STAT 432/532: Lecture 7
Parametric models

The log is the key
Probability distributions: transformation formula

If $X$ has a density $f$, and $X = h(Y)$, then, following the heuristics

$$f(x) \, dx = f(h(y)) \, h'(y) \, dy$$

we obtain that the density of $Y$ is

$$f(x) = f(h(y)) |h'(y)|$$

The absolute value is necessary to ensure that the density is positive; also, the function $h$ must be one-one.

When we need the distribution of $Z = h(X)$ instead, the easiest way is to invert $h$ and use the previous formula. A special formula for this case can be also derived using the formula for the derivative of the inverse function.
Examples

If \( X \) has a density \( f \), then the density of

\[
\begin{align*}
Y &= e^X \quad \text{is} \quad f(\log y) \\
Y &= \mu + X \quad \text{is} \quad f(y - \mu) \\
Y &= \sigma X \quad \text{is} \quad \frac{1}{\sigma} f \left( \frac{y}{\sigma} \right) \quad \text{(for } \sigma > 0) \\
Y &= \mu + \sigma X \quad \text{is} \quad \frac{1}{\sigma} f \left( \frac{y - \mu}{\sigma} \right) \quad \text{(for } \sigma > 0)
\end{align*}
\]
Transformations of survival functions

If $X$ has a survival function $S_X$, and $X = h(Y)$, where $h$ is an increasing function, then the survival function $S_Y$ of $Y$ is

$$S_Y(y) = \mathbb{P}[Y \geq y] = \mathbb{P}[h(Y) \geq h(y)] = \mathbb{P}[X \geq h(y)] = S_X(h(y))$$

Examples: the survival function of

$$Y = e^X \quad \text{is} \quad S_X(\log y)$$

$$Y = \mu + X \quad \text{is} \quad S_X(y - \mu)$$

$$Y = \sigma X \quad \text{is} \quad S_X\left(\frac{y}{\sigma}\right) \quad \text{(for } \sigma > 0)$$

$$Y = \mu + \sigma X \quad \text{is} \quad S_X\left(\frac{y - \mu}{\sigma}\right) \quad \text{(for } \sigma > 0)$$
Probability distributions: location/scale families

The family containing all distributions of the form \( \frac{1}{\sigma} f \left( \frac{y - \mu}{\sigma} \right) \)
typically with \( \mu \) arbitrary and \( \sigma > 0 \), is called a location/scale family of distributions. The distribution with \( \mu = 0 \) and \( \sigma = 1 \) is called the standard distribution of this family. Examples:

**normal:** \( \frac{1}{\sigma \sqrt{\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \)  
standard normal: \( \frac{1}{\sqrt{\pi}} e^{-\frac{x^2}{2}} \)

**extreme (minimum):** \( \frac{1}{\sigma} e^{\left( \frac{x-\mu}{\sigma} - e^{\frac{x-\mu}{\sigma}} \right)} \)  
standard extreme: \( e^{x} - e^{-x} \)

**logistic:** \( \frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma \left( 1 + e^{-\frac{x-\mu}{\sigma}} \right)^2} = \frac{1}{\sigma \left( e^{\frac{x-\mu}{2\sigma}} + e^{-\frac{x-\mu}{2\sigma}} \right)^2} \)

standard logistic: \( \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{(e^{2x} + e^{-\frac{1}{2}x})^2} \)
Caution about the names of extreme distributions

The distribution with a density
\[
\frac{1}{\sigma} e^{\left(\frac{x-\mu}{\sigma} - e^{\frac{x-\mu}{\sigma}}\right)}, \text{ with standard form } e^x - e^x, \text{ is } \text{extreme minimum}
\]

\[
\frac{1}{\sigma} e^{\left(-\frac{x-\mu}{\sigma} - e^{-\frac{x-\mu}{\sigma}}\right)}, \text{ with standard form } e^x - e^x, \text{ is } \text{extreme maximum}
\]

The alternative name *Gumbel distribution* is also used, and also for both: there is a Gumbel distribution for the maximum, and a Gumbel distribution for the minimum

In the context of survival analysis, customarily

\text{extreme } \equiv \text{extreme minimum}

the exponential of this distribution has a *Weibull distribution*

which specializes to an *exponential distribution* for \( \sigma = 1 \)

Beware: do not confuse extreme distribution with generalized extreme value distribution, which combines extreme (Gumbel), Fréchet, and Weibull into one family (Fisher and Tippett, 1928)
### Terminology and notation in R package `survival`

<table>
<thead>
<tr>
<th>T in R with log link</th>
<th>log T (location/scale family) in R with identity link</th>
</tr>
</thead>
<tbody>
<tr>
<td>exponential</td>
<td>(extreme)</td>
</tr>
<tr>
<td>special case of:</td>
<td></td>
</tr>
<tr>
<td>Weibull, gamma</td>
<td></td>
</tr>
<tr>
<td>exponential</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>extreme (minimum) (Gumbel)</td>
</tr>
<tr>
<td>weibull</td>
<td>extreme</td>
</tr>
<tr>
<td>Rayleigh (χ with 2 df)</td>
<td>(extreme)</td>
</tr>
<tr>
<td>special case of:</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td></td>
</tr>
<tr>
<td>rayleigh</td>
<td></td>
</tr>
<tr>
<td>lognormal</td>
<td>normal (gaussian)</td>
</tr>
<tr>
<td>lognormal, loggaussian</td>
<td></td>
</tr>
<tr>
<td>loglogistic</td>
<td>logistic</td>
</tr>
<tr>
<td>loglogistic</td>
<td>logistic</td>
</tr>
</tbody>
</table>
Common features of distributions fitted by `survreg`

Distributions of the survival time $T$ have (typically) two parameters:

*scale* $\lambda$ and *shape* $\alpha$

The distributions of log $T$ form a location/scale family, with

$$\log T = \mu + \sigma Z$$

*location* equal to $\mu = -\log(\lambda)$ and *scale* $\sigma = \frac{1}{\alpha}$

(note that scale in log $T$ is related to shape, not scale in $T$)

$Z$ is the standard distribution of the family, $\mu=0$, $\sigma=1$ ($\lambda=\alpha=1$)

The default for $Z$ is the extreme (minimum) distribution

which for $\sigma = 1$ means the exponential distribution for $T$

and for general $\sigma$ means the Weibull distribution for $T$

other options are given by the table
Exponential distribution

Historically the first one, especially in reliability (less in medicine)
Two possible parametrizations for its density \((\lambda, \beta > 0)\)

\[
\lambda e^{-\lambda t} \quad \frac{1}{\beta} e^{-\frac{t}{\beta}} \quad \text{(for } t \geq 0)\]

although the one with \(\lambda\) is more common
   especially in survival analysis (it is also in R)
Expected value: \(1/\lambda\)
Lack of memory \(\rightarrow\) constant hazard function (hazard rate): \(\lambda\)
Survival function (closed form): \(S(t) = e^{-\lambda t}\)
   Note: \(\log S(t)\) is linear
Quantile function (closed form): \(Q(p) = \frac{\log(1 - p)}{\lambda}\)
Thus, \(\lambda\) is the (hazard) rate and \(\beta = 1/\lambda\) the mean

*Good for classroom calculations*
Weibull distribution: parametrizations

Two (at least) possible parametrizations for its density \((\lambda, \beta, \alpha > 0)\)

\[
\alpha \lambda (\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha} \quad \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha}
\]

(for \(t \geq 0\))

conforming to the two parametrizations for the density of the exponential distribution - which is the special case for \(\alpha = 1\)

Thus the one with \(\lambda\) is again preferred in survival analysis

and thus also used by the package survival

but the R functions \{r,p,d,q\}weibull use the other one

Note: the index or shape parameter \(\alpha\) remains the same in all parametrizations

But the scale parameter \(\lambda\) can change to \(1/\beta\) - and also to \(\vartheta^{1/\alpha}\)

Parametrization ”used in medical statistics” and thus also in some survival analysis literature \(\vartheta, \alpha > 0\)

\[
\alpha \vartheta t^{\alpha-1} e^{-\vartheta t^\alpha}
\]

(for \(t \geq 0\))

Beware: letters do not indicate anything

some use \(\lambda\) or \(\beta\) for \(\vartheta\), some \(\rho\) for \(\lambda\), some \(\kappa\) for \(\alpha\)...
Weibull distribution: properties

Fisher and Tippett (1928), Weibull (1949)

Plays central rôle in parametric survival theory - among other things, all related functions can be specified in closed form

Survival function: \[ S(t) = e^{-(\lambda t)^\alpha} \]

Hazard function: \[ h(t) = \frac{\alpha \lambda (\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha}}{e^{-(\lambda t)^\alpha}} = \alpha \lambda (\lambda t)^{\alpha-1} \]

Cumulative hazard function: \[ H(t) = (\lambda t)^\alpha \]

Note: \[ \log H(t) = \alpha \log \lambda + \alpha \log t \]

that is, \[ \log H(t) \] linear in \[ \log t \] ("log-log" scale)

or also "complementary log-log scale"

as \[ \log H(t) = \log(-\log S(t)) \]

Quantile function (closed form): \[ Q(p) = \frac{(-\log(1 - p))^{\frac{1}{\alpha}}}{\lambda} \]
Weibull distribution: special cases

The special case for $\alpha = 1$: exponential

The special case for $\alpha = 2$ gives the Rayleigh distribution, customarily parametrized, however, with $\sigma = 1/(\lambda \sqrt{2})$

$$\frac{t}{\sigma^2} e^{-\frac{t^2}{2\sigma^2}} \quad \text{(for } x \geq 0\text{)}$$

If $W$ has the Weibull distribution,

then $(W/\lambda)\alpha$ has the exponential distribution with $\lambda = 1$

and $\log W$ has the extreme distribution

that is, the extreme minimum (Gumbel minimum) distribution
Weibull distribution: hazards

The hazard function of the Weibull is:

Flexibility in hazard functions for different values of parameters:

Weibull hazard function for $\lambda = 1$ and various $\alpha$

$$d\text{weibull}(x, 1/2, 1)/(1 - p\text{weibull}(x, 1/2, 1))$$

$\alpha = 0.5$
$\alpha = 1$
$\alpha = 1.5$
$\alpha = 3$
Weibull distribution: proportional hazards

Take the third parametrization of the Weibull: if we make $\vartheta = \lambda^\alpha$ an exponential function of a linear combination of covariates

$$\log \vartheta_i = \gamma_1 x_{i1} + \cdots + \gamma_p x_{ip}$$

the corresponding hazards differ only in the multiplicative term $\vartheta_i$

**model of proportional hazards**

Note that $\log \vartheta_i = \log \lambda_i^\alpha = \alpha \log \lambda$, so if we take the first parametrization instead and set

$$\log \lambda_i = \eta_1 x_{i1} + \cdots + \eta_p x_{ip}$$

then $\gamma_j = \alpha \eta_j$ - the above coefficients are merely rescaled
Weibull distribution: accelerated life model

If $T_0$ is some life duration, then $T = T_0/\lambda$ is that same life sped by $\lambda$

(example: if $T_0 = 6$ and $\lambda = 2$, then $T_0/\lambda = 3$)

We can model our survival distributions as being one $T_0$, only sped by possibly different $\lambda$:

**accelerated life model:**

*all* $T$ *have distributions such that $T\lambda$ is some fixed distribution*

(the distribution of $T_0$, say)

If $T_0$ has a distribution with a density $f$ and survival function $S$

then the distribution of $T_0/\lambda$ has

density $\lambda f(\lambda t)$ and survival function $S(\lambda t)$

and hazard function $\lambda h(\lambda t)$

This is the case whenever $\log T$ is a location/scale family (in particular, when $T$ has the Weibull distribution)

If we write $\log \lambda_i = \eta_1 x_{i1} + \cdots + \eta_p x_{ip}$

then $\log T = \log T_0 - \eta_1 x_{i1} - \cdots - \eta_p x_{ip}$
Accelerated life: location/scale families

If $\log T$ is a location/scale family
then we reparametrize: $\mu = -\log \lambda$
and have the accelerated life model
with $T_0$ having the distribution $\sigma \varepsilon$
where $\varepsilon$ has the standard distribution of the location scale family

So if $\log \lambda$ depends linearly on covariates, the model is

$$\log T_i = \mu_i + \sigma \varepsilon_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sigma \varepsilon_i$$

which is what R function `survreg` from the `survival` package fits

Beware: $\sigma$ is scale here, but in the Weibull parametrization, it pertains to $\alpha$, which is not called the scale, but shape parameter (or index)

Also,

$$T_i \lambda = T_i e^{\log \lambda_i} = \frac{T_i}{e^{-\log \lambda_i}} = \frac{T_i}{e^{\mu_i}} = \frac{T_i}{e^{\beta_1 x_{i1} + \cdots + \beta_p x_{ip}}}$$

15
Figure out the parametrization

Note:

```r
> y=rweibull(10000,5,25)
> summary(survreg(Surv(y)~1))
```

Call:
```
survreg(formula = Surv(y) ~ 1)
```
```
        Value Std. Error  z  p
(Intercept) 3.22    0.00211 1525 0
Log(scale)  -1.61   0.00781 -206 0
```

Scale= 0.2

Weibull distribution

What is $\sigma$? The meaning of other parameters?
Lognormal distribution

It is the distribution of $T$ such that $\log T$ has normal distribution.

**Parametrization:** $\mu = -\log \lambda$

- Variance usually remains $\sigma^2$; it may be, however, interesting also to reparametrize sometimes $\alpha = 1/\sigma$, to underline possible similarity to the Weibull.

- **Density:**
  \[
  \frac{\alpha}{t \sqrt{2\pi}} e^{-\frac{1}{2}(\alpha \log(t\rho))^2}
  \]

- **Survival function:** $1 - \Phi(\alpha \log(t\rho))$

**Disadvantage:** survival function not in closed algebraic form

- (and then neither hazard function)

**Modeling disadvantage:** sensitivity of small failure times

Also, while the exponential is not a special case, it is quite difficult to distinguish between it and lognormal with $\sigma \approx 0.8$. 
Log-logistic distribution

It is the distribution of $T$ such that $\log T$ has logistic distribution.

Parametrization: $\mu = -\log \rho$, $\sigma = 1/\alpha$

Density: $$\frac{\lambda \alpha (\lambda t)^{\alpha-1}}{(1 + (\lambda t)^\alpha)^2}$$

Survival function: $$\frac{1}{1 + (\lambda t)^\alpha}$$

Hazard function: $$\frac{\lambda \alpha (\lambda t)^{\alpha-1}}{1 + (\lambda t)^\alpha}$$

Logistic has slightly heavier tails than normal.
But log-logistic has both survival and hazard functions in closed form.
Other distributions used in survival analysis

Other distributions used in modeling survival times are

- Gamma (most frequently apart from the previous ones)
- Gompertz-Makeham
- Inverse Gaussian
- Pareto
- Generalized gamma

and possibly others
**Graphical check for model adequacy:** \textit{qqplot}

<table>
<thead>
<tr>
<th>distribution</th>
<th>log-distribution</th>
<th>transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>extreme</td>
<td>( \log(-\log(1-p)) )</td>
</tr>
<tr>
<td>lognormal</td>
<td>normal</td>
<td>( \Phi^{-1}(p) )</td>
</tr>
<tr>
<td>loglogistic</td>
<td>logistic</td>
<td>( \log\left(\frac{p}{1-p}\right) )</td>
</tr>
</tbody>
</table>

We plot ordered \( \log(y) \) against transformed uniformly spaced \( p \)
Examples to try with simulated data

```r
> y=rweibull(100,5,20)
> plot(sort(log(y)),log(-log(1-ppoints(length(y)))),pch=16)
> y=exp(rnorm(100,5,20))
> qqnorm(log(y))  # there is a special function for this
> y=exp(rlogis(100,5,20))
> plot(sort(log(y)),
+ log(ppoints(length(y))/(1-ppoints(length(y)))),pch=16)
```

Try a couple of times, varying parameters, to gain a feeling for it!
Application to the uncensored survival times

```r
> utime = function(obj) summary(obj)$time
> usurv = function(obj) summary(obj)$surv
> midf = function(x) (x + c(1,x[-length(x)]))/2

> aml.kmm = survfit(Surv(time,status)~x,subset=(x=="Maintained"), + data=aml,type="kaplan-meier")
> plot(log(qweibull(1-midf(usurv(aml.kmm)),1)),log(utime(aml.kmm)), + pch=16,cex=2)
> aml.wei = survreg(Surv(time,status)~x,subset=(x=="Maintained"), + data=aml,dist="weibull")
> abline(aml.wei$coef,aml.wei$scale)
> title("Maintained: Weibull qq-plot")
```
looks like linear enough to support the Weibull distribution
Try other distributions? Watch tiny details!

```r
> plot(qnorm(1-midf(usurv(aml.kmm))),log(utime(aml.kmm)),
+ pch=16,cex=2)
> aml.nor = survreg(Surv(time,status)~x,subset=(x=="Maintained"),
+ data=aml,dist="lognormal")
> abline(aml.nor$coef,aml.nor$scale)
> title("Maintained: normal qq-plot")

> plot(qlogis(1-midf(usurv(aml.kmm)),0,1),log(utime(aml.kmm)),
+ pch=16,cex=2)
> aml.log = survreg(Surv(time,status)~x,subset=(x=="Maintained"),
+ data=aml,dist="loglogistic")
> abline(aml.log$coef,aml.log$scale)
> title("Maintained: logistic qq-plot")
```
The results

Which one looks best? Hard to say...
When the favorite is picked: then what?

> summary(aml.log)

...  Value Std. Error  z   p
(Intercept) 3.515 0.306 11.48 1.65e-30
xNonmaintained 0.000 0.000 NaN NaN
Log(scale) -0.612 0.318 -1.93 5.39e-02
...

## An aesthetic flaw
...

> attach(aml)
> aml.log = survreg(Surv(time[x=="Maintained"], + status[x=="Maintained"]))~1,dist="loglogistic")
> summary(aml.log)

...  Value Std. Error  z   p
(Intercept) 3.515 0.306 11.48 1.65e-30
Log(scale) -0.612 0.318 -1.93 5.39e-02

Scale= 0.542
...
Predictions and their confidence intervals

> confi3 = function(m,s,ci=.95) c(m,m,m)+qnorm(1-(1-ci)/2)*c(-s,0,s)

On the original scale

> prd=predict(aml.log,type="response",se=TRUE)
> confi3(prd$fit[1],prd$se.fit[1])

1            1
13.44088 33.60127 53.76166

> prd=predict(aml.log,type="quantile",se=TRUE,p=0.5)
> confi3(prd$fit[1],prd$se.fit[1])

[1] 13.44088 33.60127 53.76166

Or transformed from a linear predictor:

> prd=predict(aml.log,type="link",se=TRUE)
> exp(confi3(prd$fit[1],prd$se.fit[1]))

1            1
18.44098 33.60127 61.22483

> prd=predict(aml.log,type="uquantile",p=0.5,se=TRUE)
> exp(confi3(prd$fit[1],prd$se.fit[1]))

[1] 18.44098 33.60127 61.22483

Note: both come the same, but this is a coincidence here (is it?)
Indeed

```r
> aml.exp = survreg(Surv(time[x=="Maintained"], + status[x=="Maintained"]~1,scale=1)
> prd=predict(aml.exp,type="response",se=TRUE)
> confi3(prd$fit[1],prd$se.fit[1])
   1 1 15.66328 60.42857 105.19386
> prd=predict(aml.exp,type="quantile",p=0.5,se=TRUE)
> confi3(prd$fit[1],prd$se.fit[1])
   1 1 10.85696 41.88589 72.91483
> prd=predict(aml.exp,type="link",se=TRUE)
> exp(confi3(prd$fit[1],prd$se.fit[1]))
   1 1 28.80835 60.42857 126.75534
> prd=predict(aml.exp,type="uquantile",p=0.5,se=TRUE)
> exp(confi3(prd$fit[1],prd$se.fit[1]))
   1 1 19.96843 41.88589 87.86011
```
or, yet another leukaemia dataset - with no censoring

```r
> library(MASS)
> leuk
```

<table>
<thead>
<tr>
<th>wbc</th>
<th>ag</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2300</td>
<td>present 65</td>
</tr>
<tr>
<td>2</td>
<td>750</td>
<td>present 156</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>100000</td>
<td>present 65</td>
</tr>
<tr>
<td>18</td>
<td>4400</td>
<td>absent 56</td>
</tr>
<tr>
<td>19</td>
<td>3000</td>
<td>absent 65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>100000</td>
<td>absent 43</td>
</tr>
</tbody>
</table>

We are interested in the difference in groups with different `ag` (present vs. absent)

But we have to account for the covariate `wbc`

or a certain function of it: seems like it is $\log(wbc)$
Thus, not the best way to analyze (why?)

> survdiff(Surv(time)~ag,data=leuk)

Call:
  survdiff(formula = Surv(time) ~ ag, data = leuk)

         N   Observed  Expected (O-E)^2/E (O-E)^2/V
ag=absent 16       16      9.3        4.83      8.45
ag=present 17       17     23.7        1.90      8.45

  Chisq= 8.4  on 1 degrees of freedom, p= 0.00365
Something better

```r
> leuk.full <- survreg(Surv(time)~ag*log(wbc),data=leuk)
> summary(leuk.full)
...

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.3423</td>
<td>1.680</td>
<td>2.585</td>
<td>0.00973</td>
</tr>
<tr>
<td>agpresent</td>
<td>4.1742</td>
<td>2.439</td>
<td>1.712</td>
<td>0.08696</td>
</tr>
<tr>
<td>log(wbc)</td>
<td>-0.1551</td>
<td>0.173</td>
<td>-0.898</td>
<td>0.36899</td>
</tr>
<tr>
<td>agpresent:log(wbc)</td>
<td>-0.3317</td>
<td>0.253</td>
<td>-1.310</td>
<td>0.19026</td>
</tr>
<tr>
<td>Log(scale)</td>
<td>0.0224</td>
<td>0.138</td>
<td>0.162</td>
<td>0.87150</td>
</tr>
</tbody>
</table>

Scale= 1.02

Weibull distribution
Loglik(model)= -145.6  Loglik(intercept only)= -153.6
Chisq= 15.88 on 3 degrees of freedom, p= 0.0012
...
> AIC(leuk.full)
[1] 301.2902

Seems like we can go for additivity (why?)
Additive model: Weibull

```r
> leuk.add <- survreg(Surv(time)~ag+log(wbc),data=leuk)
> summary(leuk.add)
...

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.8524</td>
<td>1.323</td>
<td>4.425</td>
<td>9.66e-06</td>
</tr>
<tr>
<td>agpresent</td>
<td>1.0206</td>
<td>0.378</td>
<td>2.699</td>
<td>6.95e-03</td>
</tr>
<tr>
<td>log(wbc)</td>
<td>-0.3103</td>
<td>0.131</td>
<td>-2.363</td>
<td>1.81e-02</td>
</tr>
<tr>
<td>Log(scale)</td>
<td>0.0399</td>
<td>0.139</td>
<td>0.287</td>
<td>7.74e-01</td>
</tr>
</tbody>
</table>

Scale= 1.04

Weibull distribution
Loglik(model)= -146.5  Loglik(intercept only)= -153.6
Chisq= 14.18 on 2 degrees of freedom, p= 0.00084
...
> AIC(leuk.add)
[1] 300.9976

Wald test: p-value for ag very low (→ coefficient significant)
Also: is the distribution exponential?
Additive model: exponential

> leuk.exp <- survreg(Surv(time)~ag+log(wbc),data=leuk, scale=1)
> summary(leuk.exp)

... Value Std. Error z p
(Intercept) 5.815 1.263 4.60 4.15e-06
agpresent  1.018 0.364 2.80 5.14e-03
log(wbc)   -0.304 0.124 -2.45 1.44e-02

Scale fixed at 1

Weibull distribution
Loglik(model)= -146.5  Loglik(intercept only)= -155.5
Chisq= 17.82 on 2 degrees of freedom, p= 0.00014
Number of Newton-Raphson Iterations: 5
n= 33

> AIC(leuk.exp)
[1] 299.081
Likelihood ratio test

```r
> leuk.none <- survreg(Surv(time)~log(wbc),data=leuk,scale=1)
> anova(leuk.none,leuk.exp)

<table>
<thead>
<tr>
<th>Terms</th>
<th>Resid. Df</th>
<th>-2*LL</th>
<th>Test Df</th>
<th>Deviance</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(wbc)</td>
<td>31</td>
<td>300.5704</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ag + log(wbc)</td>
<td>30</td>
<td>293.0810</td>
<td>= 1</td>
<td>7.489309</td>
<td>0.006206637</td>
</tr>
</tbody>
</table>

> AIC(leuk.none)
[1] 304.5704
```
Diagnostics

```r
> stimes <- leuk$time/predict(leuk.add,type="response")
> leuk.res = survfit(Surv(stimes)~1)
> plot(leuk.res,log=TRUE)
> plot(leuk.res,fun="cloglog")
> plot(log(qweibull(1-midf(usurv(leuk.res)),1)),
+ log(utime(leuk.res)),pch=16,cex=2)
> leuk.wei = survreg(Surv(stimes)~1,dist="weibull")
> abline(leuk.wei$coef,leuk.wei$scale)
```

Rationale: under the accelerated life model,

\[
\frac{T_i}{e^{\beta_1 x_{i1} + \cdots + \beta_p x_{ip}}} \quad \text{have the same distribution}
\]
approximate linearity supports the exponential error distribution
Hazard function on the log-log scale

this is to check for the general Weibull error distribution
another check for the general Weibull error distribution
Try another distribution?

```r
> leuk.log <- survreg(Surv(time) ~ ag + log(wbc), data = leuk, 
+                 + dist = "loglogistic")
> summary(leuk.log)
...

          Value Std. Error  z     p
(Intercept)  8.027     1.701 4.72 2.37e-06
agpresent   1.155     0.431 2.68 7.30e-03
log(wbc)   -0.609     0.176 -3.47 5.21e-04
Log(scale) -0.374     0.145 -2.58 9.74e-03

Scale= 0.688

Log logistic distribution
Loglik(model)= -146.6   Loglik(intercept only)= -155.4
Chisq= 17.58 on 2 degrees of freedom, p= 0.00015
Number of Newton-Raphson Iterations: 4
n= 33

> AIC(leuk.log)
[1] 301.1649
```
> leuk.glog <- survreg(Surv(time)~ag*log(wbc),data=leuk, +  + dist="loglogistic")
> summary(leuk.glog)
...

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>4.989</td>
<td>2.468</td>
<td>2.02</td>
<td>0.04321</td>
</tr>
<tr>
<td>agpresent</td>
<td>6.261</td>
<td>3.200</td>
<td>1.96</td>
<td>0.05037</td>
</tr>
<tr>
<td>log(wbc)</td>
<td>-0.292</td>
<td>0.256</td>
<td>-1.14</td>
<td>0.25495</td>
</tr>
<tr>
<td>agpresent*log(wbc)</td>
<td>-0.546</td>
<td>0.339</td>
<td>-1.61</td>
<td>0.10683</td>
</tr>
<tr>
<td>Log(scale)</td>
<td>-0.397</td>
<td>0.142</td>
<td>-2.80</td>
<td>0.00518</td>
</tr>
</tbody>
</table>

Scale= 0.673

Log logistic distribution
...

> AIC(leuk.glog)
[1] 300.6424
Let’s check...

```r
> stimes <- leuk$time/predict(leuk.glog, type="response")
> leuk.res = survfit(Surv(stimes) ~ 1)
> plot(qlogis(1-midf(usurv(leuk.res)), 0, 1), log(utime(leuk.res)),
      + pch=16, cex=2)
> leuk.rlog = survreg(Surv(stimes) ~ 1, dist="loglogistic")
> abline(leuk.rlog$coef, leuk.rlog$scale)
```
...maybe not

check for the general log-logistic error distribution
Can do also different scale parameters

```r
> leuk.gen <- survreg(Surv(time)~strata(ag)+log(wbc),data=leuk)
> summary(leuk.gen)
```

```
... Value Std. Error z p
(Intercept) 7.499 1.475 5.085 3.68e-07
log(wbc) -0.422 0.149 -2.834 4.59e-03
absent 0.152 0.221 0.688 4.92e-01
present 0.142 0.216 0.658 5.11e-01
```

Scale:
- absent
  - present
    - 1.16
    - 1.15

```
> AIC(leuk.gen)
```

Note that the fit for \( \log(\text{wbc}) \) is \textit{the same} for both \texttt{ag} present and absent; hence this is different than just fitting separate models for different levels of \texttt{ag}
Gehan data (with censoring)

```r
> gehan.wei <- survreg(Surv(time, cens) ~ treat, data = gehan)
> summary(gehan.wei)

Call:
survreg(formula = Surv(time, cens) ~ treat, data = gehan)

    Value Std. Error     z  p
(Intercept)  3.516     0.252 13.96 2.61e-44
  treatcontrol -1.267     0.311 -4.08 4.51e-05
Log(scale)   -0.312     0.147 -2.12 3.43e-02

Scale= 0.732

Weibull distribution
Loglik(model)= -106.6  Loglik(intercept only)= -116.4
Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
Number of Newton-Raphson Iterations: 5
n= 42
```
Check for the Weibull

> stimes <- gehan$time/predict(gehan.wei,type="response")
> gehan.res = survfit(Surv(stimes)~1)
> plot(log(qweibull(1-midf(usurv(gehan.res)),1)),
+ log(utime(gehan.res)),pch=16,cex=2)
> gehan.wei = survreg(Surv(stimes)~1,dist="weibull")
> abline(gehan.wei$coef,gehan.wei$scale)
Weibull qqplot of the Gehan data

log(utime(gehan.res))

log(qweibull(1 - midf(usurv(gehan.res)), 1))

not bad...
... maybe we shouldn’t try anything else...
... but why not

> summary(gehan.log)

Call:
`survreg(formula = Surv(time, cens) ~ treat, data = gehan, dist = "lognormal")`

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.1720</td>
<td>0.244</td>
<td>13.001</td>
<td>1.21e-38</td>
</tr>
<tr>
<td>treatcontrol</td>
<td>-1.3468</td>
<td>0.316</td>
<td>-4.255</td>
<td>2.09e-05</td>
</tr>
<tr>
<td>Log(scale)</td>
<td>-0.0792</td>
<td>0.132</td>
<td>-0.598</td>
<td>5.50e-01</td>
</tr>
</tbody>
</table>

Scale= 0.924

Log Normal distribution
Loglik(model)= -106.7  Loglik(intercept only)= -115.4
Chisq= 17.38 on 1 degrees of freedom, p= 3.1e-05
Number of Newton-Raphson Iterations: 4
n= 42
Finally: Life Testing of Motorettes

> plot(survfit(Surv(time,cens)~temp,data=motors),lwd=2,lty=4:1)
> legend(1000,0.4,lty=4:1,lwd=2,legend=levels(factor(motors$temp)),+
+ title="Temperature")
> title("survival of motorettes under various temperatures")
> motor.wei <- survreg(Surv(time,cens)~temp,data=motors)
> summary(motor.wei)

Call:
survreg(formula = Surv(time, cens) ~ temp, data = motors)

Value Std. Error    z     p
(Intercept) 16.3185 0.62296 26.2 3.03e-151
temp   -0.0453 0.00319 -14.2 6.74e-46
Log(scale) -1.0956 0.21480 -5.1 3.38e-07

Scale= 0.334

Weibull distribution
...
Kaplan-Meier estimates of group survival functions

survival of motorettes under various temperatures

Temperature

- 150
- 170
- 190
- 220
Predictions of survival at the temperature 130° C

> predict(motor.wei, newdata=list(temp=130),
+ type="quantile", p=c(0.5, 0.1))
[1] 29913.58 15934.59
> prd=predict(motor.wei, newdata=list(temp=130),
+ type="uquantile", p=0.5, se=TRUE)
> exp(confint(prd$fit[1], prd$se.fit[1]))
     2.5  97.5
19684.11 45459.12
> prd=predict(motor.wei, newdata=list(temp=130),
+ type="uquantile", p=0.1, se=TRUE)
> exp(confint(prd$fit[1], prd$se.fit[1]))
     2.5  97.5
10349.01 24534.83
Appendix: residual analysis posible, but not obvious

```R
> plot(resid(motor.wei, type="deviance"), pch=16, cex=2)
> plot(log(predict(motor.wei)),
+ resid(motor.wei, type="deviance"), pch=16, cex=2)
```