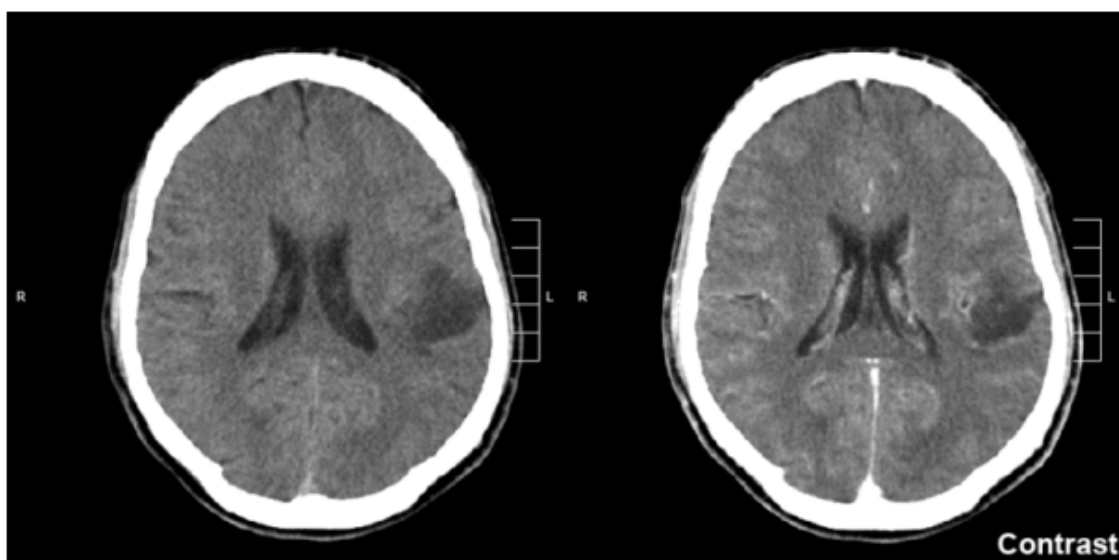




TAREA 1

María Pallares Diez

Profesor: Francisco J. Santonja



Módulo: Herramientas
matemáticas e informáticas

Asignatura: Probabilidad y
Simulación

Máster Universitario en
Bioestadística

Universidad de Valencia

INFORME DE ANÁLISIS DE SUPERVIVENCIA EN PACIENTES CON GLIOMA EN ESTADO AVANZADO

Proyecto de Investigación del Hospital Clínic Universitari de València

Autor: Equipo de Bioestadística

Fecha: octubre 2024

INTRODUCCIÓN

El objetivo de este estudio es analizar el tiempo de supervivencia de pacientes con gliomas en estado avanzado tratados con un nuevo fármaco experimental. Los gliomas son tumores malignos del sistema nervioso central, localizados principalmente en el cerebro. Su pronóstico es generalmente desfavorable, y los esfuerzos clínicos se centran en controlar el crecimiento del tumor y mejorar la calidad de vida del paciente.

En este proyecto, se analizaron los tiempos de supervivencia de 28 pacientes, medidos en meses, para determinar la efectividad del tratamiento y proporcionar información valiosa para futuras intervenciones clínicas. Los datos se han modelado utilizando diferentes distribuciones estadísticas, y se ha seleccionado el modelo con mejor ajuste para realizar los cálculos de probabilidad y las interpretaciones correspondientes.

DATOS Y METODOLOGÍA

- Variable de interés: Tiempo de supervivencia desde el diagnóstico hasta la ocurrencia del evento de interés (muerte).
- Número de pacientes: 28.
- Formato de los datos: Tiempo de supervivencia en meses.

Tiempos de Supervivencia Observados (meses):

36	47	35	22	125	15	28	90	13	230
63	28	19	28	88	85	14	9	9	12
35	19	36	22	13	18	21	67		

Se han ajustado los siguientes modelos probabilísticos a los datos de supervivencia:

- Gamma
- Exponencial
- Normal
- Chi-cuadrado
- Uniforme

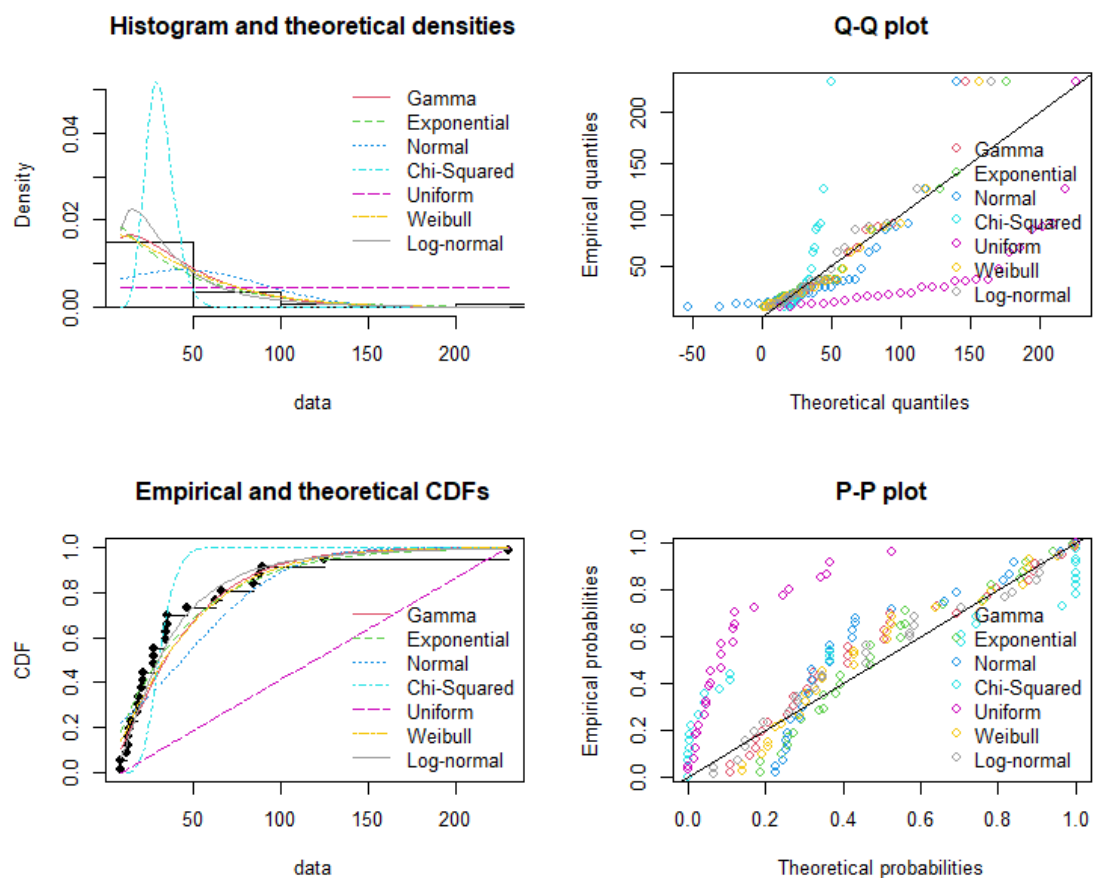
6. Weibull

7. Log-normal

Cada modelo se ha evaluado en R utilizando los criterios de información de Akaike (AIC) y Bayesiano (BIC) para determinar el mejor ajuste.

RESULTADOS DE EVALUACIÓN DE MODELOS

Representamos cada uno de los modelos para tener una comparación gráfica de diferentes distribuciones teóricas ajustadas a los datos empíricos. Este tipo de visualización es útil para evaluar cuál de las distribuciones se ajusta mejor a los datos observados.

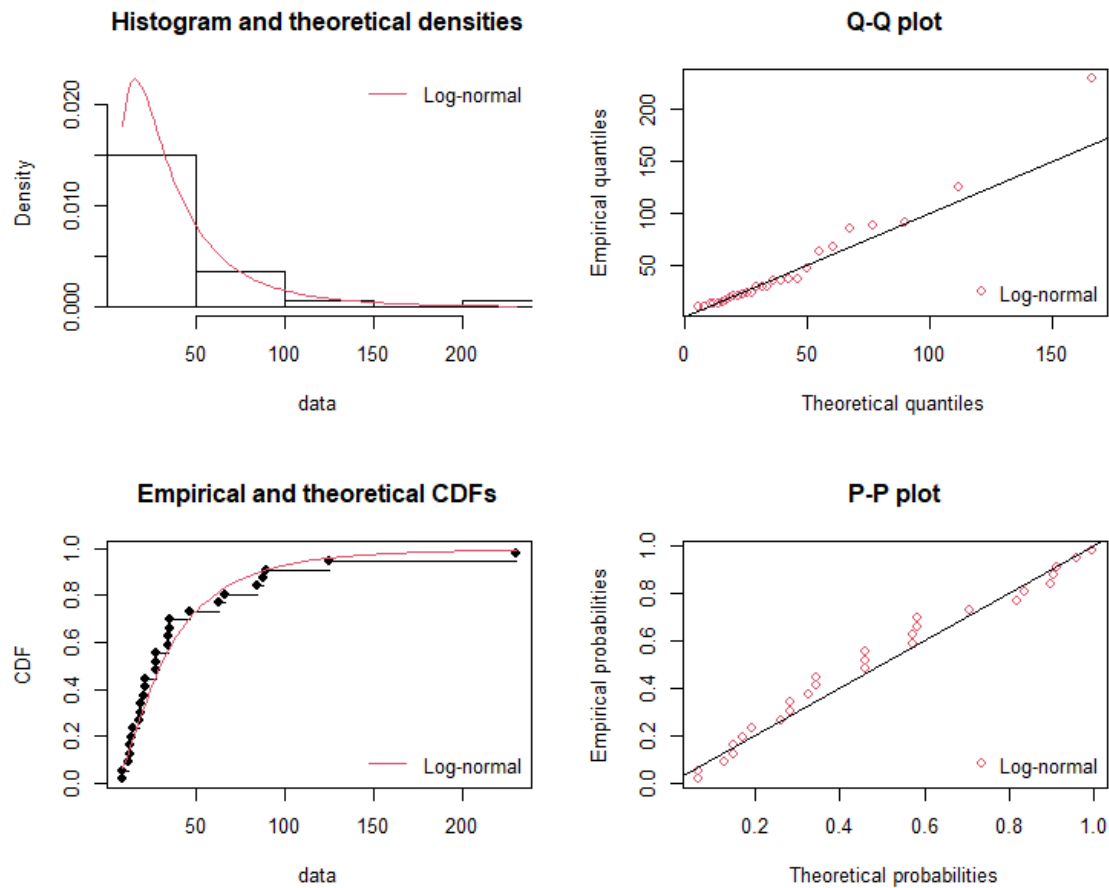


Cada uno de los gráficos proporciona una forma diferente de evaluar visualmente el ajuste de los modelos teóricos a los datos observados:

- El **histograma** y las densidades teóricas indican qué tan bien las distribuciones modelan la forma general de los datos.
- El **Q-Q plot** muestra cómo se alinean los cuantiles teóricos y empíricos.
- El **CDF** compara las distribuciones acumuladas, y el **P-P plot** compara probabilidades.

En todos estos gráficos, la distribución Log-normal destaca como la mejor, lo que valida la selección del modelo para el análisis, por ello la representamos en solitario para

comprobarlo, y vimos que en cada uno de los gráficos el modelo log-normal se ajusta perfectamente a nuestros datos.



A continuación, calculamos y observamos los valores AIC y BIC para cada uno de los modelos. Los valores AIC y BIC indican qué tan bien cada modelo se ajusta a los datos. A menor valor de AIC/BIC, mejor es el ajuste del modelo. El mejor modelo seleccionado fue la distribución **Log-normal**, que presentó los valores más bajos de AIC y BIC.

Modelo	AIC	BIC
Gamma	269.18	271.85
Exponencial	269.69	271.01
Normal	297.93	300.60
Chi-cuadrado	546.57	547.90
Uniforme	306.29	308.96
Weibull	270.74	273.41
Log-normal	262.68	265.34

Interpretación: El modelo Log-normal mostró el mejor ajuste, lo que sugiere que la distribución de los tiempos de supervivencia se asemeja a una forma logarítmica. Esta forma es consistente con situaciones en las que la tasa de supervivencia cambia de manera no lineal a lo largo del tiempo.

ANÁLISIS DE SUPERVIVENCIA

Se calcularon con la ayuda de R las probabilidades de supervivencia para dos periodos clave:

1. Probabilidad de supervivencia de al menos 12 meses:

$$P(T \geq 12) = 87.4\%$$

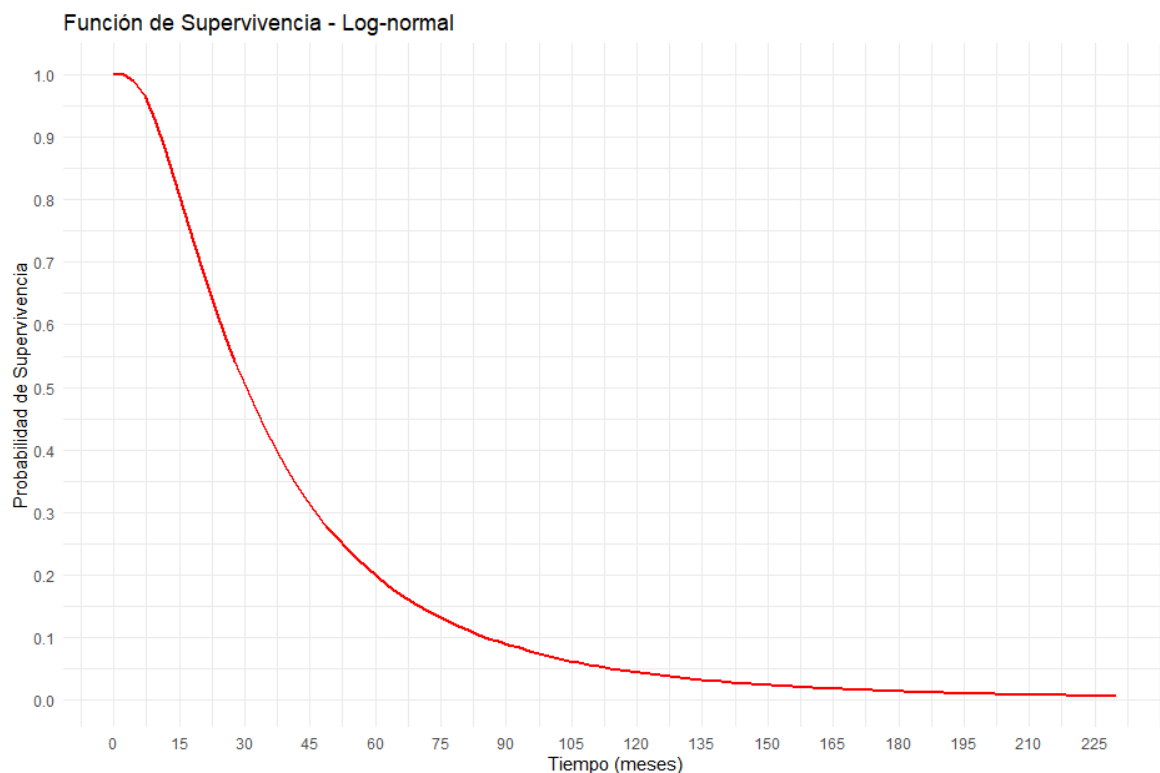
2. Probabilidad de supervivencia de al menos 24 meses:

$$P(T \geq 24) = 61.3\%$$

Estos resultados indican que más del 87% de los pacientes sobreviven al menos un año, mientras que el 61% supera los dos años de tratamiento. Esto sugiere que el fármaco podría tener un impacto positivo en la supervivencia a corto plazo, pero se observa una caída significativa en la probabilidad de supervivencia a medida que el tiempo avanza.

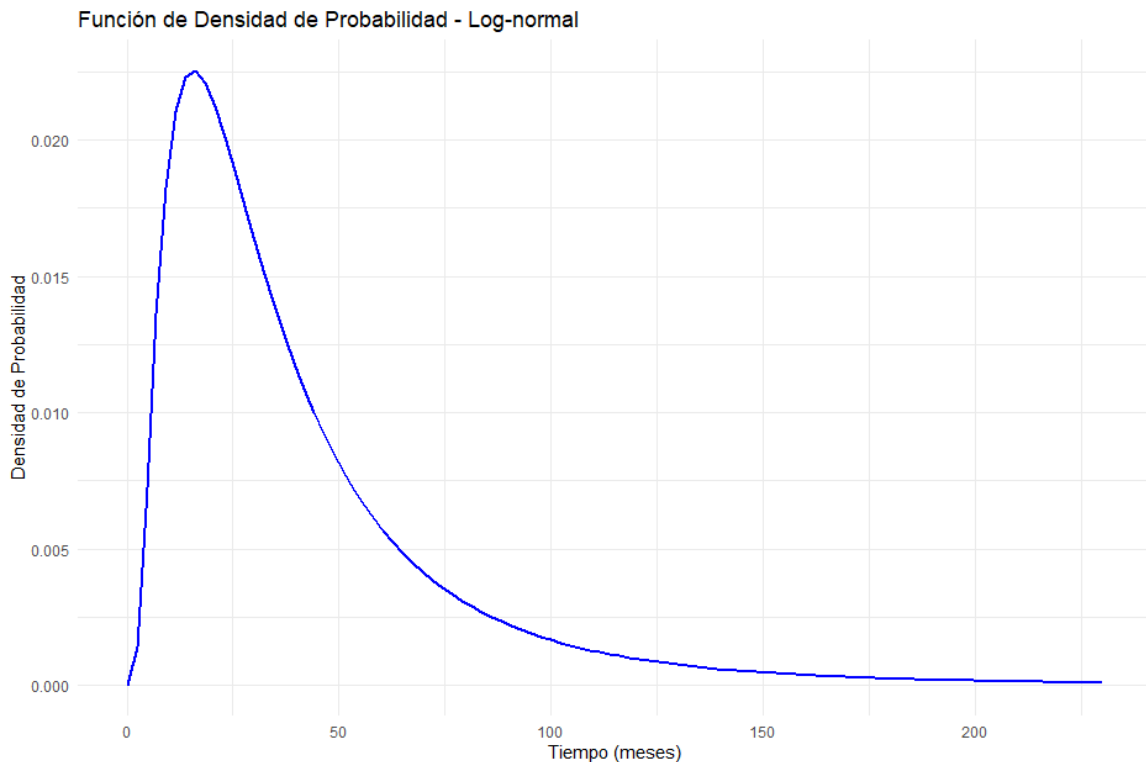
REPRESENTACIÓN DE LA FUNCIÓN DE SUPERVIVENCIA

Se construyó la función de supervivencia $S(t) = P(T \geq t)$ para el modelo Log-normal. La gráfica muestra cómo la probabilidad de supervivencia disminuye a lo largo del tiempo. La curva inicial tiene una pendiente pronunciada, lo que indica un alto riesgo en los primeros meses tras el diagnóstico, estabilizándose posteriormente en los pacientes que superan los dos años de tratamiento.



REPRESENTACIÓN DE LA FUNCIÓN DE DENSIDAD

A continuación, se representó también con la ayuda de R la función de densidad de probabilidad $f(t) = P(T \geq t)$.



La gráfica mostrada representa la Función de Densidad de Probabilidad ajustada a los tiempos de supervivencia de los pacientes utilizando la distribución Log-normal. Esta gráfica nos proporciona información sobre la probabilidad de ocurrencia del evento (por ejemplo, fallecimiento) en un instante específico a lo largo del tiempo. A continuación, se presenta un análisis detallado de la información que nos proporciona esta imagen:

Descripción General

1. Eje X ("Tiempo (meses)"):

- El eje X muestra el tiempo de supervivencia en meses, desde el diagnóstico hasta la ocurrencia del evento terminal. La gráfica cubre un rango amplio que va desde los primeros meses hasta más de 200 meses.
- Este eje nos indica en qué momento específico se calcula la probabilidad de fallecimiento para un individuo.

2. Eje Y ("Densidad de Probabilidad"):

- El eje Y representa la densidad de probabilidad de ocurrencia del evento en un tiempo específico t .
- La **densidad** no es una probabilidad acumulada, sino la **tasa de cambio de la probabilidad** en ese instante de tiempo. Valores más altos indican una mayor propensión a que el evento ocurra en ese momento específico.

3. Comportamiento de la Curva:

- La curva tiene una forma **asimétrica**, con un **pico pronunciado** en los primeros meses (alrededor de 20 meses) y luego una disminución constante.
- Este comportamiento es característico de la distribución Log-normal, donde la densidad de probabilidad inicialmente aumenta, alcanza un valor máximo y luego disminuye.

Detalles Clave y Análisis

1. Pico Principal de la Densidad:

- El **pico** de la densidad se encuentra alrededor de los **20 meses**, lo que sugiere que existe un **riesgo máximo de ocurrencia** del evento (fallecimiento) en este punto.
- A medida que el tiempo avanza, la densidad disminuye de manera pronunciada, lo que indica que después de los 20 meses, la probabilidad de fallecimiento en un instante específico comienza a disminuir.

2. Disminución Continua:

- A partir del punto de máximo riesgo (~20 meses), la curva muestra una **caída gradual**. Esto sugiere que los pacientes que superan este periodo tienen una probabilidad menor de fallecer en los meses posteriores.
- A partir de los **100 meses**, la densidad es muy baja y se mantiene cercana a cero. Esto indica que la mayoría de los eventos de interés (fallecimientos) ya habrán ocurrido para este tiempo.

3. Interpretación Clínica:

- Los datos sugieren que los primeros **20-50 meses** son críticos para los pacientes tratados con el fármaco, ya que es en este rango de tiempo donde se concentra la mayor densidad de riesgo.
- Pacientes que sobreviven más allá de este periodo tienen una probabilidad de fallecimiento significativamente menor, lo que sugiere que la efectividad del tratamiento podría estabilizar la enfermedad a largo plazo.

CONCLUSIONES

1. **El modelo Log-normal fue el mejor ajuste** para los datos de supervivencia de los pacientes con glioma en estado avanzado, según los criterios de AIC y BIC. La distribución Log-normal refleja de manera adecuada el comportamiento de los tiempos de supervivencia, ajustándose mejor que otros modelos como el Weibull o el Exponencial.
2. La **probabilidad de supervivencia a 12 meses** se estima en **87.4%**, mientras que a los **24 meses** es de **61.3%**. Esto sugiere que la mayoría de los pacientes tratados

con el fármaco logran sobrevivir el primer año tras el diagnóstico, pero la supervivencia disminuye progresivamente a partir del segundo año.

3. La **gráfica de densidad de probabilidad** muestra un **pico máximo de riesgo alrededor de los 20 meses**. Esto indica que, en promedio, el periodo de mayor riesgo de fallecimiento para los pacientes se concentra en los primeros dos años. Pasado este punto crítico, la densidad de probabilidad disminuye de manera constante, lo que sugiere que los pacientes que superan los 50 meses tienen un menor riesgo a largo plazo.
4. La **función de supervivencia** y la **densidad de probabilidad** destacan la importancia de intervenir clínicamente en los primeros 50 meses para maximizar la efectividad del tratamiento. Los resultados indican la necesidad de evaluar la progresión de la enfermedad de manera más frecuente durante este periodo.

ANEXO

Código en R:

```
# Cargar las librerías necesarias
```

```
library(readr)
```

```
library(MASS)
```

```
library(fitdistrplus)
```

```
library(ggplot2)
```

```
# Leer los datos
```

```
Super <- read.csv("C:/Users/Maria/Desktop/UV/Herramientas_matemáticas_informáticas/PROBABILIDAD/R/Super.csv", sep="")
```

```
View(Super)
```

```
head(Super)
```

```
Temps <- Super$Temps
```

```
# Ajustar las distribuciones
```

```
fgamma <- fitdist(Temps, "gamma")
```

```
fexp <- fitdist(Temps, "exp")
```

```
fnorm <- fitdist(Temps, "norm")
```

```
fchi <- fitdist(Temps, "chisq", start=list(df=2))
```

```
funif <- fitdist(Temps, "unif")
```

```
fweibull <- fitdist(Temps, "weibull")
```

```
flognorm <- fitdist(Temps, "lnorm")
```

```
# Crear el gráfico de comparación
```

```
par(mfrow=c(2,2))
```

```
plot.legend <- c("Gamma", "Exponential", "Normal", "Chi-Squared", "Uniform", "Weibull", "Log-normal")
```

```
# Comparación de distribuciones usando los gráficos de `fitdistrplus`
```

```
denscomp(list(fgamma, fexp, fnorm, fchi, funif, fweibull, flognorm),  
legendtext=plot.legend) # Comparación de densidades
```

```
qqcomp(list(fgamma, fexp, fnorm, fchi, funif, fweibull, flognorm), legendtext=plot.legend)
# Comparación QQ plot
```

```
cdfcomp(list(fgamma, fexp, fnorm, fchi, funif, fweibull, flognorm), legendtext=plot.legend)
# Comparación CDF
```

```
ppcomp(list(fgamma, fexp, fnorm, fchi, funif, fweibull, flognorm), legendtext=plot.legend)
# Comparación PP plot
```

```
#Ver los AIC y BIC de cada uno de los modelos
```

```
summary(fgamma)
```

```
summary(fexp)
```

```
summary(fnorm)
```

```
summary(fchi)
```

```
summary(funif)
```

```
summary(fweibull)
```

```
summary(flognorm)
```

```
#Representamos log-normal que es la que tiene mejores valores AIC y BIC
```

```
par(mfrow=c(2,2))
```

```
plot.legend <- c("Log-normal")
```

```
denscomp(list(flognorm), legendtext=plot.legend) # Comparación de densidades
```

```
qqcomp(list(flognorm), legendtext=plot.legend) # Comparación QQ plot
```

```
cdfcomp(list(flognorm), legendtext=plot.legend) # Comparación CDF
```

```
ppcomp(list(flognorm), legendtext=plot.legend) # Comparación PP plot
```

```
# Calcular la probabilidad de supervivencia para 12 y 24 meses con la distribución Log-normal
```

```
prob_12 <- 1 - plnorm(12, meanlog = flognorm$estimate[1], sdlog = flognorm$estimate[2])
```

```
prob_24 <- 1 - plnorm(24, meanlog = flognorm$estimate[1], sdlog = flognorm$estimate[2])
```

```
# Mostrar las probabilidades calculadas
```

```
cat("Probabilidad de supervivencia de al menos 12 meses:", prob_12, "\n")
```

```
cat("Probabilidad de supervivencia de al menos 24 meses:", prob_24, "\n")
```

```
# Crear una secuencia de tiempos para graficar
```

```

time_values <- seq(0, max(Temps), length.out = 100)

# Calcular la función de supervivencia para cada tiempo
surv_values <- 1 - plnorm(time_values,
                          meanlog = flognorm$estimate["meanlog"],
                          sdlog = flognorm$estimate["sdlog"])

#GRAFICAR  $S(t)=P(T\geq t)$ :
# Crear un dataframe para la gráfica
surv_data <- data.frame(Time = time_values, Survival_Probability = surv_values)

# Graficar la función de supervivencia
ggplot(surv_data, aes(x = Time, y = Survival_Probability)) +
  geom_line(color = "red", size = 1) +
  labs(title = "Función de Supervivencia - Log-normal",
       x = "Tiempo (meses)",
       y = "Probabilidad de Supervivencia") +
  scale_x_continuous(breaks = seq(0, max(time_values), by = 15)) + # Marcas del eje X cada 15 meses
  scale_y_continuous(breaks = seq(0, 1, by = 0.1)) + # Marcas del eje Y cada 0.1 (10%)
  theme_minimal()

#GRAFICAR  $f(t)=P(T\geq t)$ :
# Crear una secuencia de tiempos para graficar la densidad de probabilidad
time_values <- seq(0, max(Temps), length.out = 100)

# Calcular la función de densidad de probabilidad para cada tiempo
meanlog <- flognorm$estimate["meanlog"]
sdlog <- flognorm$estimate["sdlog"]

# Calcular la función de densidad de probabilidad para cada tiempo usando los parámetros
ajustados por `fitdist`
density_values <- dlnorm(time_values, meanlog = meanlog, sdlog = sdlog)

```

```
# Crear un dataframe para la gráfica
density_data <- data.frame(Time = time_values, Density = density_values)

# Graficar la función de densidad de probabilidad
ggplot(density_data, aes(x = Time, y = Density)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Función de Densidad de Probabilidad - Log-normal",
        x = "Tiempo (meses)",
        y = "Densidad de Probabilidad") +
  theme_minimal()
```