cubículo forrado de acero para el proceso de fundición de aluminio.

Finalmente se podría complementar estos análisis con estudios de predicción: Cuánto tiempo durará determinada batería? Cómo se podría construir un período de garantía? Que tiempo de absolución se podría esperar para un paciente de cáncer? Cuánto tiempo pasará hasta que un componente industrial falle? (Ejemplo 1.2)

La habilidad para predecir es a menudo dependiente del análisis de tiempo de vida de los datos.

En algunos contextos (en medicina por ejemplo) estos análisis son a menudo desarrollados sin ningún tipo de distribución. Sin embargo, en algunas aplicaciones de ingeniería, los análisis pueden depender de un modelo el cuál es arreglado según los datos, es decir un modelo adaptado o ajustado a los datos.

La precisión de las inferencias a su vez depende de la bondad del ajuste del modelo.

Por esto pueden existir tanto modelos libres de distribución como modelos basados en aproximaciones de distribuciones ya conocidas.

1.3.1. Funciones de Supervivencia.

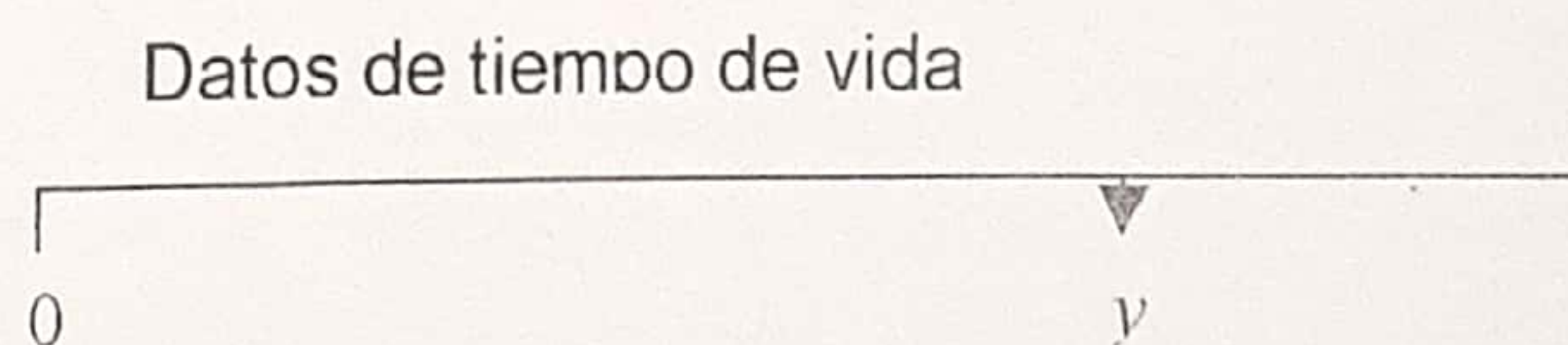
Al hablar de tiempo de vida o tiempo de supervivencia se hará referencia a un período entre el principio de la observación hasta la ocurrencia de la muerte.

La muerte es también denominada falla, por esto al tiempo de muerte se lo conoce también como tiempo de falla.

Lo que se podría interpretar en el caso biológico como muerte, en el caso de máquinas o componentes estos experimentan fallas.

Definición:

Una variable aleatoria Y es una **Variable Aleatoria de Supervivencia** si un valor observado y de Y pertenece al intervalo $[0, \infty)$.



Conjuntamente, datos de supervivencia $y_1, y_2, y_3, \dots, y_n$ son observaciones de una variable aleatoria positiva Y .

Examinaremos específicamente el caso continuo y si llegare a tomarse en cuenta el caso de una variable aleatoria discreta, este caso será expresado claramente.

Suponiendo que Y tiene una función de densidad de probabilidad f y una función de distribución acumulada F , entonces:

$$F(y) = P(Y \leq y) = \int_0^y f(u) du. \quad (1.1)$$

El estudio de funciones de supervivencia es la base de un análisis de supervivencia.

Definición:

La **Función de Supervivencia** S , para todos los valores de y , está definida por:

$$S(y) = 1 - F(y).$$

Esto es:
$$S(y) = P(Y > y) = \int_y^{\infty} f(u) du, \quad (1.2)$$

Aplicando el teorema fundamental del Cálculo, diferenciando ambos lados con respecto a la variable u :

$$f(u) = -\frac{d}{du} S(u) \quad (1.3)$$

De la ecuación dispuesta en (1.2) se puede establecer lo siguiente, tomando en cuenta que al integrar las funciones de densidad de probabilidad el resultado es igual uno:

$$S(0) = \int_0^{\infty} f(u) du = 1$$

$$S(\infty) = \lim_{y \rightarrow \infty} S(y) = \lim_{y \rightarrow \infty} \int_y^{\infty} f(u) du = 0$$

Finalmente:

$$\text{Si } a \geq b, \text{ entonces } S(b) - S(a) = \int_b^a f(u) du \geq 0$$

Lo cual quiere decir que una función de supervivencia siempre es decreciente.

De estas tres propiedades se deduce el siguiente teorema:

Teorema 1.1. La función de supervivencia S es monótona decreciente en el intervalo $[0, \infty)$. Además, S satisface $S(0) = 1$, $S(\infty) = 0$.

Esta función de supervivencia pertenece a alguna variable aleatoria de supervivencia la cual posee una función de densidad de probabilidad $f(u) = -\frac{d}{du} S(u)$.

La forma general de una función de supervivencia para una variable aleatoria continua es como sigue:

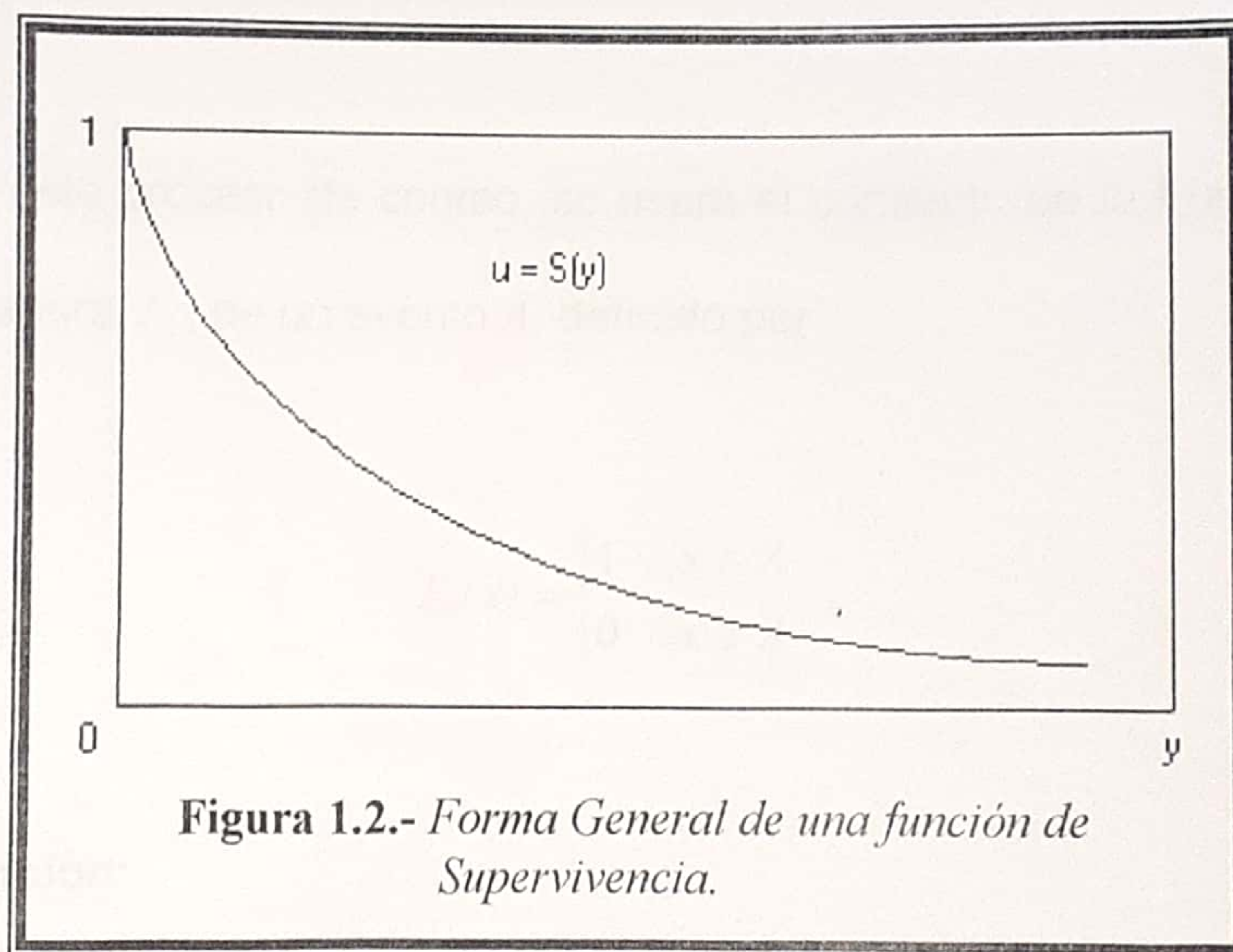


Figura 1.2.- Forma General de una función de Supervivencia.

1.3.2. Utilización de datos para estimar la Función de Supervivencia.

Asumiendo que se ha obtenido datos de una variable aleatoria de supervivencia, entonces se pueden usar estos datos para estimar la función de supervivencia $S(y)$. Se empieza asumiendo un modelo para la distribución de Y , pero tomando en cuenta lo que sugieren los datos para la forma que tendrá la función de supervivencia.

El método más simple y más obvio es a través de la *Función Empírica de Supervivencia*, la cual cuenta el número de puntos (datos) que son mayores a y .

Para este proceso de conteo, se usará el concepto de la Función Indicadora, I_A , de un evento A , definido por:

$$I_A(x) = \begin{cases} 1 & , x \in A \\ 0 & , x \notin A \end{cases}$$

Definición:

Dadas n observaciones $Y_1, Y_2, Y_3, \dots, Y_n$ independientes e idénticamente distribuidas con la misma distribución de Y , la

Función Empírica de Supervivencia S_n está definida para todos los valores de y como:

$$S_n(y) = \frac{\text{número de observaciones} > y}{n} = \frac{1}{n} \sum_{i=1}^n I_{(y, \infty)}(Y_i) \quad (1.4)$$

Además $S_n(y)$ es un estimador de la función de supervivencia $S(y) = P(Y > y)$.

La definición anterior demuestra que para valores fijos de y , $S_n(y)$ es un promedio de n variables aleatorias independientes e idénticamente distribuidas, Z_i , donde: $Z_i = I_{(y, \infty)}(Y_i)$. Por esto:

$$nS_n(y) = \sum_{i=1}^n Z_i$$

Cada Z_i posee la misma distribución de probabilidad Bernoulli (es decir que son i.i.d.), donde $P(Z_i = 1) = P(Y_i > y) = S(y)$ y $P(Z_i = 0) = P(Y_i \leq y) = 1 - S(y)$.

Nótese que ésta distribución de probabilidad tiene media $E(Z_i) = S(y)$ y varianza $Var(Z_i) = S(y)[1 - S(y)]$, dado que Z_i proviene de una distribución de Bernoulli de media p y varianza $p(1-p)$; donde $p = S(y)$.

Esto quiere decir que $nS_n(y)$ posee una distribución binomial:

$$nS_n(y) \rightarrow \text{Bin}(n, S(y)).$$

Entonces obteniendo un valor esperado para hallar la media y obteniendo a su vez la varianza del estimador $S_n(y)$ de la función de supervivencia:

$$E[S_n(y)] = \frac{1}{n}[nS(y)] = S(y) \quad (1.5)$$

$$\text{Var}[S_n(Y)] = \frac{1}{n^2} \{nS(y)[1 - S(y)]\} = \frac{S(y)[1 - S(y)]}{n} \quad (1.6)$$

Para un punto fijo $y = y^*$ se estima $S(y^*) = P(Y > y^*)$, la probabilidad de sobrevivir más allá de y^* , usando el estimador $S_n(y^*)$.

La ecuación (1.5) demuestra que $S_n(y^*)$ es un estimador insesgado de $S(y^*)$.

Para conseguir una medida de la precisión para $S_n(y^*)$, es apropiado emplear un intervalo de confianza aproximado basado en dos errores estándares.

$$S_n(y^*) \pm 2 \sqrt{\frac{S_n(y^*)[1 - S_n(y^*)]}{n}}$$

Aquí nuevamente S_n es usado para estimar S

Los cálculos se muestran con el siguiente ejemplo:

Ejemplo 1.3.

Función de Supervivencia Empírica para el efecto placebo.

Gehan, Lawless y otros han discutido datos de un experimento clínico, examinando el tiempo en semanas de remisión de esteroides inducidos para pacientes con leucemia.

A un grupo de 21 pacientes se les fue suministrado 6-MP, a un segundo grupo también de 21 pacientes se les suministró placebo.

Los datos y_1, y_2, \dots, y_{21} , los cuales son presentados en la siguiente tabla, son tiempos de remisión de este segundo grupo (al cual se le suministró placebo) (1).

(1). Peter J. Smith, 2002, Analysis of Failure and Survival Data, Primera Edición, Pag. 6.

Grupo Placebo							
Tiempo de remisión de esteroides inducidos (semanas)							
1	1	2	2	3	4	4	5
5	8	8	8	8	11	11	12
12	15	17	22	23			

Tabla III .- Cuadro de resultados del experimento del Ejemplo 1.3.

A continuación se mostrará como estimar la probabilidad de que el tiempo de remisión para los pacientes de placebo sea superior a tres meses (que es lo mismo que 12 semanas).

Usando $S_{21}(12)$ para estimar $S(12)$, la verdadera probabilidad de sobrevivir más allá de 12 semanas, se nota que en la tabla existen cuatro observaciones que caen más allá de 12. Esto significa que cuatro de las funciones indicadoras $I_{(12,\infty)}(y_i)$, para $i = 1, 2, 3, \dots, 21$ tienen valor 1 y las otras 17 poseen valor 0.

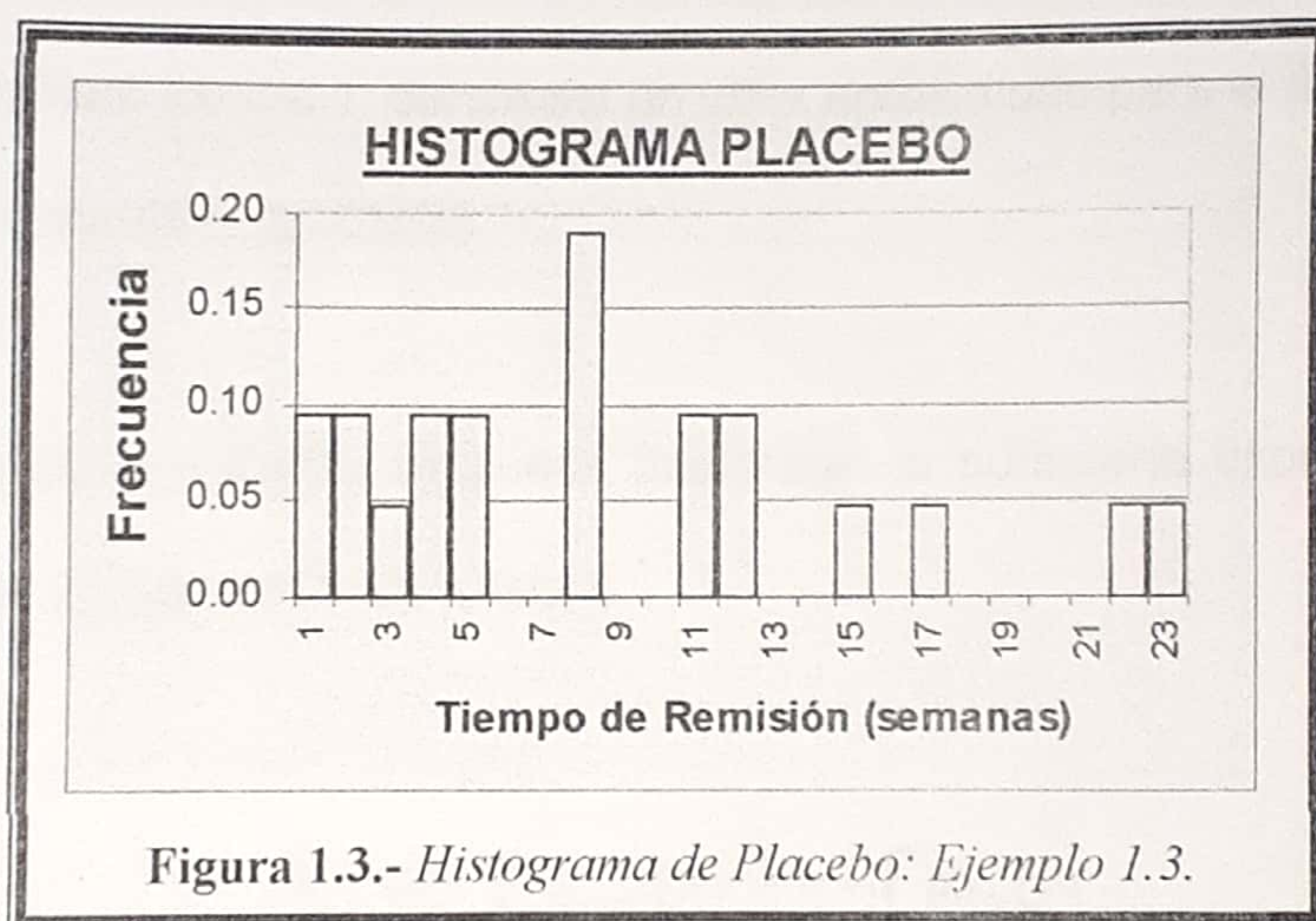
Así:

$$S_{21}(12) = \frac{1}{21} \sum I_{(12,\infty)}(y_i) = \frac{4}{21} \quad (1.7)$$

y un intervalo aproximado de dos desviaciones estándar para la verdadera probabilidad es:

$$S_{21}(12) \pm 2 \sqrt{\frac{S_{21}(12)[1 - S_{21}(12)]}{21}} = \frac{4}{21} \pm 2 \sqrt{\frac{4}{21} \left(1 - \frac{4}{21}\right)}$$

Esto se reduce a 0.19 ± 0.171 . La aparente carencia de precisión es causada por el relativamente pequeño tamaño de la muestra.



El cálculo realizado en la ecuación (1.7) muestra que $S_{21}(12)$ es hallada asignando a cada punto que cae más allá de 12 un valor de $\frac{1}{21}$, para después sumarlos y así obtener el valor del estimador.

Pero una vez construido el histograma para estos datos, con barras de un ancho igual a 1, ya que cada barra representa una semana, entonces $S_{21}(12)$ representaría el área total de las barras o cajas que se encuentran mas allá de 12 semanas.

Si se examina el histograma, existen cuatro cajas mayores a 12 y con una altura de aproximadamente 0.05 y con el ancho que se mencionó igual a 1, dando así un valor aproximado para el área de 0.2 (esto es $\approx 4(0.05 \times 1)$).

Usando el cálculo, se puede interpretar la sumatoria expresada para $S_{21}(12)$ como una integral:

$$S_{21}(12) = \sum I_{(12, \infty)}(y_i) \frac{1}{21} = - \int_{12}^{\infty} dS_{21}(y)$$

como una estimación de

$$- \int_{12}^{\infty} dS(y) = S(12)$$

1.3.3. Tiempo Promedio de Fallas.

El Teorema 1.1 indica que no es posible comparar variables aleatorias de supervivencia en base a la comparación de sus funciones de supervivencia, ya que gráficamente se puede notar que estas podrían cruzarse.

Esto es, si dos funciones de supervivencia S_1 y S_2 satisfacen $S_1(y) > S_2(y)$ para algún y , no necesariamente significa que $S_1(y) > S_2(y)$ para todo y .

Una forma de distinguir variables aleatorias de supervivencia es a través de sus medias.

Definición:

Si una variable aleatoria de supervivencia Y tiene valor esperado $E(Y)$, entonces el tiempo medio de fallas de Y , o *MTTF* (por sus siglas en inglés), está dado por:

$$MTTF = E(Y)$$

Esta definición muestra que se puede escribir al *MTTF* como la integral:

$$MTTF = \int_0^{\infty} uf(u)du$$

Aplicando la integración por partes de la forma $\int m dn = mn - \int ndm$, para la integral de la derecha de la ecuación, tomando como $n = u$ y $dm = f(u)du$, y utilizando el hecho de que $F(u) = 1 - S(u)$ y que $\int f(u)du = F(u)$; y valiéndonos además de las propiedades de las funciones de supervivencia $S(0) = 1$, $S(\infty) = 0$ al momento de evaluar los límites superior = ∞ e inferior = 0, se obtiene como resultado el *MTTF* en términos de la función de supervivencia como se expone en el Teorema 1.2.

Teorema 1.2.

$$MTTF = \int_0^{\infty} S(u)du$$

Por consiguiente, para encontrar un promedio de la duración de vida, se integra la función de supervivencia.

1.3.4. Funciones de Supervivencia Exponenciales.

Muchas aplicaciones en la práctica pueden ajustarse a funciones de supervivencia exponenciales definidas de la siguiente manera:

Si β es una constante, entonces la función S definida por:

$$S(y) = e^{-\frac{y}{\beta}}, \text{ para } y \geq 0$$

representa la función de supervivencia para alguna variable aleatoria Y , recalcando que sí cumple con las tres propiedades de estas funciones.

Integrando se tiene:

$$MTTF = \int_0^{\infty} S(u) du = \int_0^{\infty} e^{-\frac{u}{\beta}} du = \beta .$$

Para dos funciones semejantes de supervivencia $S_1(y) = e^{-\frac{y}{\beta_1}}$ y $S_2(y) = e^{-\frac{y}{\beta_2}}$, representando variables aleatorias con sus respectivos tiempos medios de falla $MTTF_1$ y $MTTF_2$, entonces:

$$\begin{aligned} S_1(y) > S_2(y) &\Leftrightarrow e^{-\frac{y}{\beta_1}} > e^{-\frac{y}{\beta_2}} \\ &\Leftrightarrow -\frac{y}{\beta_1} > -\frac{y}{\beta_2} \\ &\Leftrightarrow \beta_1 > \beta_2 \\ &\Leftrightarrow MTTF_1 > MTTF_2 \end{aligned}$$

1.3.5. Con Respecto a la Función de Riesgo.

La función de confiabilidad examina la probabilidad de que las fallas/muertes (de personas, de unidades experimentales, de sistemas computarizados, etc.) ocurran luego de un instante dado en el tiempo.

Para controlar el tiempo de vida de un componente mediante el apoyo de la distribución del tiempo de vida, se usa la **Función de Riesgo**.

La probabilidad de que un sistema falle ahora, esto es, en $(y, y + \Delta y)$, dado que este ha sobrevivido hasta ahora, está dada por:

$$P(y < Y < y + \Delta y | Y > y) = \frac{P(y < Y < y + \Delta y, Y > y)}{P(Y > y)} = \frac{P(y < Y < y + \Delta y)}{P(Y > y)}$$

Promediando esta sobre la longitud del intervalo de tiempo actual Δy da como resultado la **Tasa Promedio de Fallas** a lo ancho del intervalo:

$$\frac{P(y < Y < y + \Delta y)}{(\Delta y)P(Y > y)} \quad (1.8)$$

Tomando el límite cuando Δy tiende a cero, ($\Delta y \rightarrow 0$), la tasa promedio de fallas se transforma en una tasa instantánea de fallas

$$\lim_{\Delta y \rightarrow 0} \frac{P(y < Y < y + \Delta y)}{(\Delta y)P(Y > y)} = \lim_{\Delta y \rightarrow 0} \frac{F(y + \Delta y) - F(y)}{(\Delta y)S(y)}$$

ya que el numerador de la expresión de la izquierda es equivalente

a:

$$\begin{aligned} P(Y > y) \wedge P(Y < y + \Delta y) &= [1 - P(Y < y)][P(Y < y + \Delta y)] \\ &= P(Y < y + \Delta y) - P(Y < y + \Delta y)P(Y < y) \\ &= P(Y < y + \Delta y) - P(Y < y + \Delta y \wedge Y < y) \\ &= P(Y < y + \Delta y) - P(Y < y) \\ &= F(y + \Delta y) - F(y) \end{aligned}$$

Además se tiene que el término $\lim_{\Delta y \rightarrow 0} \frac{F(y + \Delta y) - F(y)}{\Delta y}$ de la

expresión final resultante representa la derivada de la Función de

Distribución de Probabilidad, por lo tanto se tiene que:

$$\lim_{\Delta y \rightarrow 0} \frac{F(y + \Delta y) - F(y)}{(\Delta y)S(y)} = \frac{F'(y)}{S(y)} = \frac{f(y)}{S(y)}$$

Esto motiva la siguiente definición:

Definición:

Una variable aleatoria de supervivencia Y tiene una Función de Riesgo, o Tasa de Riesgo o Fuerza de Mortalidad, h , para $y > 0$ definida por:

$$h(y) = \frac{f(y)}{S(y)}$$

Esta es la tasa instantánea de muerte (o falla) al tiempo y , dado que la supervivencia individual es mayor al tiempo y .

Las funciones de Riesgo rastrean como la tasa de fallas cambia con el tiempo.

La función de Supervivencia está expresada en términos de la función de Riesgo como lo indica el siguiente teorema:

Teorema 1.3.

$$S(y) = e^{-\int_0^y h(u) du}$$

Prueba:

A partir del hecho de que $f(u) = -S'(u)$, esto quiere decir que

$h(u) = -\frac{d}{du} \ln S(u)$, e integrando cada parte de esta ecuación entre

0 y y se tiene:

$$\ln S(u) \Big|_0^y = -\int_0^y h(u) du$$

Entonces resolviendo se obtiene,

$$\ln S(y) - \ln S(0) = -\int_0^y h(u) du$$

pero $S(0) = 1 \Rightarrow \ln S(0) = 0$, y finalmente se despeja $S(y)$, obteniéndose así la afirmación del teorema.

1.3.6. Esperanza de Vida a la Edad t .

Suponiendo que se ha rastreado el tiempo de vida de un paciente, de un diente de engranaje de una escalera eléctrica, de un software informático o de algún dispositivo en general; para una particular edad t . Esto es, la unidad experimental está aún viva y bajo observación al tiempo t . La pregunta de ahora es: ¿Cuánta vida le queda en promedio?

Si el tiempo de supervivencia es denotado por la variable aleatoria Y , y se conoce que la constante t es tal que $Y > t$, entonces la diferencia $Y - t$ es llamada Tiempo de vida Residual a la edad t .

Definición:

Para $t > 0$, el promedio del tiempo de vida Residual a la edad t , $r(t)$, está dado por:

$$r(t) = E(Y - t | Y > t).$$

Nótese que t es constante, si se evalúa el valor esperado de la definición anterior, se obtiene:

$$\begin{aligned} r(t) &= E(Y - t | Y > t) \\ &= \int_t^{\infty} (y - t) dP(Y \leq y | Y > t) \\ &= \int_t^{\infty} (y - t) d\left(\frac{S(t) - S(y)}{S(t)}\right) \\ &= \int_t^{\infty} (y - t) \left(\frac{-S'(y) dy}{S(t)}\right) \\ &= \frac{-(y - t)S(y)|_t^{\infty} + \int_t^{\infty} S(y) dy}{S(t)} \\ &= \frac{\int_t^{\infty} S(y) dy}{S(t)}. \end{aligned}$$

Definición:

En esta derivación se ha usado el resultado de que $yS(y) \rightarrow 0$ cuando $y \rightarrow \infty$.

Se ha encontrado que el promedio del tiempo de vida residual tiene la expresión integral:

$$r(t) = \frac{\int_t^{\infty} S(y)dy}{S(t)} \quad (1.9)$$

Si se evalúa esta expresión al tiempo $t = 0$, se obtiene, usando el Teorema 1.2,

$$r(0) = \frac{\int_0^{\infty} S(y)dy}{S(0)} = \int_0^{\infty} S(y)dy = E(y) = MTTF. \quad (1.10)$$

La esperanza de vida a una edad t dada es una medida de duración del tiempo de vida, dado que la edad t ya ha sido cumplida. Formalmente:

Definición:

Para una variable aleatoria de supervivencia Y y un tiempo fijo t , el promedio de vida esperado a la edad t está dado por $E(Y|Y > t)$.

Por (1.9), el promedio de vida esperado a la edad t está relacionado con el tiempo de vida residual adicionándole el tiempo t ya vivido. Esto significa que,

$$E(Y|Y > t) = t + r(t) = t + \frac{\int_t^{\infty} S(y)dy}{S(t)} \quad (1.11)$$

Ejemplo 1.4.

Comparación entre la esperanza de vida humana con la de una bombilla de luz.

A continuación se presentan dos tablas de datos, donde la tabla de la izquierda muestra la esperanza de vida a la edad t para hombres y mujeres, junto con varios valores de t reportados en "The Age", Melbourne, 4 de Marzo de 1991, de información suministrada por la Agencia Australiana de Estadísticas(1).

(1). Peter J. Smith, 2002, Analysis of Failure and Survival Data, Primera Edición, Pag. 11.

Esperanza de vida			Esperanza de vida	
Edad Actual	Masculino	Femenino	Edad Actual	Bombilla de luz
t	(años)	(años)	t	
0	73.3	79.5	0	73.3
10	74.2	80.3	10	83.3
20	74.7	80.5	20	93.3
30	75.2	80.7	30	103.3
40	75.9	81	40	113.3
50	76.7	81.6	50	123.3
60	78.3	82.8	60	133.3
70	81.5	84.8	70	143.3
80	86.5	88.4	80	153.3
90	93.5	94.1	90	163.3

Tabla IV .- *Tabla de datos del Ejemplo 1.4.*

Estos datos muestran que las distribuciones de supervivencia para hombres y mujeres son ligeramente diferentes. Cuando nacemos, el valor esperado de la variable longitud de vida, X , es 73.3 años para hombres y 79.5 años para mujeres. A medida que se vive a la edad t , (la columna de la izquierda), la esperanza de vida futura más allá de t se reduce. Esto es, una esperanza de vida para un hombre a la edad 10 es solo $74.2 - 10 = 64.2$ años más. Eso es menos que el promedio de tiempo de vida de 73.3 años cuando se está en la edad cero. Esto se mantiene declinando: la esperanza de vida a la edad 20 es $74.7 - 20 = 54.7$ años.

Por otro lado, las esperanzas de vida descritas en la tabla de datos de la derecha, son para una bombilla hipotética con tiempo de supervivencia variable aleatorio $X \sim \exp(73.3)$.

Cuando dicha bombilla se enciende, esta tiene un valor esperado de vida de 73.3 unidades de tiempo (obviamente no son años). Sin embargo, partiendo de que la distribución del tiempo de vida es exponencial, quiere decir que existe la falta de memoria. Es decir que la esperanza de vida de la bombilla de luz a la edad de 10 unidades es $83.3 - 10 = 73.3$ unidades más, que es la misma esperanza desde que se la encendió por primera vez.

Se puede observar que la distribución de supervivencia de la bombilla es la misma para cualquier período de tiempo, lo que no sucede con los humanos.

El promedio residual de la esperanza de vida $r(t)$ a la edad t está dada en la siguiente tabla tanto para hombres como para bombillas de luz. Este promedio se lo obtiene sustrayendo t de la esperanza de vida a la edad t .

Esperanza de vida futura a la edad t		
Edad Actual	Masculino	Bombilla
t	(años)	(u de tiempo)
0	73.3	73.3
10	64.2	73.3
20	54.7	73.3
30	45.2	73.3
40	35.9	73.3
50	26.7	73.3
60	18.3	73.3
70	11.5	73.3
80	6.5	73.3
90	3.5	73.3

Tabla V.- Promedio residual de la esperanza de vida a la edad t para los datos del Ejemplo 1.4.

Estos resultados demuestran que para humanos la esperanza futura de supervivencia a la edad t se reduce a medida que t incrementa, pero típicamente para bombillas de luz y ciertos tipos de componentes electrónicos esto no sucede.

Existe una expresión alternativa interesante para la ecuación (1.9), partiendo de que:

$$r(t) = \frac{\int_t^{\infty} S(u) du}{S(t)}$$

Entonces, teniendo en cuenta que $S(\infty) = 0$, se puede encontrar que:

$$\begin{aligned} -\frac{1}{r(t)} &= \frac{-S(t)}{\int_t^{\infty} S(u)du} \\ &= \frac{\frac{d}{dt} \left[\int_t^{\infty} S(u)du \right]}{\int_t^{\infty} S(u)du} \\ &= \frac{d}{dt} \left[\ln \int_t^{\infty} S(u)du \right]. \end{aligned} \quad (1.12)$$

El siguiente teorema indica que la función de supervivencia puede ser expresada en términos del promedio residual del tiempo de vida:

Teorema 1.4.

$$S(y) = \frac{r(0)}{r(y)} e^{-\int_0^y \frac{dt}{r(t)}}$$

Prueba:

Integrando cada lado de la ecuación final de (1.12) entre 0 y y se obtiene:

$$\begin{aligned}
-\int_0^y \frac{dt}{r(t)} &= \int_0^y d \left[\ln \int_t^\infty S(u) du \right] \\
&= \ln \int_y^\infty S(u) du - \ln \int_0^\infty S(u) du \\
&= \ln \int_y^\infty S(u) du - \ln r(0) \\
&= \ln \left[\frac{\int_y^\infty S(u) du}{r(0)} \right].
\end{aligned}$$

Exponenciando cada lado y usando la ecuación (1.9) dada:

$$e^{-\int_0^y \frac{dt}{r(t)}} = \frac{\int_y^\infty S(u) du}{r(0)} = \frac{r(y)S(y)}{r(0)}.$$

Finalmente, se demuestra la igualdad establecida en el teorema al multiplicar la ecuación por el factor $\frac{r(0)}{r(y)}$.

CAPITULO 2

2. MODELOS, CENSURAS Y VEROSIMILITUD PARA DATOS DE TIEMPOS DE FALLA

En el presente capítulo se expondrán algunos de los modelos empleados para los análisis de supervivencia y confiabilidad. Estos modelos poseen la característica de ser paramétricos, por tanto se basan en distribuciones teóricas dadas. Entre los modelos ha ser presentados se encuentran modelos para la obtención de la función de distribución acumulada de las fallas y para la obtención de la función de supervivencia de los datos, y finalmente modelos de riesgo.

2.1. Modelos para Procesos Continuos de Tiempos de Falla.

Como se explicó en el capítulo anterior, la medida más ampliamente usada para la confiabilidad de un producto es su distribución de tiempos de falla y los datos de confiabilidad que se

recojan contendrán información de los tiempos de falla de muestras de materiales, componentes o de sistemas completos. En este capítulo se presentarán modelos básicos para estos datos. Muchos procesos de tiempos de falla son modelados sobre una escala continua.

El símbolo T será utilizado para denotar una variable aleatoria continua, no negativa, describiendo *el tiempo de falla de una unidad o sistema*.

Cabe recalcar que a la variable T se la denomina también *variable tiempo de vida de una unidad o sistema*, ya que esta mide el tiempo de duración o de vida antes de que dicha unidad o sistema falle.

2.1.1. Funciones de Distribución de Tiempos de Falla.

La distribución de probabilidad para el tiempo de falla T puede ser caracterizado por una función de distribución acumulada, una función de densidad de probabilidad, una función de supervivencia, o por una función de riesgo.

La elección de una de estas funciones depende de la conveniencia de la especificación del modelo, interpretación o desarrollo técnico.

Función de Distribución Acumulada (fda).-

La función de distribución acumulada de T , $F(t) = P(T \leq t)$, devuelve la probabilidad de que una unidad falle antes del tiempo t . Alternativamente, $F(t)$ puede ser interpretada como la proporción de unidades en la población que fallarán antes del tiempo t .

Función de Densidad de Probabilidad (fdp).-

La función de densidad de probabilidad para una variable aleatoria continua T está definida como la derivada de $F(t)$ con respecto a t . $f(t) = dF(t)/dt$. La función de densidad de probabilidad puede ser usada para representar frecuencias relativas de fallas como una función del tiempo.

Función de Supervivencia (fs).-

La función de supervivencia, también conocida como función de confiabilidad, como se definió anteriormente es el complemento de

la función de distribución acumulada $S(t) = 1 - F(t) = \int_t^{\infty} f(x)dx$, y devuelve la probabilidad de supervivencia más allá del tiempo t .

Función de Riesgo (fr).-

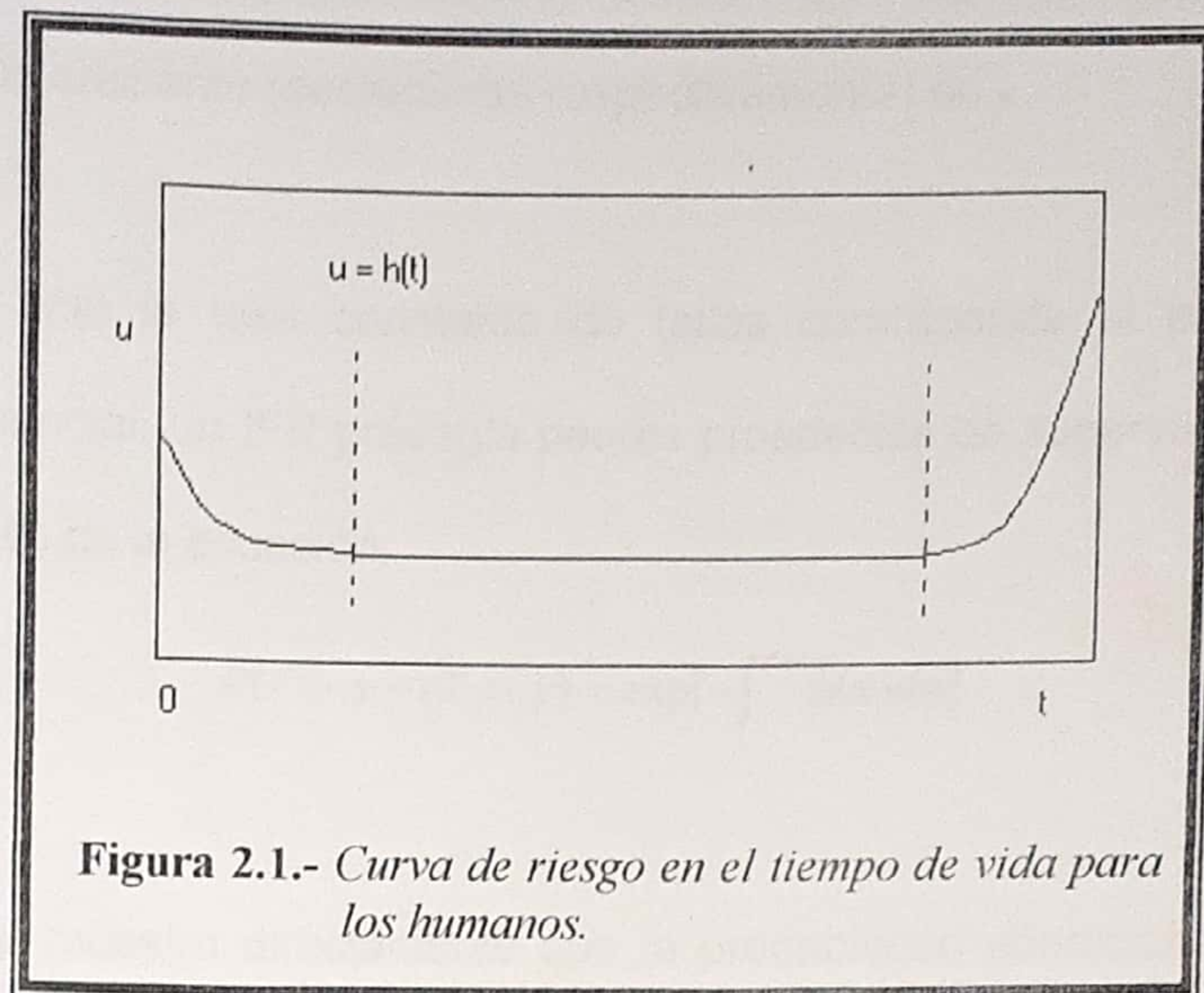
La función de riesgo, también conocida como tasa de riesgo, función de la tasa instantánea de fallas entre otros nombres, está definida por:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

La función de riesgo como se mencionó anteriormente representa la propensión de falla en el siguiente pequeño intervalo de tiempo (Δt), dado la supervivencia al tiempo t , en términos actuariales la función de riesgo representa la Fuerza de Mortalidad.

Tasa de Fallas Creciente (IFR).-

El proceso de envejecimiento es medido por la función de riesgo. Por ejemplo, los humanos poseen un inminente riesgo de muerte en gran medida creciente a medida que estos alcanzan edades mayores y esto corresponde a una función de riesgo creciente.



La figura muestra el gráfico de un riesgo característico del tiempo de vida en los humanos, el cual tiene una forma de tina de baño. Para los recién nacidos el riesgo inminente de muerte disminuye con el tiempo, cuando alcanzan edades medias el riesgo es relativamente constante; finalmente la tasa de riesgo aumenta durante las edades de madurez o vejez.

Definición:

Una variable aleatoria Y tiene una **tasa de fallas creciente** o **IFR**, por sus siglas en inglés; (respectivamente **tasa de fallas**

decreciente o **DFR**) si $h(y)$, la función de riesgo de Y , es una función creciente (decreciente respectivamente) de y .

Dado que la tasa constante de fallas corresponde al modelo exponencial, un IFR presagia peores prospectos de supervivencia a través de la ecuación:

$$P(Y > y + t | Y > y) = \exp\left[-\int_y^{y+t} h(u) du\right],$$

la cual muestra directamente que la probabilidad condicional de supervivencia para la edad y , decrece como una función de y para cada $t > 0$ cuando $h(y)$ es creciente.

Los humanos en la época de infancia experimentan una DFR debido al obstáculo inicial de nacimiento. La mortalidad infantil decrece con el tiempo durante el período de inmadurez. Similarmente, el proceso de manufactura de algún producto puede experimentar un período de salvedad. La adultez es caracterizada por un razonable riesgo constante, el cual se incrementa luego de la edad de 40 años (la fase de "desgaste o destrucción"). Todo esto crea en conjunto un *riesgo en forma de Tina de Baño*, como se aprecia en la figura, a lo largo de la distribución del tiempo de vida.

Función de Riesgo Acumulada.-

Para algunos propósitos resulta de gran ayuda recurrir a la Función de Riesgo Acumulada, definida por:

$$H(t) = \int_0^t h(x) dx$$

La función de distribución acumulada o función de supervivencia para T puede ser obtenida de la función de riesgo. Por ejemplo, para alguna distribución continua

$$F(t) = 1 - \exp[-H(t)] = 1 - \exp\left[-\int_0^t h(x) dx\right]$$

Adicionalmente, en el capítulo anterior se descubrió que las distribuciones de supervivencia pueden ser equivalentemente especificadas por: f , la función de densidad de probabilidad; F , la función de distribución acumulada; S , la función de Supervivencia; h , la función de riesgo, o r , el promedio residual del tiempo de vida.

Esto significa que si una de estas representaciones es especificada, entonces las otras pueden obtenerse a partir de esta.

Por consiguiente la discusión acerca de las distribuciones de tiempos de vida sólo necesitan concentrarse en una de estas expresiones.

Las funciones concernientes al riesgo juegan un papel muy importante dentro de la supervivencia y confiabilidad de unidades o sistemas, brindando mucha información al momento de realizar un análisis de tiempos de vida, es por esto que junto con modelos que involucran a la Función de Distribución Acumulada para la misma variable (tiempos de falla T), serán también tomados en cuenta algunos modelos de riesgo.

2.1.2. Modelos de Riesgo.

A menudo, bajo estas alternativas, la función de riesgo es la modelada y especificada debido a su interpretación directa como inminente riesgo.

Esto puede ayudar incluso a identificar, más efectivamente que la función de supervivencia, el mecanismo mediante el cual ocurren las fallas.

Los investigadores probablemente pueden tener una noción o conocimiento previo acerca de cómo el riesgo inminente cambia con el tiempo para los tiempos de vida. Por ejemplo las bombillas de luz tienden a detenerse inesperadamente y mantienen su esperanza de vida constante, mientras que los humanos envejecen y se van "destruyendo" reduciendo así su esperanza de vida futura a medida que pasa el tiempo.

Se esperaría que las formas de las funciones de riesgo de las bombillas y de las personas sean diferentes.

Por tanto la experiencia de la gente incrementa el riesgo, mientras que las bombillas tienden a mostrar un riesgo constante.

2.1.2.1. Modelo de Riesgo Constante.

Suponiendo que $\beta > 0$ es una constante y $\omega = 1/\beta$. Un **modelo de riesgo constante** es usualmente planteado donde un riesgo inminente de fallas no cambia con el tiempo.

Modelo de Riesgo Constante:

$$h(t) = \omega \text{ para } y > 0. \quad (2.1)$$

Por el Teorema 1.3, la correspondiente función de supervivencia para $t > 0$ es:

$$\begin{aligned} S(t) &= e^{-\int_0^t h(u) du} \\ &= e^{-\int_0^t \frac{1}{\beta} du} \\ &= e^{-\frac{1}{\beta} \int_0^t 1 du} \\ &= e^{-\frac{t}{\beta}} \end{aligned}$$

con función de densidad de probabilidad $f(t) = -S'(t) = \frac{1}{\beta} e^{-\frac{t}{\beta}}$, media $E(T) = \beta$ y varianza $Var(T) = \beta^2$. Esto caracteriza el modelo de probabilidad exponencial.

Definición:

La variable tiempo de falla T sigue un modelo de probabilidad exponencial con media $\beta > 0$ y se escribe $T \sim \exp(\beta)$ cuando el riesgo es constante con $h(t) = \frac{1}{\beta}$.

Este es uno de los modelos de probabilidades más comúnmente usados para simular tiempos de vida de componentes; el modelo exponencial brinda a menudo un buen ajuste a los datos de supervivencia.

Una posible explicación es que los tiempos entre eventos en un proceso de Poisson son distribuidos exponencialmente. Suponiendo que se esté modelando el tiempo de vida de un sistema. Pueden presentarse eventos, tales como shocks tóxicos para el sistema, que ocurren acorde a un proceso de Poisson con tasa ω por unidad de tiempo. Entonces X_t , el número de shocks ocurridos en un intervalo de longitud y satisface $X_t \sim \text{Poisson}(\omega t)$. Si el primero de estos shocks es fatal (el sistema falla), entonces T , el tiempo de falla del sistema, satisface

$$S(t) = P(T > t) = P(X_t = 0) = \frac{(\omega t)^0 e^{-\omega t}}{0!} = e^{-\omega t},$$

la cual es la función de supervivencia exponencial para una variable aleatoria exponencial con media $\beta = \frac{1}{\omega}$.

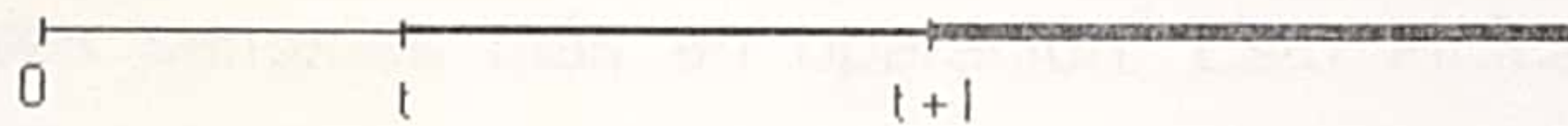
Una característica importante de la distribución exponencial es la "*propiedad de pérdida de memoria*": la distribución futura del tiempo de vida de un componente no depende del pasado, sino que esta depende solo del presente. Esto es, alcanzando alguna edad determinada, la probabilidad de sobrevivir t unidades más de tiempo es la misma que cuando esta tenía la edad cero.

Teorema 2.1. Si $T \sim \exp(\beta)$, entonces para algún $t > 0$ y $l > 0$, se cumple que:

$$P(T > t+l | T > t) = P(T > l).$$

Prueba:

La probabilidad $P(T > t+l | T > t)$ puede ser interpretada como una probabilidad condicional $P(A|B) = P(A \cap B)P(B)^{-1}$, donde los eventos serían: $A = \{T > t+l\}$ y $B = \{T > t\}$. Por tanto dado que $l > 0$, la intersección de estos dos eventos sería simplemente el conjunto de valores del evento A .



En el gráfico esta intersección se la representa como el área doblemente sombreada, se puede notar que el conjunto intersección resultante está formado por $A = \{T > t+l\}$

Finalmente, se tiene que $P(T > r) = e^{-\frac{r}{\beta}}$ para algún valor positivo de

r .

$$\begin{aligned}
 P(T > t+l | T > t) &= \frac{P(\{T > t+l\} \cap \{T > t\})}{P(\{T > t\})} \\
 &= \frac{P(\{T > t+l\})}{P(\{T > t\})} \\
 &= \frac{e^{-\frac{t+l}{\beta}}}{e^{-\frac{t}{\beta}}} \\
 &= e^{-\frac{l}{\beta}} \\
 &= S(l).
 \end{aligned}$$

El Teorema 2.1 afirma que para un componente con un tiempo de vida distribuido exponencialmente, la probabilidad de que un componente con 3 meses de uso dure tres semanas más en operación, es la misma probabilidad de que uno con 6 meses de uso dure tres semanas más en operación. Esto indica que el tiempo de vida del componente no pasa a través de un período de envejecimiento, en el cual existe un incremento del riesgo de mortalidad.

En este sentido un tiempo de vida exponencial es diferente al tiempo de vida de un humano, donde la distribución de supervivencia se reduce conforme aumenta la edad.

Los datos del ejemplo 2.1. tienen una media actual de 107.47 y una desviación estándar de 106.13 (1), reflejando que la media y la desviación estándar son el mismo valor para el modelo exponencial de tiempo de vida.

Sin embargo en esta sección, esto resulta importante para notar la forma general del esquema, su extremado sesgo hacia la derecha y la presencia de valores altos aberrantes, que exceden 5 veces la media.

La columna de la derecha junto al esquema, muestra estimaciones simples de la función de riesgo basada en el conteo de los datos. Usando una estimación intuitiva del riesgo basada en la fórmula

$$\frac{P(t < T < t + \Delta t)}{(\Delta t)P(T > t)}$$

para la tasa promedio de fallas a lo ancho de intervalos de tamaño $\Delta t = 20$, se obtiene, por ejemplo, en la primera línea, un riesgo estimado de:

(1). Peter J. Smith, 2002, Analysis of Failure and Survival Data, Primera Edición, Pag. 21.

$$\frac{\frac{31}{200}}{20 \frac{200}{200}} = 0.00775$$

Hay 31 observaciones en la primera línea del esquema y 200 observaciones caen dentro o más allá de la primera línea del esquema. Por consiguiente, para $t = 0$, se ha tomado $\frac{31}{200}$ como un

estimador de $P(t < T < t + \Delta t)$ y $\frac{200}{200}$ como un estimador $P(T > t)$.

Los cálculos son similares para las otras líneas del diagrama. Por ejemplo en la segunda línea del diagrama, el riesgo estimado es:

$$\frac{\frac{30}{200}}{20 \frac{169}{200}} = 0.00888.$$

Esta es una elemental forma de estimación de la función de riesgo a partir de un diagrama de Tallos-Hojas.

Nótese que todas las estimaciones de la función de riesgo en la columna de la derecha están cercanamente a 0.01(para dos

lugares decimales), sugiriendo una función de riesgo $h(t) = \frac{1}{\beta}$ para

$\beta = 100$.

El parámetro β es muy importantemente un parámetro de escala.

Con esto se puede establecer el siguiente teorema:

Teorema 2.2. Si $T \sim \exp(\beta)$, entonces $\frac{T}{\beta} \sim \exp(1)$.

En particular este teorema sugiere que si se multiplica o divide tiempos de vida exponenciales por una constante, entonces el *MTTF* correspondiente se lo multiplica o divide por la misma constante.

2.1.2.2. Modelo de Riesgo Potencia.

Suponiendo que $\alpha > 0$ y $\beta > 0$ son constantes. Un **modelo de riesgo potencia** es planteado usualmente con un riesgo inminente de fallas que se incrementa rápidamente con el tiempo.

$$\text{Modelo de Riesgo Potencia: } h(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1} \text{ para } t > 0. \quad (2.2)$$

Nótese que la forma de la ecuación (2.2) es simplemente $h(t) = (\text{constante}) t^{\text{constante}}$; sin embargo la elección de los parámetros

de este modelo en términos de α y β se la realiza claramente cuando la función de supervivencia está determinada. Por el Teorema 1.3, la correspondiente función de supervivencia es, para $t > 0$,

$$\begin{aligned} S(t) &= e^{-\int_0^t h(u) du} \\ &= e^{-\int_0^t \frac{\alpha}{\beta^\alpha} u^{\alpha-1} du} \\ &= e^{-\frac{1}{\beta^\alpha} \int_0^t \alpha u^{\alpha-1} du} \\ &= e^{-\left(\frac{t}{\beta}\right)^\alpha} \end{aligned}$$

Por consiguiente la función de supervivencia para una variable con modelo de riesgo potencia tiene una forma simple particular donde la potencia en sí es trasladada al exponente:

$$S(t) = e^{-\left(\frac{t}{\beta}\right)^\alpha}$$

La función de densidad de probabilidad $f(t) = -S'(t)$ es

$$f(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha}, \quad t > 0$$

Definición:

La variable aleatoria tiempo de vida o tiempo de falla T sigue un modelo de probabilidad de Weibull con parámetros $\alpha > 0$ y $\beta > 0$, y se escribe $T \sim Weibull(\alpha, \beta)$ cuando T tiene una potencia de

$$\text{riesgo de la forma } h(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1}.$$

Como en el caso exponencial, el parámetro β es un parámetro de escala. El exponente de la función de supervivencia contiene el resultado de t dividido para β . Con esto se puede establecer lo que sostiene el siguiente teorema:

Teorema 2.3. Si $T \sim Weibull(\alpha, \beta)$, entonces $\frac{T}{\beta} \sim Weibull(\alpha, 1)$.

El modelo de Weibull tiene una función de riesgo muy simple como también una forma simple para la función de supervivencia. Estas junto con sus dos parámetros flexibles, hacen al modelo muy útil en algunos contextos de ingeniería. Sin embargo, la media y varianza de la distribución resultan un tanto más difícil de determinar. Para esto se necesita la función Gamma.

Definición:

La **Función Gamma**, $\Gamma(\alpha)$, está definida, para todo $\alpha > 0$, por la integral

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

La función gamma cuenta con algunas propiedades interesantes y características especiales, las cuales son bosquejadas en el siguiente teorema:

Teorema 2.4. Propiedades de la Función Gamma

1. $\Gamma(1) = 1$
2. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
3. $\Gamma(x+1) = x\Gamma(x)$, para x que sea número real positivo.
4. $\Gamma(n+1) = n!$, para $n = 1, 2, 3, \dots$, entero positivo.

La función gamma es usada para determinar la media y varianza del modelo de probabilidad de Weibull.

Primero, se saca provecho de un cambio de escala: usando la distribución de $X \sim Weibull(\alpha, 1)$, los momentos alrededor de cero están dados por

$$\begin{aligned} E(X^r) &= \int_0^{\infty} x^r \alpha x^{\alpha-1} e^{-x^\alpha} dx \\ &= \int_0^{\infty} u^{\frac{r}{\alpha}} e^{-u} du \quad (\text{sustitución } u = x^\alpha) \\ &= \Gamma\left(1 + \frac{r}{\alpha}\right). \end{aligned}$$

Ahora si $T = \beta X \sim Weibull(\alpha, \beta)$, entonces el r -ésimo momento alrededor de cero está dado por:

$$E(T^r) = E(\beta^r X^r) = \beta^r E(X^r) = \beta^r \Gamma\left(1 + \frac{r}{\alpha}\right)$$

El primer y segundo momento alrededor de cero son usados para determinar la media $E(T)$ y la varianza $E(T^2) - E(T)^2$ de T .

Teorema 2.5. Si $T \sim Weibull(\alpha, \beta)$, entonces $E(T) = \beta \Gamma\left(1 + \frac{1}{\alpha}\right)$ y

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

Cabe recalcar que existen modelos que pueden ser construidos a partir de distintos valores del parámetro α .

Por ejemplo cuando $\alpha = 1$, el modelo de Weibull se reduce a un modelo exponencial con riesgo constante; esto es $Weibull(1, \beta) \equiv \exp(\beta)$. Si $\alpha > 1$, entonces la función de riesgo de Weibull es creciente. Similarmente si $\alpha < 1$ entonces la función de riesgo de Weibull es decreciente. Esto hace a la función de Weibull un modelo muy flexible en una amplia variedad de situaciones: Riesgos crecientes, riesgos decrecientes y riesgos constantes.

2.1.2.3. Modelos de Riesgo Polinomiales.

Cuando los modelos de potencia son sumados, se tienen los denominados modelos polinomiales, suponiendo que a_0, a_1, \dots, a_k son constantes.

Modelo de Riesgo de Rayleigh:

$$h(t) = a_0 + a_1 y + a_2 y^2 + \dots + a_k y^k \quad (2.3)$$

Tales riesgos han sido modelados útilmente para $k = 1$. Para $k = 2$, el coeficiente podría ser elegido de tal manera que $h(t)$ siga una curva cuadrática o una forma de tina de baño (con $h(t) > 0$ para todo $t > 0$) como el bosquejo de la Figura 2.1.

2.1.2.4. Riesgos Exponenciales.

Anteriormente se discutió formas simples para la función de riesgo, incluyendo constantes, potenciales y lineales. Los riesgos exponenciales proveen de una extensión natural, especialmente dado que la función de supervivencia aún permanece fácil de calcular con una fórmula explícita.

Suponiendo que $b > 0$ y $-\infty < u < \infty$ son constantes.

Modelo de Riesgo Exponencial:

$$h(t) = \frac{1}{b} \exp\left(\frac{t-u}{b}\right) \text{ para } 0 < t < \infty. \quad (2.4)$$

Nótese que la forma de (2.4) es simplemente $h(t) = (\text{constante})e^{\text{constante}}$. Un modelo de riesgo exponencial ocurre frecuentemente en la ciencia actuarial para modelar tiempos de vida en humanos. Por el Teorema 1.3, la función de supervivencia es, para $t > 0$,

$$\begin{aligned} S(y) &= e^{-\int_0^t h(y) dy} \\ &= \exp\left[-\int_0^t \frac{1}{b} e^{\frac{y-u}{b}} dy\right] \\ &= \exp\left[-e^{\frac{y-u}{b}} \Big|_0^t\right] \\ &= \exp\left[-\exp\left(\frac{t-u}{b}\right) + \exp\left(-\frac{u}{b}\right)\right] \end{aligned}$$

Si esta función de supervivencia es cuidadosamente reparametrizada, esta puede ser expresada en la forma:

$$S(t) = \exp[\theta(1 - e^{\alpha t})],$$

conocida como *Modelo de Supervivencia de Gompertz*. Las relaciones entre los parámetros (α, θ) y (u, b) son:

$$\alpha = \frac{1}{b} \quad \text{y} \quad \theta = e^{-\frac{u}{b}}$$

Nótese que la función de densidad de probabilidad está dada por

$$f(t) = -S'(t) = \frac{1}{b} \exp\left[\frac{t-u}{b} - \exp\left(\frac{t-u}{b}\right)\right] = \frac{1}{b} \exp\left(\frac{t-u}{b}\right) S(t).$$

Definición:

La variable aleatoria de tiempo de vida Y sigue un **modelo de probabilidad de Gompertz** con parámetros $b > 0$ y $-\infty < u < \infty$,

y se puede escribir $T \sim Gompertz(u, b)$ cuando T tiene un riesgo

exponencial de la forma $h(t) = \frac{1}{b} \exp\left(\frac{t-u}{b}\right)$ para $0 < t < \infty$.

Una vez que se han tomado en cuenta algunos modelos de riesgo interesantes y aplicables, es preciso tomar en cuenta también algunos modelos los cuales están propuestos directamente con la función de distribución acumulada y que al igual que los modelos de riesgo, brindan mucha información en el análisis de fallas o de supervivencia.

2.2. Modelos para Datos Discretos de un Proceso Continuo.

Muchos procesos de tiempos de falla son modelados sobre una escala continua.

Debido a las inherentes limitaciones en la precisión de mediciones, sin embargo, los datos de tiempo de falla son siempre discretos.

Las limitaciones en la observación o en la habilidad de los instrumentos de medida, pueden causar el surgimiento de datos censurados o truncados.

A continuación se presentarán estructuras generales para modelar estos datos, luego se describirá diferentes tipos de observaciones que surgen en análisis de confiabilidad de datos y finalmente se mostrará como calcular la verosimilitud (o "probabilidad de los datos").

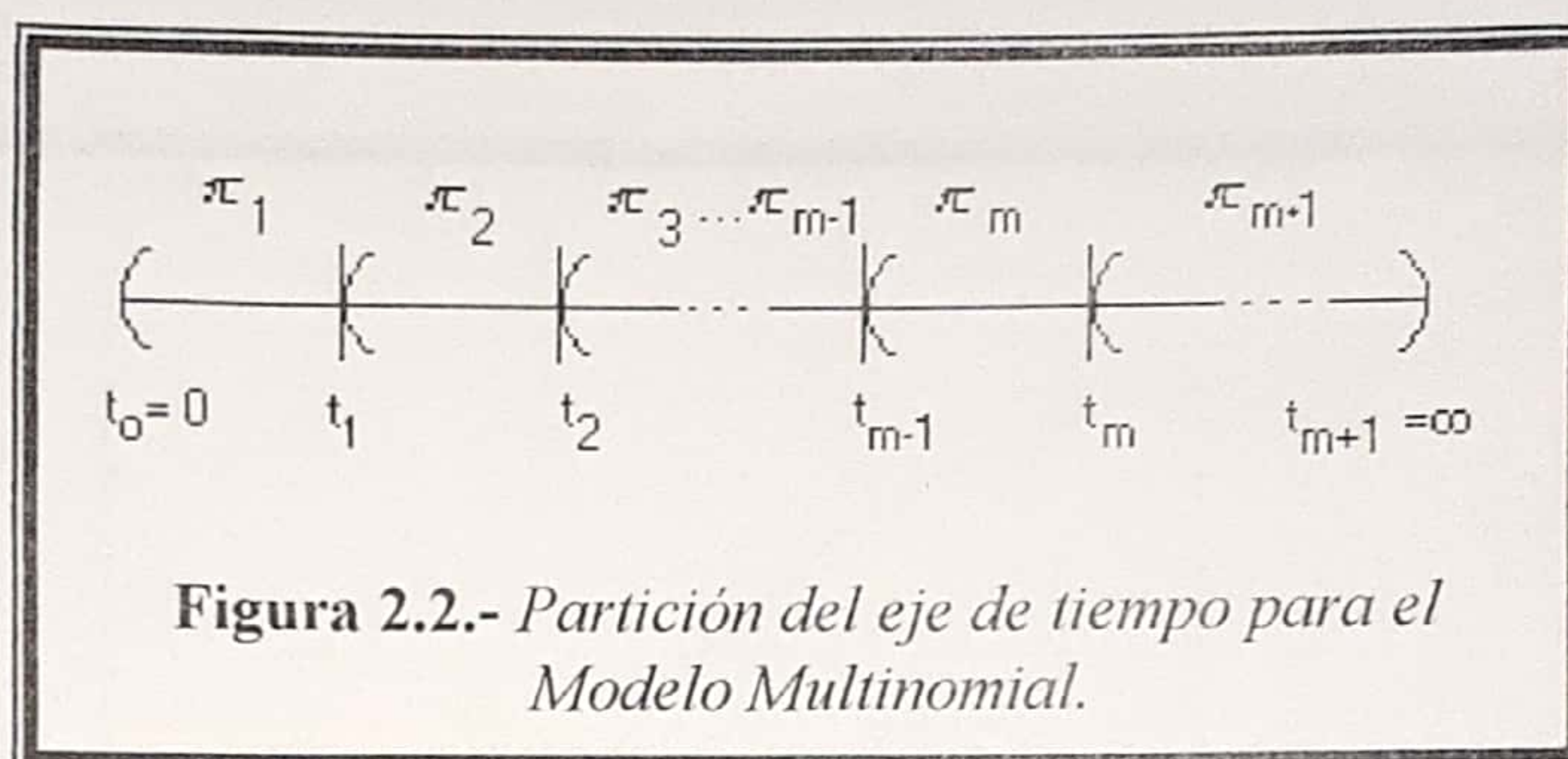
2.2.1. Modelo de Tiempo de Falla Multinomial.

Debido a que todos los datos son discretos, resulta conveniente particionar el eje o línea de tiempo $(0, \infty)$ en $m+1$ intervalos de observación. La partición depende de los tiempos de inspección o

de la precisión de las medidas; esto puede ser expresado como sigue:

$$(t_0, t_1], (t_1, t_2], \dots, (t_{m-1}, t_m], (t_m, t_{m+1}) \quad (2.5)$$

donde $t_0 = 0$ y $t_{m+1} = \infty$. Esta partición se la ilustra en el siguiente gráfico:



Por ejemplo, si tiempos de falla son registrados alrededor de una hora, entonces cada intervalo será de una longitud de una hora, incrementándose hasta llegar a t_m , el cual es el último registro. En general estos intervalos no requieren ser de igual longitud, nótese que el último intervalo es de longitud infinita.

Se define:

$$\pi_i = P(t_{i-1} < T \leq t_i) = F(t_i) - F(t_{i-1}) \quad (2.6)$$

como la probabilidad multinomial de que una unidad falle en el intervalo i . Nótese que $\pi_i \geq 0$ y $\sum_{j=1}^{m+1} \pi_j = 1$. La función de supervivencia evaluada en t_i es :

$$S(t_i) = P(T > t_i) = 1 - F(t_i) = \sum_{j=i+1}^{m+1} \pi_j$$

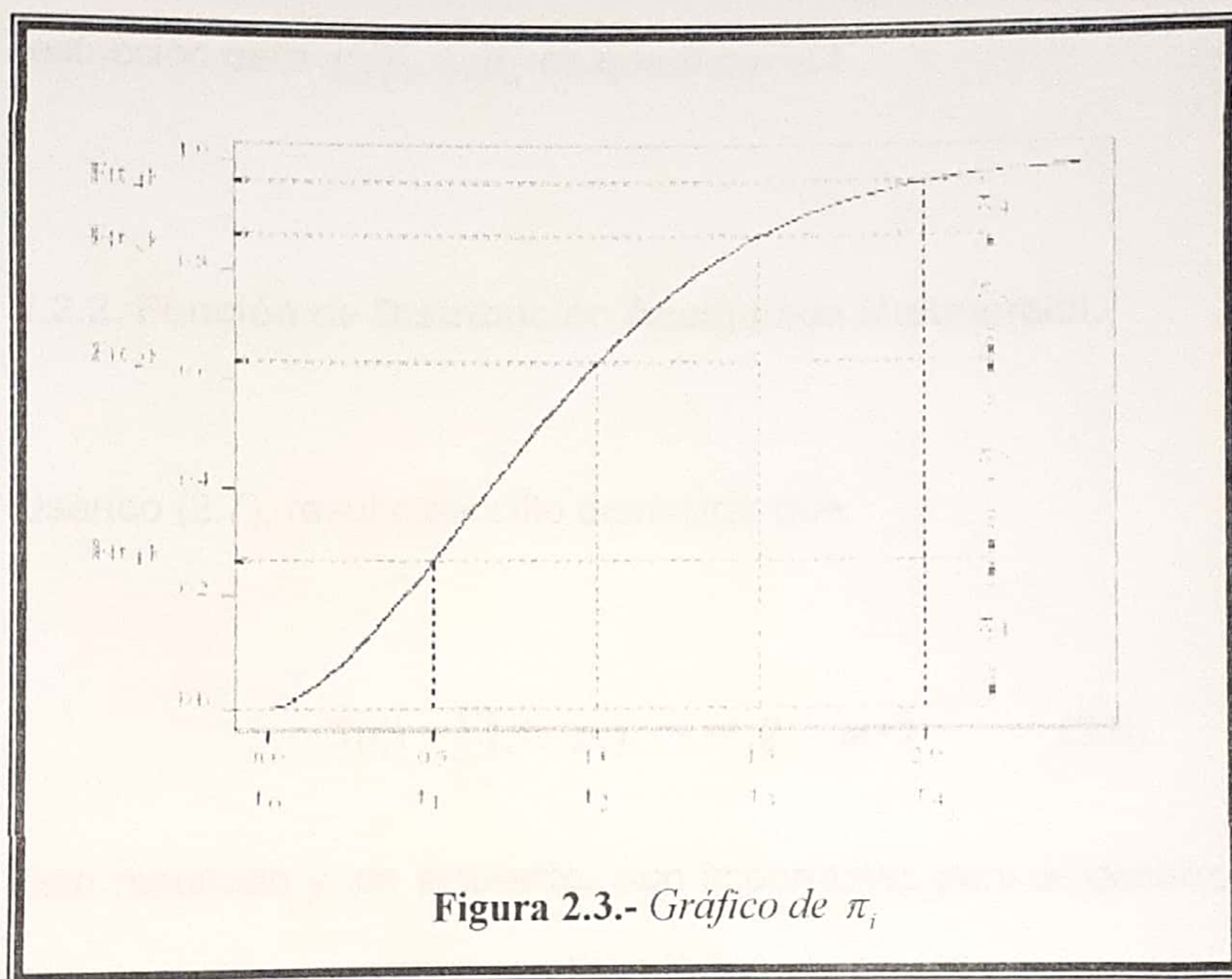


Figura 2.3.- Gráfico de π_i

Entonces:

$$p_i = P(t_{i-1} < T \leq t_i | T > t_{i-1}) = \frac{F(t_i) - F(t_{i-1})}{1 - F(t_{i-1})} = \frac{\pi_i}{S(t_{i-1})} \quad (2.7)$$

es la probabilidad condicional de que una unidad falle en el intervalo i , dado que una unidad estuvo aún operando al principio del intervalo i . De esta manera $p_{m+1} = 1$, mientras que la única restricción para p_1, p_2, \dots, p_m es que $0 \leq p_i \leq 1$.

2.2.2. Función de Distribución Acumulada Multinomial.

Usando (2.7), resulta sencillo demostrar que:

$$S(t_i) = \prod_{j=1}^i [1 - p_j], \quad i = 1, 2, \dots, m+1 \quad (2.8)$$

Este resultado y los siguientes son importantes para el desarrollo de métodos para análisis de datos. Luego la función de distribución acumulada de T , evaluada en t_i , puede ser expresada como:

$$F(t_i) = 1 - \prod_{j=1}^i [1 - p_j], \quad i = 1, 2, \dots, m+1$$

o alternativamente como:

$$F(t_i) = \sum_{j=1}^i \pi_j, \quad i = 1, 2, \dots, m+1 \quad (2.9)$$

De esta manera $\pi = (\pi_1, \pi_2, \dots, \pi_{m+1})$ o $p = (p_1, p_2, \dots, p_m)$ son alternativamente conjuntos de parámetros básicos para modelos discretos de datos de tiempos de falla.

2.3. Censuras.

2.3.1. Mecanismos de censura.

La Censura restringe la habilidad para observar con exactitud tiempos de falla. Es decir que las censuras ocurren cuando no se está en la capacidad para observar ***los resultados de la variable de interés***. Por ejemplo, en un experimento clínico controlando la remisión de pacientes de cáncer, algunos pacientes pueden aún estar en remisión al cierre del experimento, así que para estos pacientes, se sabe solo que sus períodos reales de remisión fueron más largos que la duración del experimento. Tal como un tiempo observado es un *tiempo de vida censurado*.

En otro ejemplo, el paciente puede retirarse del experimento y mudarse al extranjero, tales pacientes son *retiradas* y su estado es descrito como *perdido por retirada*, lo que significa que su progreso no puede ser seguido o controlado; otra vez, su real remisión resulta mayor de lo que puede ser observada. Tales pacientes se encuentran **censurados a la derecha**, dado que sobre una línea de tiempo sus tiempos de vida reales están a la derecha de sus tiempos de censura observados.

Por otro lado, consideremos experimentos de sacrificio en animales sometidos a prueba (ratones por ejemplo). De interés es el tiempo desde la inyección de un agente cancerígeno hasta el instante en el cual aparece un tumor. Un intento para medir este tiempo es realizado mediante el sacrificio del animal, es decir que al instante de muerte se observa si es que existe la presencia de tumores, de haber tumores, el tiempo de aparición requerido es menor que el tiempo de censura realmente observado y la observación se encuentra **censurada a la izquierda**. Si los tumores están ausentes, el tiempo de aparición del tumor es censurado a la derecha. Esto demuestra como ambos datos con censura a la izquierda y a la derecha pueden ocurrir en el mismo conjunto de datos.

Estos son ejemplos de los denominados **Datos con Censura Tipo**

I. Para tales datos censurados, el número actual de puntos censurados es desconocido, y por tanto es una variable aleatoria. Por ejemplo, no se puede conocer al principio del experimento cuántos ratones mostrarán tumores al tiempo de sacrificio.

Los datos del Ejemplo 2.3 muestran un ejemplo gráfico de la censura de Tipo I.

Ejemplo 2.3.

Rehenes occidentales mantenidos por más de 6 meses desde 1984.

Un artículo de "The Australian", en Noviembre de 1991 reportó datos sobre "rehenes occidentales mantenidos por más de 6 meses desde 1984". Pasado 1984, y por un período duradero dentro de la década de los 90's, más de 30 occidentales fueron secuestrados por varios grupos conservadores en el Medio Este; cada grupo aparentemente tenía su propio itinerario, pero algunos operaban bajo las políticas globales del grupo *Hezbollah*.

Algunos de los aspectos de estos datos son reproducidos en el diagrama presentado en el ANEXO No.1. Historias de casos individuales son representadas en líneas de tiempo. De interés es la variable "Tiempo de liberación". La censura a la derecha ocurrió cuando algunos rehenes no habían sido liberados para el tiempo del reporte, o habían muerto por otra causa (1).

(1). Peter J. Smith, 2002, *Analysis of Failure and Survival Data*, Primera Edición, Pag. 73.

2.3.1.1. Censura Tipo I.

En este tipo de censura el número de observaciones no censuradas es una variable aleatoria, mientras que en el Tipo II, el número de observaciones no censuradas es fijado con anticipación. Estos conceptos se encuentran especificados claramente por la notación y definiciones siguientes.

Supóngase que $Y_1, Y_2, Y_3, \dots, Y_n$ son variables de supervivencia independientes y que además se encuentran idénticamente distribuidas.

Definición:

Las variables de supervivencia $Y_1, Y_2, Y_3, \dots, Y_n$ se encuentran censurados hacia la derecha por constantes fijas $t_1, t_2, t_3, \dots, t_n$, si la muestra observada consiste en pares ordenados (Z_i, δ_i) para $i = 1, 2, \dots, n$, donde cada i : $Z_i = \min\{Y_i, t_i\}$,

$$\delta_i = \begin{cases} 1 & \text{si } Y_i \leq t_i \quad (\text{no_censurados}) \\ 0 & \text{si } Y_i > t_i \quad (\text{censurados}) \end{cases}$$

donde t_i es el tiempo fijo de censura y δ_i el indicador de censura para Y_i .

Supuesto: $Y_1, Y_2, Y_3, \dots, Y_n$ se sobreentienden independientes del mecanismo de creación de $t_1, t_2, t_3, \dots, t_n$.

Note que de acuerdo con esta definición, t_i es observado cuando Y_i es censurada. Si Y_i no está censurada, y corresponde a un tiempo real de muerte observado, entonces t_i con frecuencia permanece desconocido.

Para los datos con censura a la izquierda, los tiempos de vida observados son $Z_i = \max\{Y_i, l_i\}$, donde l_i es el tiempo censurado (a la izquierda) asociado con Y_i . Dado que, para datos censurados a la izquierda,

$$-Z_i = \min\{-Y_i, -l_i\}$$

quiere decir que la censura a la izquierda es un caso especial de la censura a la derecha con el eje de tiempo reverso.

2.3.1.2. Censura Tipo II.

Supóngase que, para una muestra dada $Y_1, Y_2, Y_3, \dots, Y_n$ de tamaño n , solo los primeros $r < n$ tiempos de vida son observados. El valor de r es fijado antes de que los datos de supervivencia sean observados. Esto significa que los datos observados consisten de las r observaciones más pequeñas.

En términos de variables aleatorias, esto puede ser expresado utilizando estadísticos de orden: De los posibles resultados $Y_1, Y_2, Y_3, \dots, Y_n$ se observa sólo los primeros r resultados clasificados $Y_{(1)}, Y_{(2)}, \dots, Y_{(r)}$. Es decir $y_{(1)} \leq y_{(2)} \leq y_{(3)} \leq \dots \leq y_{(r-1)} \leq y_{(r)}$.

En la práctica no es muy utilizada la técnica de manipulación de datos con Censura de Tipo II.

Luego de profundizar un poco en el tema concerniente a la censura, se puede especificar los mecanismos que provocan este tipo de restricción, la cual es bastante común en los análisis de confiabilidad de datos:

- Generalmente existen restricciones sobre la longitud de las pruebas de vida o sobre otros estudios de confiabilidad y, como

resultado, los datos tienen que ser analizados antes de que todas las unidades hayan fallado.

Cuando en una prueba, dado un tiempo predeterminado, quedan algunas unidades que no han experimentado falla; dicho tiempo es conocido como "*tiempo de censura*" y se dice que existe una "*Censura Tipo I*". Las unidades pueden ser probadas simultáneamente o en secuencia.

- Una prueba de vida que concluye luego de un número especificado de fallas da lugar a una "*censura de fallas*", y se dice que existe una "*Censura Tipo II*". A pesar de que las propiedades estadísticas de los estimadores de fallas censuradas son más simples que las correspondientes propiedades para los de tiempo de censura, las pruebas con fallas censuradas son menos comunes en la práctica.

- En muchas pruebas de vida, las fallas son descubiertas solo en el instante de inspección. Los intervalos de observaciones censuradas consisten de un límite superior e inferior sobre un tiempo de falla, tales datos son además conocidos como datos de inspección o datos agrupados.

Si una unidad ya ha fallado al momento de su primera inspección, esta constituye una observación censurada a la izquierda.

Si una unidad no ha fallado al instante de su última inspección, esta constituye una observación censurada a la derecha, siendo ∞ el punto final del intervalo.

- Algunos productos tienen más de una causa de falla. Si el principal interés está enfocado en una causa particular de falla, fallas originadas por otras causas pueden, en algunas situaciones, ser vistas como una forma de censura aleatoria a la derecha.

- En algunas situaciones las unidades se encuentran introducidas en el campo de actividad o puestas a prueba en diferentes tiempos. Esto es conocido como "acceso tambaleante". Si los datos son analizados en un punto o instante de tiempo cuando no todas las unidades han fallado, los datos contendrán múltiples censuras a la derecha, con algunos tiempos de falla excediendo los tiempos en marcha.

2.3.2. Supuestos Importantes en los Mecanismos de Censura.

El uso de muchos modelos y métodos para analizar datos censurados, implica importantes supuestos acerca de la naturaleza de la censura y su relación con el proceso de fallas.

Un tiempo de censura, tiempo en el cual se detiene el proceso de observación para una unidad que no ha fallado, puede ser aleatorio o predeterminado.

Para que los métodos estándar de análisis de datos censurados sean válidos, es necesario que el tiempo de censura de una unidad dependa solo de la historia del proceso observado de tiempos de falla.

Este supuesto para la causa de censura sería violado, por ejemplo, si las unidades fueran retiradas de la prueba antes de la falla existente, pero en respuesta de algún precursor para una futura falla.

Para los mecanismos estándar de censura mencionados al principio, los tiempos de detención dependen solo de la historia de los procesos observados de tiempo de fallas.

Independientemente, los métodos estándar de análisis de censura de datos requieren el supuesto de que la censura no es informativa; esto implica que los tiempos de censura de unidades no proveen información acerca de la distribución de los tiempos de falla.

2.4. Verosimilitud.

2.4.1. Verosimilitud basada en Métodos Estadísticos.

La idea general de la inferencia por verosimilitud es la de ajustar los datos por medio de combinaciones de modelos paramétricos para las cuales la probabilidad es grande. Las combinaciones de modelos paramétricos con relativamente altas probabilidades, son más verosímiles que las combinaciones con baja probabilidad. Los métodos de verosimilitud proveen herramientas generales para ajustar modelos para los datos.

Los métodos pueden ser aplicados en una amplia variedad de modelos paramétricos y no paramétricos con datos censurados, en intervalos y truncados.

Además los métodos de verosimilitud permiten además ajustar modelos con variables explicatorias como es el caso del análisis de regresión.

Existe una teoría bien desarrollada de verosimilitud para muestras grandes para modelos regulares que proveen métodos directos para ajustar modelos de datos. La teoría garantiza que estos métodos son, en muestras grandes, estadísticamente eficientes (los estimadores son más precisos).

Estas propiedades son solo aproximadas para el caso de medianos y pequeños tamaños de muestra, y varios estudios han demostrado que los métodos de verosimilitud generalmente se desempeñan tan bien como otros métodos disponibles.

Con datos censurados, "muestras grandes" realmente quiere decir "gran número de fallas" donde la típica línea guía para tamaños grandes es de 20 o más, pero esto realmente depende del problema en cuestión.

La Teoría de Verosimilitud puede ser extendida a modelos irregulares más complicados, pero los conceptos básicos son similares.

2.4.2. Especificando la Función de Verosimilitud.

La función de verosimilitud es igual o aproximadamente proporcional a la probabilidad de los datos.

A continuación se presentará un método general para el cálculo de la probabilidad de un conjunto de datos dado.

Luego, para un conjunto de datos dado y un modelo especificado, la verosimilitud es vista como una función de los parámetros desconocidos del modelo; aquí se puede utilizar los valores de π_i o los valores de p_i del modelo multinomial mencionado anteriormente.

La forma de la función de verosimilitud dependerá de factores como:

- Modelo de probabilidad asumido
- Forma de los datos disponibles (censurados, intervalos censurados, etc.)

- El foco de estudio. Esto incluye asuntos relacionados a la identificación de parámetros.

La verosimilitud total puede ser escrita como la probabilidad conjunta de los datos. Asumiendo n observaciones independientes, la verosimilitud de la muestra es:

$$L(p) = L(p; \text{DATOS}) = C \prod_{i=1}^n L_i(p; \text{dato}_i), \quad (2.10)$$

donde $L_i(p; \text{dato}_i)$ es la probabilidad de la observación i , y p es el vector de parámetros a ser estimado. Para estimar p de los datos disponibles, se encuentran los valores de p que maximicen $L(p)$. En las situaciones usuales, donde la constante C no depende de p , se puede simplemente tomar $C = 1$ por propósitos de estimación de p . La verosimilitud descrita en (2.10) además puede ser escrita como una función del vector de probabilidades multinomial π .

Vale la pena recalcar que de aquí en adelante se utilizará la estimación directa de p .

2.4.3. Contribución a la Función de Verosimilitud.

Tomando en cuenta los intervalos de ambigüedad, como son: la censura a la izquierda, intervalos censurados y censura a la derecha; las contribuciones de verosimilitud para cada uno de estos casos, la cual se describirá en la Tabla VII, es simplemente la probabilidad de ocurrencia de falla en el correspondiente intervalo de incertidumbre.

Tipo de Censura	Rango	Verosimilitud
d_i observaciones de intervalo censurado entre t_{i-1} y t_i	$t_{i-1} < T \leq t_i$	$[F(t_i) - F(t_{i-1})]^{d_i}$
l_i observaciones con censura a la izquierda en t_i	$T \leq t_i$	$[F(t_i)]^{l_i}$
r_i observaciones con censura a la derecha en t_i	$T > t_i$	$[1 - F(t_i)]^{r_i}$

Tabla VII .- *Contribución de la Censura a la Función de Verosimilitud*

Observaciones en intervalos censurados.

Si un tiempo de falla de una unidad se sabe que ha ocurrido entre los instantes t_{i-1} y t_i , la probabilidad de este evento es:

$$L_i(p) = \int_{t_{i-1}}^{t_i} f(t)dt = F(t_i) - F(t_{i-1}) \quad (2.11)$$

A pesar de que muchos datos surgen de la observación de un proceso de tiempo continuo puede ser sin embargo como habiendo ocurrido en intervalos similares a (t_{i-1}, t_i)

Observaciones con censura a la izquierda.

Las observaciones con censura a la izquierda ocurren en las aplicaciones de pruebas de vida cuando una unidad ha fallado al tiempo de su primera inspección, lo único que se conoce es que la unidad falló antes de ser inspeccionada. En otras situaciones, las observaciones con censura a la izquierda surgen cuando el valor exacto de un resultado no ha sido observado y se tiene, en lugar de eso, un límite superior en ese resultado.

Para este caso la probabilidad y contribución de verosimilitud es como sigue:

$$L_i(p) = \int_0^{t_i} f(t)dt = F(t_i) - F(0) = F(t_i) \quad (2.12)$$

La ecuación (2.7) demuestra como L_i puede ser escrita como función de p . Alternativamente, (2.8) demuestra como L_i puede ser escrita como función de π . Nótese que una observación con censura a la izquierda puede además ser considerada como una observación de intervalo censurado entre 0 y t_i .

Observaciones con censura a la derecha.

La censura a la derecha es común en análisis de confiabilidad de datos. Por ejemplo si una unidad continua operando luego de terminar el tiempo plazo de inspección o prueba de vida, esta constituye una observación con censura a la derecha.

Si existe un límite inferior t_i para el i -ésimo tiempo de falla, el tiempo de falla está en algún lugar del intervalo (t_i, ∞) . Entonces la probabilidad y la contribución a la verosimilitud para esta observación con censura a la derecha es:

$$L_i(p) = \int_{t_i}^{\infty} f(t)dt = F(\infty) - F(t_i) = 1 - F(t_i) \quad (2.13)$$

Verosimilitud Total.

La verosimilitud total, o probabilidad conjunta de los datos, para n observaciones independientes es:

$$L(p, DATOS) = \prod_{i=1}^n L_i(p, dato_i) \quad (2.14)$$

$$= \prod_{i=1}^{m+1} [F(t_i)]^{l_i} [F(t_i) - F(t_{i-1})]^{d_i} [1 - F(t_i)]^{r_i} [f(t_i)]^{u_i}$$

donde $n = \sum_{j=1}^{m+1} (d_j + r_j + l_j + u_j)$, además se consideró a la constante

$C = 1$ debido a que no depende de p .

Nótese que en la fórmula se toma en cuenta el número de fallas que ocurren en el instante de inspección, lo que se debe a que pueden ocurrir fallas en el tiempo de inspección a pesar de que sea muy poco probable.

Se desea encontrar el vector p para el cual la función $L(p)$ es máxima. El vector p que maximiza $L(p)$ provee un estimador de máxima verosimilitud de $F(t)$. Para algunos problemas será más conveniente escribir la verosimilitud y hacer la optimización en términos de π .

CAPÍTULO 3

3. ESTIMACION NO PARAMÉTRICA

En el presente capítulo se presenta una forma alternativa para la estimación de la función de supervivencia, el cual es uno de los objetivos de este análisis. Aquí se detallaran las técnicas estadísticas no paramétricas para realizar dicha estimación y además la construcción de los intervalos de confianza para los estimadores calculados.

3.1. Introducción.

Un análisis de datos, cualquiera que este sea y sin importar su objetivo, debería empezar mediante el uso de herramientas analíticas y gráficas las cuales no requieran la suposición fuerte de modelos predeterminados. Tales métodos permiten que los datos sean interpretados sin ninguna distorsión que pudiera ser causada por asumir un modelo inadecuado.

Por tanto se describirán métodos para calcular estimaciones no paramétricas e intervalos de confianza para $F(t)$, que como se mencionó anteriormente guarda una estrecha relación con la función de supervivencia $S(t)$.

En algunos casos, tales estimaciones serán requeridas para un análisis simple, mientras que en otras situaciones un análisis no paramétrico provee un paso importante para la consecución de un modelo más estructurado que permite mayor precisión en las estimaciones, así como también inferencias más amplias.

3.2. Estimación individual de intervalos de datos censurados.

En esta sección se muestra la forma de calcular una estimación no paramétrica de una función de distribución acumulada de intervalos de datos censurados, para el caso en que todas las unidades fallen, o para el caso de las observaciones con censura a la derecha encontradas en un punto al final del intervalo en estudio.

Muchos análisis y estudios involucran una inspección que empieza en un instante cero con una muestra inicial de unidades. Luego, la

información del estado de las unidades está disponible al final de cada intervalo de tiempo.

Sea n el número inicial de unidades, es decir el tamaño de la muestra, y sea d_i el número de unidades que han muerto o fallado en el i -ésimo intervalo $(t_{i-1}, t_i]$. El estimador no paramétrico $\hat{F}(t_i)$, basado en la distribución binomial, está dado por:

$$\hat{F}(t_i) = \frac{\# \text{ fallas hasta } t_i}{n} = \frac{\sum_{j=1}^i d_j}{n}$$

En general, este estimador no paramétrico $\hat{F}(t_i)$ está definido para todos los valores de t_i (puntos finales de todos los intervalos). Adicionalmente, si se conoce que el intervalo i no tiene fallas, entonces $\hat{F}(t) = \hat{F}(t_{i-1})$ para $t_{i-1} \leq t \leq t_i$. Si se conoce que el intervalo i contiene una o más fallas, $\hat{F}(t)$ incrementa de $\hat{F}(t_{i-1})$ a $\hat{F}(t_i)$ en el intervalo $(t_{i-1}, t_i]$. En este caso $\hat{F}(t)$, está indefinido para $t_{i-1} \leq t \leq t_i$.

Intuitivamente, esto es porque no se conoce la ubicación exacta de las fallas dentro del intervalo $(t_{i-1}, t_i]$ y de esta manera no se tiene

información acerca de cómo $\hat{F}(t)$ se está incrementando en el intervalo.

Se puede demostrar mediante el uso de la distribución binomial que $\hat{F}(t_i)$ es el estimador de máxima verosimilitud de $F(t)$.

3.3. Estimación de múltiples datos censurados.

Ahora, para esta sección se mostrará la manera de calcular una estimación no paramétrica de una función de distribución acumulada de datos con múltiple censura a la derecha, es decir cuando existen fallas que ocurren después de que algunas unidades han sido censuradas.

Supóngase que una muestra inicial de n unidades empieza operando en el instante cero, si una unidad no ha fallado en el intervalo i , esta es censurada al final del intervalo i o continua en el intervalo $i + 1$. La información acerca del estado de las unidades se encuentra disponible al final de cada intervalo. Los intervalos pueden ser grandes o pequeños y no necesitan ser de igual longitud, teniendo en cuenta que no debe existir intervalos sobrepuestos.

Sea d_i el número de unidades que mueren o fallan en el i -ésimo intervalo $(t_{i-1}, t_i]$. Además sea r_i el número de unidades que sobreviven al intervalo i y poseen censura a la derecha al tiempo t_i . Las unidades que se encuentran con vida al principio del intervalo i se las denomina "conjunto en riesgo" para el intervalo i , es decir que estas se encuentran en riesgo de falla, y el tamaño del conjunto de riesgo al principio del intervalo i es:

$$n_i = n - \sum_{j=0}^{i-1} d_j - \sum_{j=0}^{i-1} r_j, \quad i = 1, 2, \dots, m \quad (3.1)$$

donde m es el número de intervalos y además se sobreentiende que $d_0 = 0$ y $r_0 = 0$. Un estimador de la probabilidad condicional de falla en el intervalo i , dado que una unidad ingresa en el intervalo, es la proporción de fallas:

$$\hat{p}_i = \frac{d_i}{n_i}, \quad i = 1, 2, \dots, m$$

Sustituyendo esto en la ecuación (2.8) provee un estimador de la función de supervivencia:

$$\hat{S}(t_i) = \prod_{j=1}^i [1 - p_j], \quad i = 1, 2, \dots, m \quad (3.2)$$

Luego el correspondiente estimador no paramétrico de $F(t_i)$ es:

$$\hat{F}(t_i) = 1 - \hat{S}(t_i), \quad i = 1, 2, \dots, m \quad (3.3)$$

Aquí \hat{p}_i es el estimador de máxima verosimilitud de la probabilidad condicional p_i de (2.7). Esto implica que $\hat{F}(t_i)$ es el estimador de máxima verosimilitud de $F(t_i)$. El estimador no paramétrico $\hat{F}(t_i)$ está definido para todos los valores de t_i que son los puntos finales de cada intervalo. Adicionalmente, si se conoce que el intervalo i tiene cero fallas entonces $\hat{F}(t_i) = \hat{F}(t_{i-1})$ para $t_{i-1} \leq t \leq t_i$. Si se conoce que el intervalo i contiene una o más fallas, $\hat{F}(t)$ incrementa de $\hat{F}(t_{i-1})$ a $\hat{F}(t_i)$ en el intervalo $(t_{i-1}, t_i]$. Como en el caso anterior $\hat{F}(t)$ no está definida para $t_{i-1} \leq t \leq t_i$.

3.4. Estimación de múltiples datos censurados con tiempos exactos de falla.

Existen casos en que las fallas son reportadas en el tiempo exacto de ocurrencia. En tales casos, los tiempos que se reportan son denotados por t_i .

Los tiempos de falla exactos surgen de un proceso de inspección continua, o en ocasiones se debe al uso de una gran cantidad de intervalos de inspección de longitud muy pequeña. En el límite, así como el número de inspecciones aumenta y la amplitud de los intervalos de inspección tiende a cero, las fallas se encuentran concentradas en una relativamente pequeña cantidad de intervalos. Es decir que muchos intervalos no contendrán fallas. $\hat{F}(t)$ permanece constante sobre todos los intervalos que no contiene fallas.

De esta manera con intervalos de longitud pequeña, \hat{F} pasa a ser una función escalón con brechas en los intervalos donde se encuentran las fallas y con saltos en el límite superior de estos intervalos. En el límite, mientras la amplitud de los intervalos se aproxima a cero, el tamaño de las brechas se aproxima a cero y la

función escalón se incrementa en los tiempos de falla reportados. El estimador no paramétrico basado en esta delimitación es generalmente conocido como el Estimador de Kaplan-Meier.

3.5. Tiempos de censura indefinidos.

Los métodos presentados anteriormente asumen que tanto las observaciones censuradas a la izquierda como aquellas que se encuentran censuradas a la derecha, se dan en el punto superior de todos los intervalos.

Sin embargo, si los tiempos de censura sólo se los puede especificar dentro de intervalos determinados, el conjunto en riesgo es decreciente bajo los intervalos de tal manera que no pueden ser especificados con precisión.

Se sabe que el estimador de la proporción de fallas está dado por:

$$\hat{p}_i = \frac{\text{número de muertes en } I_i}{\text{número potencial de muerte en } I_i} \quad (3.4)$$

Un problema surge con la estimación del denominador de la fracción (3.4) ya que las muertes potenciales en el intervalo dependen de la distribución de las observaciones censuradas dentro del intervalo. Ciertamente, si todas las censuras ocurren inmediatamente al principio del intervalo I_i , entonces el número en riesgo (con el potencial de muerte en el intervalo) es esencialmente $n_i - r_i$. Por otro lado, si la censura ocurre justo antes del final del intervalo, entonces las unidades censuradas estuvieron esencialmente en riesgo para la duración del intervalo y el número en riesgo es n_i obtenido por la interpretación de $r_i = 0$.

Por esta razón existen dos métodos extremos para el manejo de observaciones censuradas en los intervalos:

- Asumiendo que todas las observaciones censuradas son trasladadas al instante t_i , el punto final superior del intervalo, el estimador $\hat{p}_i = d_i/n_i$, que es el estimador de la proporción de fallas que se propuso en secciones anteriores.
- Asumiendo que todas las observaciones censuradas son trasladadas al instante t_{i-1} , el punto inferior del intervalo, en este

caso el valor para el estimador de la proporción de fallas es

$$\hat{p}_i = d_i / (n_i - r_i).$$

Para los propósitos de la estimación se promedian los dos extremos del tamaño de la muestra,

$$\frac{(n_i - r_i) + (n_i)}{2} = n_i - \frac{1}{2}r_i,$$

promedio que define el número efectivo en riesgo, n'_i ,

$$n'_i = n_i - \frac{1}{2}r_i \quad (3.5)$$

Esta consideración del número efectivo en riesgo en un intervalo dado, esencialmente asume que la censura ocurre uniformemente a lo largo del intervalo.

A este estimador se lo conoce como *media armónica de los estimadores extremos*:

$$\hat{p}_i = \frac{d_i}{n'_i} = \frac{d_i}{(n_i - \frac{r_i}{2})} \quad (3.6)$$

Este ajuste efectuado al estimador puede ser sustituido en las ecuaciones (3.2) y en las siguientes ecuaciones las cuales involucren la estimación de p , para registrar los valores en la tabla actuarial o "tabla de vida" correspondientes a los estimadores no paramétricos de $F(t)$ y del error estándar $\hat{\Delta}_{NEF}$.

3.6. Cálculo de Tablas de Vida.

Para el cálculo de estas tablas de vida, que no son otra cosa que cuadros o esquemas que permiten registrar toda la información obtenida de los datos de tiempos de falla, resulta conveniente tomar en cuenta los datos referentes a la supervivencia. Es decir que en algunas ecuaciones se tomará en cuenta la proporción q que sobrevive a algún instante dado de tiempo, basándose en p la proporción de muertes (o de fallas): $q = 1 - p$

Supóngase que $Y_1, Y_2, Y_3, \dots, Y_n$ son variables aleatorias independientes e idénticamente distribuidas con función de supervivencia S y función de densidad de probabilidad f .

A continuación se presenta un método, denominado *Método de Tabla de Vida Cómplice*, para estimar la probabilidad de supervivencia más allá de un punto fijado en la presencia de datos censurados. Este es uno de los más viejos y más ampliamente utilizados métodos para el análisis no paramétrico y presentación de datos censurados a la derecha. El método de tabla de vida ha crecido tradicionalmente de los contextos médicos y actuariales.

Este método es especialmente útil para la estimación con gran cantidad de datos. Cuando no existe censura o incluso con censura Tipo II, la función de supervivencia empírica es fácilmente esquematizada para proveer información sobre la distribución del tiempo de vida (tiempo de falla) y sobre la forma de S . Para el caso de censura Tipo I, se sigue la aproximación de Lawless (1982) para demostrar como una tabla de vida es realizada usando software estadísticos. Primero, se desarrolla la estructura de la tabla de vida y la notación.

Tres pasos para el análisis de una Tabla de Vida

1. Dividir el eje del tiempo de vida en intervalos fijos disjuntos
2. Estimar la probabilidad condicional de supervivencia a lo largo de cada intervalo
3. Estimar S en el intervalo de puntos finales.

Ahora se profundizará acerca de cómo cada uno de estos tres pasos es logrado.

Paso 1: Empezando por imaginar n pacientes con vida y bajo observación a medida que transcurre el tiempo, y los registros del tiempo de supervivencia (tiempo de muerte) o tiempo de censura son almacenados.

Para almacenar estas medidas, estos registros son mantenidos en formato intervalo. Se empieza la construcción dividiendo $[0, \infty)$, el soporte de la distribución del tiempo de vida, dentro de intervalos disjuntos fijos.

$$I_i = (t_{i-1}, t_i], \quad i = 1, 2, \dots, k+1, \text{ teniendo que } t_0 = 0 \text{ y } t_{k+1} = \infty.$$

Los t_i 's son puntos finales de los intervalos elegidos por los analistas de datos, los cuales en muchos casos son elegidos de igual manera que los intervalos para los histogramas. Una elección de muy pocos intervalos resulta en una precisión limitada.

Teniendo contruidos los intervalos, se monitorean los datos (tiempos de censura y tiempos de muerte) ocurridos en los intervalos. Uno de los más importantes conceptos de tablas de vida es acerca del *número en riesgo* en un intervalo dado.

Definición:

En una tabla de vida, el número en riesgo en algún intervalo es el número de vidas bajo observación y no censuradas al principio del intervalo.

En términos de notación para el i -ésimo intervalo $I_i = [t_{i-1}, t_i)$, se escribe:

n_i como el **número en riesgo** en I_i ;

d_i como el **número de muertes o fallas** observadas en I_i ;

r_i como el **número de censuras** en I_i ;

Aquellas censuras en I_i son además denominadas retiradas o pérdidas por retirada en I_i . Estas definiciones implican que $n_1 = n$, dado que la muestra completa se encuentra inicialmente en riesgo. Claramente, y por lo definido en la ecuación (3.1)

$n_i = n_{i-1} - d_{i-1} - r_{i-1}$ demuestra como esta cantidad en riesgo en el $(i - 1)$ ésimo intervalo se propaga dentro del i -ésimo intervalo. Esta propagación es el corazón del análisis de las tablas de vida, haciendo a la tabla de vida un gráfico que muestra la oportunidad de supervivencia de pacientes a lo largo de un intervalo dado, debido a que estos están con vida al comienzo del intervalo. Tales aproximaciones poseen una estructura de probabilidad condicional. Las probabilidades condicionales de supervivencia a lo largo de I_i , dado que hay vida al principio de I_i , son:

$$\begin{aligned}
 q_i &= (\text{Sobreviva a lo largo de } I_i \mid \text{ sigue operando al principio de } I_i) \\
 &= P(T > t_i \mid T > t_{i-1}) \\
 &= \frac{S(t_i)}{S(t_{i-1})}
 \end{aligned} \tag{3.7}$$

Entonces, dado que $S(0) = 1$.

$$q_1 = \frac{S(t_1)}{S(t_0)} = \frac{S(t_1)}{S(0)} = S(t_1)$$

Esto significa que las probabilidades de supervivencia en los puntos finales del intervalo puede ser escrito como un producto de términos abreviado:

$$S(t_i) = [S(t_1)] \left[\frac{S(t_2)}{S(t_1)} \right] \left[\frac{S(t_3)}{S(t_2)} \right] \dots \left[\frac{S(t_{i-1})}{S(t_{i-2})} \right] \left[\frac{S(t_i)}{S(t_{i-1})} \right] \quad (3.8)$$

Y estos términos a la derecha de la ecuación (2.6) son simplemente las probabilidades condicionales de supervivencia determinadas por la ecuación (2.5). Esto demuestra que la probabilidad de supervivencia más allá de un intervalo de puntos finales es un producto de probabilidades condicionales como lo declara el siguiente teorema.

Teorema 3.1. En una tabla de vida construida por intervalos $I_i = [t_{i-1}, t_i)$, para $i = 1, 2, \dots, k+1$,

$$S(t_i) = q_1 q_2 q_3 \dots q_i,$$

donde q_i es la probabilidad condicional de supervivencia a lo largo del intervalo I_i , dado que existe vida o la unidad sigue operando al principio de este intervalo.

En la estimación de tablas de vida, se utiliza el Teorema 3.1 para estimar supervivencias en el intervalo fijo de puntos finales.

Para el Paso 2 en la construcción de las tablas de vida, claramente, la probabilidad q_i necesita ser estimada, así que S puede ser estimada en t_i . El estimador más obvio a usarse es el clásico estimador insesgado para una proporción de una distribución binomial:

$$\text{Estimación de } q_i = 1 - \frac{\text{número de muertes en } I_i}{\text{número potencial de muerte en } I_i} \quad (3.9)$$

Esta estimación trabaja también para la proporción de fallas p_i , ya que cuando se estima $1 - q_i$, se está estimando la proporción que muere o falla en el intervalo dado que ellos estuvieron vivos y bajo observación al principio del intervalo. Por consiguiente, como una simple proporción binomial, esta se estima como el número de muertes en el intervalo dividido para el número en riesgo.

Siguiendo la ecuación (3.9), y suponiendo $n'_i > 0$, la estimación actuarial \hat{q}_i de q_i es tomada como:

$$\hat{q}_i = 1 - \frac{d_i}{n'_i}, \quad i = 1, 2, 3, \dots, k+1. \quad (3.10)$$

Si, para un intervalo dado, $n'_i = 0$, entonces la estimación actuarial de \hat{q}_i es tomada como $\hat{q}_i = 0$.

Finalmente para el Paso 3 en la construcción de la tabla de vida,

$$\hat{S}(t_i) = \hat{q}_1 \hat{q}_2 \hat{q}_3 \dots \hat{q}_i, \quad i = 1, 2, 3, \dots, k+1, \quad (3.11)$$

es la estimación actuarial de supervivencia en

$$0 = t_0, t_1, t_2, \dots, t_k, t_{k+1} = \infty.$$

La naturaleza del producto de la estimación en la ecuación (3.11) surge de la estructura del producto de $\hat{S}(t_i)$ en el Teorema 3.1. Claramente, si el i -ésimo intervalo cae más allá del soporte de los tiempos de vida observados, no existen observaciones individuales en riesgo al principio del intervalo, entonces $n_i = n'_i = 0$, $\hat{q}_i = 0$ y $\hat{S}(t_i) = 0$, evidentemente asegurando que la supervivencia estimada sea cero.

Derivando la estimación en la ecuación (3.11), se ha asumido que las censuras individuales tienen la misma distribución de tiempo de vida que las observaciones no censuradas (q_i es la misma para cada observación individual); esta puede no ser la misma en un experimento médico. Pacientes en un experimento clínico pueden ser considerados como pérdidas por retirada debido a efectos laterales experimentados en su grupo de tratamiento.

Se puede clasificar dos supuestos básicos para la construcción de tablas de vida:

Supuesto 1: El mecanismo generador de los tiempos de censura es independiente de los tiempos de vida observados. Este supuesto es requerido para asegurar que la probabilidad q_i es la misma para cada observación individual.

Supuesto 2: Tanto los tiempos de muerte (tiempos de falla) y los tiempos de censura en un intervalo dado son asumidos con una distribución uniforme a través del intervalo. Este supuesto es necesario en la construcción del número efectivo en riesgo en un intervalo dado.

La estructura de una tabla de vida varía a través de diferentes tipos de presentaciones de software. Esencialmente, se puede estimar una tabla de vida como una generalización de un histograma de frecuencias que comprende datos censurados a la derecha.

La tabla generalmente incluiría listas de $I_j, n_j, d_j, w_j, \tilde{q}_j = 1 - \tilde{p}_j$, y $\tilde{S}(a_j)$.

3.7. Intervalos de confianza para datos con censura simple.

3.7.1. Intervalo de confianza de punto centrado con aproximación normal para $F(t_i)$.

Para un valor especificado de t_i , un intervalo simple con $100(1 - \alpha)\%$ de confianza para $F(t_i)$ viene dado por:

$$[\underline{F}(t_i), \bar{F}(t_i)] = \hat{F}(t_i) \pm z_{(1-\alpha)/2} \overset{\wedge}{SE}_{\hat{F}} \quad (3.12)$$

donde $z_{(p)}$ es el cuantil p de la distribución normal estándar y:

$$\Delta se_{\hat{F}} = \sqrt{\frac{\hat{F}(t_i)[1-\hat{F}(t_i)]}{n}} \quad (3.13)$$

es un estimador del error estándar de $\hat{F}(t_i)$. Este intervalo de confianza está basado en el supuesto de que:

$$Z_F = \frac{\hat{F}(t_i) - F(t_i)}{\Delta se_{\hat{F}}}$$

puede ser adecuadamente aproximada a una distribución normal estándar $N(0,1)$. Para que esta aproximación sea adecuada, $n\hat{F}(t_i)$ debería ser por lo menos 5 o 10 y no más de $n-5$ o $n-10$.

3.8. Intervalos de confianza de punto centrado para datos con censura múltiple.

3.8.1. Fórmula de Greenwood.

Se han realizado varias derivaciones de la fórmula del error estándar, pero el análisis se concentrará en aquella que aproxima a $\hat{S}(t_i)$ mediante un producto de proporciones binomiales independientes para los intervalos precedentes a t_i . Esto requerirá tamaños de muestra positivos (número en riesgo) en cada uno de los intervalos concernientes.

Si se condiciona a dicha expresión, el resultado es más preciso; incondicionalmente, el error estándar es sólo una aproximación.

Dado que $\hat{F}(t_i) = 1 - \hat{S}(t_i)$, entonces se tiene que

$$Var[\hat{F}(t_i)] = Var[\hat{S}(t_i)].$$

Se sabe que $q_j = 1 - p_j$, por consiguiente cuando se tome en cuenta la proporción q_j se estará hablando acerca de la probabilidad de supervivencia de alguna unidad a través del

intervalo j , dado que esta sigue operando al principio de este intervalo.

Debido que los valores de q_j son aproximadamente proporciones binomiales no correlacionadas, se tiene que:

$$Var[\hat{F}(t_i)] = Var[\hat{S}(t_i)] \approx [S(t_i)]^2 \sum_{j=1}^i \frac{p_j}{n_j(1-p_j)} \quad (3.14)$$

Esta igualdad se la puede probar mediante el uso de una cadena de teoremas, los cuales serán probados individualmente para demostrar finalmente esta ecuación (3.14).

Teorema 3.2. $E[\hat{S}(t_j)] \approx q_1 q_2 q_3 \dots q_j = S(t_j)$, $j = 1, 2, \dots, m+1$

Este teorema se lo prueba usando el hecho de que $\hat{p}_i = \frac{d_i}{n_i}$ la probabilidad condicional de que una unidad falle en el i -ésimo intervalo dado que sigue operando al inicio de este intervalo, y por lo tanto q_i es la probabilidad condicional de que una unidad sobreviva al i -ésimo intervalo dado que dicha unidad sigue operando a al inicio del mismo.

Si se tiene para $i < j$ y $n_i > 0$, luego si se condiciona $n_i > 0$ entonces se tiene contadores binomiales para d_i :

$$d_i | n_i \sim \text{Bin}(n_i, p_i)$$

Además se deben tomar en cuenta algunas definiciones y teoremas auxiliares:

Teorema 3.2.1. $E(Y_1) = E[E(Y_1|Y_2)]$ donde Y_1 y Y_2 son dos variables aleatorias.

Para la demostración de este teorema se supone que Y_1 y Y_2 tienen la función de densidad conjunta $f(y_1, y_2)$ y las densidades marginales $f_1(y_1)$ y $f_2(y_2)$, respectivamente.

Además se conoce que:

- $f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2$ (de forma análoga para la densidad marginal $f_2(y_2)$).

- Por definición para $i \neq j$: $f(y_i|y_j) = \begin{cases} \frac{f(y_i, y_j)}{f_j(y_j)}, & f_j(y_j) > 0 \\ 0, & \text{si no} \end{cases}$
- Por definición $E(Y_1|Y_2) = \int_{-\infty}^{\infty} y_1 f(y_1|y_2) dy_1$

Entonces:

$$\begin{aligned}
 E(Y_1) &= \int_{-\infty}^{\infty} y_1 f_1(y_1) dy_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2) dy_1 dy_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1|y_2) f_2(y_2) dy_1 dy_2 \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y_1 f(y_1|y_2) dy_1 \right] f_2(y_2) dy_2 \\
 &= \int_{-\infty}^{\infty} E(Y_1|Y_2) f_2(y_2) dy_2 = E[E(Y_1|Y_2)]
 \end{aligned}$$

Usando esto se tiene:

$$\begin{aligned}
 E[1 - \hat{q}_i] &= E\left[\frac{d_i}{n_i}\right] \\
 &= E\left\{E\left[\frac{d_i}{n_i} \mid n_i\right]\right\} \\
 &= E\left\{\frac{1}{n_i} E[d_i \mid n_i]\right\} \\
 &= E\left\{\frac{1}{n_i} [n_i(1 - q_i)]\right\} \\
 &= E\{(1 - q_i)\} \\
 &= (1 - q_i) = p_i
 \end{aligned}$$

Incondicionalmente, $E[1 - \hat{q}_i] \approx 1 - q_i$ y por lo tanto $E[\hat{q}_i] \approx q_i$.

Similarmente, con la condición $n_i > 0$, para $l < i$ se tiene que

$E[\hat{q}_l \hat{q}_i] \approx q_l q_i$. En general:

$$E[\hat{S}(t_i)] = E[\hat{q}_1 \hat{q}_2 \hat{q}_3 \dots \hat{q}_i] \approx q_1 q_2 q_3 \dots q_i = S(t_i), \quad i = 1, 2, \dots, m+1$$

Con lo que queda demostrado el Teorema 3.2.

Una vez que se han demostrado estos teoremas auxiliares, se puede hacer uso de ellos y demostrar la igualdad de la varianza

$$Var[\hat{S}(t_i)] = Var[S(t_i)]^2 \sum_{j=1}^i \frac{p_j}{n_j(1-p_j)}.$$

Dado que las fallas observadas siguen una distribución binomial y por los argumentos del Teorema 3.1, el valor de la varianza para \hat{q}_i está dado por:

$$Var[\hat{q}_i] = \frac{q_i(1-q_i)}{n_i},$$

Así, incondicionalmente, el segundo momento puede ser escrito como:

$$\begin{aligned} E[\hat{q}_i^2] &\approx Var[\hat{q}_i] + E^2[\hat{q}_i] \\ &\approx \frac{q_i(1-q_i)}{n_i} + q_i^2 = q_i^2 \left(1 + \frac{p_i}{q_i n_i} \right) \end{aligned}$$

Combinando estos resultados y basándose en el valor esperado $E[\hat{S}(t_i)]$, se tiene que:

$$E[\hat{S}(t_i)^2] = E[\hat{q}_1^2 \hat{q}_2^2 \hat{q}_3^2 \dots \hat{q}_i^2] \approx E[\hat{q}_1^2] E[\hat{q}_2^2] E[\hat{q}_3^2] \dots E[\hat{q}_i^2].$$

Por lo tanto:

$$\begin{aligned} \text{Var}[\hat{S}(t_i)] &= E[\hat{S}(t_i)^2] - E^2[\hat{S}(t_i)] \\ &\approx \prod_{j=1}^i q_j^2 \left(1 + \frac{p_j}{q_j n_j} \right) - S(t_i)^2 \\ &= S(t_i)^2 \left[\prod_{j=1}^i \left(1 + \frac{p_j}{q_j n_j} \right) - 1 \right] \\ &\approx S(t_i)^2 \left[1 + \sum_{j=1}^i \frac{p_j}{q_j n_j} - 1 \right] \end{aligned}$$

la aproximación en la última línea se debe a:

$$\prod_{j=1}^i (1 + a_j) \approx 1 + \sum_{j=1}^i a_j \quad \text{para} \quad a_j = \frac{p_j}{q_j n_j}$$

Finalmente sustituyendo \hat{p}_j por p_j y $\hat{S}(t_i)$ por $S(t_i)$, se obtiene el estimador de la varianza, conocido también como la Fórmula de Greenwood:

$$\hat{\text{Var}}[\hat{F}(t_i)] = \hat{\text{Var}}[\hat{S}(t_i)] \approx [\hat{S}(t_i)]^2 \sum_{j=1}^i \frac{\hat{p}_j}{n_j (1 - \hat{p}_j)} \quad (3.15)$$

Un estimador del error estándar de $\hat{F}(t_i)$ es:

$$\overset{\wedge}{se}_{\hat{F}} = \sqrt{\overset{\wedge}{Var}[\hat{F}(t_i)]} \quad (3.16)$$

3.8.2. Intervalo de confianza de punto centrado con aproximación normal para $F(t_i)$.

Debido a que $\hat{F}(t)$ está definida sólo en el extremo final de cada intervalo que contiene fallas, $\hat{F}(t)$ es generalmente estimada solo en tales puntos; es decir que si algún intervalo no tuviere fallas, $\hat{F}(t)$ se mantiene constante sobre ese intervalo. Para un punto superior específico t_i , en el cual se requiere un estimador de $F(t)$, una aproximación normal con un intervalo de $100(1-\alpha)\%$ de confianza para $F(t_i)$ es:

$$[\underline{F}(t_i), \bar{F}(t_i)] = \hat{F}(t_i) \pm z_{(1-\alpha/2)} \overset{\wedge}{se}_{\hat{F}} \quad (3.17)$$

donde $z_{(p)}$ es el cuantil p de la distribución normal estándar. En general una aproximación de una sola cola puede ser obtenida reemplazando $z_{(1-\alpha/2)}$ por $z_{(1-\alpha)}$ y utilizando un punto final apropiado del intervalo de confianza de dos colas.

Cabe recalcar que esta aproximación de los intervalos de confianza se basa también en que la distribución de

$$Z_{\hat{F}} = \frac{\hat{F}(t_i) - F(t_i)}{\overset{\wedge}{se}_{\hat{F}}} \quad (3.18)$$

puede ser aproximada adecuadamente por una distribución normal estándar. Entonces

$$P[z_{(\alpha/2)} < Z_{\hat{F}} \leq z_{(1-\alpha/2)}] \approx 1 - \alpha \quad (3.19)$$

que a su vez implica:

$$P[\hat{F}(t_i) - z_{(1-\alpha/2)} \overset{\wedge}{se}_{\hat{F}} < F(t_i) \leq \hat{F}(t_i) - z_{(\alpha/2)} \overset{\wedge}{se}_{\hat{F}}] \approx 1 - \alpha \quad (3.20)$$

La ecuación (3.20) brinda una aproximación de la cobertura de probabilidad para intervalos calculados con el procedimiento (3.17).

Note que en la ecuación (3.20), $F(t_i)$ es fijo, mientras que $\hat{F}(t_i)$ y

$\overset{\wedge}{se}_{\hat{F}}$ son aleatorios.

La aproximación en la ecuación (3.20) es una aproximación para muestras grandes, por tanto esta se mejora cuando el tamaño muestral se incrementa.

Cuando el tamaño de la muestra no es grande, sin embargo, la distribución de $Z_{\hat{F}}$ puede estar sesgada en mala forma y la

distribución normal puede que no provea una aproximación adecuada, particularmente en las colas de la distribución (donde $\hat{F}(t)$ está cercana a 0 o 1).

Por ejemplo, es posible que la ecuación (3.17) de cómo resultado $\underline{E}(t) < 0$ o $\bar{F}(t) > 1$, resultados que están fuera del rango posible para $F(t)$. Generalmente una mejor aproximación podría ser obtenida mediante el uso de la transformación $\text{logit}(p) = \log[p/(1-p)]$ y basando los intervalos de confianza en la distribución de:

$$Z_{\text{logit}(\hat{F})} = \frac{\text{logit}[\hat{F}(t_i)] - \text{logit}[F(t_i)]}{\frac{\Lambda}{SE_{\text{logit}(\hat{F})}}} \quad (3.21)$$

Debido que $\text{logit}[\hat{F}(t_i)]$, como una variable aleatoria normal estándar, no posee restricción con respecto al rango $(-\infty, \infty)$, la expresión (3.21) puede esperarse de que se aproxime a una distribución normal estándar $N(0,1)$ mucho más que la expresión (3.18) para la distribución de Z .

El intervalo hacia el cual nos dirige la expresión anterior, con $100(1 - \alpha)\%$ de confianza, con una aproximación de dos colas es:

$$[F(t_i), \bar{F}(t_i)] = \left[\frac{\hat{F}}{\hat{F} + (1 - \hat{F})w}, \frac{\hat{F}}{\hat{F} + (1 - \hat{F})/w} \right] \quad (3.22)$$

donde

$$w = \exp \left\{ z_{(1-\alpha/2)} \frac{\hat{F}}{[\hat{F}(1 - \hat{F})]} \right\}$$

Los puntos finales de este intervalo siempre caerán entre 0 y 1. Una aproximación de una sola cola para un intervalo de confianza de $100(1 - \alpha)\%$ puede ser obtenido reemplazando $z_{(1-\alpha/2)}$ por $z_{(1-\alpha)}$ y utilizando un punto final apropiado del intervalo de confianza de dos colas.

3.9. Bandas de confianza simultáneas.

Los intervalos de confianza de punto centrado definidos anteriormente son útiles para obtener el estado de $F(t_i)$ en un valor especificado de t_i . Sin embargo, en algunas aplicaciones es necesario cuantificar la incertidumbre muestral, de forma simultánea, sobre un rango de valores de t . Para realizar esto, se

puede hacer uso de las bandas simultáneas de confianza para $F(t)$.

La completa cobertura de probabilidad para el conjunto de intervalos de punto centrado es generalmente menor que para algún intervalo individual.

3.9.1. Bandas de Confianza Simultáneas para $F(t)$ a partir de tamaños de muestras grandes.

Las bandas simultáneas para $F(t_i)$ con $100(1-\alpha)\%$ de confianza pueden ser obtenidas de:

$$\left[\underline{F}(t), \overline{F}(t) \right] = \hat{F}(t) \pm e_{(a,b,1-\alpha/2)} \wedge se_{\hat{F}(t)} \text{ para todo } t \in [t_L(a), t_U(b)] \quad (3.23)$$

donde el rango $[t_L(a), t_U(b)]$ toma la forma de una función compleja del patrón de censura en los datos, lo que será explicado más adelante. Sin datos censurados el rango de t está dado por los valores de t para los cuales $a \leq \hat{F}(t) \leq b$.

Los factores aproximados $e_{(a,b,1-\alpha/2)}$ se encuentran ya tabulados.

Debido a que los factores $e_{(a,b,1-\alpha/2)}$ poseen los mismos valores para todos los valores de t , esta familia de bandas es conocida como bandas de confianza simultáneas de "Igual Precisión".

Los factores $e_{(a,b,1-\alpha/2)}$ son mayores que los correspondientes valores de los $z_{(1-\alpha/2)}$ de una aproximación normal con intervalos de confianza de punto centrado.

De esta manera el ancho de las bandas simultáneas, en algún punto dado t , es más amplio que el ancho del correspondiente intervalo de confianza de punto centrado en tal punto. Este resultado es necesario para la explicación de la naturaleza simultánea de las bandas de la ecuación (3.15).

Bandas simultáneas de confianza como las definidas en la ecuación (3.23) están basadas en la distribución aproximada de:

$$Z_{\max \hat{F}} = \max_{t \in [t_L(a), t_U(b)]} \left[\frac{\hat{F}(t) - F(t)}{\frac{\Lambda}{se_{\hat{F}(t)}}} \right] \quad (3.24)$$

De igual manera como se explicó en la sección anterior para los intervalos de confianza de punto centrado, generalmente resulta

mejor calcular las bandas de confianza simultáneas basadas en la transformación *logit* de \hat{F} . Esto se lo calcula como sigue:

$$[\underline{F}(t_i), \bar{F}(t_i)] = \left[\frac{\hat{F}(t)}{\hat{F}(t) + [1 - \hat{F}(t)]w}, \frac{\hat{F}(t)}{\hat{F}(t) + [1 - \hat{F}(t)]/w} \right] \quad (3.25)$$

donde

$$w = \exp \left\{ e_{(a,b,1-\alpha/2)}^{\Lambda} \frac{se_{\hat{F}}}{[\hat{F}(1 - \hat{F})]} \right\}$$

Los puntos finales de estas bandas siempre caerán entre 0 y 1.

Las bandas calculadas a partir de esta ecuación están basadas en la distribución aproximada de la función aleatoria:

$$Z_{\max_logit(\hat{F})} = \max_{t \in [t_L(a), t_U(b)]} \left[\frac{\logit[\hat{F}(t)] - \logit[F(t)]}{se_{\logit[\hat{F}(t)]}} \right] \quad (3.26)$$

3.9.2. Determinación del Rango de Tiempo para las Bandas de Confianza Simultáneas para $F(t)$.

Cuando se especifican las cantidades a y b se determina el rango $[t_L(a), t_U(b)]$ sobre el cual las bandas de confianza simultáneas para $F(t)$ están definidas. Sea:

$$\hat{\sigma}(t) = n \sum_{j:t_j \leq t} \frac{d_j}{n_j(n_j - d_j)} \quad \text{y} \quad \hat{K}(t) = \frac{\hat{\sigma}(t)}{1 + \hat{\sigma}(t)}$$

donde la sumatoria está definida sobre j tal que $t_j \leq t$. Entonces las bandas simultáneas de confianza tienen un rango que cubre todos los valores de t tales que $a \leq \hat{K}(t) \leq b$. La función $\hat{K}(t)$ se comporta como un estimador no paramétrico $\hat{F}(t)$; ya que no es decreciente, $0 \leq \hat{K}(t) \leq 1$, y además $\hat{K}(t) = \hat{F}(t)$ cuando no existe censura.

3.10. Censura Arbitraria.

El estimador no paramétrico de la ecuación (3.2) trabaja sólo para algunos tipos de patrones de censura (tales como múltiple censura a la derecha e intervalos de censura disjuntos).

Cuando la censura es más complicada, se requiere una alternativa. Esta necesidad surge ya que, con censura complicada, no se pueden conocer los valores de n_i en la ecuación (3.1).

Existe un estimador de Peto Turnbull que provee la generalización requerida para la estimación no paramétrica de máxima verosimilitud, el cual puede ser aplicado en los siguientes casos:

- Datos con censura arbitraria (tales como combinaciones de censuras a la izquierda, a la derecha e intervalos de censura los cuales están superpuestos)
- Datos Truncados

La idea básica es la de escribir la verosimilitud como en la ecuación (2.10) maximizándola para estimar el vector p o π para de esta manera calcular un estimador de $F(t)$.

CAPÍTULO 4

4. ESTIMACIÓN Y PREDICCIÓN PARA LOS DATOS OBTENIDOS EN UNA EMPRESA QUE COMERCIALIZA EQUIPOS DE IMPRESIÓN Y FOTOCOPIADO, Y EN UNA INDUSTRIA MANUFACTURERA DE LLANTAS.

En el presente capítulo se realizará un análisis de los tiempos de falla que ocurren en una empresa que produce equipos de fotocopiado e impresión con tecnología avanzada, y luego se realizará un estudio estadístico para la proporción de fallas existentes dentro de una industria manufacturera de producción de llantas.

4.1. Análisis de los tiempos de falla en los equipos comercializados por una empresa de tecnología.

La empresa que provee los equipos para la impresión y fotocopiado brinda también un servicio post – venta, el cual permite

a los clientes que adquieren sus equipos realizar consultas acerca del manejo de alguno de ellos. Además el cliente tiene la potestad de comunicarse con la empresa para comunicarle algún fallo que detecte y que no permita el correcto funcionamiento del equipo.

Estas llamadas realizadas por los clientes son almacenadas en una base de datos, en la cual se registra primero al cliente, luego el número de serie del equipo, la fecha y hora de la llamada, y por último la falla que se encontró.

4.1.1. Alcance del estudio.

Como se explicó en el capítulo 1, es necesario dejar bien establecido el alcance del estudio, los modos de falla y el período de tiempo de observación.

Alcance y Modos de falla.

Las unidades que van a ser muestreadas son los equipos de dos líneas de la compañía, los cuales serán analizados como un todo sin tomar en cuenta la estructura de las partes o componentes, es

decir que el estudio se enfocará en la funcionalidad total del equipo.

Por tanto lo que se plantea es un Análisis de Supervivencia para medir la confiabilidad de los equipos pertenecientes a estas dos líneas. De esta manera, el análisis será similar al que se realiza en equipos no reparables, teniendo como principal indicador de la confiabilidad al Tiempo Promedio De Falla (*MTTF*), descartando los conceptos de mantenibilidad y disponibilidad.

Tiempo de Medición.

Para el estudio resultan de interés los datos de las fechas de la instalación del equipo en la empresa cliente y la fecha de la llamada para reportar la primera falla del mismo. El tiempo desde la instalación hasta la primera falla que se presenta, será medido en días, tomando en cuenta la cantidad exacta de días que tiene cada mes. El tiempo de límite de observación, según lo que sugieren los datos obtenidos, será de 150 días para la Línea de Equipo 1, mientras que para la Línea de Equipo 2 será de 75 días.

Además existen casos en que los equipos son devueltos por el cliente, debido a la finalización del contrato leasing, o porque estos deciden cambiarlos por un equipo con otras funciones y con mejor tecnología. Los equipos pueden ser también retirados cuando existe incumplimiento de pago por parte de la empresa cliente. Dentro del análisis por realizar, a estos escenarios se los conoce como "*pérdidas o retiradas*", las cuales constituyen parte de la ya definida censura a la derecha.

Se tomó la muestra de las dos líneas de equipos que poseen mayor demanda por parte de las empresas clientes. La línea de equipos más vendida dentro del país pertenece a un equipo multifunción mientras que la segunda línea pertenece a una máquina copiadora. Para ambas el modelo y codificación son proporcionados por la empresa proveedora.

Los datos fueron obtenidos en el formato anteriormente mencionado, en donde se recogieron los registros para 20 equipos distintos para cada línea de producto. La fecha de instalación, por obvias razones no es igual para todos los equipos aunque pertenezcan a la misma línea, sin embargo para cada uno se tomó como día cero el día de instalación y se contabilizaron los días hasta la fecha de falla.

Línea de Equipo 1			Línea de Equipo 2		
instalación	1 falla	cant. días	instalación	1 falla	cant. días
23/09/01	05/11/01	43	25/01/00	28/01/00	3
20/11/02	20/01/03	61	14/06/00	16/06/00	2
07/08/02	16/09/02	40	24/08/00	07/11/00	68
29/01/01	19/03/01	51	08/06/00	03/07/00	25
05/12/00	04/10/01	303	05/03/01	19/03/01	14
13/02/02	02/10/02	229	22/08/00	15/09/00	24
22/05/01	25/06/01	34	06/11/01	27/11/01	21
30/04/01	06/06/01	37	14/04/00	24/04/00	10
30/03/01	19/09/01	204	31/10/00	08/11/00	8
08/06/01	01/10/01	126	28/02/00	12/04/00	43
20/03/03	22/04/03	33	13/12/00	08/01/01	26
06/07/01	26/10/01	112	05/03/01	19/03/01	14
18/01/02	23/08/02	215	07/08/00	25/10/00	79
03/07/01	26/11/01	145	14/09/00	30/10/00	46
23/05/03	28/07/03	66	15/06/00	07/07/00	22
30/06/02	17/09/02	79	16/03/02	18/05/02	63
20/05/02	08/07/02	49	28/11/02	08/12/02	10
20/03/01	10/05/01	51	14/06/01	09/09/01	86
20/01/02	13/02/02	24	03/10/02	27/12/02	84
16/04/01	11/01/02	250	31/05/01	02/07/01	32

Tabla VIII .- Datos de fechas de instalación y de primera llamada para los equipos de tecnología.

El tiempo hasta la falla de cada equipo es entonces igual al número de días T hasta el primer mal funcionamiento o descompostura.

Generalmente se reporta en la teoría estadística que este tipo de variables aleatorias siguen una distribución geométrica con parámetro p , el cual representa la probabilidad de éxito para esta distribución.

Paradójicamente el éxito en el estudio se produce cuando el equipo falla.

Esta hipótesis se verificará luego a través de la prueba de bondad de ajuste correspondiente.

Dado que para los equipos muestreados los tiempos hasta que ocurra la primera falla están en días, resulta conveniente estimar primero el promedio de días hasta que ocurra la primera falla, para así poder realizar la estimación de la probabilidad de éxito p .

Se conoce que una variable Y que sigue una distribución geométrica tiene como valor esperado:

$$E[Y] = \frac{1}{p}$$

Por lo tanto al estimar la media de los datos en días, \bar{t} , se obtiene el valor esperado $E[T]$ para reemplazarlo en la fórmula y de esta manera estimar el parámetro p :

$$\hat{p} = \frac{1}{\bar{t}}$$

Calculando la media para cada muestra se tiene:

$$\hat{\mu}_1 = \bar{t}_1 = 107.6 \qquad \hat{\mu}_2 = \bar{t}_2 = 36.6$$

Por tanto los parámetros estimados son:

$$\hat{p}_1 = \frac{1}{\hat{\mu}_1} = 0.0093 \approx 0.01 \quad \hat{p}_2 = \frac{1}{\hat{\mu}_2} = 0.0273 \approx 0.028$$

Una vez estimado el parámetro p para ambas líneas, se puede ampliar el tamaño de muestra realizando una simulación para cada línea, generando valores que sigan la misma distribución a partir de las probabilidades \hat{p}_1 y \hat{p}_2 halladas.

Para esta simulación, se generaron los números aleatorios implementando la combinación de dos algoritmos congruenciales MLCG (Multiplicative Linear Congruential Generator), que dan como resultado números aleatorios distribuidos uniformemente entre 0 y 1. Luego, a partir de estos números aleatorios y de los parámetros estimados, se generan los valores de las dos variables aleatorias geométricas.

Para cada caso se generaron 80 valores, aumentando así el tamaño de muestra a 100 en cada línea.

A continuación se presenta el algoritmo para la generación de estos valores para cada una de las variables aleatorias geométricas:

Algoritmo para la obtención de valores de una variable aleatoria geométrica con parámetro p .

Hay que tomar en cuenta que para la implementación del algoritmo es preciso interactuar con un archivo de base de datos en el cual se encontrará el valor de las semillas, el mismo que a su vez almacenará valores para nuevas semillas luego de cada generación de números aleatorios mediante la combinación de 2 MLCG's.

Por tanto los pasos para el algoritmo son los siguientes:

1. Se asigna el valor del parámetro p de la distribución geométrica, probabilidad de que el equipo falle, y se inicializa un contador en uno. En este caso el contador se refiere a la cantidad de días hasta la primera falla.
2. Se toma la semilla del archivo independiente de base de datos para la generación de un número aleatorio u entre 0 y 1.
3. Se genera el valor para el experimento de Bernoulli mediante la comparación del número aleatorio con el parámetro p fijado:

Sea u el número pseudo-aleatorio generado y x el valor del experimento Bernoulli 0 ó 1, donde el 1 representa la falla del equipo y el 0 quiere decir que el equipo se mantiene estable sin falla; se tiene entonces:

Si $0 \leq u \leq 1-p$, entonces $x = 0$, se incrementa el contador en uno, el algoritmo MLCG almacena automáticamente el nuevo valor para la semilla y se vuelve al paso 2.

Por otro lado, si $1-p \leq u \leq 1$, entonces $x = 1$, lo que significa que el equipo falló, tomando el valor del contador para obtener así el valor de la variable aleatoria geométrica.

Es necesario indicar que para la generación de la censura a la derecha para los equipos de ambas líneas, dado que son independientes del mecanismo de generación de los tiempos de vida de los equipos, primero se calculó la proporción de los equipos que conforman las unidades muestrales en cada línea sin importar el tiempo en que se "retiraron", para luego ubicar en el orden de generación las censuras a cada uno de los equipos.

Es decir que si en la simulación del mecanismo de censura, el primer valor obtiene censura, el primer valor generado será tomado en cuenta como censura a la derecha o como retirada.

El algoritmo para esta simulación es bastante parecido al anterior, ya que el mecanismo de censura para este análisis sigue también un experimento de Bernoulli con dos valores: 0 para indicar que existe censura (a la derecha) y 1 para indicar que el tiempo registrado es por falla y no por retirada o censura.

4.1.2. Contrastes de Hipótesis para pruebas de Bondad de Ajuste.

Para tener consistencia estadística de que los datos obtenidos en la empresa junto con los generados siguen una distribución geométrica con los parámetros establecidos, se realizará una prueba de Bondad de Ajuste.

Existen varias pruebas para analizar la bondad de ajuste, las dos más usadas son la prueba de bondad de ajuste basada en la distribución Ji cuadrado (χ^2) empleada para las poblaciones

discretas y continuas; y la prueba no paramétrica de Kolmogorv – Smirnov que funciona mejor que la anterior para poblaciones continuas.

Ambas pruebas se construyen a partir de un nivel de significancia α o error tipo I, que puede ser especificado también por el nivel de confianza con el que se trabaje, $100(1 - \alpha)\%$.

Para cada prueba de hipótesis se debe contar con una muestra aleatoria y con un valor p asociado. Este valor p es el nivel de significancia mínimo para rechazar la hipótesis nula, y según teoría estadística existe el criterio para aceptar lo postulado por la hipótesis nula cuando el valor p es mayor que 0.01

El contraste basado en la distribución Ji cuadrado se lo plantea de la siguiente manera:

H_0 : La distribución de la población es $F_0(x)$

H_1 : La distribución de la población NO es $F_0(x)$

$$EP: \chi^2 = \frac{\sum_{i=1}^m (n_i - np_{i0})^2}{np_{i0}}$$

RR: $\chi^2 > \chi^2_{\alpha}$ con $(m - 1 - 1)$ grados de libertad.

Donde m es el número de casillas, l es el número de parámetros a estimar.

En ambas líneas de producto se trabajará con 2 grados de libertad ya que se tuvo para las dos líneas $m = 4$ casillas y un parámetro a estimar. En el primer caso se tuvieron 4 intervalos (casillas) de longitud 50 y para el segundo caso 4 intervalos de longitud de 25 días.

Se obtuvieron los siguientes resultados:

H_0 : La distribución de los tiempos de falla para la línea de equipo 1 es Geométrica con $p = 0.01$

$H_1: \neg H_0$

Prueba de bondad de ajuste	Ji Cuadrado
Estadístico de prueba	0.3784
Grados de libertad	2
Valor p	0.8276

Tabla IX .- *Bondad de Ajuste Geométrico para los datos de la Línea de Equipo 1.*

Se puede notar que el valor p es muy elevado, cercano a 1, lo que indica que además de aceptar la hipótesis nula estos datos se ajustan muy bien a una distribución geométrica con el parámetro especificado.

Ahora, para la segunda línea de equipo se tiene:

H_0 : La distribución de los tiempos de falla para la línea de equipo 2 es Geométrica con $p = 0.028$

$H_1: \neg H_0$

Prueba de bondad de ajuste	Ji Cuadrado
Estadístico de prueba	3.6114
Grados de libertad	2
Valor p	0.1644

Tabla X .- *Bondad de Ajuste Geométrico para los datos de la Línea de de Equipo 2.*

En este caso, a pesar de que el valor p no es muy grande, se tiene que aplicando el criterio que este valor debe ser mayor a 0.01, se acepta la hipótesis de que los datos provienen de una distribución geométrica.

Para cada caso, se puede observar que con un 99% de confianza existe evidencia estadística para afirmar que los datos provienen de una distribución geométrica con los parámetros definidos anteriormente.

Estos resultados obtenidos sirven de ayuda para la obtención de la función de supervivencia y de riesgo de los datos.

4.1.3. Función de Supervivencia y Función de Riesgo.

Luego de tener consistencia estadística para afirmar que los datos se ajustan a una distribución geométrica, se pueden estimar las funciones de supervivencia y de riesgo a partir de su distribución acumulada.

Se tiene que la distribución de probabilidad de una variable aleatoria geométrica está dada por: $p(y) = q^{y-1}p$, donde $q = 1 - p$.

Por lo tanto la función de distribución acumulada viene dada por:

$$F(y) = \sum_{i=1}^y q^{i-1} p$$

Encontrando una expresión para esta función de distribución acumulada se tiene:

$$F(y) = \sum_{i=1}^y q^{i-1} p = p \sum_{i=1}^y q^{i-1} = p \sum_{i=0}^{y-1} q^i$$

Resolviendo la sumatoria:

$$\sum_{i=0}^{y-1} q^i = \sum_{i=0}^{\infty} q^i - \sum_{i=y}^{\infty} q^i$$

Ya que q es una probabilidad, cumple la restricción: $|q| < 1$,

entonces la primera sumatoria $\sum_{i=0}^{\infty} q^i$ converge a $\frac{1}{1-q}$. Ahora

trabajando la segunda sumatoria, y efectuando el cambio de variable: $u = i - y$

$$\sum_{i=y}^{\infty} q^i = \sum_{u+y=y}^{\infty} q^{u+y} = q^y \sum_{u=0}^{\infty} q^u = q^y \left(\frac{1}{1-q} \right)$$

Por tanto la expresión para $F(Y)$ queda de la siguiente manera:

$$F(y) = p \left[\left(\frac{1}{1-q} \right) - q^y \left(\frac{1}{1-q} \right) \right]$$

$$= p \left[\left(\frac{1}{1-q} \right) (1 - q^y) \right]$$

$$= p \left(\frac{1}{p} \right) (1 - q^y)$$

$$= 1 - q^y$$

Hallada esta expresión en términos de la variable aleatoria, se puede definir ya la función de supervivencia para el caso de las variables en estudio; teniendo en cuenta que:

$$S(y) = 1 - F(y) = 1 - (1 - q^y) = q^y$$

Puede notarse que la función de supervivencia para una variable aleatoria geométrica es simplemente una potencia de la probabilidad de fracaso.

Por otro lado, la función de riesgo viene dada por:

$$h(t) = \frac{p(t)}{1 - F(t)} = \frac{q^{t-1}p}{q^t} = q^{-1}p$$

Como se observa, se tiene una función de riesgo constante, ya que esta no depende del tiempo t .

	Línea de Equipo 1	Línea de Equipo 2
Función de Supervivencia	$S_1(t) = (0.99)^t$	$S_2(t) = (0.972)^t$
Función de Riesgo	$h_1(t) = (0.99)^{-1}(0.01)$	$h_2(t) = (0.972)^{-1}(0.028)$

Tabla XI .- *Funciones de Supervivencia y de Riesgo a partir de la distribución Geométrica*

Un hecho interesante resulta cuando se plantea la siguiente premisa:

Dado que la función de riesgo es constante, y que los datos fueron obtenidos de forma discreta, se puede construir un modelo de riesgo constante a partir de $h(t) = w = (0.99)^{-1}(0.01) = 0.0101 = \frac{1}{\beta}$

teniéndose así que $\beta = 100$ aproximadamente, para la línea de equipo 1 y $h(t) = w = (0.972)^{-1}(0.028) = 0.0288 = \frac{1}{\beta}$ con aproximadamente $\beta = 35$ para la línea de equipo 2.

Como se expuso en el capítulo 2, cuando se tiene un riesgo constante de la forma $h(y) = \frac{1}{\beta}$, la variable tiempo de falla T sigue un modelo de probabilidad exponencial.

Aquí surge la pregunta si los datos ajustados por una distribución discreta (geométrica), pueden ajustarse también a una distribución continua como lo es la exponencial. Para realizar una prueba de bondad de ajuste con una distribución continua resulta conveniente emplear la prueba no paramétrica de Kolmogorov – Smirnov (K-S).

En este contraste la hipótesis nula (H_0) y la hipótesis alternativa (H_1) son planteadas de igual forma que en el contraste basado en la distribución Ji Cuadrado, pero el estadístico de prueba y la región de rechazo toman una nueva forma:

H_0 : La distribución de la población es $F_0(x)$

H_1 : La distribución de la población NO es $F_0(x)$

EP: $D = \max|F_0(x) - F_n(x)|$

RR: $D > D_0(\alpha)$

Donde $D_0(\alpha)$ son valores tabulados y $F_n(x)$ es la distribución empírica de la muestra ordenada.

Realizando la prueba para ambas líneas, teniendo en cuenta los parámetros estimados a partir de la función de riesgo constante se obtuvo lo siguiente:

H_0 : La distribución de los tiempos de falla para la línea de equipo 1 es Exponencial con $\beta = 100$

$H_1: \neg H_0$

Prueba de bondad de ajuste	K - S
Estadístico de prueba	0.0494
Valor p	0.9676

Tabla XII .- *Bondad de Ajuste Exponencial para los datos de la Línea de Equipo 1.*

H_0 : La distribución de los tiempos de falla para la línea de equipo 2 es Exponencial con $\beta = 35$.

$H_1: \neg H_0$

Prueba de bondad de ajuste	K - S
Estadístico de prueba	0.1207
Valor p	0.1084

Tabla XIII .- *Bondad de Ajuste Exponencial para los datos de la Línea de Equipo 2.*

Como se puede observar, para la primera línea el valor p es muy elevado, incluso más que para el ajuste a la distribución geométrica, lo que confirma que existe evidencia estadística para afirmar que los datos provienen de una distribución exponencial. A pesar de que el valor p para la segunda línea apenas supera el

valor crítico del 0.01, se puede afirmar estadísticamente que los datos para este caso siguen una distribución exponencial con el parámetro especificado.

Esto tiene sentido ya que los valores de una variable discreta están contenidos dentro de los de una variable continua, y además la prueba no paramétrica de Kolmogorov - Smirnov trabaja con la máxima diferencia entre el valor observado y el valor esperado.

Al tener ya consistencia estadística de que se puede contar con un modelo de probabilidad exponencial, el indicador de confiabilidad en cada caso sería igual a $MTTF = \beta$, y los modelos de supervivencia y de riesgo para la variable aleatoria T pueden ser escritos también como:

	Línea de Equipo 1	Línea de Equipo 2
Función de Densidad	$f_1(t) = \frac{1}{100} e^{-\frac{t}{100}}$	$f_2(t) = \frac{1}{35} e^{-\frac{t}{35}}$
Función de Supervivencia	$S_1(t) = e^{-\frac{t}{100}}$	$S_2(t) = e^{-\frac{t}{35}}$
Función de Riesgo	$h_1(t) = \frac{1}{100}$	$h_2(t) = \frac{1}{35}$

Tabla XIV.- Funciones de Densidad, de Supervivencia y de Riesgo a partir de la distribución Exponencial

4.1.4. Modelo Multinomial.

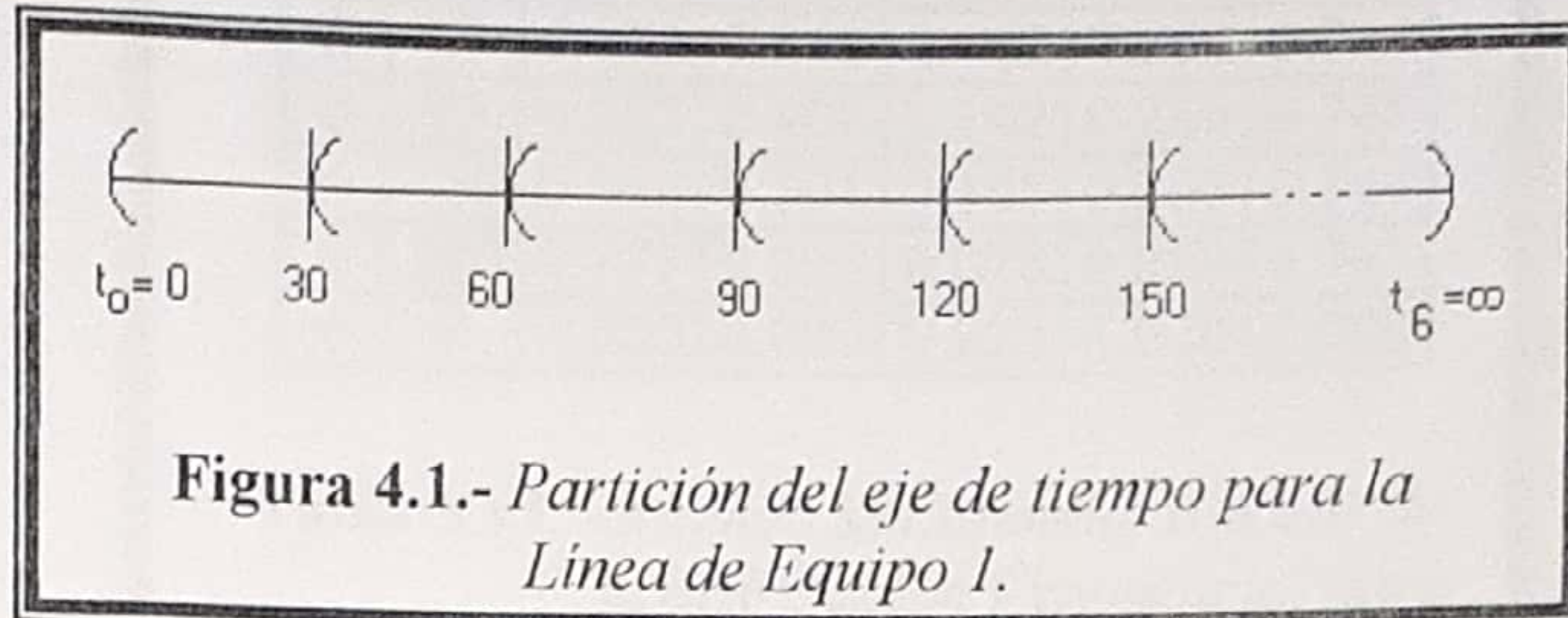
Para obtener el modelo multinomial de los datos de tiempos de falla es necesario particionar el eje de tiempo el cual está dado en días.

La partición para la línea 1 estará dada en seis intervalos de treinta días, con excepción del último intervalo que tendrá una longitud infinita. Para la línea 2 se tendrán igualmente cinco intervalos pero de quince días y un sexto de longitud infinita.

Una vez particionado el eje se procede a calcular las correspondientes probabilidades del modelo y la respectiva estimación de la función de supervivencia para cada intervalo.

Con el fin de poder realizar una comparación entre las dos funciones de distribución acumulada, se realizará el cálculo correspondiente tanto para la distribución geométrica como para la distribución exponencial.

Para la Línea de Equipo 1



Con la distribución Geométrica ($p = 0.01$)

i	t_i	$F(t_i)$	$S(t_i)$	π_i	p_i	$1-p_i$
0	0	0.000	1.000	-	-	-
1	30	0.260	0.739	0.260	0.260	0.740
2	60	0.453	0.547	0.193	0.260	0.740
3	90	0.595	0.405	0.142	0.260	0.740
4	120	0.701	0.299	0.105	0.260	0.740
5	150	0.778	0.221	0.078	0.260	0.740
6	∞	1.000	0.000	0.221	1.000	0.000

Tabla XV.- Modelo Multinomial a partir de la distribución Geométrica para la Línea de Equipo 1.

Con la distribución Exponencial ($\beta = 100$)

i	t_i	$F(t_i)$	$S(t_i)$	π_i	p_i	$1 - p_i$
0	0	0.000	1.000	-	-	-
1	30	0.259	0.741	0.259	0.259	0.741
2	60	0.451	0.549	0.192	0.259	0.741
3	90	0.593	0.407	0.142	0.259	0.741
4	120	0.699	0.301	0.105	0.259	0.741
5	150	0.777	0.223	0.078	0.259	0.741
6	∞	1.000	0.000	0.223	1.000	0.000

Tabla XVI. - Modelo Multinomial a partir de la distribución Exponencial para la Línea de Equipo 1.

Como se podría haber esperado, existe gran similitud entre ambas funciones de supervivencia para cada intervalo, ya que la diferencia existente entre estos valores es de milésimas.

Para la Línea de Equipo 2

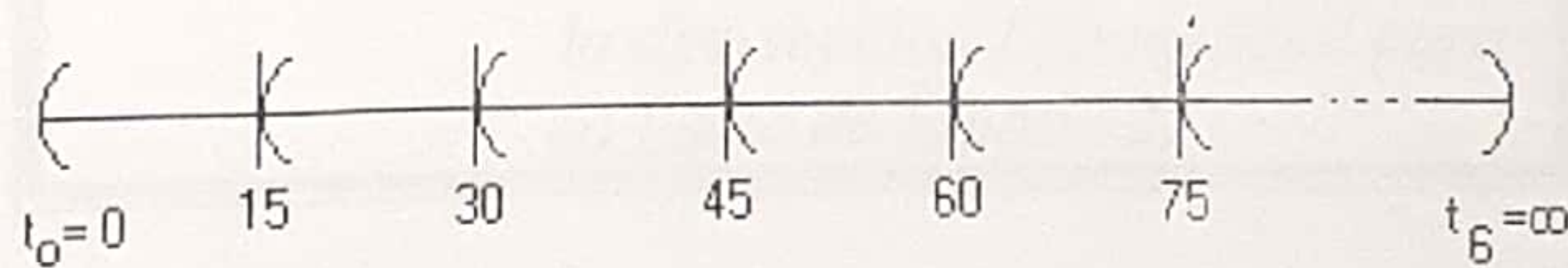


Figura 4.2. - Partición del eje de tiempo para la Línea de Equipo 2.

Con la distribución Geométrica ($p = 0.028$)

i	t_i	$F(t_i)$	$S(t_i)$	π_i	p_i	$1 - p_i$
0	0	0.000	1.000	-	-	-
1	15	0.347	0.653	0.347	0.347	0.653
2	30	0.573	0.426	0.227	0.347	0.653
3	45	0.721	0.279	0.148	0.347	0.653
4	60	0.818	0.182	0.097	0.347	0.653
5	75	0.881	0.119	0.063	0.347	0.653
6	∞	1.000	0.000	0.112	1.000	0.000

Tabla XVII .- Modelo Multinomial a partir de la distribución Geométrica para la Línea de Equipo 2.

Con la distribución Exponencial ($\beta = 34.7$)

i	t_i	$F(t_i)$	$S(t_i)$	π_i	p_i	$1 - p_i$
0	0	0.000	1.000	-	-	-
1	15	0.351	0.649	0.351	0.351	0.649
2	30	0.579	0.421	0.228	0.351	0.649
3	45	0.727	0.273	0.148	0.351	0.649
4	60	0.823	0.177	0.096	0.351	0.649
5	75	0.885	0.115	0.062	0.351	0.649
6	∞	1.000	0.000	0.112	1.000	0.000

Tabla XVIII .- Modelo Multinomial a partir de la distribución Exponencial para la Línea de Equipo 2.

Al igual que en el caso anterior, la diferencia para los valores de la función de supervivencia es mínima.

Nótese además que en todos casos, la probabilidad condicional de que un equipo falle en el intervalo i dado que sobrevivió al intervalo $i - 1$, (p_i) , es la misma para cada intervalo, debido no solo a la longitud uniforme de los intervalos sino también a la forma que adopta cada una de las funciones de distribución acumulada.

Cabe recalcar que como se dijo en capítulos anteriores, la sumatoria de los π_i debe ser igual a 1, como en efecto sucede.

Ahora, surge otro hecho importante, para estos dos casos paramétricos no se han tomado en cuenta las censuras como tales, sino más bien como tiempos de falla, por esto la conveniencia de obtener un modelo que explique mejor el fenómeno que presentan los datos.

Para este es necesario valerse de un paquete de software estadístico como lo es MINITAB, el cual a partir de una estimación no paramétrica diagnostica el fenómeno y brinda un modelo paramétrico que mejor se ajusta a los datos.

A continuación se presentarán los métodos y estimadores no paramétricos de confiabilidad para luego presentar un modelo

paramétrico que por lo general sigue uno de Weibull, ya que este es el que explica un 80 o 90% de los casos.

4.1.5. Estimación No Paramétrica.

Una vez realizadas las respectivas estimaciones paramétricas, es necesario hallar estimadores para la función de supervivencia que no tengan que sujetarse a supuestos ni restricciones de cálculos paramétricos.

En esta sección se realizará una estimación no paramétrica de la función de supervivencia para las dos líneas de equipos, calculando las Tablas de vida para cada caso.

Como se mencionó anteriormente, el análisis se basará en 5 inspecciones, para el equipo 1 cada treinta días y para el equipo 2 cada quince. De esta manera existirá censura a la derecha para equipos que no fallen durante estos 150 días y 75 días respectivamente.

Cabe indicar que los intervalos para la función de supervivencia evaluada en los distintos tiempos, fueron hallados basándose en la fórmula de Greenwood, con una aproximación normal y con un 95% de confianza.

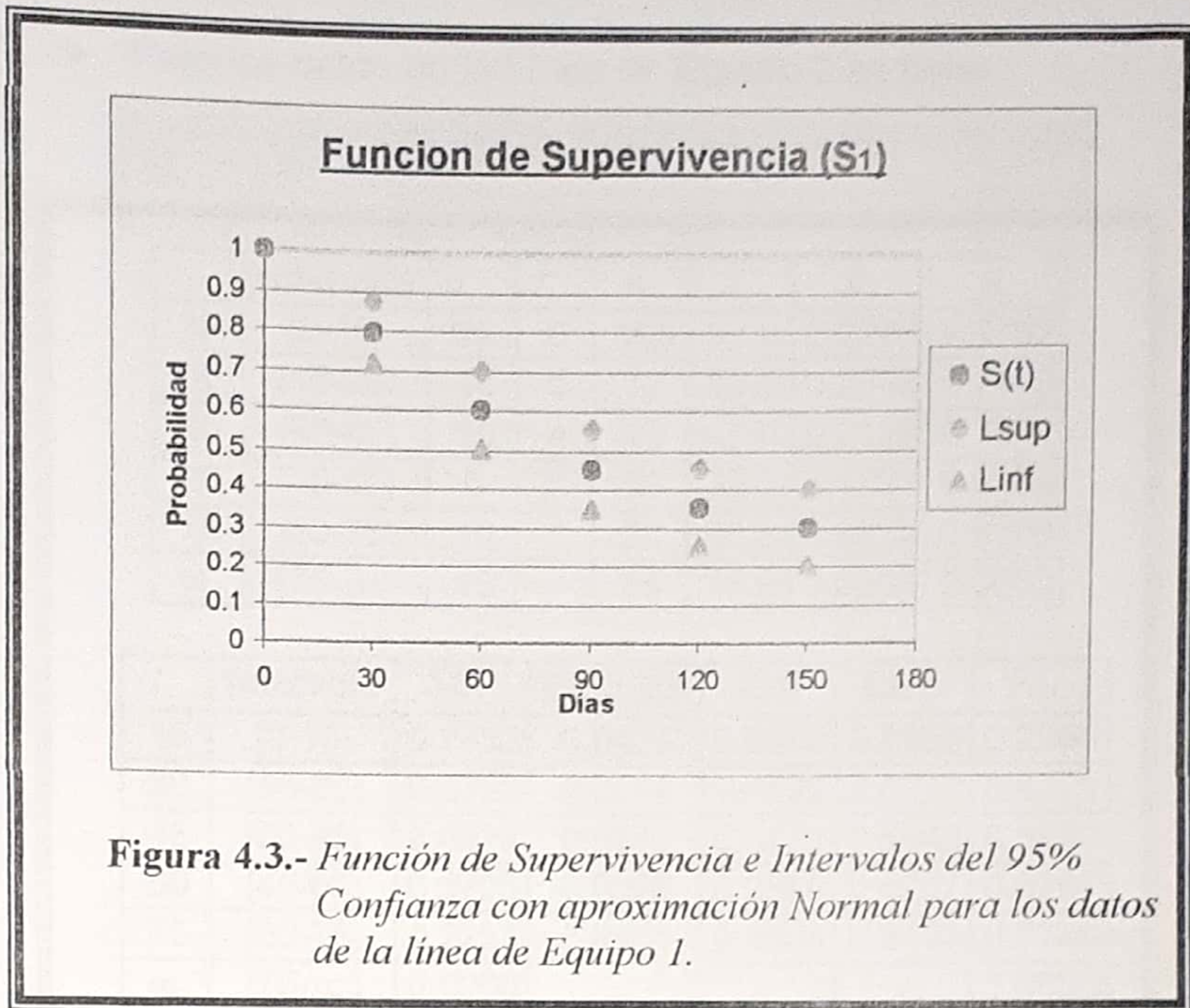
A continuación se presentan las tablas de vida para cada línea de equipo:

- Para los datos de la Línea de Equipo 1 se tiene:

t_i	Intervalo	d_i	r_i	n_i	n_i'	\hat{p}_i	\hat{q}_i
30	[0-30)	20	4	100	98.0	0.2041	0.7959
60	[30-60)	18	5	76	73.5	0.2449	0.7551
90	[60-90)	13	2	53	52.0	0.2500	0.7500
120	[90-120)	8	3	38	36.5	0.2192	0.7808
150	[120-150)	4	2	27	26.0	0.1538	0.8462
∞	[150- ∞)	21	-	21	21.0	1.0000	0.0000

t_i	Intervalo	$\hat{S}(t_i)$	Error Est	Linf	Lsup	$\hat{F}(t_i)$
30	[0-30)	0.7959	0.0407	0.7161	0.8757	0.2041
60	[30-60)	0.6010	0.0504	0.5022	0.6998	0.3990
90	[60-90)	0.4507	0.0523	0.3483	0.5532	0.5493
120	[90-120)	0.3520	0.0512	0.2517	0.4522	0.6480
150	[120-150)	0.2978	0.0499	0.1999	0.3957	0.7022
∞	[150- ∞)	0.0000	-	-	-	1.0000

Tabla XIX .- *Tabla de Vida para los datos de la Línea de Equipo 1.*



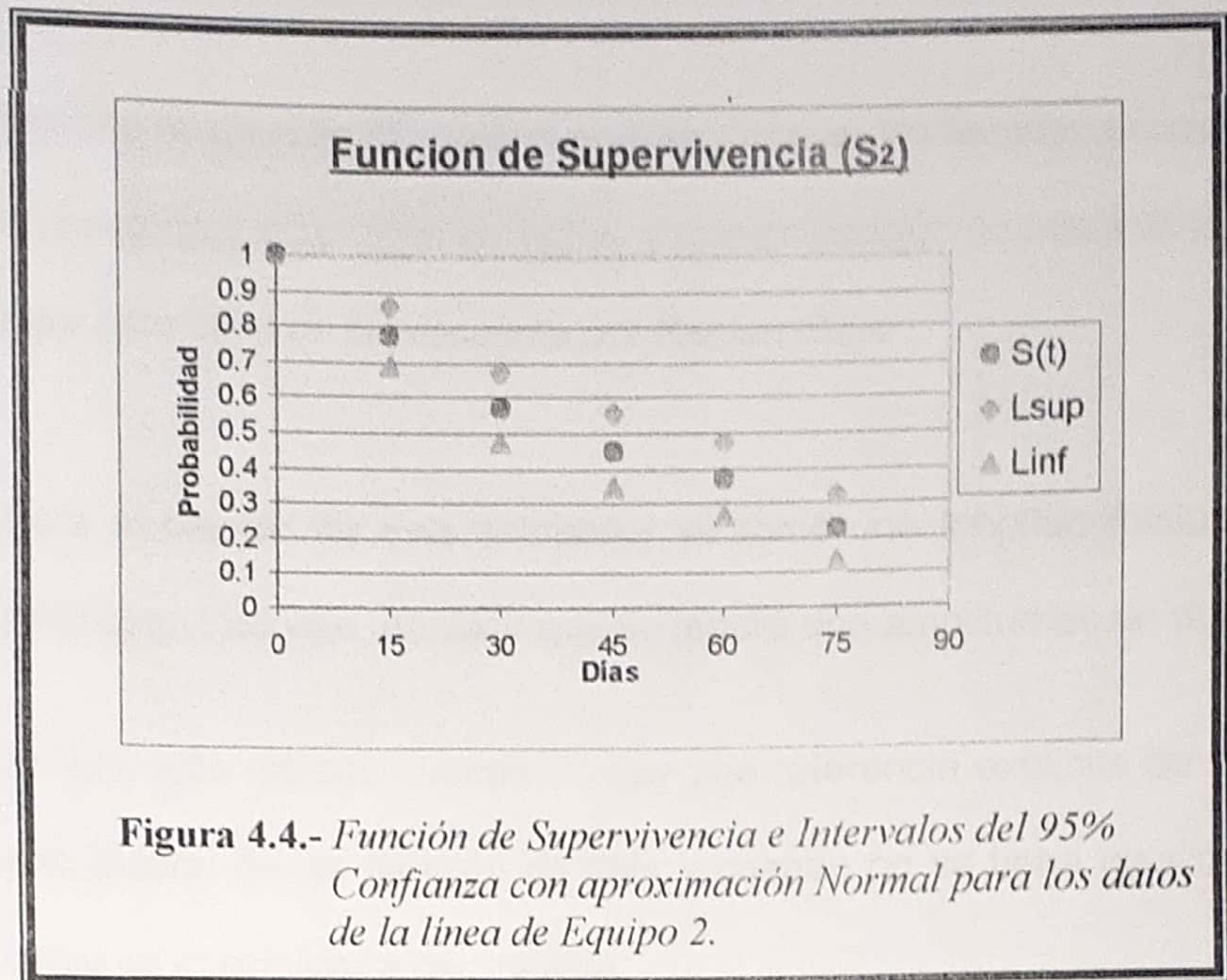
En el gráfico se muestra la estimación no paramétrica de la función de supervivencia en cada instante de inspección t_i de la primera línea con sus respectivos intervalos de confianza del 95%. Dado que se utilizó la aproximación normal para los intervalos de confianza, estos son simétricos.

- Para los datos de la Línea de Equipo 2 se tiene:

t_i	Intervalo	d_i	r_i	n_i	n_i'	\hat{p}_i	\hat{q}_i
15	[0-15)	22	6	100	97.0	0.2268	0.7732
30	[15-30)	18	5	72	69.5	0.2590	0.7410
45	[30-45)	10	4	49	47.0	0.2128	0.7872
60	[45-60)	6	3	35	33.5	0.1791	0.8209
75	[60-75)	10	1	26	25.5	0.3922	0.6078
∞	[75- ∞)	15	-	15	15.0	1.0000	0.0000

t_i	Intervalo	$\hat{S}(t_i)$	Error Est	Linf	Lsup	$\hat{F}(t_i)$
15	[0-15)	0.7732	0.0425	0.6899	0.8565	0.2268
30	[15-30)	0.5729	0.0514	0.4722	0.6737	0.4271
45	[30-45)	0.4510	0.0530	0.3472	0.5549	0.5490
60	[45-60)	0.3703	0.0528	0.2668	0.4737	0.6297
75	[60-75)	0.2251	0.0481	0.1308	0.3193	0.7749
∞	[75- ∞)	0.0000	-	-	-	1.0000

Tabla XX .- *Tabla de Vida para los datos de la Línea de Equipo 2.*



En el gráfico se muestra la estimación no paramétrica de la función de supervivencia en cada instante de inspección t_i de la primera línea con sus respectivos intervalos de confianza del 95%. Dado que se utilizó la aproximación normal para los intervalos de confianza, estos son simétricos.

4.1.5.1. Estimación de Kaplan-Meier.

Resulta de gran ayuda para el análisis conocer los tiempos exactos en los cuales ocurrieron las fallas, y una estimación no paramétrica para este caso es la propuesta por Kaplan-Meier.

Para el cálculo de este estimador se tomó una longitud mínima para cada intervalo, es decir que se tendrá una amplitud de un día.

Se usa este método cuando no hay una referencia explícita de la distribución de los tiempos de falla y cuando no se tiene idea de cómo es la estructura del sistema.

La aplicación de este método requiere lo siguiente:

- Los datos generados por cada equipo es el período de tiempo transcurrido desde el momento de origen (encendido del equipo, inicio de la navegación, etc) hasta el momento de fin (se presento la falla)
- Se usan datos censurados, es decir, si n sistemas de los cuales fallaron K en el intervalo $(0, t)$, existe censura para los demás $n-k$

sistemas que no fallaron, es decir que sobrevivieron hasta el termino del período de medición.

- Se considera que el tiempo t^* de censura es fijo

No se usa el número en riesgo efectivo, ya que el método plantea la reducción máxima conveniente para la amplitud de los intervalos.

Como resultado del análisis se tiene la curva de supervivencia o la curva estimada de confiabilidad. Precisamente a partir del método de estimación de Kaplan – Meier se puede obtener un modelo paramétrico de Weibull, modelo que explica la mayoría de los casos, el mismo que contemple las censuras y el fenómeno en sí de los datos obtenidos.

A continuación se presentan los resultados obtenidos para el estimador de Kaplan – Meier:

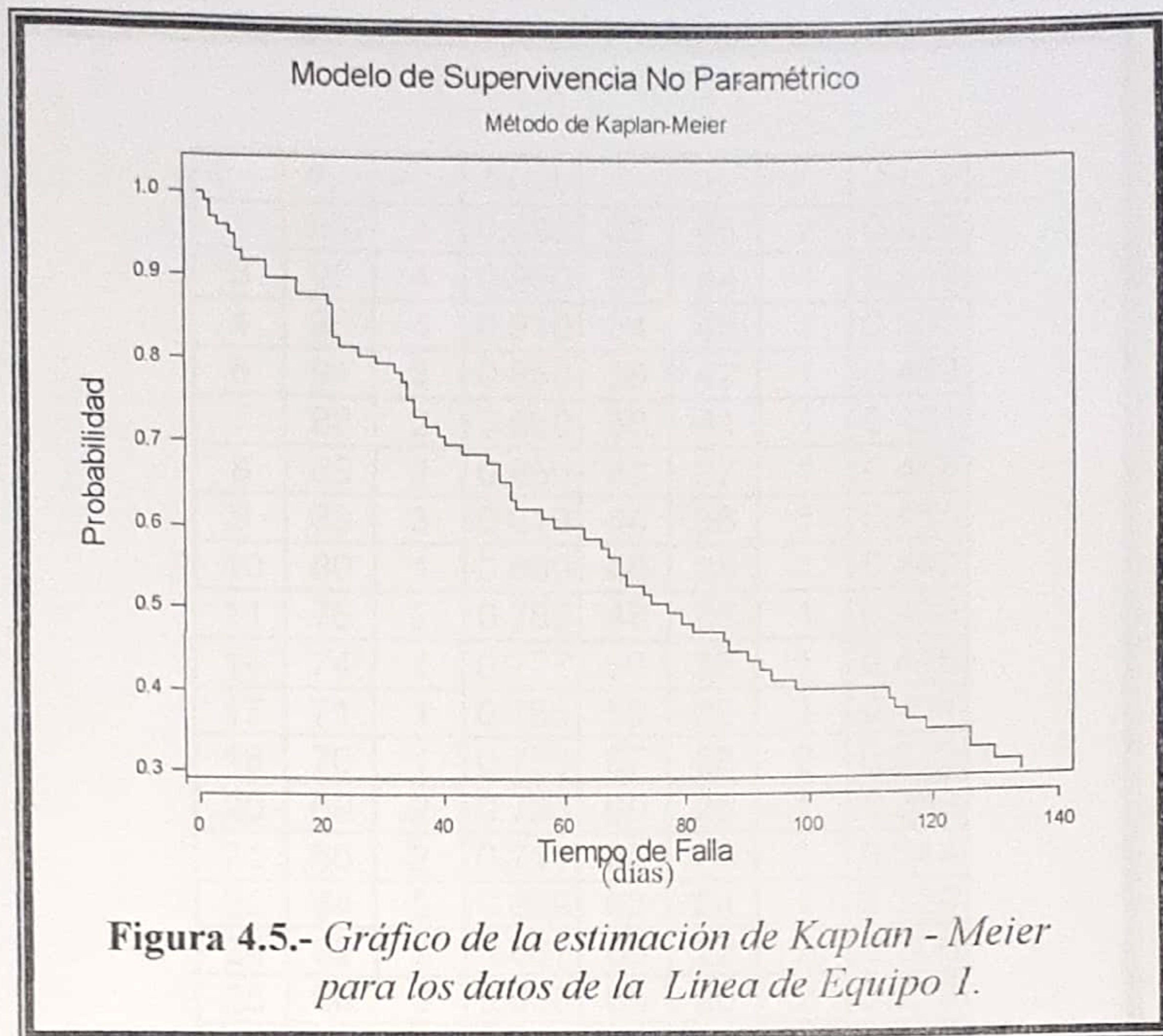
- Para los datos de la Línea de Equipo 1 se tiene:

t_i	n_i	d_i	$S(t_i)$	t_i	n_i	d_i	$S(t_i)$
1	100	1	0.990	56	55	1	0.610
2	99	2	0.970	58	54	1	0.599
3	97	1	0.960	63	52	1	0.587
5	95	1	0.950	66	51	1	0.578
6	93	2	0.929	67	50	1	0.564
7	91	1	0.919	69	49	2	0.541
11	90	2	0.899	70	47	1	0.530
16	88	2	0.878	73	46	1	0.518
21	86	1	0.868	74	44	1	0.508
22	85	4	0.827	77	43	1	0.495
23	81	1	0.817	79	42	1	0.483
26	79	1	0.807	81	41	1	0.471
29	77	1	0.796	86	40	1	0.459
32	75	1	0.786	87	39	1	0.448
33	74	1	0.775	90	38	1	0.436
34	72	2	0.754	92	37	1	0.424
35	70	2	0.732	94	36	1	0.412
37	68	1	0.721	98	35	1	0.400
39	67	1	0.710	113	32	1	0.388
40	66	1	0.700	114	31	1	0.375
43	63	1	0.689	116	30	1	0.363
47	62	1	0.677	119	28	1	0.350
49	61	2	0.655	126	27	2	0.324
51	58	2	0.633	130	24	1	0.310
52	56	1	0.621	134	23	1	0.297

$MTTF$	Error Estándar	Intervalo Confianza 95%	
		Inferior	Superior
79.3953	4.9701	69.65	89.14

Tabla XXI.- Estimación de Kaplan - Meier para los datos de la Línea de Equipo 1.

Como se observa el indicador $MTTF$, el tiempo promedio de falla es aproximadamente 79.4



En la gráfica, se puede notar claramente los tramos en los cuales la función de supervivencia es constante.

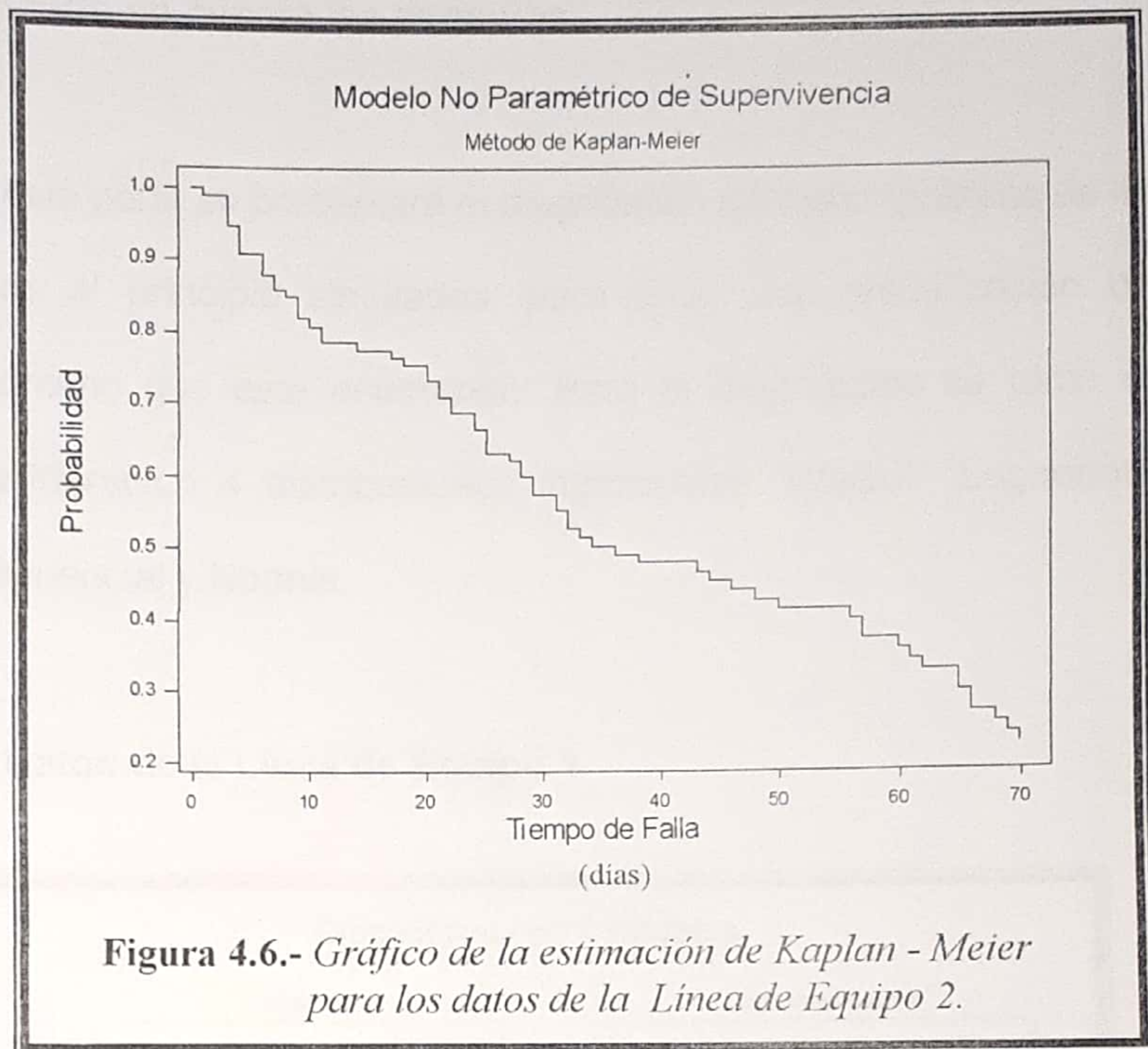
- Para los datos de la Línea de Equipo 2 se tiene:

t_i	n_i	d_i	$S(t_i)$	t_i	n_i	d_i	$S(t_i)$
1	100	1	0.990	32	46	2	0.529
3	99	4	0.950	33	44	1	0.517
4	95	4	0.910	34	43	1	0.505
6	91	3	0.880	36	42	1	0.493
7	88	2	0.860	38	41	1	0.481
8	85	1	0.850	43	37	1	0.468
9	83	3	0.819	44	36	1	0.455
10	80	1	0.809	46	35	1	0.442
11	76	2	0.788	48	33	1	0.428
14	74	1	0.777	50	32	1	0.415
17	71	1	0.766	56	30	1	0.401
18	70	1	0.755	57	28	2	0.372
20	69	2	0.733	60	26	1	0.358
21	66	2	0.711	61	25	1	0.344
22	64	2	0.689	62	24	1	0.329
24	62	2	0.667	65	22	2	0.300
25	60	3	0.633	66	20	2	0.270
27	56	1	0.622	68	18	1	0.255
28	55	2	0.599	69	17	1	0.240
29	52	2	0.576	70	16	1	0.225
31	49	2	0.553	-	-	-	-

$MTTF$	Error Estándar	Intervalo Confianza 95%	
		Inferior	Superior
40.21	2.6243	35.07	45.35

Tabla XXII .- Estimación de Kaplan - Meier para los datos de la Línea de Equipo 2.

Para la segunda línea, el indicador para el tiempo promedio de falla es aproximadamente 40.21



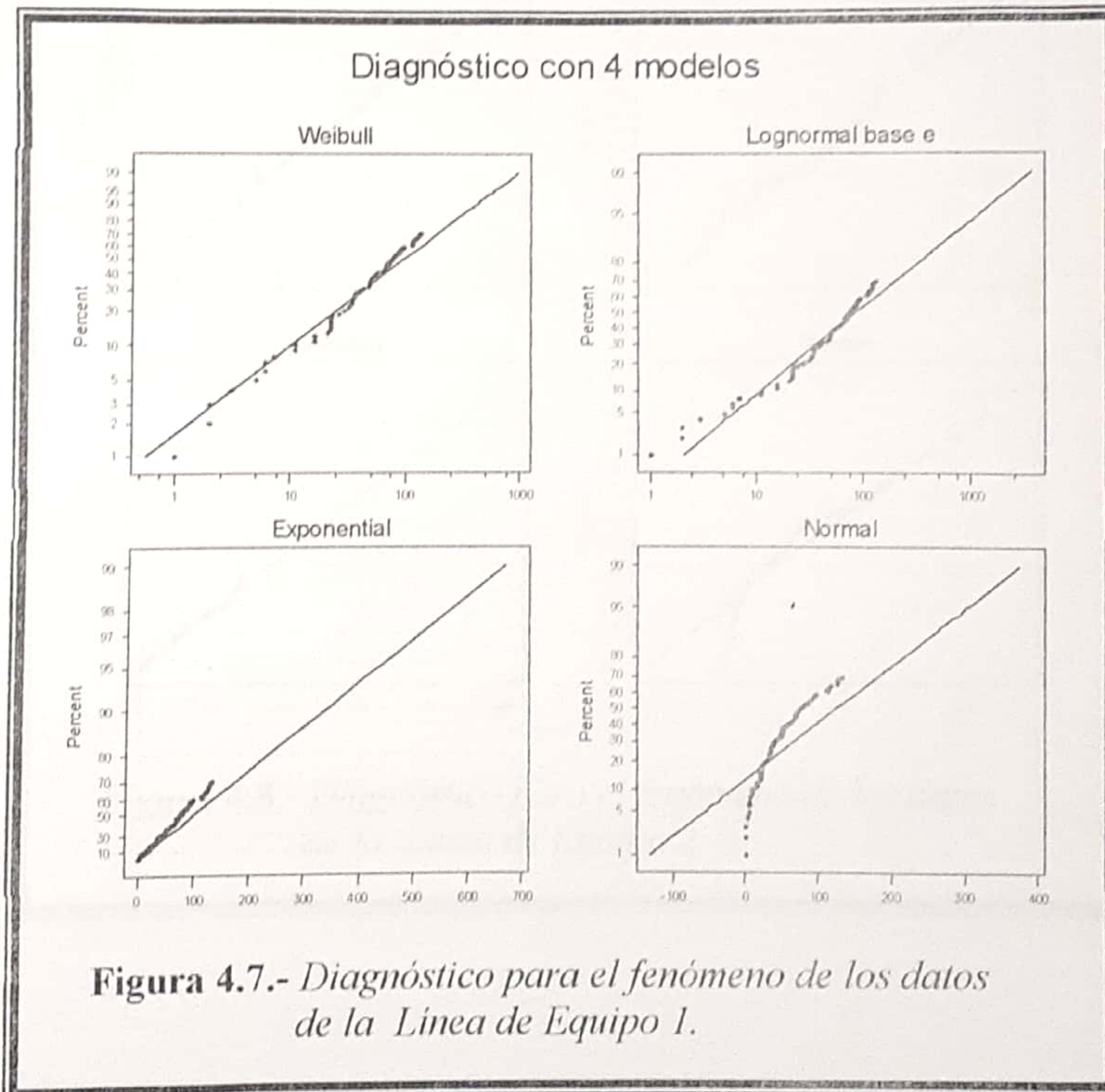
Para esta estimación basta trabajar solo con el número en riesgo, ya que los intervalos son de mínima longitud (igual a un día) y la ubicación de las unidades censuradas dentro de cada intervalo no influyen mayormente en el resultado del estimador.

Una vez analizado el estimador de Kaplan - Meier, se puede adoptar a partir del mismo un modelo paramétrico que se ajuste a los datos.

4.1.6. Diagnóstico para ajuste a un modelo paramétrico tomando en cuenta las censuras.

En esta parte se presentará el diagnóstico mediante gráficos de los datos al principio simulados, para tener una visualización del fenómeno que está ocurriendo. Para el diagnóstico se tomó en consideración 4 distribuciones importantes: Weibull, Lognormal, Exponencial y Normal.

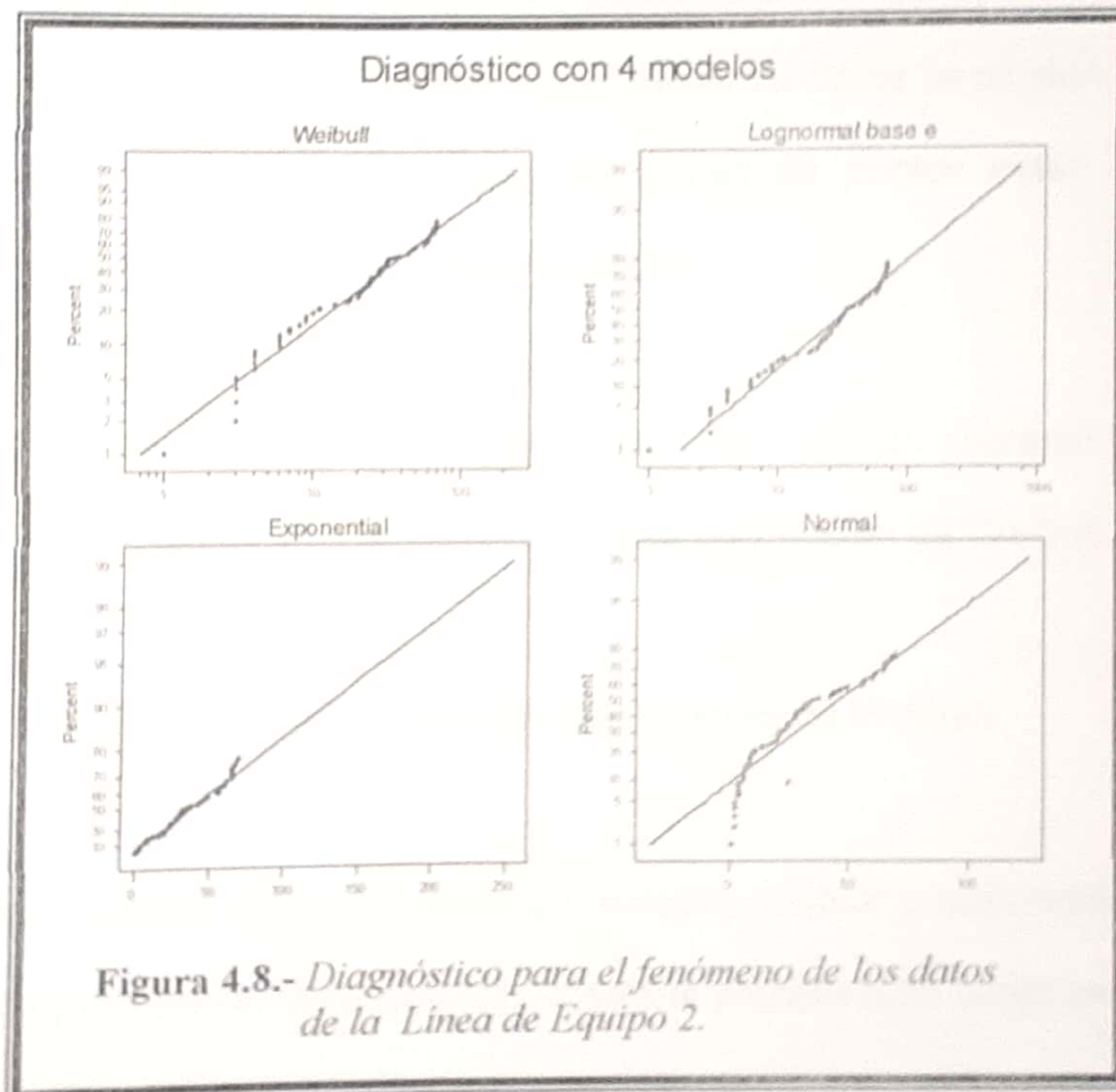
☞ Datos de la Línea de Equipo 1



Distribución	MTTF	Error Estándar	Intervalo Confianza 95%	
			Inferior	Superior
Weibull	165.900	29.830	116.626	235.994
Lognormal	317.837	108.532	162.759	620.676
Exponencial	144.952	18.262	113.236	185.552
Normal	122.087	12.070	98.431	145.743

Tabla XXIII.-Valores de los Tiempos Medios de Falla para las 4 Distribuciones con los datos de la Línea de Equipo 1.

☞ **Datos de la Línea de Equipo 2**



Distribución	MTTF	Error Estándar	Intervalo Confianza 95%	
			Inferior	Superior
Weibull	54.367	6.901	42.392	69.725
Lognormal	82.701	19.328	52.310	130.750
Exponencial	55.803	6.869	43.840	71.029
Normal	46.274	3.764	38.890	53.652

Tabla XXIV.- Valores de los Tiempos Medios de Falla para las 4 Distribuciones con los datos de la Línea de Equipo2.

Como se dijo anteriormente, se puede observar que, para ambos casos y tomando en cuenta las censuras, los datos se ajustan muy bien a una Weibull, ya que las nubes de puntos están muy cercanas a la curva de esta distribución.

Una vez obtenido el diagnóstico, es preciso obtener los estimadores de los parámetros de esta distribución de Weibull.

Estimación de los parámetros del modelo de Weibull.

Para realizar esta estimación se necesita el valor o estimación de $S(t)$ y los valores de los tiempos hasta la primera falla; luego se usa la regresión de Weibull para la estimación de los parámetros α y β .

El estimador de supervivencia de Kaplan – Meier es utilizado para la estimación de $S(t)$.

Luego, partiendo del modelo de confiabilidad o supervivencia de Weibull, el cual viene dado por:

$$S(t) = \exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right]$$

se obtiene por despeje la siguiente igualdad:

$$\ln(-\ln(S(t))) = \alpha \ln(TTF) - \alpha \ln \beta$$

donde TTF representa los tiempos de falla.

Se puede observar que la forma de la ecuación está estructurada de manera similar al de un modelo de regresión lineal simple

$Y = \alpha X + C$, donde:

$$Y = \ln(-\ln R(t))$$

$$X = \ln(TTF_i)$$

$$C = -\alpha \ln \beta$$

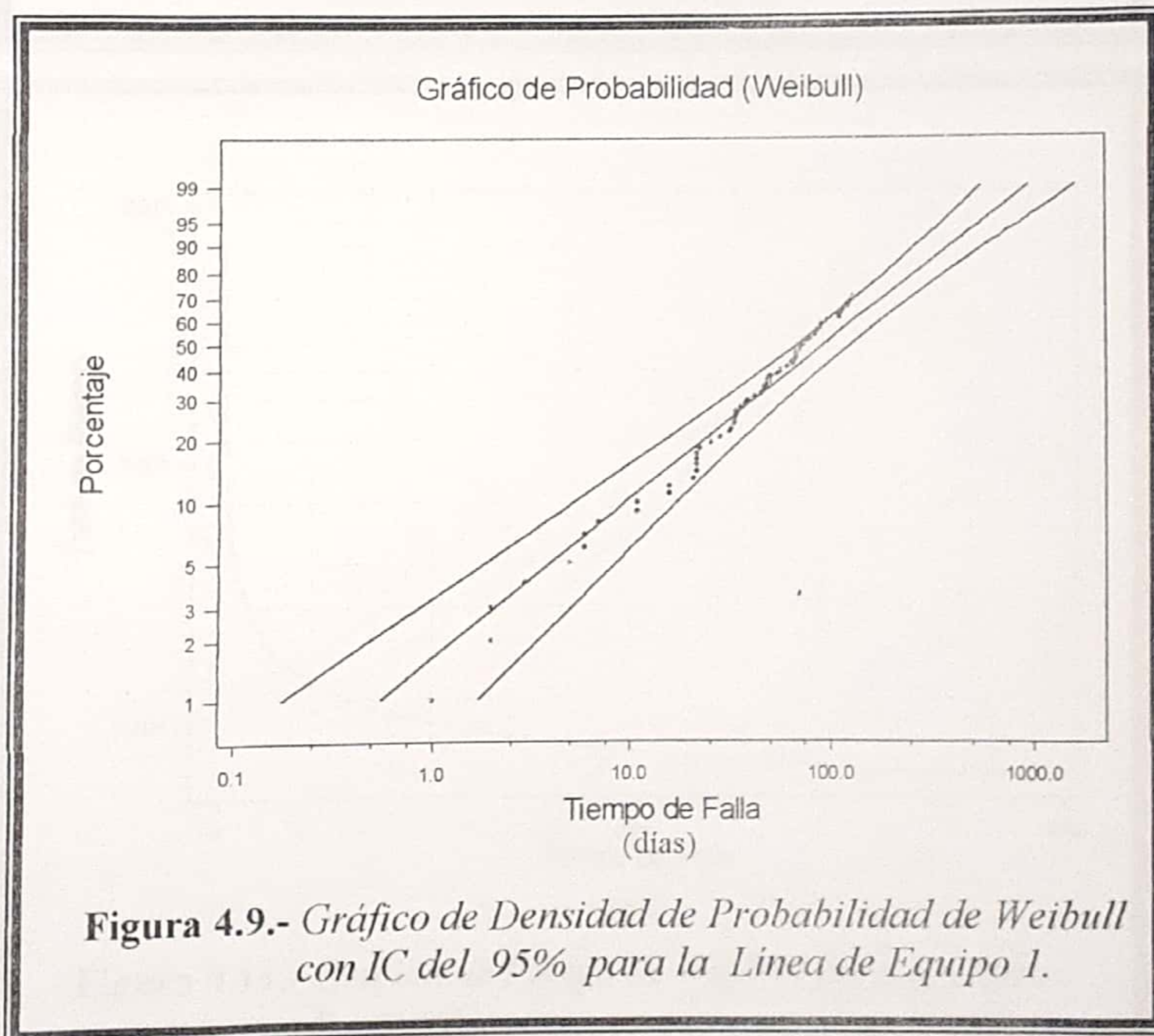
Una vez que se estimados estos parámetros, se procederá a graficar las funciones de riesgo y de supervivencia para cada caso.

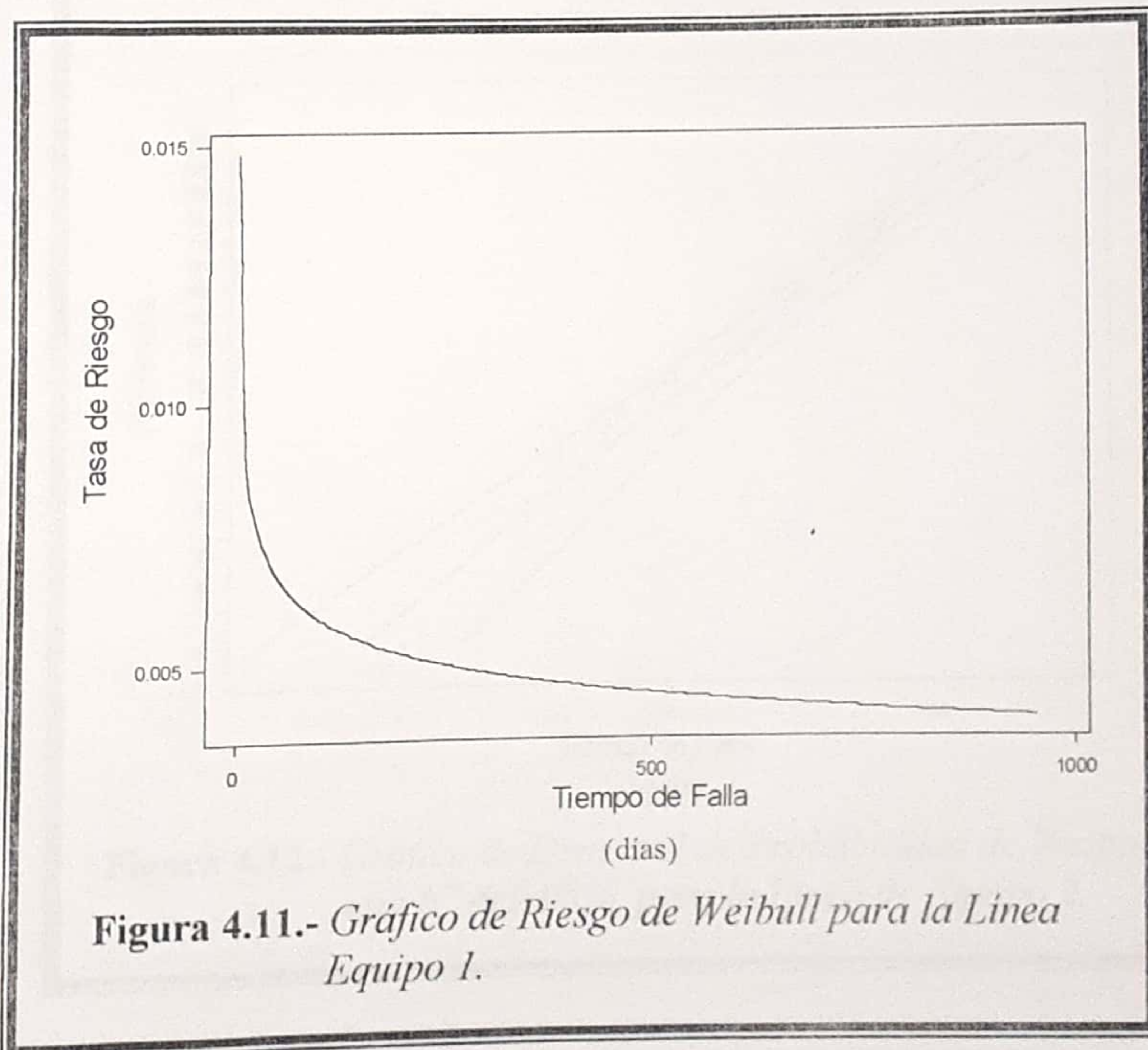
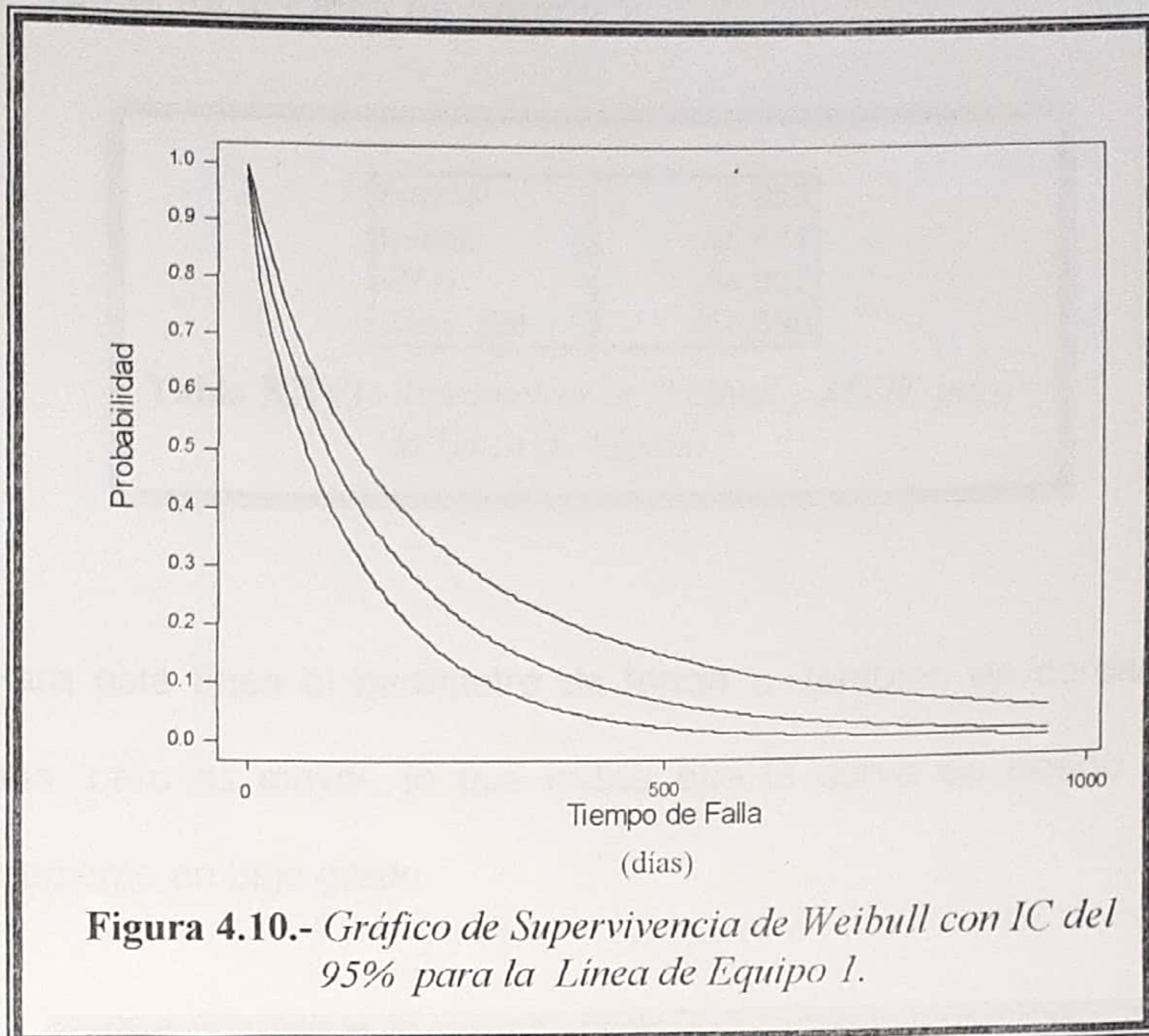
☞ Datos de la Línea de Equipo 1.

Forma	0.82
Escala	149.32
<i>MTTF</i>	165.90
Desv. Est.	202.87

Tabla XXV.- *Parámetros de Weibull y MTTF para la Línea de Equipo 1.*

El parámetro de forma α es cercano a uno, lo que indica que no se aleja mucho de una distribución exponencial. Sin embargo el *MTTF* es muy superior al encontrado con la distribución exponencial.



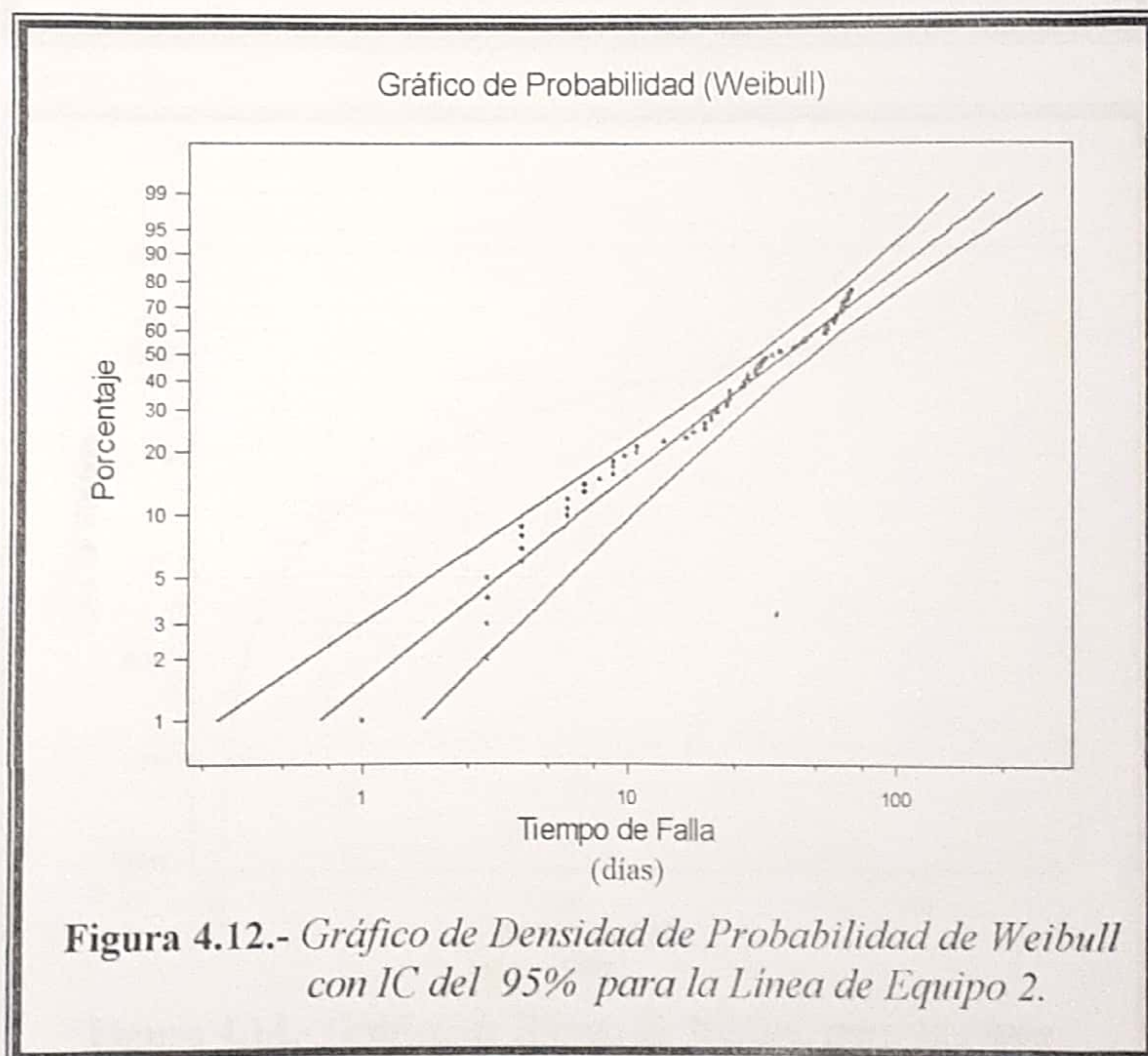


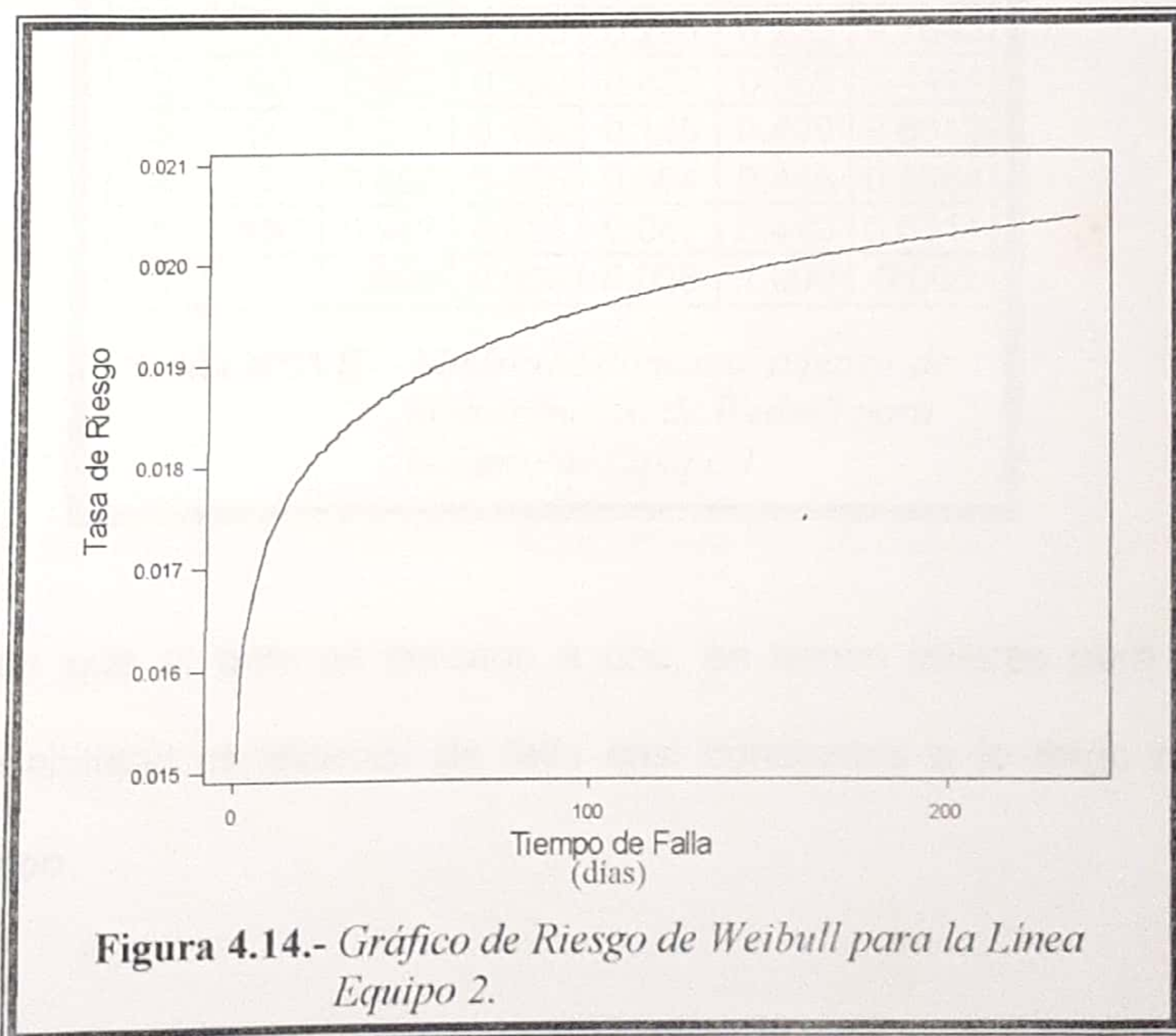
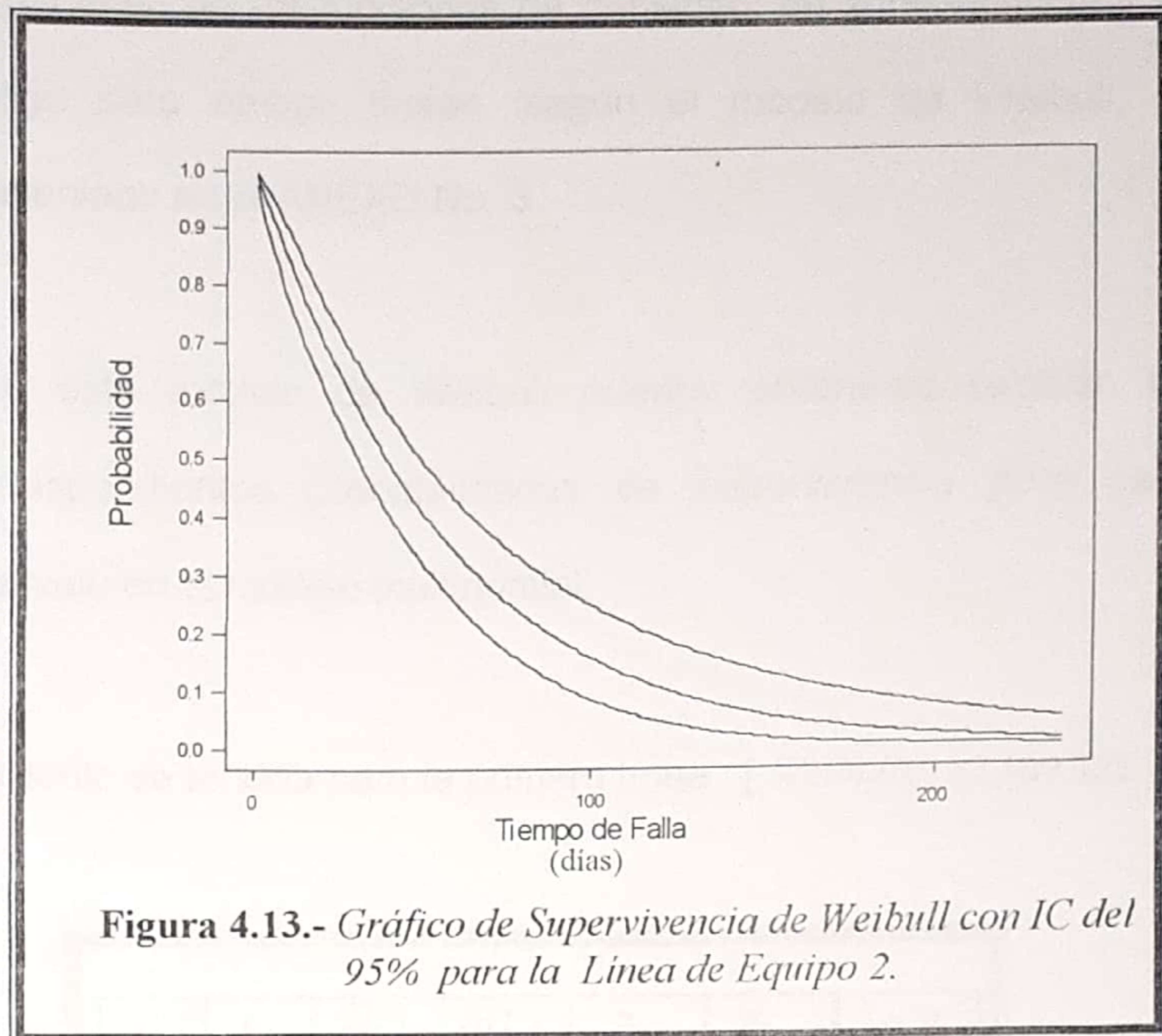
☞ Datos de la Línea de Equipo 2.

Forma	1.052
Escala	55.471
MTTF	54.367
Desv. Est.	51.699

Tabla XXVI.- *Parámetros de Weibull y MTTF para la Línea de Equipo 2.*

Para esta línea el parámetro de forma α también es cercano a uno, pero es mayor, lo que indica que la curva de riesgo será creciente en bajo grado.





El resumen de las funciones de densidad, de supervivencia y de riesgo para ambas líneas, según el modelo de Weibull, es presentado en el ANEXO No. 3.

Con este modelo de Weibull pueden obtenerse también las correspondientes probabilidades de supervivencia para cada intervalo en el modelo multinomial.

Por tanto se tendría para la primera línea, [*Weibull*(0.82,149.32)]

i	t_i	$F(t_i)$	$S(t_i)$	π_i	p_i	$1 - p_i$
0	0	0	1	-	-	-
1	30	0.235	0.765	0.235	0.235	0.7648
2	60	0.662	0.338	0.427	0.559	0.4414
3	90	0.811	0.189	0.148	0.439	0.5612
4	120	0.895	0.105	0.084	0.445	0.5554
5	150	0.942	0.058	0.047	0.449	0.5511
6	∞	1.000	0.000	0.058	1.000	0.000

Tabla XXVII .- *Modelo Multinomial a partir de la distribución de Weibull para la Línea de Equipo 1.*

Dado que el beta es cercano a uno, se tienen valores para la probabilidad condicional de falla casi constantes a lo largo del tiempo.

Así mismo para la segunda línea, [*Weibull*(1.052,55.471)]

i	t_i	$F(t_i)$	$S(t_i)$	π_i	p_i	$1 - p_i$
0	0	0.000	1.000	-	-	-
1	15	0.223	0.777	0.223	0.223	0.7768
2	30	0.408	0.592	0.184	0.238	0.7625
3	45	0.552	0.448	0.144	0.243	0.7568
4	60	0.662	0.338	0.111	0.247	0.7531
5	75	0.747	0.253	0.084	0.25	0.7502
6	∞	1.000	0.000	0.253	1.000	0.000

Tabla XXVIII .- *Modelo Multinomial a partir de la distribución de Weibull para la Línea de Equipo 2.*

Como se puede notar en la tabla, la probabilidad de supervivencia para cada intervalo también se mantiene un tanto constante, decreciendo de manera lenta. Esto se debe que el riesgo es creciente pero no acelerado.

No hay que olvidar que el análisis se lo realizó a partir de una muestra generada a partir de una distribución geométrica, por lo que los valores obtenidos representan fallas sin tomar en cuenta censuras.

Fijándose en este supuesto, es necesario realizar un diagnóstico de los 20 datos originales para establecer una comparación y tener una idea de que tan lejos del fenómeno original se encuentran los datos simulados.

	Línea 1		Línea 2	
	Weibull	K - M	Weibull	K - M
Forma	0.9965		1.169	
Escala	215.5200		55.239	
<i>MTTF</i>	215.8400	87.568	52.330	39.623
Desv. Est.	46.6100		44.930	

Tabla XXIX .- *Comparación entre los modelos de Weibull y Kaplan-Meier para los 20 datos originales de cada línea de equipo.*

Como se puede observar, en ambas líneas, los valores obtenidos para el *MTTF* por el método de Kaplan – Meier son muy distantes de los obtenidos por la distribución de Weibull. Este suceso puede surgir debido a que la muestra es muy pequeña para asumir un modelo paramétrico, o simplemente el software no utiliza el método de Kaplan – Meier para la estimación de los parámetros de la distribución de Weibull.

Es decir que, cuando no se posee mayor cantidad de datos ($n < 30$), es posible apoyarse en la teoría estadística, como en este caso asumiendo un modelo geométrico, en el cual se generan datos de tiempos hasta que ocurra el primer éxito. Entonces el éxito estaría representado en dos eventos:

1. En que el equipo falle, o
2. Que exista una retirada o técnicamente hablando una censura a la derecha.

4.2. Análisis estadístico de las fallas en la producción de una empresa de manufactura.

Dado que el concepto de confiabilidad es muy amplio, este se ve involucrado además al momento de realizar un análisis estadístico tanto para la producción como para la proporción de fallas dentro de una empresa dedicada a la manufactura. Es decir que se puede tener una buena estimación de la confiabilidad de las máquinas, sistemas o componentes que empleen este tipo de industrias, a partir de un estudio estadístico de las fallas que ocurren en su producción.

Para el estudio se escogió una empresa grande de manufactura, dedicada a la producción de llantas, para la cual su producción y distribución de fallas serán el objeto de estudio.

La industria de llantas es muy grande y posee gran variedad en los tipos de llantas, hay que recordar que estas son producidas según el tamaño del vehículo y según otras características mucho más específicas; sin embargo el estudio a realizar estará basado en los 3 tipos de llantas que poseen mayor producción en la empresa de llantas seleccionada.

Antes de iniciar con el análisis resulta necesario conocer, por lo menos a breves rasgos, cuál es el procedimiento para la fabricación de las llantas, para así de esta manera poder tener una idea de las máquinas empleadas y de las posibles fallas que pudieren existir dentro del proceso.

Proceso de Elaboración de las llantas

El proceso para la elaboración de las llantas consiste en 8 pasos, tomados desde el ingreso y mezcla de la materia prima, hasta la inspección del producto elaborado final.

Paso 1: **Molino.**- El caucho que se utilizará para la elaboración de las llantas pasa por un proceso de extrusión.

Paso 2: **Mezclas.**- Cerca de siete elementos son combinados con un peso exacto.

Paso 3: **Armado.**- Las llantas una vez pasado el proceso de enfriamiento del rodamiento y calandrado, van a ensamblaje.

Paso 4: **Ensamblaje.**- El proceso de ensamblaje de carcasa y laterales difiere según el tipo de llanta.

Paso 5: **Vulcanización.**- Las máquinas de vulcanización trabajan a una temperatura de 240 grados. Según la llanta, el proceso demora de 15 minutos a una hora.

Paso 6: **Traslado.**- Los neumáticos son inspeccionados y los productos "scrap" o basura son retirados.

Paso 7: **Rebarbeo.**- El producto es objeto de una limpieza o "maquillaje" de caucho sobrante y revisión de todas las fugas.

Paso 8: **Concluido.**- Finalmente se revisan los laterales y se da el visto bueno al producto para su salida al mercado (1).

Una vez que se tiene una idea de las máquinas que son utilizadas en el proceso de fabricación, se puede comenzar con el estudio de medición de la confiabilidad de estas máquinas y por ende de la confiabilidad del proceso como tal.

A los productos que no pasan la inspección final para ponerlos al mercado, se les denomina productos basura o "scraps". Entre las fallas más comunes se encuentran: soplos entre pliegues, deformidad, desgarres, nudosidades, etc.

(1). Diario Expreso.- Sección de Economía: Página 8: Martes, 31 de Diciembre de 2002

La empresa fabricante objeto de estudio, lleva sus registros en una amplia base de datos, en la cual se encuentran codificados los tipos de llanta con su respectiva producción y porcentaje de scraps para un determinado período de tiempo.

4.2.1. Selección de los tipos de llantas para la aplicación del análisis estadístico.

Como se mencionó anteriormente, existen varios tipos de llanta según características específicas, como por ejemplo las radiales, es por esto que se han seleccionado 3 tipos de llantas, las cuales están dentro de las que poseen mayor cantidad de producción y que además constituyen tres grandes gremios, es decir que poseen varios subtipos de llantas, y por tanto estos difieren en los procesos de elaboración. Estas diferencias incluso pueden significar distintas máquinas para su fabricación.

Cabe recalcar que la codificación y clasificación de las llantas son realizadas por la empresa fabricante.

A continuación se presenta una tabla con los tipos de llanta seleccionados, junto con su codificación:

Código	Descripción
31	Camion
21	Cmt. Bias
22	Cmt. Radial

Tabla XXX .- *Codificación de los tipos de llantas seleccionados.*

Con el fin de tener datos más actuales, se estableció la toma de datos de la producción empezando el primer día del año anterior, es decir 1 de enero de 2002, hasta el 31 de Mayo de 2003.

Estos datos fueron obtenidos en períodos mensuales y para cada gremio o tipo de llanta se escogieron 5 subtipos con la mayor cantidad de unidades producidas, es decir que se tiene una muestra de tamaño $n = 85$ (17 meses * 5 subtipos) para cada tipo de llanta seleccionado.

Tabla de Datos

En la recopilación de datos por tipo de llanta se incluyó, la cantidad de llantas producidas, la cantidad de llantas que fueron desechas por fallas o no conformidades, y por último el porcentaje que representan estas unidades defectuosas dentro de la producción total.

El formato de la hoja para la recopilación de los datos es como el que se presenta en la Tabla XXIV, en la cual se muestran los datos de los diecisiete meses para un sólo subtipo de llanta perteneciente al gremio *Cmt. Bias*.

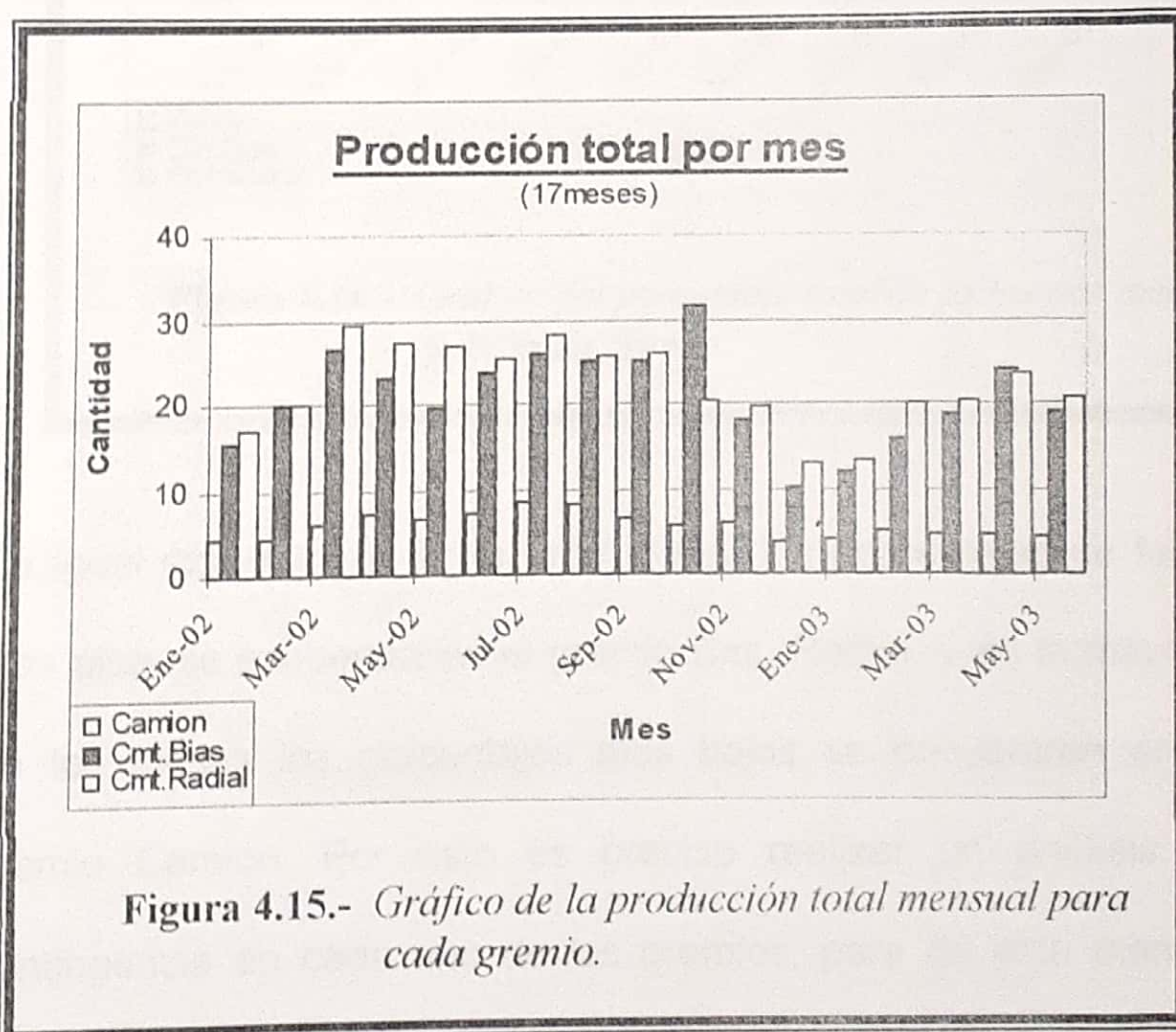
Cmt. Bias		
Cantidad Producida	Cantidad scraps (fallas)	Porcentaje (%)
4272	24	0.56
6946	65	0.94
9152	29	0.32
6666	31	0.47
3288	33	1
6556	33	0.5
8049	34	0.42
6912	35	0.51
7882	31	0.39
12884	46	0.36
5100	47	0.92
3482	66	1.9
3290	49	1.49
3933	16	0.41
6266	13	0.21
7822	48	0.61
5613	45	0.8

Tabla XXXI.- *Tabla de la toma de datos.*

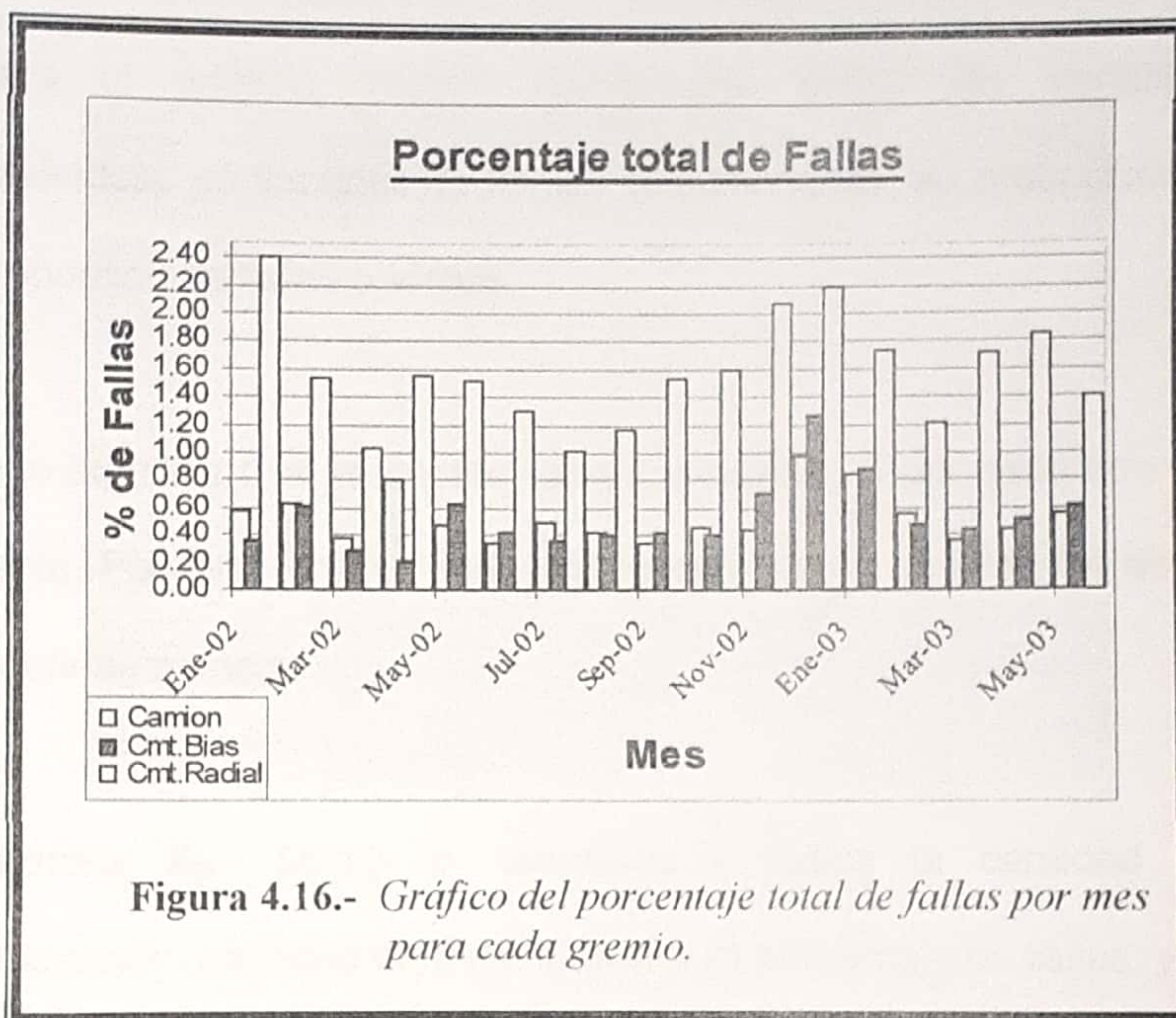
Los datos completos de todos los subtipos para los tres gremios son presentados en el ANEXO No. 2.

Por otro lado para obtener la producción total por mes para cada gremio, se puede sumar la cantidad mensual de unidades producidas en cada subtipo, teniéndose así la producción, cantidad y porcentaje total de fallas para cada gremio.

A continuación se presentarán los gráficos correspondientes a la cantidad de producción total en los diecisiete meses para cada gremio, junto con el porcentaje de fallas respecto a esta producción total.



Se puede observar que para todos los meses, excepto para octubre de 2002, el gremio Cmt. Radial posee la mayor cantidad de unidades producidas, mientras que el gremio Camion es el que posee menor cantidad para todos los meses en estudio.



De igual manera que en la producción, los porcentajes de fallas más altos se encuentran en el gremio Cmt. Radial, y en la mayoría de los meses los porcentajes más bajos se encuentran en el gremio Camion. Por esto es preciso realizar un análisis de contingencia en cada uno de los gremios, para de esta manera

medir el grado de independencia de la variable cantidad de unidades producidas versus la cantidad de unidades defectuosas.

4.2.2. Definición de Variables.

Para el análisis, resultó conveniente definir las variables basándose en los tipos de llantas seleccionadas; su producción y proporción de fallas o scraps.

Esto significa que se contará con 9 variables, 3 por cada tipo de llanta. Por tanto resulta conveniente definir las y codificarlas de la siguiente manera:

Variable X_{jk} .- Donde el subíndice k indica la cantidad de producción, cantidad de defectuosas o el porcentaje de fallas; y el subíndice j indica el tipo de llanta. Es decir que tanto k como j varían entre 1 y 3.

Valores de j	Tipo de llanta
1	Camion
2	Cmt. Bias
3	Cmt. Radial

Valores de k	Descripción
1	Cantidad de unidades producidas
2	Cantidad de unidades defectuosas
3	Porcentaje de fallas

Tabla XXXII.- *Definición de Variables según tipo de llanta.*

Por tanto si se tiene la variable X_{21} , esta se estará refiriendo a la cantidad de unidades producidas para el tipo de llanta Cmt. Bias para cada uno de los subtipos dentro de los diecisiete meses.

4.2.3. Análisis Univariado.

Una vez definidas las variables, se procederá como primer paso a realizar un análisis univariado para cada una de ellas, tomando en cuenta el tamaño de muestra $n = 85$ (diecisiete meses por cinco subtipos). Es decir que se realizará un análisis por corrida de producción.

Para este análisis univariado se obtendrá la estadística descriptiva de cada variable y se efectuarán gráficos que serán de ayuda para visualizar de una manera más específica la dispersión y tendencia de los datos.

Además se realizará la prueba de Bondad de Ajuste para saber a que distribución teórica se aproximan los datos de la cantidad de unidades con fallas al igual que los datos de porcentaje de unidades defectuosas.

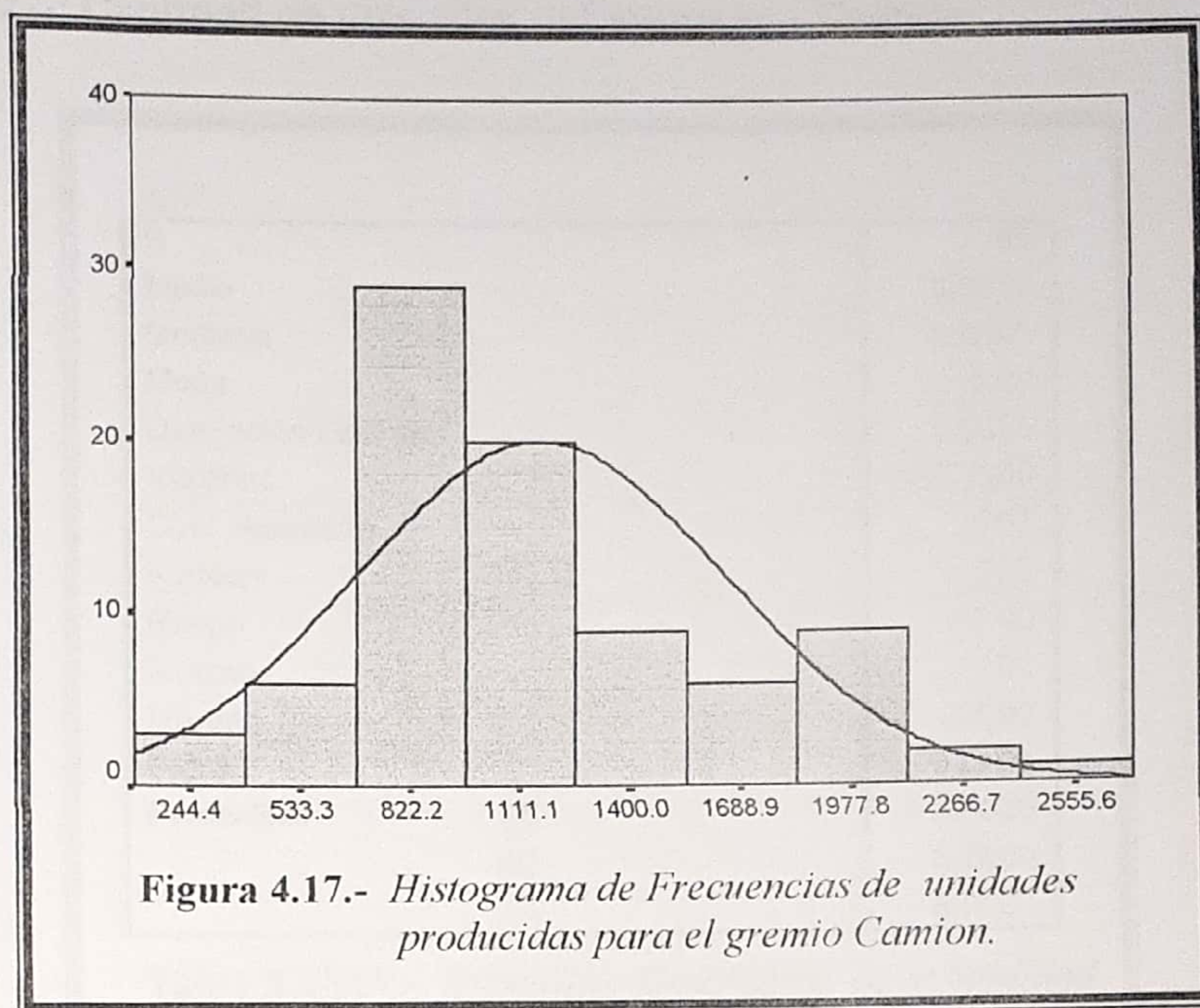
Para la media de la variable cantidad de unidades defectuosas se realizará un contraste de hipótesis por cada gremio. Luego se construirán intervalos de confianza para las medias determinadas en cada uno de los contrastes.

4.2.3.1. Estadística Descriptiva.

X₁₁: Cantidad de unidades producidas - *Camion*

X11		
N		85
Media		1160.6941
Mediana		1040.0000
Moda		548.00
Desviación Estándar		487.6435
Varianza		237796.2
Coef. Asimetría		.735
Kurtosis		.277
Rango		2405.00
Mínimo		178.00
Máximo		2583.00
Suma		98659.00
Percentiles	25	849.0000
	50	1040.0000
	75	1355.0000

Tabla XXXIII .-*Estadística Descriptiva de la cantidad de unidades producidas para el gremio Camion.*



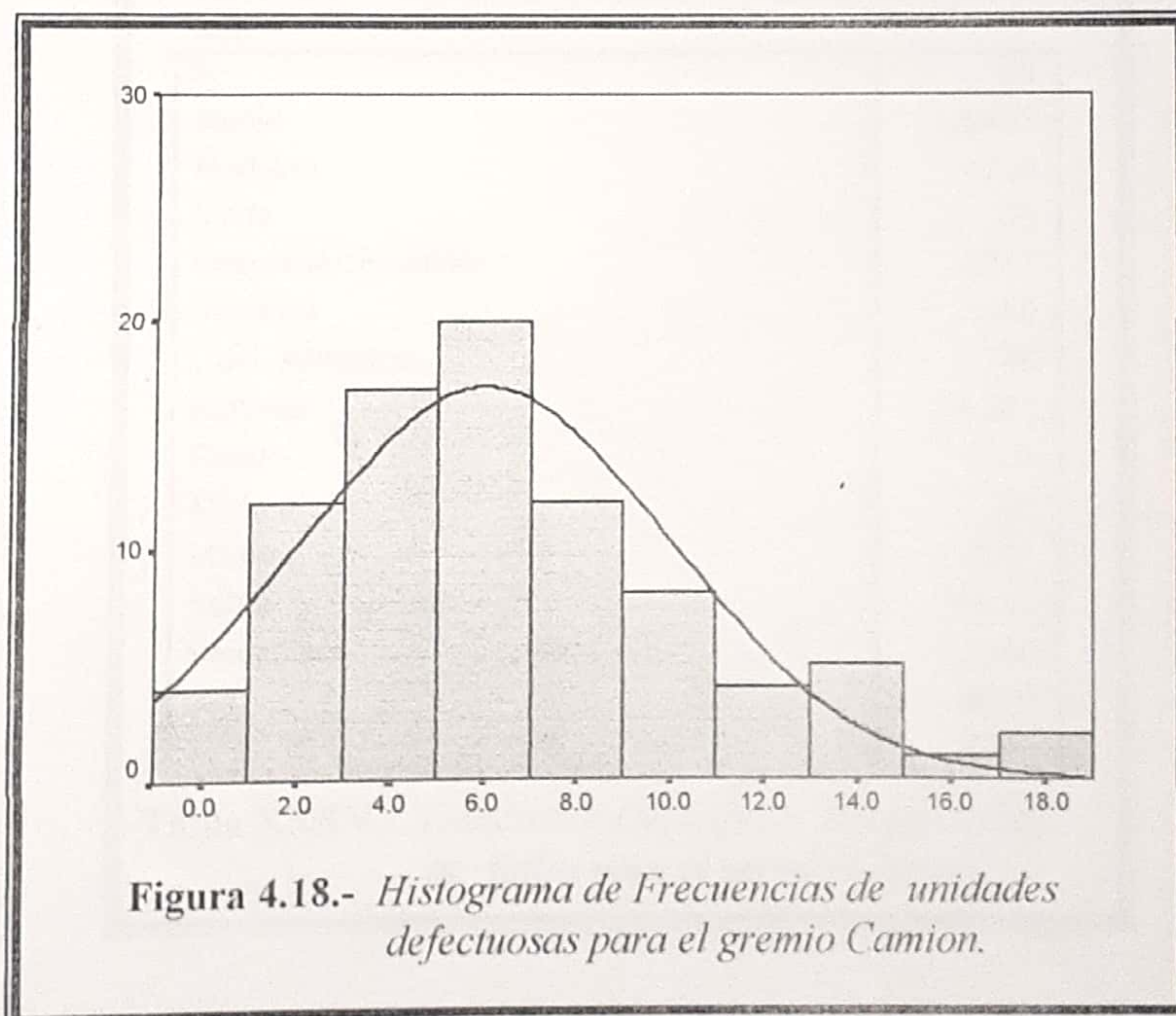
La cantidad de unidades producidas para este gremio posee una gran variabilidad (varianza = 237796.2), ya que como se puede observar el mínimo es 178 y el máximo de unidades producidas es de 2583. Se tiene además que el promedio de producción para este gremio es de 1160 unidades.

El coeficiente de kurtosis es de 0.277, lo que indica que la curva es platicúrtica. Para esta variable se tiene que los datos están concentrados en el centro con un pequeño sesgo positivo de 0.735.

X₁₂: Cantidad de unidades defectuosas - Camion

X12		
N		85
Media		6.0353
Mediana		5.0000
Moda		5.00
Desviación Estándar		3.9655
Varianza		15.7249
Coef. Asimetría		.767
Kurtosis		.222
Rango		17.00
Mínimo		.00
Máximo		17.00
Suma		513.00
Percentiles	25	3.0000
	50	5.0000
	75	8.0000

Tabla XXXIV.- Estadística Descriptiva de la cantidad de unidades defectuosas para el gremio Camión.



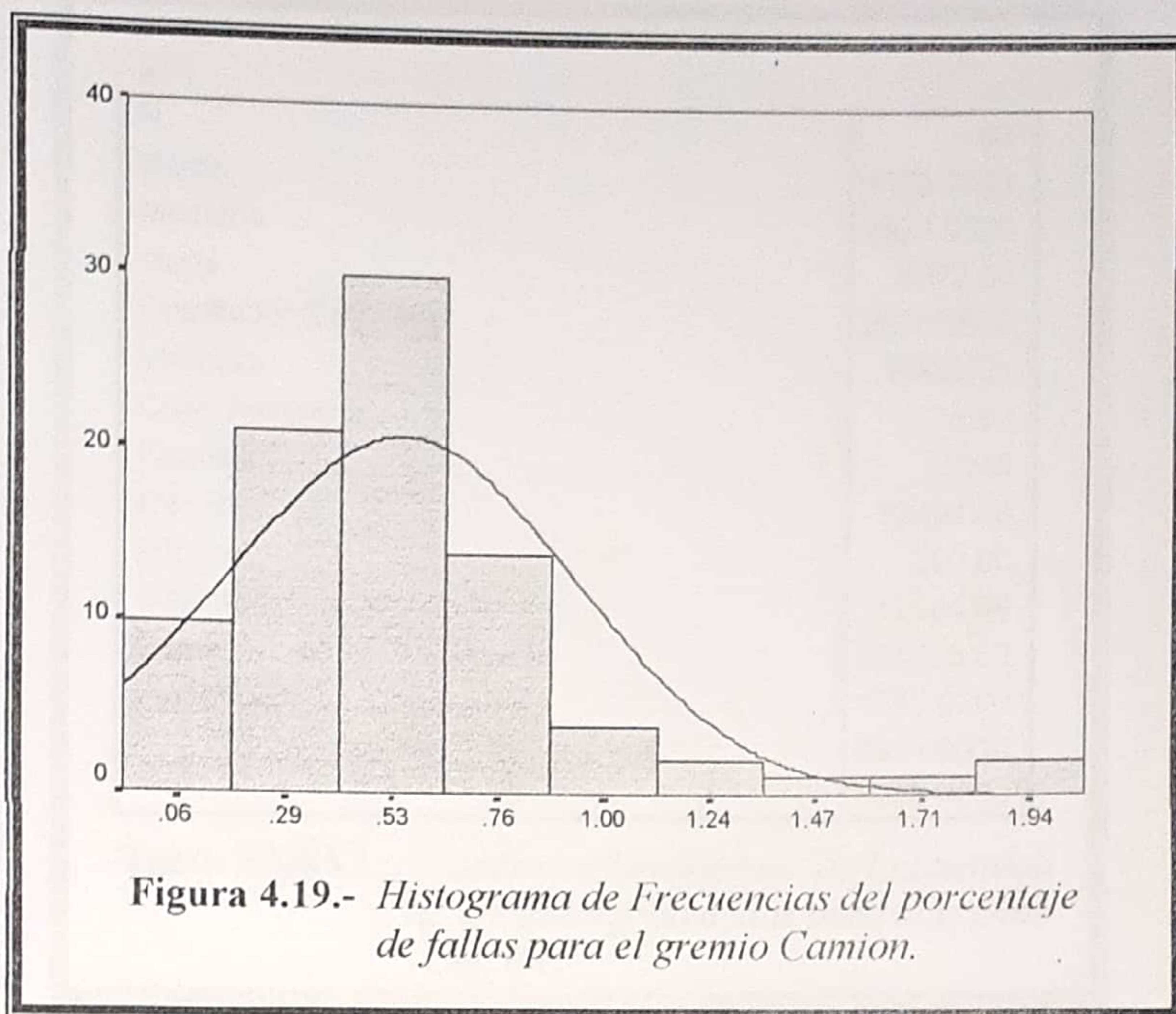
Para la variable cantidad de unidades defectuosas se puede observar que la variabilidad no es muy grande, es decir que la varianza y por tanto la desviación estándar poseen valores pequeños de 15.72 y 3.97 respectivamente. El promedio de unidades defectuosas es para este gremio es de 6.04.

La curva es platicúrtica ya que el coeficiente de kurtosis es menor que 3. Se puede notar que existe un pequeño sesgo positivo de 0.767.

X₁₃: Porcentaje de fallas - *Camion*

X ₁₃		
N		85
Media		.5453
Mediana		.4700
Moda		.00
Desviación Estándar		.3847
Varianza		.1480
Coef. Asimetría		1.746
Kurtosis		4.361
Rango		2.01
Mínimo		.00
Máximo		2.01
Suma		46.35
Percentiles	25	.3150
	50	.4700
	75	.6700

Tabla XXXV.- *Estadística Descriptiva del porcentaje de fallas para el gremio Camion.*



Para esta variable se tiene que el porcentaje promedio de fallas en la producción es de 0.545 y la desviación estándar es de 0.3847. La curva tiene una forma leptocúrtica ya que su valor es mayor que 3, y se tiene un sesgo positivo medido por el coeficiente de asimetría igual a 1.746.

X₂₁: Cantidad de unidades producidas - Cmt. Bias

X ₂₁		
N		85
Media		4175.3529
Mediana		3924.0000
Moda		6072.00
Desviación Estándar		2638.6590
Varianza		6962521
Coef. Asimetría		.682
Kurtosis		.046
Rango		12494.00
Mínimo		390.00
Máximo		12884.00
Suma		354905.00
Percentiles	25	1745.0000
	50	3924.0000
	75	6088.0000

Tabla XXXVI.- Estadística Descriptiva de la cantidad de unidades producidas para el gremio Cmt.Bias.

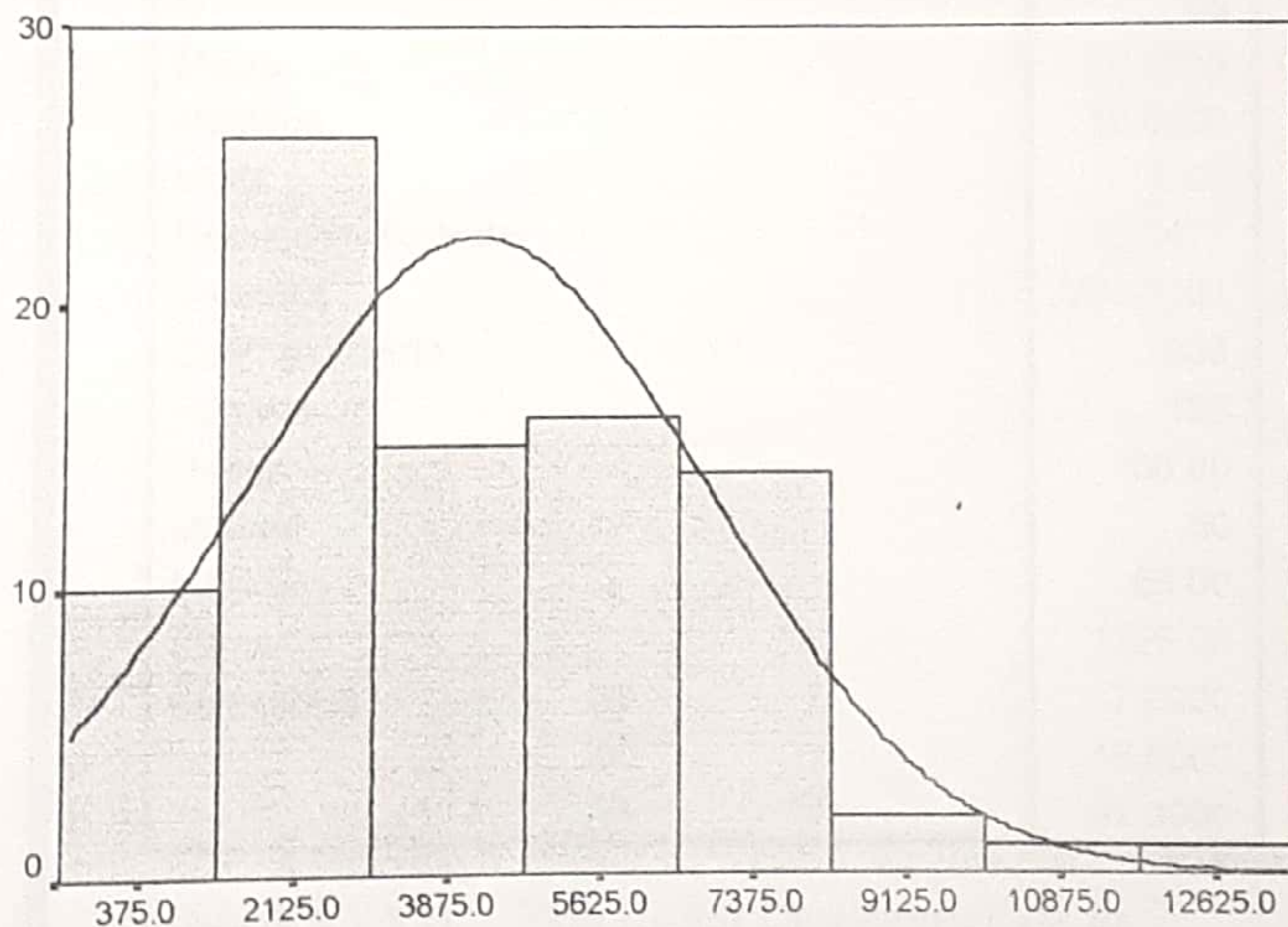


Figura 4.20.- Histograma de Frecuencias de unidades producidas para el gremio Cmt.Bias

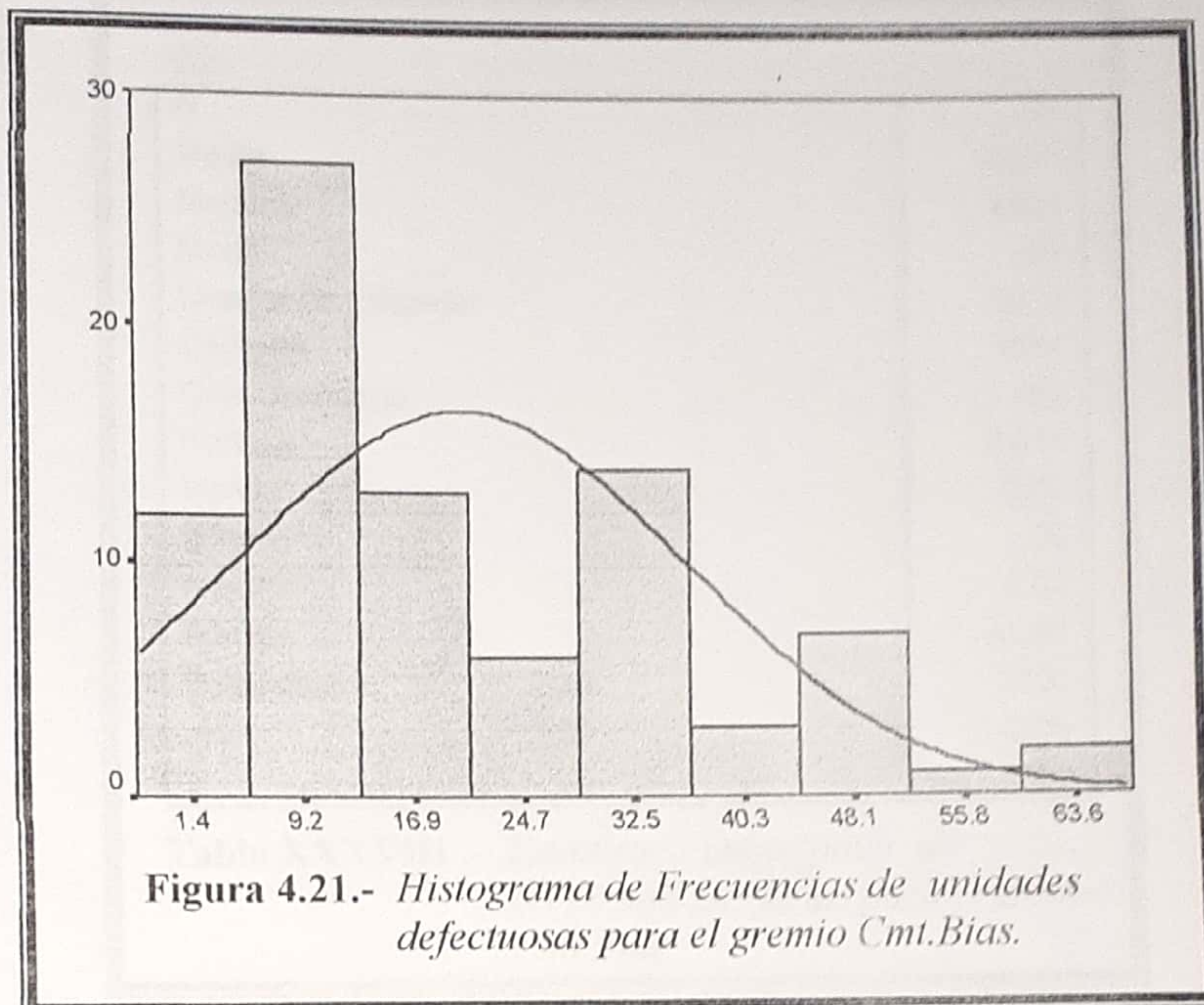
Para este gremio la cantidad de unidades producidas posee una variabilidad demasiado grande, con una desviación estándar igual a 2638.66, ya que como se puede observar se tiene un rango de 12494 unidades. El promedio de producción para este gremio es de 4175.35 unidades.

Para esta variable se tiene que los datos están concentrados mayormente en el centro, ya que el coeficiente de asimetría es de 0.682.

X₂₂: Cantidad de unidades defectuosas - Cmt. Bias

X22		
N		85
Media		20.3059
Mediana		16.0000
Moda		6.00
Desviación Estándar		15.9471
Varianza		254.3101
Coef. Asimetría		.936
Kurtosis		.166
Rango		66.00
Mínimo		.00
Máximo		66.00
Suma		1726.00
Percentiles	25	7.0000
	50	16.0000
	75	31.0000

Tabla XXXVII .- Estadística Descriptiva de la cantidad de unidades defectuosas para el gremio Cmt. Bias.



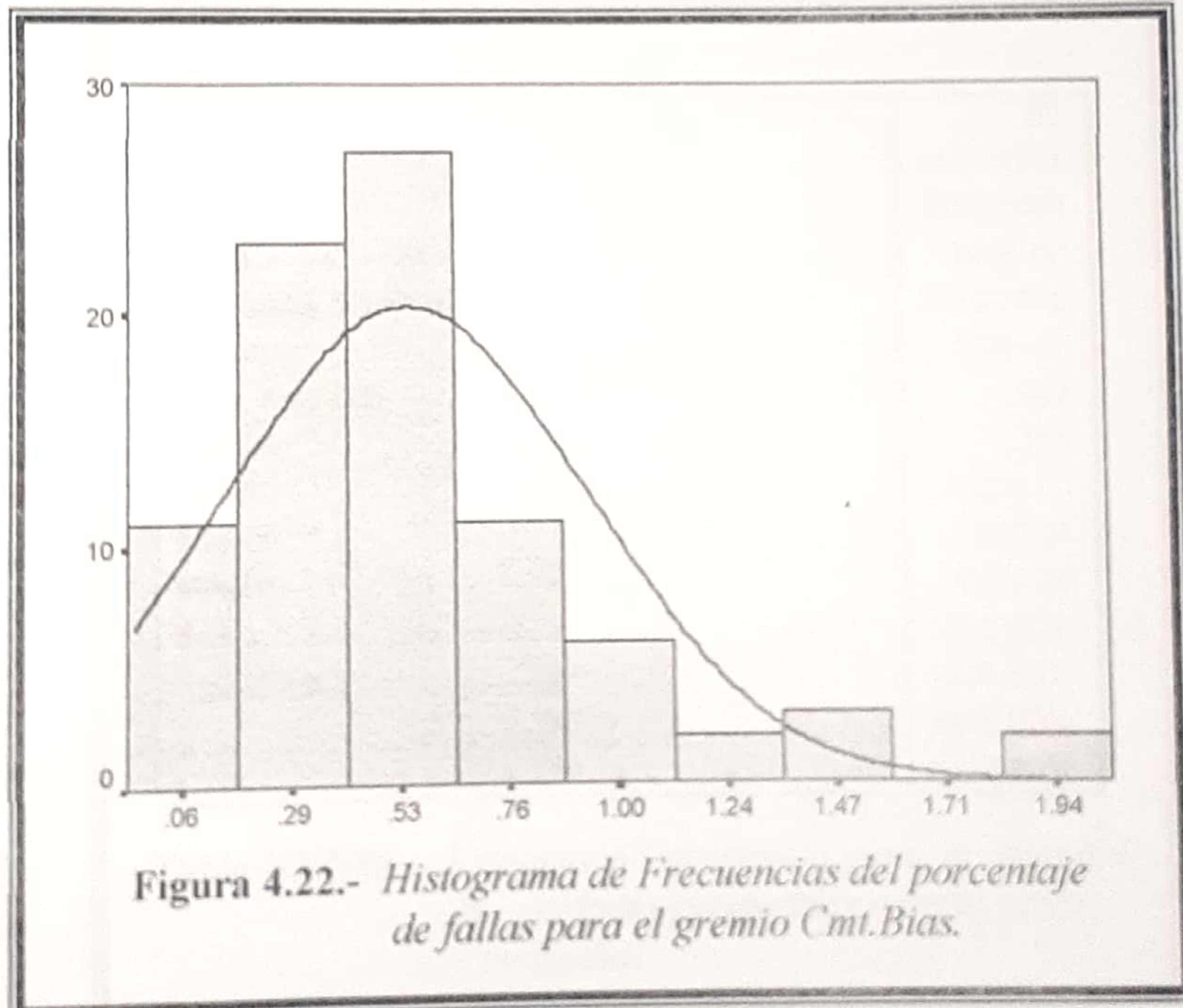
En este gremio, la variable cantidad de unidades defectuosas posee una variabilidad relativamente alta, donde la desviación estándar es de 15.94 y la varianza es igual a 254.31. El promedio de unidades defectuosas para este gremio es de aproximadamente 20.31.

El coeficiente de kurtosis es igual a 0.166, lo que indica que la curva es aplanada. El coeficiente de asimetría indica un sesgo positivo de 0.936.

X₂₃: Porcentaje de fallas - Cmt. Bias

X23		
N		85
Media		.5398
Mediana		.4900
Moda		.08
Desviación Estándar		.3929
Varianza		.1544
Coef. Asimetría		1.483
Kurtosis		3.018
Rango		2.02
Mínimo		.00
Máximo		2.02
Suma		45.88
Percentiles	25	.2850
	50	.4900
	75	.7050

Tabla XXXVIII .- Estadística Descriptiva del porcentaje de fallas para el gremio Cmt.Bias.



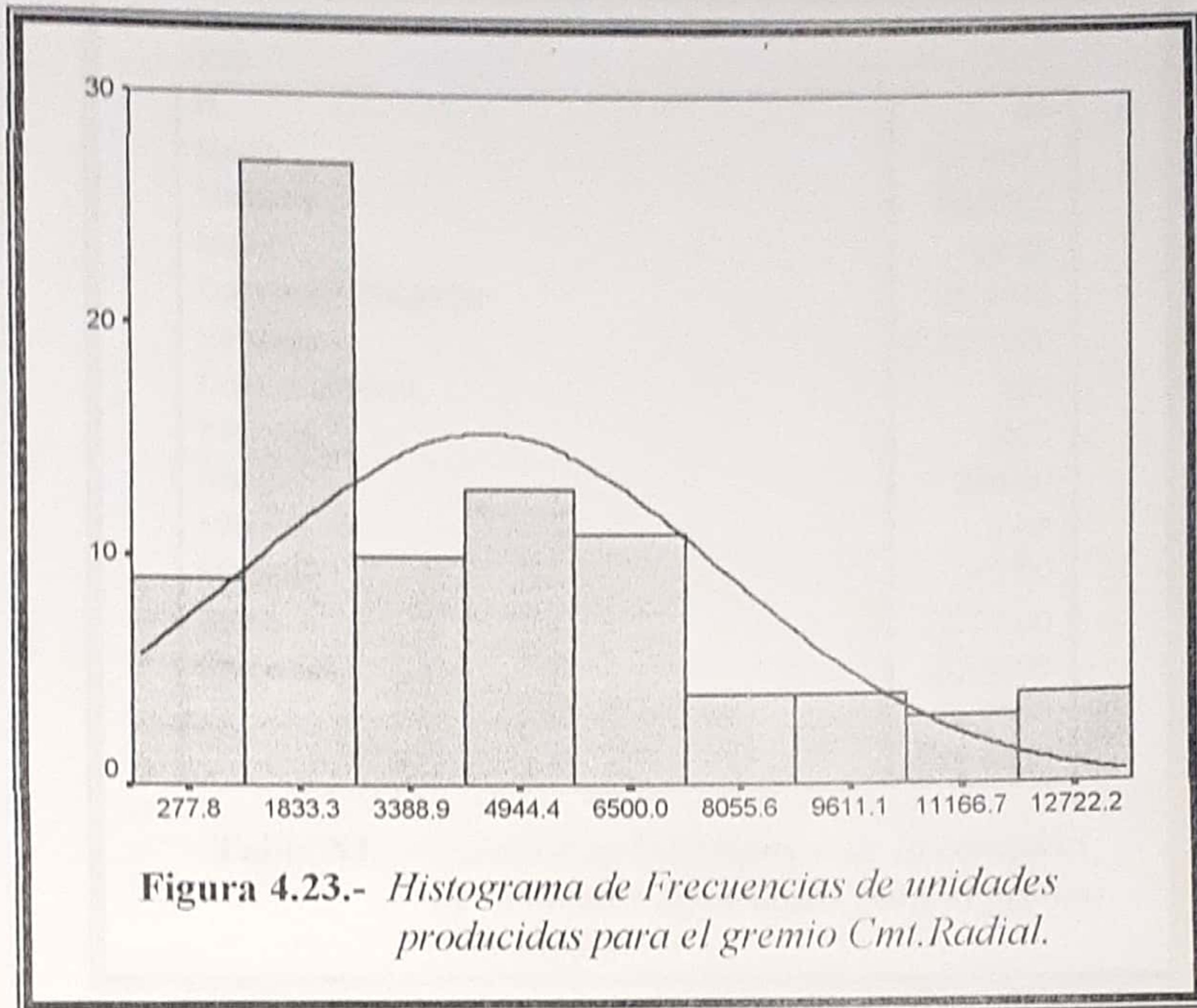
La variable porcentaje de fallas en este gremio posee un rango similar en comparación con el gremio anterior, por esto posee una dispersión también similar, reflejada en la desviación estándar de 0.3929.

El porcentaje promedio de fallas en la producción es de 0.539. La curva tiene una forma mesocúrtica ya que el coeficiente de kurtosis es aproximadamente igual a 3. Se tiene un sesgo positivo ya que el coeficiente de asimetría igual a 1.483.

X₃₁: Cantidad de unidades producidas - Cmt. Radial

X31		
N		85
Media		4435.4941
Mediana		3856.0000
Moda		6338.00
Desviación Estándar		3412.7515
Varianza		1.2E+07
Coef. Asimetría		.942
Kurtosis		.105
Rango		13238.00
Mínimo		220.00
Máximo		13458.00
Suma		377017.00
Percentiles	25	1579.0000
	50	3856.0000
	75	6338.0000

Tabla XXXIX .- Estadística Descriptiva para la cantidad de unidades producidas para el gremio Cmt.Radial.



La variable cantidad de producción posee un promedio de 4435.49 unidades, con una alta dispersión reflejada en el valor de la desviación estándar igual a 3412.75.

Esta variable posee una curva platicúrtica con una kurtosis de 0.105, y un sesgo pequeño con un coeficiente de asimetría de 0.942 positivo.

X₃₂: Cantidad de unidades defectuosas - Cmt. Radial

X32		
N		85
Media		67.4941
Mediana		55.0000
Moda		32.00
Desviación Estándar		45.9416
Varianza		2110.6339
Coef. Asimetría		.923
Kurtosis		.500
Rango		209.00
Mínimo		4.00
Máximo		213.00
Suma		5737.00
Percentiles	25	32.0000
	50	55.0000
	75	99.5000

Tabla XL.- Estadística Descriptiva de la cantidad de unidades defectuosas para el gremio Cmt.Radial.

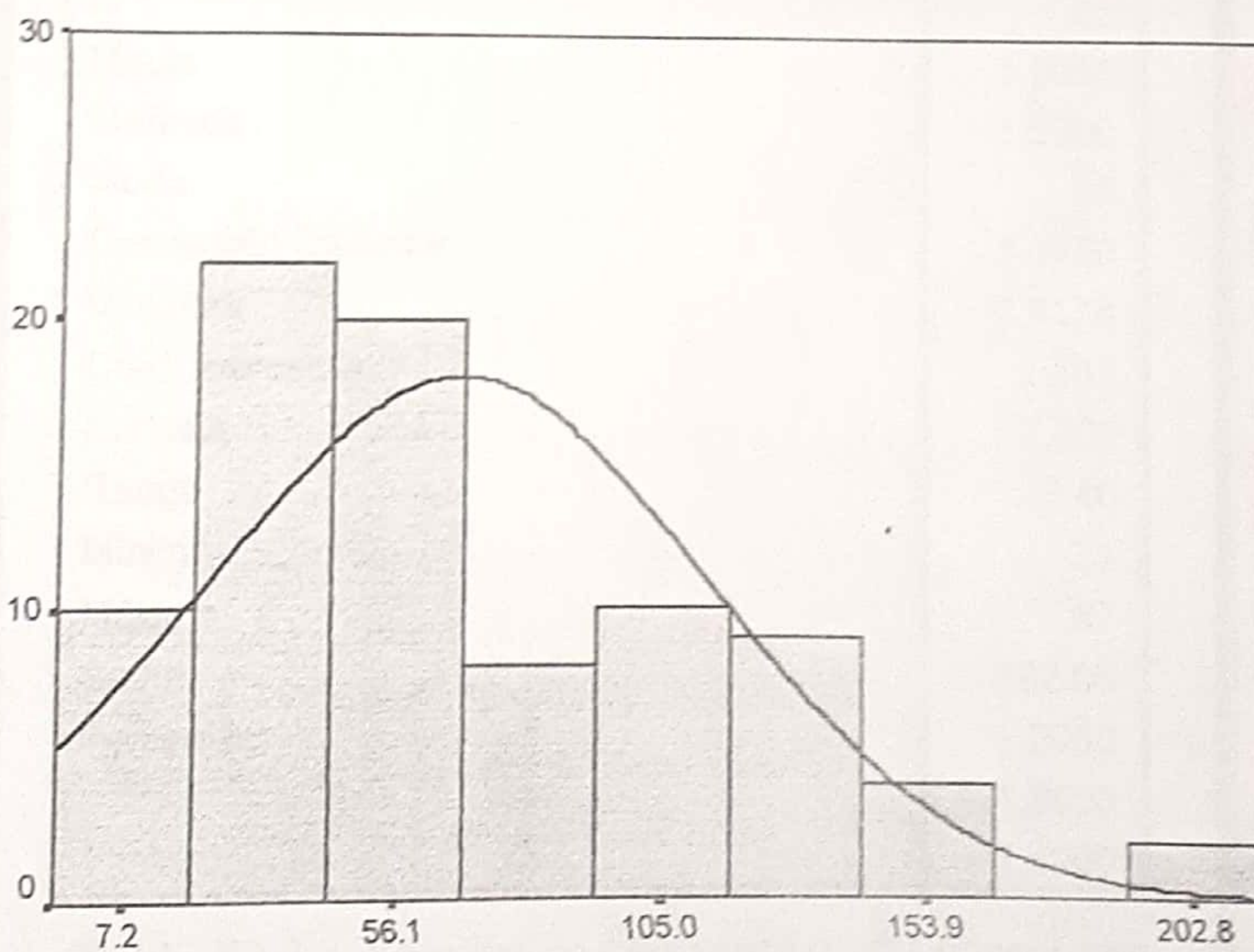


Figura 4.24.- Histograma de Frecuencias de unidades defectuosas para el gremio Cmt.Radial.

En el gremio *Cmt. Radial*, la variable cantidad de unidades defectuosas posee una variabilidad muy alta en comparación con los gremios anteriores. Se tiene que la desviación estándar es de 45.94 y la varianza es igual a 2110.63. El rango es de 209 unidades.

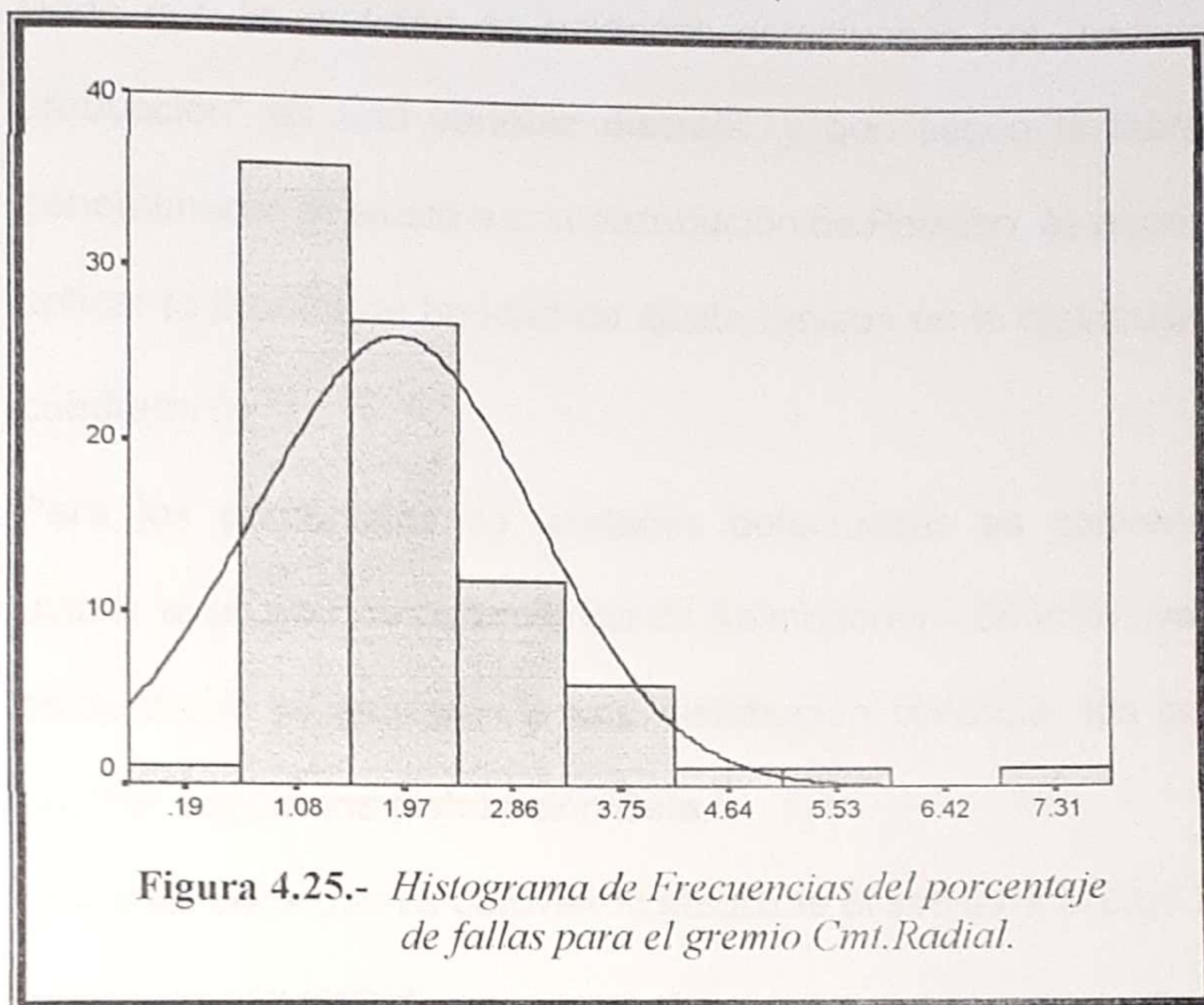
El promedio de unidades defectuosas es para este gremio es de aproximadamente 67.49.

El coeficiente de kurtosis es igual a 0.5, mientras que el coeficiente de asimetría es cercano a 1.

X₃₃: Porcentaje de fallas - *Cmt. Radial*

X ₃₃		
N		85
Media		1.9066
Mediana		1.5900
Moda		1.05
Desviación Estándar		1.1480
Varianza		1.3178
Coef. Asimetría		2.161
Kurtosis		7.309
Rango		7.46
Mínimo		.21
Máximo		7.67
Suma		162.06
Percentiles	25	1.2050
	50	1.5900
	75	2.4000

Tabla XLI .- *Estadística Descriptiva del porcentaje de fallas para el gremio Cmt.Radial.*



La variable porcentaje de fallas para este gremio posee un rango mayor que el de los gremios anteriores, por tanto posee una dispersión más alta, reflejada en la varianza y desviación estándar de 1.32 y 1.15 respectivamente.

El porcentaje promedio de fallas en la producción es de 1.91. La curva tiene una forma leptocúrtica ya que el coeficiente de kurtosis es relativamente alto con un valor de 7.31. El coeficiente de asimetría es también relativamente grande con un valor de 2.16.

4.2.3.2. Pruebas de bondad de ajuste.

Dado que la cantidad de unidades defectuosas por "corrida de producción" es una variable discreta, y que según la literatura generalmente se ajusta a una distribución de Poisson, es necesario aplicar la prueba de bondad de ajuste basada en la distribución Ji cuadrado (χ^2).

Para los porcentajes de unidades defectuosas es conveniente utilizar la prueba no paramétrica de Kolmogorov - Smirnov, ya que estos datos se ajustarán a una distribución continua, los cuales podrían seguir una distribución Beta.

Los resultados que se obtuvieron mediante el software S-plus 2000 fueron los siguientes:

H ₀ : La distribución de la población es F ₀ (X)						
H ₁ : La distribución de la población no es F ₀ (X)						
Variable	Cant. unid. defectuosas			Porcentaje de fallas		
Contraste	Ji Cuadrado			K - S		
Gremio	Camion	Cmt. Bias	Cmt. Radial	Camion	Cmt. Bias	Cmt. Radial
Distrib. Teórica	Poisson (λ)			Beta $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$		
Param.	6.0	20.3	67.5	2.02 365.25	1.88 346.27	2.69 138.15
Estad. Prueba	18.6778	9383.56	488754	0.1017	0.0641	0.0938
Valor p	0.0001	0	0	0.3425	0.8765	0.4431
Estado	R	R	R	A	A	A

Tabla XLII .- Pruebas de Bondad de ajuste para la cantidad de unid. defectuosas y para el porcentaje de fallas para cada gremio.

La fila denominada Estado, se refiere a la aceptación (A) o rechazo (R) de la hipótesis nula. Vale la pena recalcar que en los contrastes basados en la distribución Ji Cuadrado se trabajó con 2 grados de libertad para todos los casos ($4-1-1=2$ gl), ya que se escogió 4 casillas donde los puntos de corte son los cuartiles, y además la distribución de Poisson requiere la estimación de un parámetro λ . El estimador de máxima verosimilitud para λ es la media muestral.

Por otro lado se eligió el ajuste a la distribución Beta, debido a que los valores de los porcentajes están entre 0 y 1; por tanto si en la tabla de datos se tiene 1.5%, el valor real de la variable aleatoria es de 0.015. Además para la estimación de los parámetros α, β de esta distribución, se tuvieron que utilizar los valores de la media y varianza muestrales, ya que la media y la varianza de la Beta están dados por:

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \text{y} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Guiándose del gráfico de la distribución de los datos, se fijaron valores para α , y teniendo ya un estimador para la media y varianza, lo que se debía hallar era un valor para β que satisfaga las dos ecuaciones de la media y varianza.

Es decir que de la primera ecuación, que es la correspondiente a la media, se obtuvo la siguiente relación luego de despejar:

$$\beta = \frac{\alpha(1 - \hat{\mu})}{\hat{\mu}}$$

Entonces dándole valores a alfa, según lo que el gráfico sugiere, se iban obteniendo valores para beta, luego estos dos valores eran reemplazados en la ecuación de la varianza y a su vez este valor era comparado con el estimador de la varianza obtenido de la muestra. Se realizó un proceso iterativo con una tolerancia de $1 \cdot 10^{-8}$ ya que los valores de la varianza para esta variable aleatoria son bajísimos. Hay que recordar que en la estadística descriptiva se tiene la varianza de los datos dados en porcentaje, por eso es preciso tener en cuenta que los valores de la estadística descriptiva están dados para una variable X , pero los valores reales están dados por $Y = \frac{X}{100}$. Esto significa que el valor esperado y la varianza de los valores reales se transforman en:

$$E[Y] = E\left[\frac{X}{100}\right] = \frac{1}{100} E[X]$$

$$Var[Y] = Var\left[\frac{X}{100}\right] = \frac{1}{100^2} Var[X]$$

Como se puede notar, las tres pruebas para las distribuciones continuas el valor p es mayor a 10%, significando que para cada caso, existe evidencia estadística para decir que los datos provienen de una distribución Beta con los parámetros especificados.

Por otro lado las tres pruebas para el ajuste a la variable discreta de Poisson obtuvieron el rechazo de la hipótesis nula, es decir que según el criterio para el valor p existe evidencia estadística para afirmar que los datos para las correspondientes variables no se ajustan o no siguen una distribución de Poisson con los parámetros especificados. Por esta razón se consideró conveniente realizar pruebas de hipótesis para la media poblacional de los datos para estas variables discretas, junto con los respectivos intervalos de confianza.

4.2.3.3. Pruebas de Hipótesis e Intervalos de confianza para la media de la variable Cantidad de unidades defectuosas para los tres tipos.

Lo que se desea probar mediante la hipótesis alternativa, es que la media de las poblaciones es inferior a un valor fijado en la hipótesis nula, en base al estimador obtenido de la media; es por esto que se plantearán contrastes de cola inferior.

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0$$

$$EP: z = \frac{\bar{y} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

$$RR: z < -z_{\alpha}$$

Donde el tamaño de muestra $n = 85$ es igual en todos los casos.

Una vez obtenidos los valores para las medias poblacionales se construirán los intervalos del 95% de confianza. La estructura de estos intervalos es la siguiente:

$$\hat{\mu} \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

Los valores de los estimadores de la media y desviación estándar ya fueron obtenidos en la estadística descriptiva para cada una de las variables.

A continuación se presentarán los respectivos resultados obtenidos de los contrastes y la construcción de los respectivos intervalos de confianza:

- Para la media poblacional de la cantidad de unidades defectuosas del gremio *Camion*.

Media muestral	6.03
Estimador de la desviación estándar	3.965
μ_0	6.5
EP (z)	-1.093
Valor p	0.1379

Tabla XLIII .- *Contraste de hipótesis para la media poblacional de la Cant Unid. Defect del gremio Camion.*

El valor del estadístico de prueba genera un valor p igual a 0.1379, por lo tanto se acepta la hipótesis nula para el valor de la media de 6.5 unidades defectuosas por corrida de producción.

Intervalo de confianza: $5.6571 \leq \mu \leq 7.3429$

- Para la media poblacional de la cantidad de unidades defectuosas del gremio *Cmt. Bias*.

Media muestral	20.31
Estimador de la desviación estándar	15.947
μ_0	22
EP (z)	-0.977
Valor p	0.166

Tabla XLIV .- *Contraste de hipótesis para la Media poblacional de la Cant. Unid. Defect del gremio Cmt. Bias.*

En este caso el valor p también es superior al 10%, por lo tanto se acepta lo postulado en la hipótesis nula, teniéndose una media poblacional para este gremio de 22 unidades defectuosas por corrida de producción.

Intervalo de confianza: $18.6097 \leq \hat{\mu} \leq 25.3902$

- Para la media poblacional de la cantidad de unidades defectuosas del gremio *Cmt. Radial*.

Media muestral	67.49
Estimador de la desviación estándar	45.941
μ_0	72
EP (z)	-0.9051
Valor p	0.1827

Tabla XLV .- *Contraste de hipótesis para la media poblacional de la Cant Unid. Defect del gremio Cmt.Radial.*

Para el gremio *Cmt. Radial*, al igual que para los anteriores, se fijó un valor para la media a ser postulado en la hipótesis nula según la dispersión existente en los datos. En este gremio, según el criterio del 10%, se aceptó la hipótesis nula para la media poblacional.

Intervalo de confianza: $62.2333 \leq \mu \leq 81.7667$

4.2.4. Análisis Bivariado

Para el análisis bivariado se tomarán en cuenta básicamente estudios de correlación e independencia entre los pares de variables: Cantidad de unidades producidas con la variable cantidad de unidades defectuosas, para cada tipo de llanta.

El análisis de independencia se lo realizará a través de las denominadas Tablas de Contingencia, y lo que se intentará probar en cada una de ellas es lo siguiente:

Dado que todos los tipos de llanta en estudio poseen una cantidad no fija de unidades producidas por mes, se desea probar si la cantidad de unidades defectuosas depende o no directamente de la producción. Es decir que mediante este análisis se podría concluir por ejemplo que a mayor producción, existe un mayor recalentamiento de las máquinas lo que genera mayor cantidad de unidades defectuosas.

El contraste de hipótesis que se plantea está basado en la distribución Ji cuadrado (χ^2) y se lo construye también a partir de un nivel de significancia α . El contraste es el siguiente:

H_0 : Los dos criterios son independientes

H_1 : Los dos criterios NO son independientes

Para que exista una buena aproximación a la distribución Ji cuadrado, es preciso que cada casilla o celda contenga por lo menos 4 observaciones, es por esto que para cada tabla se definirán primero los intervalos según sea el caso.

4.2.4.1. Tablas de Contingencia.

Primero se realizará un análisis para medir la dependencia entre los tipos de llanta con el porcentaje de fallas que presente cada uno de estos, por corrida de producción.

Luego, para cada tipo de llanta se realizará una tabla de contingencia, en la cual los valores de la variable cantidad de unidades producidas estarán ubicados en las filas, mientras que los valores de la variable cantidad de unidades defectuosas estarán situados en las columnas de la tabla.

A continuación se presenta la primera tabla de contingencia: Tipo de llanta – Porcentaje de fallas por corrida de producción.

Para cumplir con la frecuencia de las casillas se optó por clasificar a los porcentajes de fallas en tres categorías:

- Bajo: [0 – 0.75%]
- Medio: [0.76 – 1.20%]
- Alto: [1.21 – 7.0%]

		Porcentaje			Total
		Bajo	Medio	Alto	
Tipo	Camion	71	8	6	85
	Cmt.Bias	67	12	6	85
	Cmt.Radial	5	16	64	85
Total		143	36	76	255

Chi-Square Tests

	Valor	Grados libertad	Asymp. Sig. (2-sided)
Pearson Chi-Square	148.647	4	.000

Tabla XLVI .- *Tabla de Contingencia Tipo de llanta - Porcentaje de fallas.*

Como se puede observar el estadístico de prueba es muy grande, lo que origina un valor p cercano a cero, y por tanto el rechazo de la hipótesis nula de independencia. Este hecho resulta un tanto fácil de visualizar, ya que para el gremio Camion existe mayor frecuencia en el porcentaje bajo, mientras que en el gremio Cmt. Radial la frecuencia más alta está en el porcentaje alto.

- Para el gremio *Camion*.

Para cumplir con el requerimiento de la distribución Ji cuadrado, se debió clasificar la cantidad de unidades defectuosas en solo dos categorías, mientras que la variable cantidad de unidades producidas se eligieron convenientemente tres categorías:

Cantidad de unidades producidas:

- Baja: [178 – 980]
- Media: [981 – 1782]
- Alta: [1783 – 52583]

Cantidad de unidades defectuosas:

- Baja: [0 – 7]
- Alta: [8 – 17]

		Cantidad Defectuosa		Total
		Baja	Alta	
Prod.	Baja	34	4	38
	Media	21	12	33
	Alta	4	10	14
Total		59	26	85

Chi-Square Tests

	Valor	Grados libertad	Asymp. Sig. (2-sided)
Pearson Chi-Square	18.720	2	.000

Tabla XLVII .- *Tabla de Contingencia Cant. Unid. Producidas – Cant. Unid. Defectuosas Gremio Camion.*

En este caso la tabla de contingencia muestra que existe mayor cantidad de unidades defectuosas mientras la producción aumenta, lo que nos lleva por intuición a concluir que existe una dependencia lineal entre las dos variables. Luego, se puede notar que el valor p de la prueba cercano a cero, lo que confirma nuestra intuición.

Por tanto la prueba de la Ji cuadrado muestra que según el criterio del 10% para el valor p, se puede afirmar que existe dependencia entre ambas variables, es decir que la cantidad de unidades defectuosas depende de la cantidad de unidades producidas para este gremio.

- Para el gremio *Cmt. Bias*.

En este gremio de igual manera se debió clasificar la cantidad de unidades defectuosas en solo dos categorías, y tres para la variable cantidad de unidades producidas:

Cantidad de unidades producidas:

- Baja: [390 – 4555]
- Media: [4556 – 7215]
- Alta: [7216 – 12884]

Cantidad de unidades defectuosas:

- Baja: [0 – 32]
- Alta: [33 – 66]

		Cantidad Defectuosa		Total
		Baja	Alta	
Prod.	Baja	45	5	50
	Media	14	9	23
	Alta	6	6	12
Total		65	20	85

Chi-Square Tests

	Valor	Grados libertad	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.871	2	.002

Tabla XLVIII .- *Tabla de Contingencia Cant. Unid. Producidas – Cant. Unid. Defectuosas Gremio Cmt. Bias.*

En este gremio, se obtuvo un valor p mayor a cero pero igualmente muy pequeño, menor que el 10%, por tanto se puede afirmar que existe dependencia entre la cantidad de unidades con defectos y la producción.

- Para el gremio *Cmt.Radial*.

Para este gremio se tienen los siguientes intervalos para cada categoría:

Cantidad de unidades producidas:

- Baja: [220 – 4071]
- Media: [4072 – 6332]
- Alta: [6333 – 13458]

Cantidad de unidades defectuosas:

- Baja: [4 – 65]
- Alta: [66 – 213]

		Cantidad Defectuosa		Total
		Baja	Alta	
Prod.	Baja	38	5	43
	Media	8	12	20
	Alta	4	18	22
Total		50	35	85

Chi-Square Tests

	Valor	Grados libertad	Asymp. Sig. (2-sided)
Pearson Chi-Square	33.429	2	.000

Tabla XLIX.- Tabla de Contingencia Cant. Unid. Producidas – Cant. Unid. Defectuosas Gremio Cmt. Radial.

La tabla de contingencia para este gremio muestra de una forma más clara que las anteriores el nivel alto de dependencia existente, ya que se puede notar que en la primera casilla, es decir la intersección del primer intervalo de la cantidad de unidades producidas [220 - 4071] con el primero de la cantidad de defectuosas [4 - 65], hay mayor frecuencia que en la segunda casilla que es la intersección con el segundo intervalo de las defectuosas. Luego se puede observar fácilmente que la cantidad de fallas aumenta conforme se incrementa la cantidad producida.

Finalmente el valor p de la prueba es cercano a cero, lo que determina estadísticamente que no existe independencia entre las dos variables en estudio.

4.2.4.2. Análisis de correlación.

Luego de obtener como resultado la dependencia entre los pares de variables cantidad de unidades producidas y cantidad de unidades defectuosas, para cada gremio, es preciso realizar un análisis de correlación entre estos mismos pares de variables con el fin de medir el grado de correlación existente.

Se procederá entonces a estimar el coeficiente de correlación lineal r_{ij} entre estas variables. Se tiene que:

$$r_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j}$$

donde $\hat{\sigma}_{ij}$ representa el estimador de la covarianza entre la variable i y la variable j , y $\hat{\sigma}_i, \hat{\sigma}_j$ representan los estimadores de la desviación estándar para la variable i y j respectivamente.

Si retomamos la codificación para estas variables, se tiene que el subíndice 1 está destinado para la cantidad de unidades producidas y el subíndice 2 para la cantidad de unidades defectuosas. Por tanto se tiene:

	Camion	Cmt. Bias	Cmt. Radial
$\hat{\sigma}_{12}$	1028.51	23773.35	120363.84
$\hat{\sigma}_1$	487.64	2638.66	3412.75
$\hat{\sigma}_2$	3.97	15.95	45.94
r_{12}	0.532	0.565	0.768

Tabla L .- *Análisis de Correlación entre las las variables Cant. Unid. Producidas y Cant. Unid. Defectuosas para cada gremio.*

Como era de esperarse existe una correlación positiva en los tres casos. Se puede también observar que los tres valores del estimador de correlación son relativamente altos, lo que sugiere que si existe una dependencia "fuerte" entre cada par de variables, especialmente para el último gremio que posee el valor más cercano a 1.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

En el presente capítulo se expondrán las debidas conclusiones del trabajo luego de realizado el análisis de supervivencia, en el caso de la empresa que comercializa equipos con tecnología; y el análisis de fallas en el proceso de producción para el caso de una empresa de manufactura de llantas. Cada una de las variables involucradas dentro de ambos estudios, fueron cuidadosamente analizadas, tratando de presentar respuestas a los planteamientos y dar una idea concreta de la situación en cuanto a porcentaje de fallas y confiabilidad según el caso.

Luego de las conclusiones presentadas se establecerán las debidas recomendaciones con la finalidad de contribuir a encontrar medidas más eficientes para el mejor modelamiento de la Confiabilidad de los sistemas. Las recomendaciones estarán planteadas como una forma de mejorar indicadores de

confiabilidad, mantenibilidad y disponibilidad, ya sea de equipos o sistemas completos.

5.1. Conclusiones.

5.1.1. Conclusiones para el estudio de la Confiabilidad de una empresa que comercializa equipos con tecnología.

1. El inicio o tiempo cero del análisis fue tomado en cuenta desde la fecha de instalación de los equipos en la empresa cliente, con el fin de simplificar el análisis y poder realizar así la respectiva simulación a partir de los datos entregados por la empresa comercializadora.
2. Se realizó, según la teoría estadística, un ajuste a la variable aleatoria discreta geométrica, ya que los datos de los registros de la primera falla de los equipos estaban dados en días enteros.
3. Bajo este supuesto de que la variable Tiempo de vida o Tiempo de falla T sigue una distribución geométrica, se generaron datos

a partir de la misma, lo cual fue comprobado estadísticamente mediante una prueba de bondad de ajuste.

4. A partir del modelo paramétrico obtenido, se obtuvo un riesgo o una función de riesgo constante para los equipos de ambas líneas de producción; lo que permitió modelar la supervivencia en una escala continua siguiendo una función de densidad exponencial.

5. Así mismo, bajo el supuesto paramétrico de la densidad exponencial para los tiempos de falla, y al tener un modelo de riesgo constante, se tiene una probabilidad relativamente alta de supervivencia de aproximadamente el 74%, para todos los intervalos de 30 días, en el caso de la primera línea de equipo. Por tanto el indicador $MTTF_1 = \beta_1$ es igual a 100, que representa el tiempo promedio de fallas para el modelo exponencial. En la segunda línea se obtuvo una probabilidad de supervivencia para cada intervalo de aproximadamente 35%. Es decir que la probabilidad de que un equipo de la línea 2 registre la primera falla entre un intervalo de 15 días es de aproximadamente 35%, probabilidad un tanto alta considerando el plazo reducido de días en cuestión. El indicador del tiempo promedio de fallas

aproximadamente igual a 35 días sugiere que no se podría poner una garantía mayor a un mes y medio para el caso de estos equipos.

6. Las probabilidades y resultados obtenidos del análisis de supervivencia paramétrico realizado a partir de la distribución geométrica, para ambas líneas de equipo, fueron muy similares a los obtenidos en la distribución exponencial, ya que la máxima diferencia entre las funciones de supervivencia evaluadas en los puntos extremos de cada intervalo, fue del 1%.

7. En el caso de la línea de equipo 1, al regirse a lo que los datos sugieren y no a supuestos paramétricos, se tienen, con los intervalos de tiempo de 30 días, probabilidades de supervivencia cíclicas; ya que en los primeros 30 días la probabilidad de supervivencia es cercana a un 80% y va decreciendo hasta el tercer intervalo (90 días), una vez que ha superado estos 90 días, la probabilidad de supervivencia se incrementa hasta un 85% hasta llegar a los 150 días. Es decir que los equipos de esta línea registran menor cantidad de fallas al poco tiempo de ser instalados al igual que cuando superan los tres meses de operación.

8. Así mismo en el caso de la línea de equipo 2, se observan las probabilidades más altas de supervivencia en los primeros intervalos de 15 días. Pero la probabilidad de supervivencia en el último intervalo de longitud finita decae a un 61% aproximadamente, sugiriendo de esta manera que luego de 2 meses o 60 días de operación del equipo, la probabilidad de que este falle se incrementa de 0.1791 a 0.3922.

9. Con el modelo paramétrico de Weibull, obtenido a partir de una estimación no paramétrica, se obtuvo que el indicador del tiempo promedio de fallas *MTTF* es aproximadamente igual a 165.9, el cual es superior al encontrado en el modelo exponencial planteado al principio.

10. Al realizar una comparación de los indicadores del Tiempo Promedio de fallas entre los modelos paramétrico y no paramétrico (Weibull y Kaplan - Meier), aproximadamente 165.9 y 79.4 respectivamente, se puede notar que existe una diferencia un tanto amplia que conlleva a concluir que este modelo paramétrico no se ajusta muy bien al fenómeno que ocurre en los datos.

5.1.2. Conclusiones para el estudio de las fallas y confiabilidad dentro del proceso de producción de una empresa manufacturera de llantas.

11. Del estudio estadístico de las fallas ocurridas en la producción de llantas, sólo se puede obtener una estimación "indirecta" de la confiabilidad de las máquinas empleadas y por tanto del sistema de producción en sí.
12. El tipo de llanta con mayor cantidad de unidades producidas es el gremio *Cmt. Radial*, con un promedio de aproximadamente 4435 unidades por corrida de producción, seguido del gremio *Cmt. Bias* con un promedio de 4175 unidades producidas aproximadamente, y finalmente el gremio *Camion* con un promedio de producción de aproximadamente 1160 unidades, el cual es muy inferior al promedio de los dos gremios anteriores.
13. Así como el gremio *Cmt. Radial* es el de mayor cantidad de unidades producidas, es el que posee mayor cantidad de unidades defectuosas. Además cabe recalcar que en cada

corrida la producción varía de centenas hasta decenas de miles de unidades.

14. Del análisis univariado, se obtuvo que el porcentaje promedio de fallas por corrida de producción, es de aproximadamente 0.54% para el gremio *Camion* igual que para el gremio *Cmt. Bias*, y para el gremio *Cmt. Radial* es de 1.91%. Es decir que de cada 200 unidades producidas, se registra aproximadamente una falla, o lo que es lo mismo para el análisis, un producto defectuoso, para los casos de los tipos de llanta *Camion* y *Cmt. Bias*; mientras que para el gremio *Cmt. Radial*, se tiene que por cada 100 unidades producidas se obtienen aproximadamente 2 llantas o unidades defectuosas.

15. Del análisis de independencia se obtuvieron las siguientes conclusiones:

- Los porcentajes de unidades defectuosas por corrida de producción si dependen del gremio o tipo de llanta.
- Así mismo para los 3 tipos de llanta, se obtuvo la dependencia de la cantidad de unidades defectuosas con respecto de la cantidad de unidades fabricadas por

corrida de producción. Es decir que se puede deducir, aunque de manera empírica, que a mayor cantidad de producción, las máquinas se recalientan o pierden precisión en el proceso y por tanto producen mayor cantidad de fallas.

16. Del análisis de correlación se puede concluir que para los 3 tipos de llanta, existe una proporción directa entre estos mismos pares de variables como son la cantidad de unidades producidas y la cantidad de unidades defectuosas, siendo la correlación más fuerte la que se origina en el gremio *Cmt. Radial* con un coeficiente de 0.768, el cual es el más cercano a uno.

5.2. Recomendaciones.

5.2.1. Recomendaciones para el estudio de la Confiabilidad de una empresa que comercializa equipos con tecnología.

1. La primera recomendación para este tipo de análisis es la de empezar definiendo que tipo de equipos o máquinas son las unidades a observar. Luego definir el alcance del estudio junto con los modos de falla y criticidad de las mismas.
2. Dado que los equipos en estudio son reparables, una métrica más adecuada para medir la confiabilidad sería a través del "Tiempo Medio entre Fallas".
3. Como diseñador o fabricante de los equipos, es importante obtener el tiempo hasta la primera falla, similar al caso de dispositivos no reparables. Pero si se toma en cuenta el caso de la empresa cliente, es de mayor interés el tiempo medio entre fallas. Se recomienda entonces, recoger los datos de todas las fallas que se registren y aplicar un modelo basado en el Tiempo Medio entre Fallas, ya que este también brinda información acerca de la mantenibilidad y disponibilidad de los

equipos debido a que este modelo incluye los tiempos fuera de servicio y los tiempos de operación entre falla y falla.

4. De gran utilidad sería también, obtener información acerca de la situación y condiciones del ambiente donde el equipo de tecnología está operando y los niveles de producción que el mismo esté generando.
5. Para el análisis resulta más conveniente empezar por un estudio de un modelo no paramétrico, con la finalidad de obtener a partir del mismo, un diagnóstico para modelos paramétricos.
6. Se recomienda que los cálculos paramétricos de supervivencia, se los realice bajo una distribución continua, ya que brinda mejores estimaciones debido a que el eje de tiempo está dado sobre una escala continua.

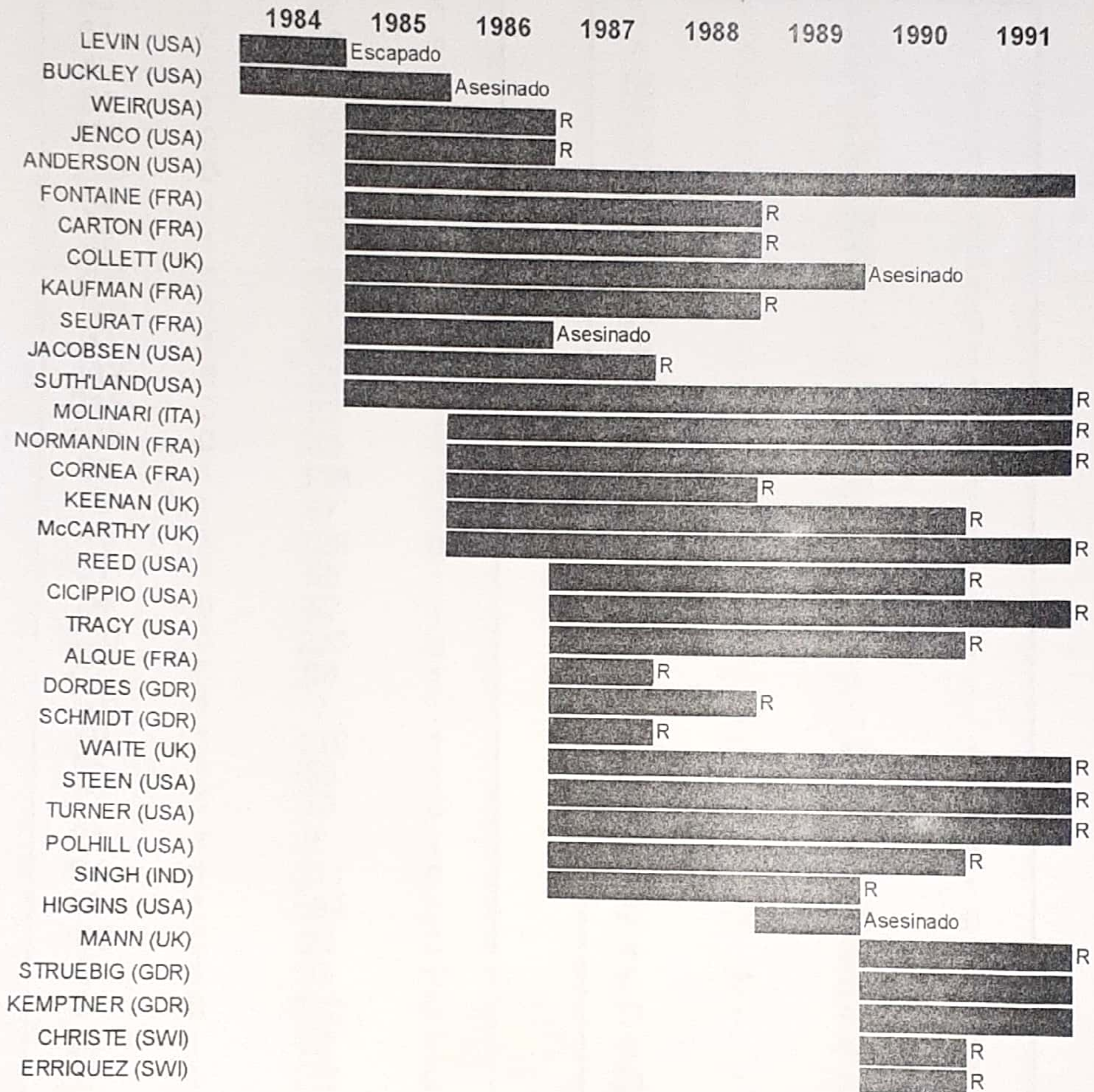
5.2.2. Recomendaciones para el estudio de las fallas y confiabilidad dentro del proceso de producción de una empresa manufacturera de llantas.

7. Para aprovechar de una mejor manera el estudio de la confiabilidad en los sistemas de producción industrial, es preciso obtener el modelamiento de supervivencia del sistema completo en sí. Es decir, según el alcance, medir los tiempos de falla para las máquinas que conforman el sistema de elaboración de llantas, tomando en cuenta si estos componentes se encuentran en serie o en paralelo.
8. El análisis de fallas por corrida de producción podría ser más detallado o directo, obteniendo primero una muestra de la producción en períodos menores a un mes, y a su vez especificando sólo por subtipos de llanta sin tener que agruparlos en gremios o tipos generales.

ANEXO

ANEXO 1

Figura del Ejemplo 2.3.
Situación de Rehenes Occidentales (Censura Tipo I)



Líneas de tiempo desde la captura inicial hasta su liberación, denotada por 'R', o tiempo de censura causado por muerte o no liberación, cerca de 1991 para 34 rehenes occidentales.
Fuente: Peter J. Smith, 2002, *Analysis of Failure and Survival Data*, pág 74.

ANEXO 2

Datos para los tiempos de primera falla

Línea de Equipo 1

Datos			
Día	Pérdida	Día	Pérdida
43	1	47	1
61	0	39	1
40	0	63	1
51	1	22	1
303	0	87	1
229	0	11	1
34	1	198	0
37	1	49	0
204	0	22	1
126	1	77	1
33	1	134	1
112	0	41	0
215	0	22	1
145	0	49	1
66	1	158	0
79	1	67	1
49	1	129	0
51	1	58	1
24	0	52	1
250	0	236	0
261	0	98	1
126	1	7	1
70	1	119	1
35	1	86	1
73	1	73	0
21	1	5	1
94	1	311	0
69	1	184	0
105	0	28	0
2	1	35	1
3	0	202	0
90	1	23	1
32	1	116	0
315	0	22	1
1	1	74	1
250	0	114	1
6	1	5	0
113	1	197	0
116	1	29	1
253	0	40	1
2	1	26	1
69	1	16	1
130	1	215	0
16	1	33	0
81	1	203	0
34	1	56	1
260	0	152	0
92	1	11	1
180	0	3	1
6	1	31	0

Línea de Equipo 2

Datos			
Día	Pérdida	Día	Pérdida
3	1	20	1
8	0	32	1
68	1	16	0
25	1	44	1
14	0	20	1
24	1	81	0
21	1	24	1
56	1	29	0
8	1	3	1
43	1	86	0
26	0	69	1
14	1	22	1
79	0	29	1
46	1	40	0
22	1	38	1
63	0	46	0
10	1	31	1
86	0	18	1
84	0	25	1
32	1	4	1
28	1	4	1
29	1	27	1
56	0	17	1
31	0	4	1
7	0	10	0
9	1	79	0
1	1	33	1
42	0	10	0
4	1	28	1
6	1	61	1
11	1	57	1
9	1	96	0
21	1	81	0
3	1	7	1
10	0	36	1
80	0	60	1
11	1	34	1
20	0	6	1
50	1	66	1
7	1	92	0
70	1	6	1
66	1	48	1
28	0	65	1
3	1	65	1
78	0	75	0
39	0	55	0
96	0	9	1
62	1	31	1
77	0	76	0
57	1	25	1

0 = censura
1 = falla

ANEXO 3

Funciones de Densidad, de Supervivencia y de Riesgo a partir de la distribución de Weibull

	Línea de Equipo 1	Línea de Equipo 2
Función de Densidad	$f_1(t) = \frac{0.82}{(149.32)^{0.82}} t^{-0.18} e^{-\left(\frac{t}{149.32}\right)^{0.82}}$	$f_2(t) = \frac{1.052}{(55.471)^{1.052}} t^{0.052} e^{-\left(\frac{t}{55.471}\right)^{1.052}}$
Función de Supervivencia	$S_1(t) = e^{-\left(\frac{t}{149.32}\right)^{0.82}}$	$S_2(t) = e^{-\left(\frac{t}{55.471}\right)^{1.052}}$
Función de Riesgo	$h_1(t) = \frac{0.82}{149.32^{0.82}} t^{-0.18}$	$h_2(t) = \frac{1.052}{(55.471)^{1.052}} t^{-0.052}$

BIBLIOGRAFÍA

1. Peter J. Smith, 2002, Analysis of Failure and Survival Data, Primera Edición, Chapman & Hall/CRC, U.S.A.
2. W.Q. Meeker and L.A. Escobar, 1998, Statistical Methods for Reliability Data, Primera Edición, Wiley.
3. Brian S. Everitt, 2002, Statistical Analyses using S-Plus, Segunda Edición, Chapman & Hall/CRC, U.S.A.
4. Mendenhall William, 1994, Estadística Matemática con Aplicaciones, Segunda Edición, Grupo Editorial Iberoamérica S.A. de C.V., México.
5. 2003, <http://www.udl.es> , España
6. 2002, <http://www.psicomecánica.com> , España
7. 2001, <http://www.math.montana.edu> , Estados Unidos