In the previous lectures, we did not make any assumptions on the distribution of the lifetime variable T

Since this random variable is only defined for positive value and often is highly skewed, we intuitively feel that a normal distribution is not suited to describe this distribution.

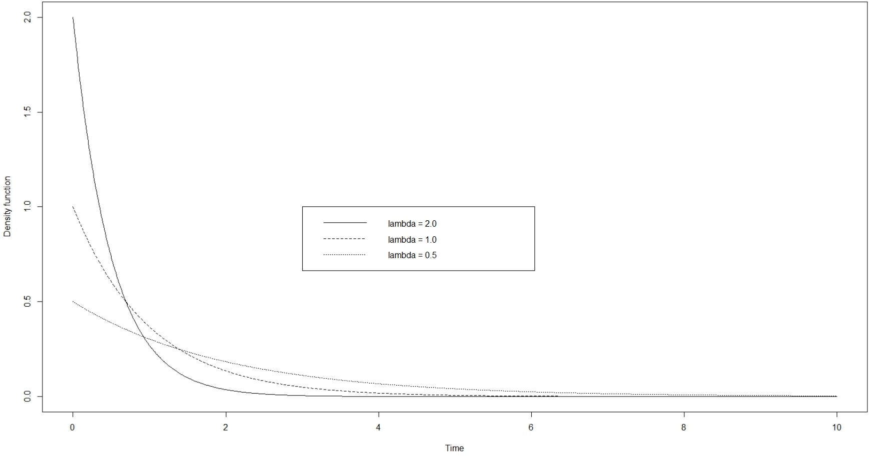
Therefore we propose some popular survival distributions.

As a first popular parametric distribution, we consider the exponential distribution.

A lifetime T has an exponential distribution if for $\lambda > 0$,

$$
\begin{aligned}
f(t) &= \lambda \exp(-\lambda t), \quad t > 0 \\
S(t) &= \exp(-\lambda t), \quad t > 0 \\
\lambda(t) &= \lambda, \quad t > 0
\end{aligned}
$$

This is the only distribution with a constant hazard.


2.0
1.5
1.0
0.5
0.0
Density function
lambda = 2.0
lambda = 1.0
lambda = 0.5
0
2
4
6
8
10
Time

Furthermore we can show that

$$E[T] = \frac{1}{\lambda} < +\infty.$$

The median $t_{50\%}$ is the time point at which 50% has failed,

$$S(t_{50\%}) = 0.5 \Leftrightarrow t_{50\%} = \frac{1}{\lambda} \log 2.$$

More generally, the p-th percentile of the survival distribution is given by

$$S(t_p) = 1 - p \Leftrightarrow t_p = -\frac{\log(1-p)}{\lambda}.$$

The exponential distribution has a lack of memory property, given by

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

In reality, most processes rarely follow an exponential distribution.

This is the result of having a hazard function which depends on time and age of an individual.

If we have events which tend to occur constantly over time, we still can use the exponential distribution.

For the mean residual lifetime, we get

$$r(t) = E[T - t|T > t] = E[T] = \frac{1}{\lambda}.$$

This means that the expected future lifetime for a system or individual is not influenced by the time t that it has lived so far.

It is assumed that there is no ageing effect.

Also if we look a small time intervals, the exponential distribution gives good approximations.

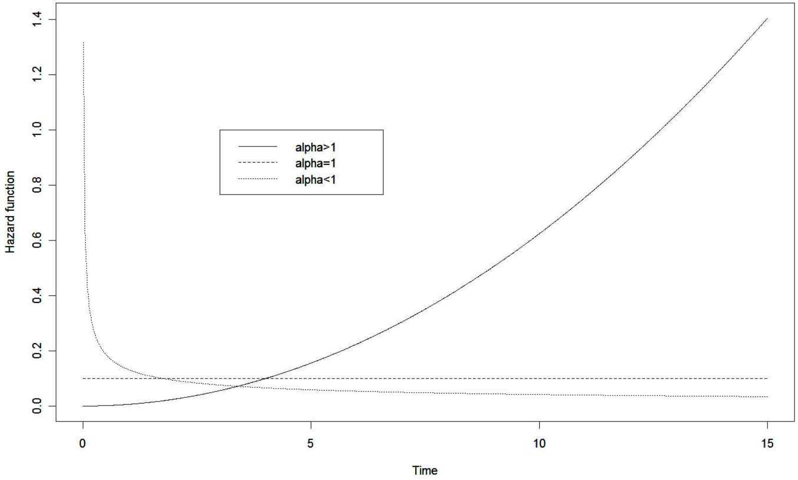
Weibull distribution

A lifetime T has a Weibull distribution if for $\lambda > 0$ and $\alpha > 0$,

$$\begin{aligned} f(t) &= \alpha\lambda \exp(-\lambda t^{\alpha-1}), \quad t > 0 \\ S(t) &= \exp(-\lambda t^{\alpha}), \quad t > 0 \\ \lambda(t) &= \lambda\alpha t^{\alpha-1}, \quad t > 0 \end{aligned}$$

where λ is a scale parameter while α is a shape parameter.

We note that the exponential distribution is a special case of the Weibull ($\alpha = 1$).


alpha>1
alpha=1
alpha<1
Hazard function
Time
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
5
10
15

As for the exponential distribution, we can show that

$$E[T] = \lambda^{-1/\alpha}\Gamma(\alpha^{-1} - 1)$$

with

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du.$$

The median $t_{50\%}$ is the time point at which 50% has failed,

$$S(t_{50\%}) = 0.5 \Leftrightarrow t_{50\%} = \left(\frac{1}{\lambda}\log 2\right)^{1/\alpha}.$$

More generally, the p-th percentile of the survival distribution is given by

$$S(t_p) = 1 - p \Leftrightarrow t_p = \left(-\frac{\log(1-p)}{\lambda}\right)^{1/\alpha}.$$

Extreme value distribution

Sometimes it is more useful to work with the logarithm of the lifetimes.

If $T \sim \text{Weibull}(\lambda, \alpha)$, we get that

$$Y = \log(T) = \mu + \sigma E$$

with $\mu = (-\log \lambda)/\alpha$, $\sigma = 1/\alpha$ and E the standard extreme value distribution

$$f_E(w) = \exp(w - e^w), \quad w \in \mathbb{R}.$$

We note that a reparametrization of the parameters gives $\mu \in \mathbb{R}$ and $\sigma > 0$.

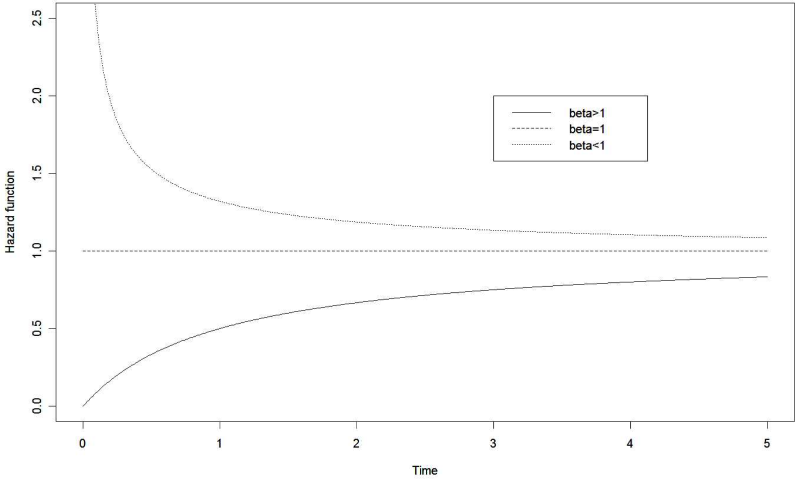
A lifetime T has a gamma distribution if for $\lambda > 0$ and $\beta > 0$,

$$\begin{aligned} f(t) &= \frac{\lambda^{\beta} t^{\beta-1}}{\Gamma(\beta)} \exp(-\lambda t), \quad t > 0 \\ S(t) &= 1 - \frac{1}{\Gamma(\beta)} \int_0^{\lambda t} u^{\beta-1} \exp(-u) du, \quad t > 0 \end{aligned}$$

where λ is a scale parameter while β is a shape parameter.

We note that

$$E[T] = \frac{\beta}{\lambda} \qquad \text{and} \qquad \mathrm{Var}(T) = \frac{\beta}{\lambda^2}.$$

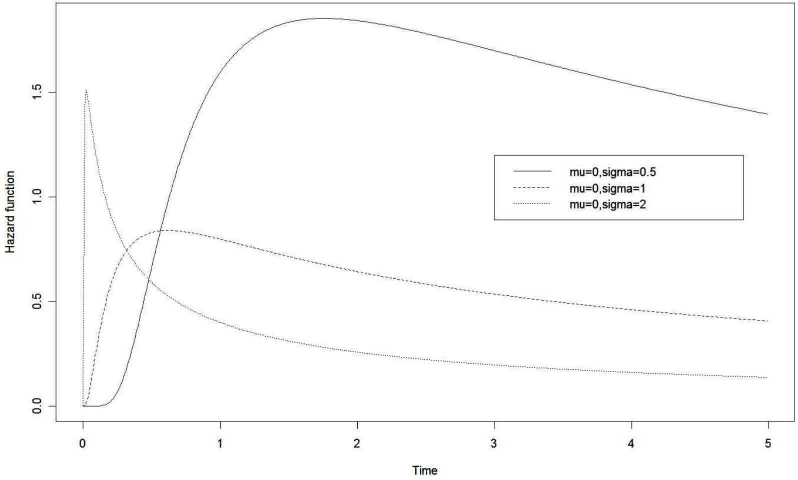

beta>1
beta=1
beta<1
Hazard function
Time
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5

A lifetime T has a lognormal distribution if $\log(T)$ has a normal distribution.

This gives

$$
\begin{aligned}
f(t) &= \frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right), \quad t > 0 \\
&= \frac{1}{t}\phi\left(\frac{\log t - \mu}{\sigma}\right), \quad t > 0 \\
S(t) &= 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right), \quad t > 0
\end{aligned}
$$

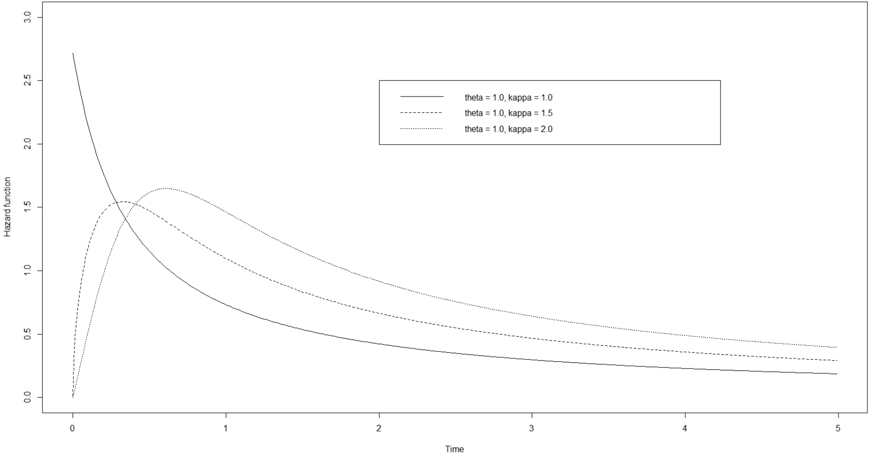
where $\mu \in \mathbb{R}$ and $\sigma > 0$.


mu=0,sigma=0.5
mu=0,sigma=1
mu=0,sigma=2
Hazard function
0.0
0.5
1.0
1.5
Time
0
1
2
3
4
5

A lifetime T has a log-logistic distribution when for $\kappa > 0$, $\theta \in \mathbb{R}$,

$$
\begin{aligned}
f(t) &= \frac{e^{\theta}\kappa t^{\kappa-1}}{(1+e^{\theta}t^{\kappa})^2} \\
S(t) &= [1+e^{\theta}t^{\kappa}]^{-1} \\
\lambda(t) &= \frac{e^{\theta}\kappa t^{\kappa-1}}{1+e^{\theta}t^{\kappa}}
\end{aligned}
$$

As for the lognormal distribution, we note that T has a log-logistic distribution if $\log(T)$ has a logistic distribution.


theta = 1.0, kappa = 1.0
theta = 1.0, kappa = 1.5
theta = 1.0, kappa = 2.0
Hazard function
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time
0
1
2
3
4
5

A lifetime T has a Gompertz distribution if

$$
\begin{aligned}
f(t) &= \theta e^{\alpha t} \exp\left(\frac{\theta}{\alpha}(1 - e^{\alpha t})\right), \quad t > 0 \\
S(t) &= \exp\left(\frac{\theta}{\alpha}(1 - e^{\alpha t})\right), \quad t > 0 \\
\lambda(t) &= \theta e^{\alpha t}, \quad t > 0
\end{aligned}
$$

where $\alpha > 0$ and $\theta > 0$.

We note that the hazard function is increasing in t and that $\lim_{t \to 0} \lambda(t) = \theta \Rightarrow \theta$ is background hazard.

Generalized Gamma distribution

We extend the Gamma distribution and define

$$f(t) = \frac{\theta \lambda^{\rho\theta} t^{\rho\theta - 1} \exp[-(\lambda t)^{\theta}]}{\Gamma(\rho)}.$$

- If $\theta = \rho = 1 \Rightarrow$ Exponential
- If $\theta = 1 \Rightarrow$ Gamma
- If $\rho = 1 \Rightarrow$ Weibull
- If $\rho \to +\infty \Rightarrow$ Lognormal

Pareto distribution

A lifetime T has a Pareto distribution if

$$\begin{aligned} f(t) &= \frac{\theta\lambda^\theta}{t^{\theta+1}}, \quad t > \lambda \\ S(t) &= \frac{\lambda^\theta}{t^\theta}, \quad t > \lambda \\ \lambda(t) &= \frac{\theta}{t}, \quad t > \lambda \end{aligned}$$

where $\lambda > 0$ and $\theta > 0$.

Inverse Gaussian distribution

A lifetime T has an Inverse Gaussian distribution if

$$
\begin{aligned}
f(t) &= \sqrt{\frac{\lambda}{2\pi t^3}}\exp\left(\frac{\lambda(t-\mu^2)}{2\mu^2 t}\right), \quad t>0 \\
S(t) &= \Phi\left(\sqrt{\frac{\lambda}{t}}\left(1-\frac{t}{\mu}\right)\right) - e^{\frac{2\lambda}{\mu}}\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(1+\frac{t}{\mu}\right)\right), \quad t>0
\end{aligned}
$$

where $\lambda > 0$.

The hazard of this distribution has a complicated form!

For some parametric distributions, we can easily verify whether they are a good candidate.

- If $T \sim \text{Exp}(\lambda)$, we have that

$$\begin{aligned} S(t) &= \exp(-\lambda t) \\ \log(S(t)) &= -\lambda t. \end{aligned}$$

$\Rightarrow$ Plot $\log(\hat{S}(t))$ vs t.

- If $T \sim \text{Weibull}(\lambda, \alpha)$, we have that

$$\begin{aligned} S(t) &= \exp(-\lambda t^{\alpha}) \\ \log(S(t)) &= -\lambda t^{\alpha} \\ \log(-\log(S(t))) &= \log(\lambda) + \alpha \log(t). \end{aligned}$$

$\Rightarrow$ Plot $\log(-\log(\hat{S}(t)))$ vs $\log(t)$.

- If T has a log-logistic distribution, we have that

$$\begin{aligned}\frac{S(t)}{1-S(t)} &= e^{-\theta}t^{-\kappa} \\ \log\left(\frac{S(t)}{1-S(t)}\right) &= -\theta-\kappa\log(t).\end{aligned}$$

$\Rightarrow$ Plot $\log\left(\frac{\hat{S}(t)}{1-\hat{S}(t)}\right)$ vs $\log(t)$.

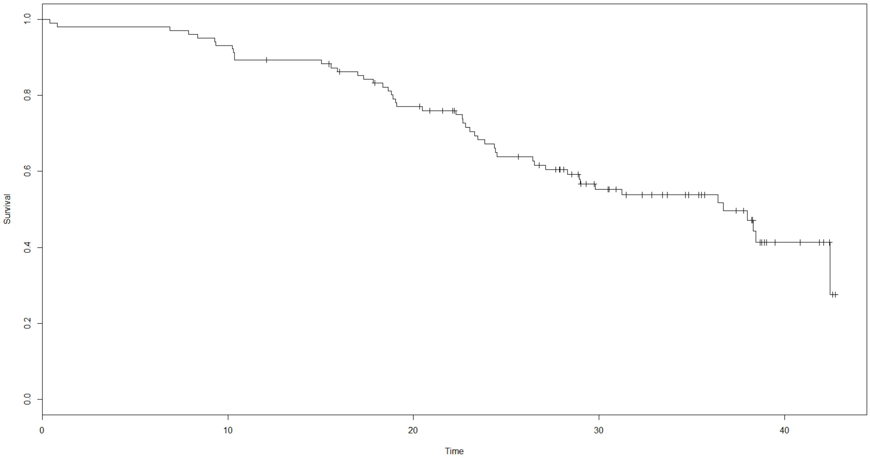
- If T has a lognormal distribution, we have that

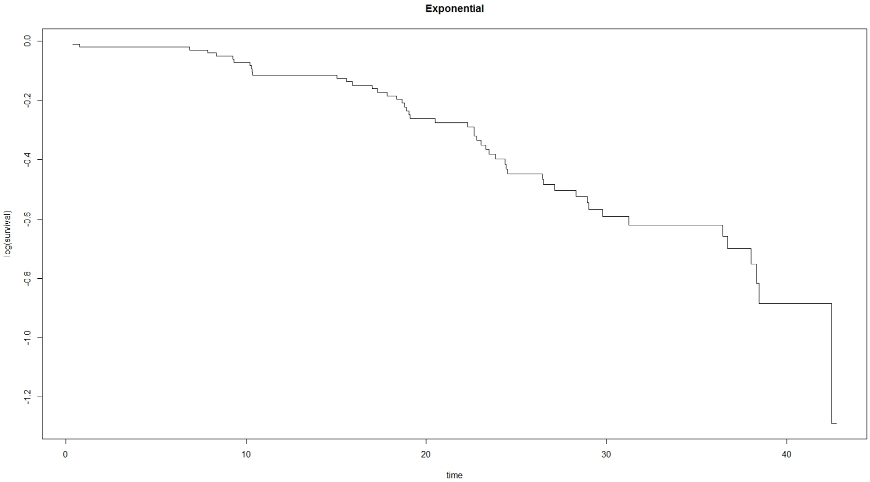
$$\begin{aligned}S(t) &= 1-\Phi\left(\frac{\log t-\mu}{\sigma}\right) \\ \Phi^{-1}(1-S(t)) &= \frac{\log(t)-\mu}{\sigma}.\end{aligned}$$

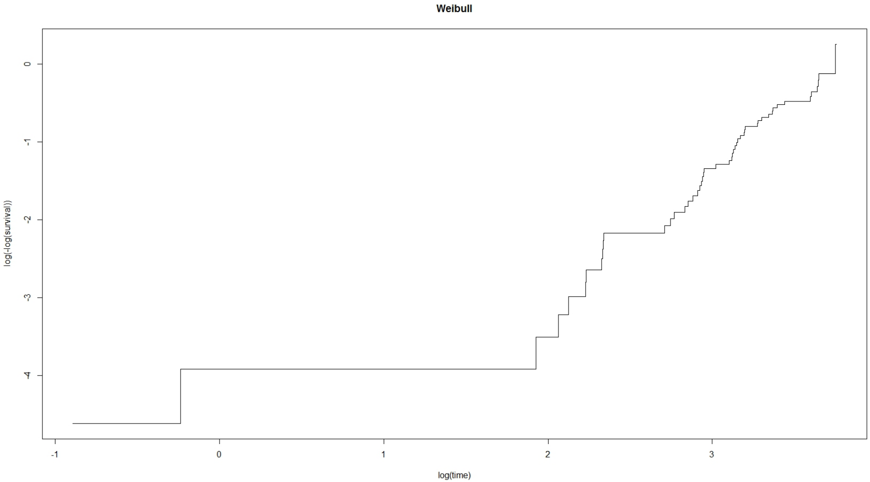
$\Rightarrow$ Plot $\Phi^{-1}(1-\hat{S}(t))$ vs $\log(t)$.

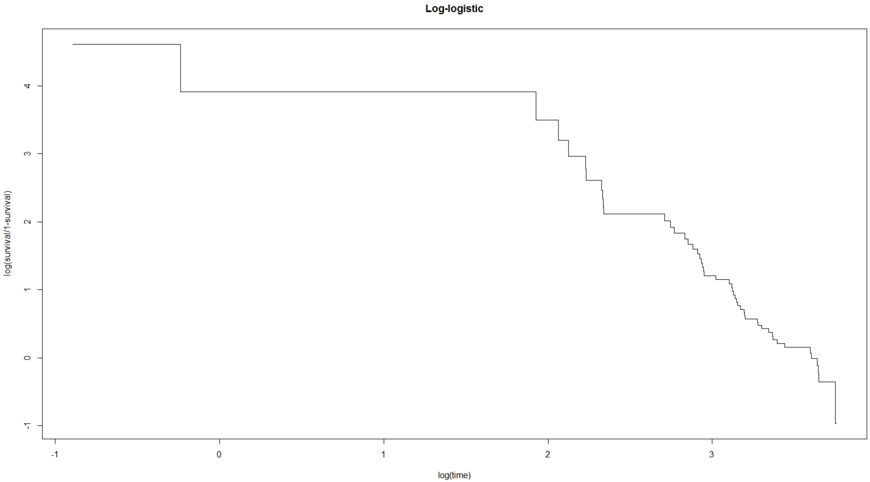
Example: NSCLC

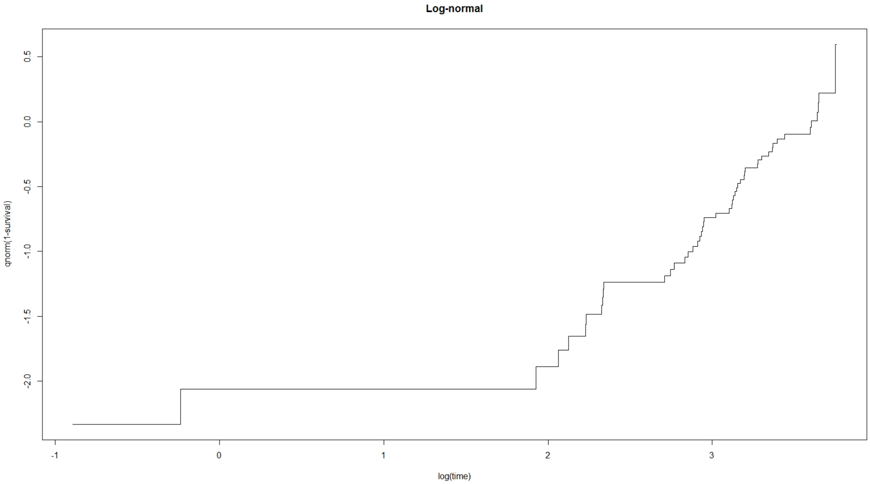
```
NSCLC<-read.table("C:/werk/Roel/Onderwijs/Theorie/G0B67AStatAnalReliaSurvData/Cursus/NSCLC.txt",
header=T,sep="\t")
fit<-survfit(Surv(survtime,survind)~1,data=NSCLC)
summary(fit)
plot(survfit(Surv(survtime,survind)~1,conf.type="none",data=NSCLC),xlab="Time",ylab="Survival")

#Exponential
#--------------
plot(fit$time,log(fit$surv),type="s",xlab="time",ylab="log(survival)",main="Exponential")

#Weibull
#--------------
plot(log(fit$time),log(-log(fit$surv)),type="s",xlab="log(time)",ylab="log(-log(survival))",
main="Weibull")

#log-logistic
#--------------
plot(log(fit$time),log(fit$surv/(1-fit$surv)),type="s",xlab="log(time)",
ylab="log(survival/1-survival)",main="Log-logistic")

#log-normal
#--------------
plot(log(fit$time),qnorm(1-fit$surv),type="s",xlab="log(time)",ylab="qnorm(1-survival)",
main="Log-normal")
```


Survival
Time
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40


Exponential
log(survival)
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
0
10
20
30
40
time


Weibull
log(-log(survival))
0
-1
-2
-3
-4
-1
0
1
2
3
log(time)


Log-logistic
log(survival/1-survival)
log(time)
-1
0
1
2
3
4
-1
0
1
2
3


Log-normal
qnorm(1-survival)
log(time)
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-1
0
1
2
3