Survival Models in SAS

Part 3: PROC LIFEREG – Part 1

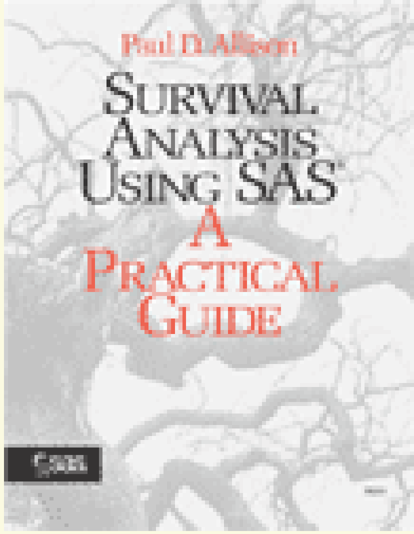
December 19, 2007

Charlie Hallahan

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG

These talks are based on the book "**Survival Analysis Using the SAS System: A Practical Guide**" (1995) by Paul Allison.

The book is part of the SAS Books-by-Users series and can be found at http://www.sas.com/apps/pubscat/bookdetails.jsp?catid=1&pc=55233



Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG

This series of talks will cover

Chapter 1: Introduction

Chapter 2: Basic Concepts of Survival Analysis

Chapter 3: Estimating and Comparing Survival Curves with PROC LIFETEST

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG

Chapter 5: Estimating Cox Regression Models with PROC PHREG

Chapter 6: Competing Risks

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG

Topics in Chapter 4:

Introduction
The Accelerated Failure Time Model
Alternative Distributions
Categorical Variables and the CLASS Statement
Maximum Likelihood Estimation
Hypothesis tests
Goodness-of-Fit Tests with the Likelihood-Ratio Statistic
Graphical Methods of Evaluating Model Fit
Left Censoring and Interval Censoring
The Piecewise Exponential Model
Generating Predictions and Hazard Functions

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: Introduction

LIFEREG estimates parametric survival models by *maximum likelihood*.

PHREG estimates semiparametric survival models using the *partial likelihood*.

Pros and cons of **LIFEREG** relative to **PHREG**:

Pros:

- **LIFEREG** accommodates left censoring and interval censoring, while **PHREG** only handles right censoring.
- **LIFEREG** can test hypotheses about the shape of the hazard function. The nonparametric estimates of the hazard function produced by **PHREG** can be difficult to interpret.
- **LIFEREG** produces more efficient estimates than **PHREG** if the shape of the survivor function is known.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: Introduction

- **LIFEREG** has a **CLASS** statement, while **PHREG** does not.

Cons:

- **LIFEREG** does not handle time-varying covariates, which is the forte of **PHREG**.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Accelerated Failure Time Model

LIFEREG fits a class of models known as *accelerated failure time* (**AFT**) models.

For these models, for any two subjects i and j, their survivor functions are related as follows: $S_j(t) = S_i(\varphi_{ij} t)$ for all t where φ_{ij} is a constant specific to the pair (i, j).

i.e., people differ by the rate at which they age.

Let T_i be a random variable denoting the event time for the i^{th} individual and let $x_{1i}, \ldots, x_{ki}$ be the values of k covariates for that individual.

LIFEREG estimates the model: $\log T_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \sigma\varepsilon_i$
where ε_i is a random disturbance term.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Accelerated Failure Time Model

Exponentiating both sides: $T_i = \exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \sigma\varepsilon_i)$
gives an alternative form of the model.

The variance of ε_i is fixed at 1 and the error variance is estimated via σ.

The assumed distribution of ε_i leads to different members of the **AFT** family.

For example, $\varepsilon_i \sim normal$ is the **log - normal model**.
(T_i~log-normal $\Leftrightarrow$ log(T_i)~normal)

Without any censored observations, the model can be estimated with **PROC REG** to produce BLUE estimates.

However, most survival data has many censored observations, and OLS doesn't handle censoring well.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Accelerated Failure Time Model

An example with the recidivism dataset (432 inmates followed one year after release).

WEEK contains the week of first arrest or censoring. ARREST = 1 if WEEK is uncensored and 0 if censored.

```
title "Estimation of the log-normal model with the recidivism
data";
proc lifereg data=survival.recid;
       model week*arrest(0)=fin age race wexp mar paro prio/
            dist=lnormal;
run;
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Accelerated Failure Time Model

```
Estimation of the log-normal model with the recidivism data

                The LIFEREG Procedure

                 Model Information

Data Set                              SURVIVAL.RECID
Dependent Variable                         Log(week)
Censoring Variable                            arrest
Censoring Value(s)                                 0
Number of Observations                           432
Noncensored Values                               114
Right Censored Values                            318
Left Censored Values                               0
Interval Censored Values                           0
Name of Distribution                       Lognormal
Log Likelihood                          -322.6945851


 Number of Observations Read         432
 Number of Observations Used         432


Algorithm converged.
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Accelerated Failure Time Model

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
fin	1	4.3657	0.0367
age	1	2.9806	0.0843
race	1	1.8824	0.1701
wexp	1	2.2466	0.1339
mar	1	2.4328	0.1188
paro	1	0.1092	0.7411
prio	1	5.8489	0.0156

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.2677	0.4617	3.3628	5.1726	85.44	<.0001
fin	1	0.3428	0.1641	0.0212	0.6645	4.37	**0.0367**
age	1	0.0272	0.0158	-0.0037	0.0581	2.98	**0.0843**
race	1	-0.3632	0.2647	-0.8819	0.1556	1.88	0.1701
wexp	1	0.2681	0.1789	-0.0825	0.6187	2.25	0.1339
mar	1	0.4604	0.2951	-0.1181	1.0388	2.43	0.1188
paro	1	0.0559	0.1691	-0.2756	0.3873	0.11	0.7411
prio	1	-0.0655	0.0271	-0.1186	-0.0124	5.85	**0.0156**
Scale	1	1.2946	0.0990	1.1145	1.5038		

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Accelerated Failure Time Model

Interpretation:

The *positive* estimate for **FIN** means that those who received financial aid had *longer* times to arrest than those who did not.

The *negative* estimate for **PRIO** means that those who had prior convictions had *shorter* times to arrest than those who did not.

While the parameters themselves are not directly interpretable, a simple transformation does have meaning.

For a dummy variable like **FIN**, e^{β} is the estimated ratio of the expected (mean) survival time for the two groups. So, $e^{.3428}$ = 1.41 means that, controlling for the other covariates, the expected time to arrest for those who received financial aid is 41% greater than those who did not recieve aid.

For a continuous variable like **PRIO**, 100(e^{β}-1) is the percent increase in the expected survival time for each one-unit increase in the variable. So, an increase of one prior conviction results in a 100($e^{-0.0655}$-1)=-6.34% decrease in the expected time to arrest.

Note: $\hat{\sigma}$=1.295 is listed as **SCALE**.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: Alternative Distributions

Distribution of ε	Distribution of T
extreme value (2 par.)	Weibull
extreme value (1 par.)	exponential
log-gamma	gamma
logistic	log-logistic
normal	log-normal

Note that all AFT models are named for the distribution of T rather than the distribution of ε or log T.

The choice of distribution affects the shape of the hazard function and, thus, the substantive interpretations.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Exponential model

Invoked by **DIST = EXPONENTIAL** in the **MODEL** statement.

Assuming ε has a *standard extreme value distribution* with the constraint σ=1.

Thus T has an *exponential distribution* with pdf $f(T) = \lambda e^{-\lambda T}$ and constant hazard function $h(T) = \lambda$.

The exponential model can be expressed in two forms:

(1) $\log h(T_i) = \beta_0^* + \beta_1^* x_{i1} + \ldots + \beta_k^* x_{ik}$ in terms of the **hazard function** of T.

(2) $\log T_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \sigma\varepsilon_i$ in terms of the **time-to-event variable** T.

For the exponential model, $\beta_j = -\beta_j^*$ for all j.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Exponential model

Intuition for the change in signs: $\beta_j = -\beta_j^*$ for all j.

$\beta_j > 0 \Rightarrow$ survival time increases as x_{ij} increases $\Rightarrow$ the hazard decreases as x_{ij} increases and vice versa for $\beta_j < 0$.

Note: **PHREG** estimates equation (1) and so reports β_j^* while **LIFEREG** estimates equation (2) and so reports β_j. This must be kept in mind when comparing output from the two PROCs.

Note that equation (1) has no error term since the uncertainty is built into the hazard function (much like a logit or probit model).

Equation (2) shows that while two individuals with the same covariate values will have the same hazard value $h(T_i)$, they will have different time-to-event values T_i (due to the presence of the error term ε_i).

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Exponential model

```
                    The LIFEREG Procedure

                      Model Information

        Data Set                             SURVIVAL.RECID
        Dependent Variable                        Log(week)
        Censoring Variable                           arrest
        Censoring Value(s)                                0
        Number of Observations                          432
        Noncensored Values                              114
        Right Censored Values                           318
        Left Censored Values                              0
        Interval Censored Values                          0
        Name of Distribution                    Exponential
        Log Likelihood                         -325.8259007


          Number of Observations Read          432
          Number of Observations Used          432


Algorithm converged.
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Exponential model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.0507	0.5860	2.9021	5.1993	47.78	<.0001
fin	**1**	**0.3663**	**0.1911**	**-0.0083**	**0.7408**	**3.67**	**0.0553**
age	**1**	**0.0556**	**0.0218**	**0.0128**	**0.0984**	**6.48**	**0.0109**
race	1	-0.3049	0.3079	-0.9085	0.2986	0.98	0.3220
wexp	1	0.1467	0.2117	-0.2682	0.5617	0.48	0.4882
mar	1	0.4270	0.3814	-0.3205	1.1745	1.25	0.2629
paro	1	0.0826	0.1956	-0.3007	0.4660	0.18	0.6726
prio	**1**	**-0.0857**	**0.0283**	**-0.1412**	**-0.0302**	**9.15**	**0.0025**
Scale	0	1.0000	0.0000	1.0000	1.0000		
Weibull Shape	0	1.0000	0.0000	1.0000	1.0000		

(compare highlighted values with results for the log-normal model on p. 8)

Lagrange Multiplier Statistics

Parameter	**Chi-Square**	**Pr > ChiSq**
Scale	**24.9302**	**<.0001**

Note that the scale parameter σ has been set to 1 and the test of the null hypothesis that σ = 1 is firmly rejected.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Weibull model

Invoked by **DIST = WEIBULL** in the **MODEL** statement.

Assuming ε has a *standard extreme value distribution* without the constraint σ=1.

Thus T has a *Weibull distribution* (conditional on the covariates).

The value of σ determines the shape of the hazard function.

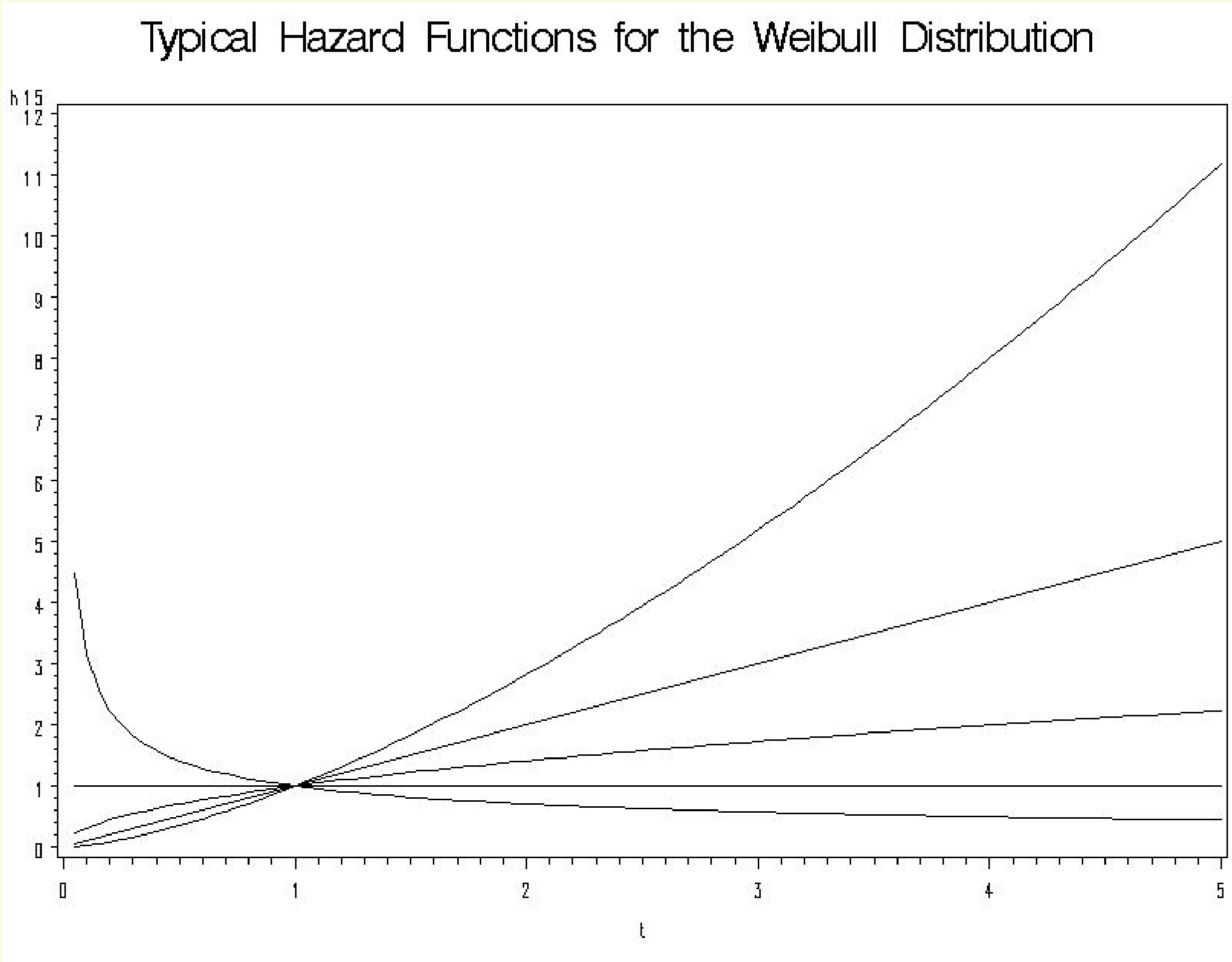
$\sigma > 1 \Rightarrow h(t)$ decreases with time

0.5< $\sigma < 1 \Rightarrow h(t)$ increases at a decreasing rate

0< $\sigma < 0.5 \Rightarrow h(t)$ increases at an increasing rate

$\sigma = 0.5 \Rightarrow h(t)$ is an increasing straight line through the origin.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Weibull model



α = 1.5, σ = 0.4

α = 1.0 , σ = 0.5

α = 0.5 , σ = 0.67

α = 0 , σ = 1.0

α = -0.5 , σ = 2.0

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Weibull model

As with the exponential model, we can express the model in two equivalent forms:

(1) $\log h(T_i) \;=\; \alpha \log t + \beta_0^* + \beta_1^* x_{i1} + \ldots + \beta_k^* x_{ik}$ in terms of the **hazard function** of T.

(2) $\log T_i \;=\; \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \sigma\varepsilon_i$ in terms of the **time-to-event variable** T.

For the Weibull model, $\beta_j^* = \dfrac{-\beta_j}{\sigma}$ for all j and $\alpha = 1/\sigma - 1$.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Weibull model

```
                The LIFEREG Procedure

                   Model Information

Data Set                              SURVIVAL.RECID
Dependent Variable                         Log(week)
Censoring Variable                            arrest
Censoring Value(s)                                 0
Number of Observations                           432
Noncensored Values                               114
Right Censored Values                            318
Left Censored Values                               0
Interval Censored Values                           0
Name of Distribution                         Weibull
Log Likelihood                          -319.3765238


 Number of Observations Read          432
 Number of Observations Used          432


Algorithm converged.
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Weibull model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.9901	0.4191	3.1687	4.8115	90.65	<.0001
fin	**1**	**0.2722**	**0.1380**	**0.0018**	**0.5426**	**3.89**	**0.0485**
age	**1**	**0.0407**	**0.0160**	**0.0093**	**0.0721**	**6.47**	**0.0110**
race	1	-0.2248	0.2202	-0.6563	0.2067	1.04	0.3072
wexp	1	0.1066	0.1515	-0.1905	0.4036	0.49	0.4820
mar	1	0.3113	0.2733	-0.2244	0.8469	1.30	0.2547
paro	1	0.0588	0.1396	-0.2149	0.3325	0.18	0.6735
prio	**1**	**-0.0658**	**0.0209**	**-0.1069**	**-0.0248**	**9.88**	**0.0017**
Scale	1	0.7124	0.0634	0.5983	0.8482		
Weibull Shape	1	1.4037	0.1250	1.1789	1.6713		

(the same variables are significant as the exponential model)

Recall the **LIFEREG** reports the log-survival-time estimates β_j.

We can calculate $\hat{\beta}_j^* = \dfrac{-\hat{\beta}_j}{\hat{\sigma}}$ where $\hat{\sigma} = 0.7124$ and $\hat{\alpha}=1/\hat{\sigma}$ - 1 = 0.4037.

Since α is the coefficient of log t in the log-hazard model, we can interpret $\hat{\alpha}$ as follows: a 1-percent increase in time since release produces a 0.40-percent increase in the hazard for arrest.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Normal model

Invoked by **DIST = LOGNORMAL** in the **MODEL** statement.

Assuming ε has a *normal distribution.*

Thus T has an *log - normal distribution* with pdf $f(T) = \dfrac{\exp\left(-\frac{1}{2}\left(\frac{\ln(T)-\mu}{\sigma}\right)^2\right)}{T(2\pi)^{1/2}\sigma}$,

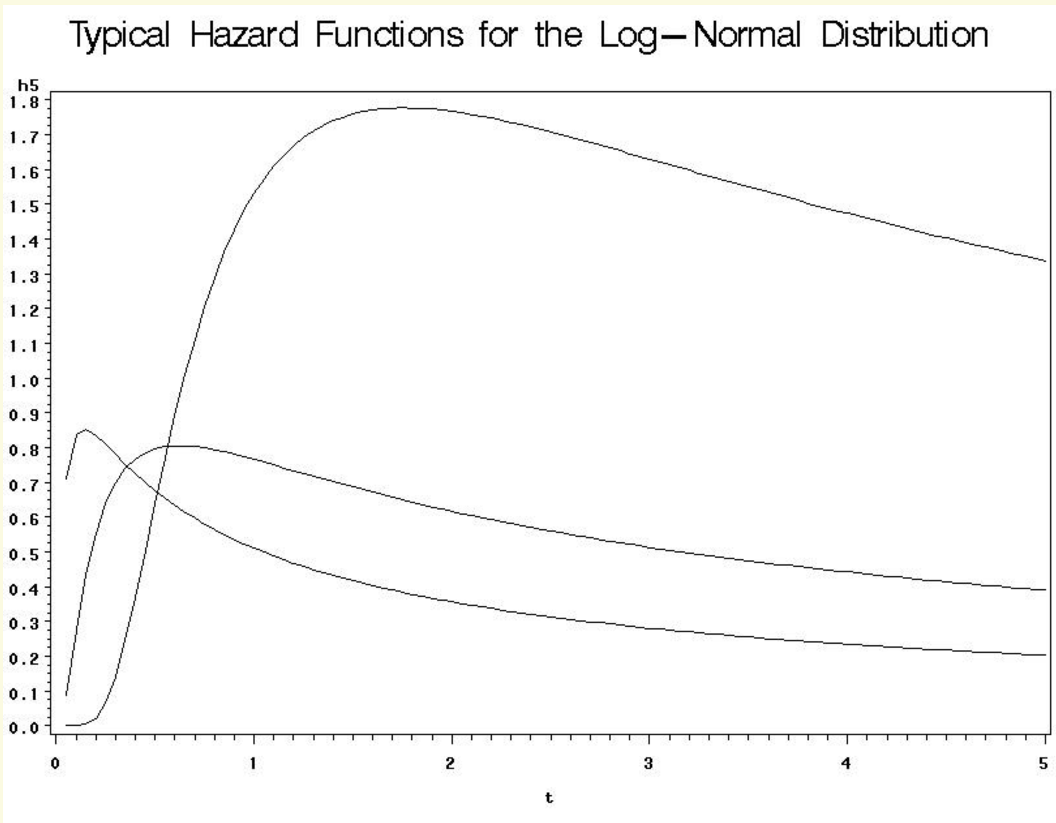
survivor function $S(T) = 1 - \Phi\left(\dfrac{\ln(T) - \mu}{\sigma}\right)$ and *hazard function* $h(T) = \dfrac{f(T)}{S(T)}$.

The log-normal model is **not** a proportional hazard model.

As a **regression model**: $\log h(T) = \log h_0\left(Te^{-\boldsymbol{\beta}\mathbf{x}}\right) - \boldsymbol{\beta}\mathbf{x}$ where $h_0(.)$ is the hazard function for an individual with $\mathbf{x} = \mathbf{0}$.

With a different function $h_0(.)$, this equation also applies to the **log - logistic** and **gamma** models.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Normal model



σ = 0.5

The U-shaped hazard is common with repeatable events – such as a residential move.

σ = 1.0

σ = 1.5

As σ increases, the log-normal hazard function closely resembles that of the Weibull and log-logistic, which have an infinite hazard when *t = 0*.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Logistic model

Invoked by **DIST = LLogistic** in the **MODEL** statement.

Assuming ε has a *logistic distribution* with pdf $f(\varepsilon) = \dfrac{e^{\varepsilon}}{\left(1+e^{\varepsilon}\right)^2}$, a symmetric distribution with zero mean .

It can be shown that log T will also have a logistic distribution.

Thus T has an *log - logistic distribution* and the log-logistic *hazard function* is

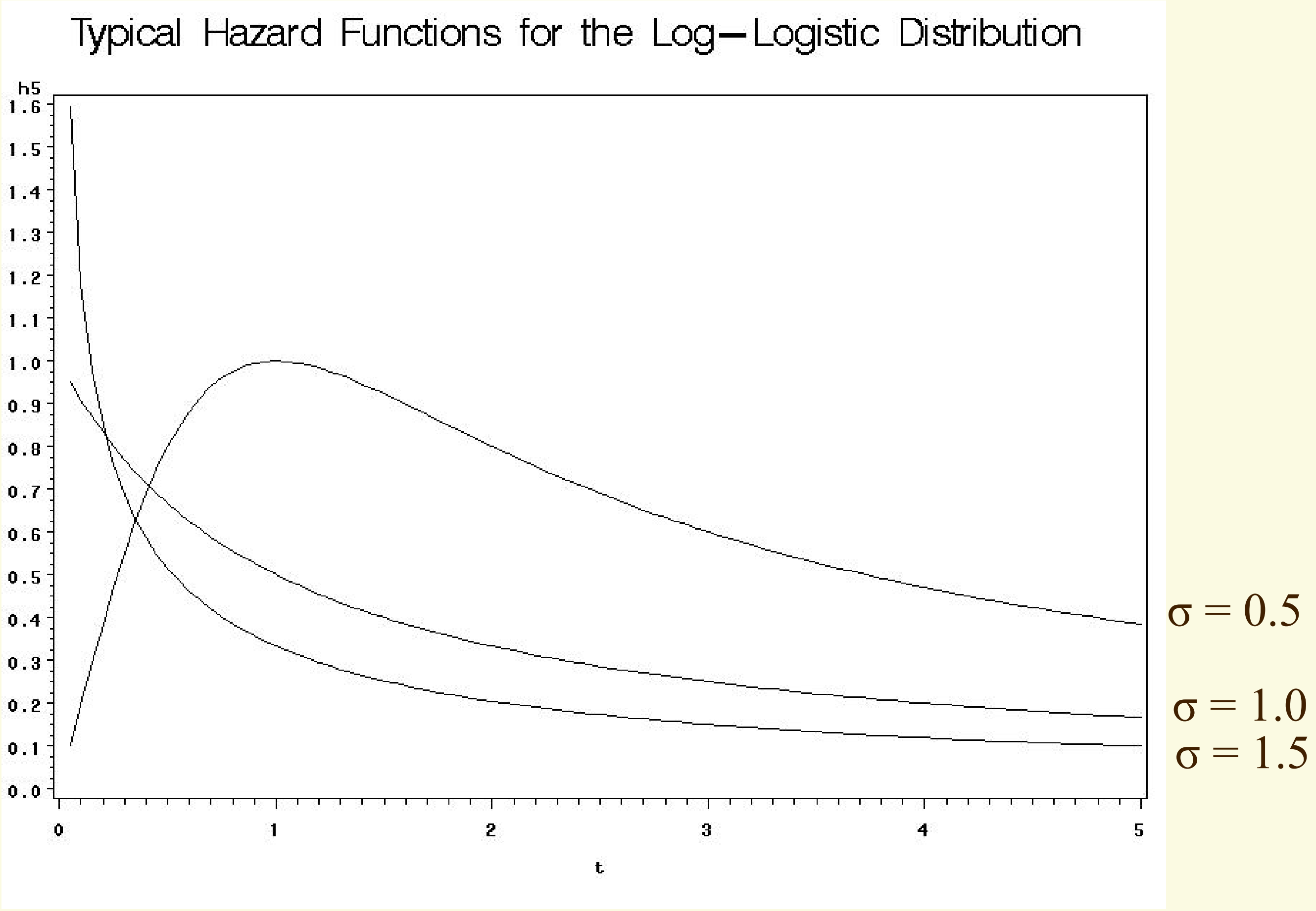
$$h(t) = \frac{\lambda\gamma\left(\lambda t\right)^{\gamma-1}}{1+\left(\lambda t\right)^{\gamma}} \quad \text{where } \gamma=1/\sigma \text{ and } \lambda=\exp\left(-\left(\beta_0+\beta_1 x_1+\ldots+\beta_k x_k\right)\right).$$

For $\sigma<1$, the log-logistic hazard is similar to the log-normal hazard in that it starts at 0 and rises to a peak.

For $\sigma > 1$, the log-logistic hazard is similar to the decreasing Weibull hazard in that it starts at infinity and declines toward 0.

For $\sigma = 1$, the log-logistic hazard has a value of λ at $t = 0$ and then declines toward 0 as t goes to infinity.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Logistic model



Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Logistic model

log - logistic survivor function: $S(t) \; = \; \dfrac{1}{1+(\lambda t)^{\gamma}}$ where, as before, γ=1/σ and

$\lambda = \exp\left(-\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)\right)$.

It follows that $\log\left[\dfrac{S(t)}{1-S(t)}\right] = \beta_0^* + \beta_1^* x_1 + \ldots + \beta_k^* x_k - \gamma \log t$ where $\beta_i^* = \beta_i / \sigma$ for $i = 1, \ldots, k$.

With t fixed at 52 weeks, the constant $\gamma \log t$ gets absorbed into the constant term and the above equation is just a logit model where $S(t)$ = the probability of not being arrested during the year.

Thus we could estimate the parameters β_i^* via the logit model. This would be less efficient than using a survival method since logit does not take into account the timing of the event (being arrested in this example).

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Logistic model

Since $\frac{S(t)}{1-S(t)}$, the odds of not being arrested up to time t, equals $\exp(\mathbf{x}\boldsymbol{\beta}^*)$, it follows that for two individuals i and j,

$$\frac{S_i(t)}{1-S_i(t)} = \phi_{ij} \frac{S_j(t)}{1-S_j(t)} \quad \text{where } \phi_{ij} = \exp((\mathbf{x}_j - \mathbf{x}_i)\boldsymbol{\beta}^*)$$

The log-logistic model is the only **AFT** model which is also a **proportional odds model**.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Logistic model

```
                    The LIFEREG Procedure

                      Model Information

        Data Set                          SURVIVAL.RECID
        Dependent Variable                     Log(week)
        Censoring Variable                        arrest
        Censoring Value(s)                             0
        Number of Observations                       432
        Noncensored Values                           114
        Right Censored Values                        318
        Left Censored Values                           0
        Interval Censored Values                       0
        Name of Distribution                   LLogistic
        Log Likelihood                      -319.3983709


         Number of Observations Read         432
         Number of Observations Used         432


Algorithm converged.
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Log-Logistic model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.9183	0.4274	3.0805	4.7561	84.03	<.0001
fin	1	0.2889	0.1456	0.0035	0.5742	**3.94**	**0.0472**
age	1	0.0364	0.0156	0.0058	0.0669	**5.45**	**0.0195**
race	1	-0.2791	0.2297	-0.7293	0.1710	1.48	0.2242
wexp	1	0.1784	0.1572	-0.1297	0.4865	1.29	0.2563
mar	1	0.3473	0.2697	-0.1812	0.8758	1.66	0.1978
paro	1	0.0508	0.1496	-0.2424	0.3440	0.12	0.7341
prio	1	-0.0692	0.0227	-0.1138	-0.0246	**9.25**	**0.0023**
Scale	**1**	**0.6471**	**0.0559**	**0.5463**	**0.7666**		

Note that $\hat{\sigma} = 0.65 \ < \ 1 \ \Rightarrow$ that the estimated hazard function follows an inverted U-shape (see p. 23).

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

LIFEREG supports the generalized (3-parameter) gamma model.

Good news: Includes as special cases the exponential, Weibull, standard (2-parameter) gamma, and the log-normal models (but not the log-logistic model). Thus, the hazard functions for this class of models can assume a large variety of shapes.

Bad news: Since the hazard function and survival function for the generalized gamma model involve both the gamma and incomplete gamma functions (which require the evaluation of integrals) , the computations are much more complicated.

One possible hazard function shape is the so-called *bathtub* or U-shape. This shape makes a lot of sense for human mortality in which there is an initial high hazard which then decreases, but again increases as time (age) increases.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

Invoked by **DIST = GAMMA** in the **MODEL** statement.

On the output, the estimate of σ is, as before, labeled **SCALE**.
The location parameter μ is represented by the linear form **xβ**.
The third parameter, δ, (see the **LIFEREG** documentation) is labeled **SHAPE** in the output.

$\delta = 0 \Rightarrow$ the log-normal distribution.
$\delta = 1 \Rightarrow$ the Weibull distribution.
$\delta = \sigma \Rightarrow$ the standard (2-parameter) gamma distribution.

In the output below, $\hat{\delta}$=0.994, indicating an approximate Weibull distribution.
With $\hat{\sigma}$=0.715, the standard gamma distribution is also plausible.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

```
                The LIFEREG Procedure

                   Model Information

Data Set                              SURVIVAL.RECID
Dependent Variable                         Log(week)
Censoring Variable                            arrest
Censoring Value(s)                                 0
Number of Observations                           432
Noncensored Values                               114
Right Censored Values                            318
Left Censored Values                               0
Interval Censored Values                           0
Name of Distribution                           Gamma
Log Likelihood                          -319.3764549


 Number of Observations Read         432
 Number of Observations Used         432


Algorithm converged.
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.9915	0.4349	3.1391	4.8439	84.23	<.0001
fin	1	0.2724	0.1401	-0.0022	0.5471	**3.78**	**0.0518**
age	1	0.0407	0.0165	0.0082	0.0731	**6.04**	**0.0140**
race	1	-0.2255	0.2280	-0.6723	0.2213	0.98	0.3226
wexp	1	0.1073	0.1659	-0.2179	0.4326	0.42	0.5177
mar	1	0.3118	0.2769	-0.2309	0.8545	1.27	0.2602
paro	1	0.0588	0.1398	-0.2152	0.3328	0.18	0.6741
prio	1	-0.0659	0.0213	-0.1076	-0.0241	**9.56**	**0.0020**
Scale	**1**	**0.7151**	**0.2396**	**0.3708**	**1.3790**		
Shape	**1**	**0.9943**	**0.4849**	**0.0439**	**1.9446**		

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

There is no direct way to fit the standard gamma distribution in **LIFEREG**.

The option **SHAPE=SCALE** *does not exist*.

It is possible to fix the values for SCALE and/or SHAPE.

The author did an experimental grid search to find the common value for **SHAPE** and **SCALE** to maximize the likelihood function.

To fit the standard gamma model with $\sigma = \delta = 0.811$, use the following code:

```
title "Estimation of the Standard Gamma model with the recidivism data";
proc lifereg data=survival.recid;
        model week*arrest(0)=fin age race wexp mar paro prio/ dist=gamma
                                 noshape1 shape1=0.811 noscale scale=0.811;
run;
```

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

```
                  The LIFEREG Procedure

                       Model Information

            Data Set                           SURVIVAL.RECID
            Dependent Variable                     Log(week)
            Censoring Variable                        arrest
            Censoring Value(s)                             0
            Number of Observations                       432
            Noncensored Values                           114
            Right Censored Values                        318
            Left Censored Values                           0
            Interval Censored Values                       0
            Name of Distribution                       Gamma
            Log Likelihood                      -319.4636775


             Number of Observations Read           432
             Number of Observations Used           432


   Algorithm converged.
```

Note that the maximized Log Likelihood value is slightly smaller than that for the generalized gamma on p. 30.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.0397	0.4269	3.2030	4.8764	89.55	<.0001
fin	1	0.2832	0.1411	0.0067	0.5597	**4.03**	**0.0447**
age	1	0.0390	0.0157	0.0083	0.0697	**6.21**	**0.0127**
race	1	-0.2498	0.2275	-0.6957	0.1960	1.21	0.2721
wexp	1	0.1348	0.1558	-0.1706	0.4403	0.75	0.3870
mar	1	0.3323	0.2756	-0.2078	0.8724	1.45	0.2279
paro	1	0.0580	0.1450	-0.2262	0.3422	0.16	0.6893
prio	1	-0.0673	0.0215	-0.1094	-0.0252	**9.83**	**0.0017**
Scale	**0**	**0.8110**	**0.0000**	**0.8110**	**0.8110**		
Shape	**0**	**0.8110**	**0.0000**	**0.8110**	**0.8110**		

The parameter estimates for the standard gamma are very close to those of the generalized gamma on p. 31 (not surprising since the log likelihood values are also very close).

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model

The *pdf* for the survival time T for the standard gamma distribution is:

$$f(t) = \frac{\lambda(\lambda t)^{K-1} e^{-\lambda t}}{\Gamma(K)} \quad \text{where } K = 1/\delta^2 \ (\delta \text{ is the shape parameter})$$

and $\lambda = \exp\left(-\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)\right)$. Note that for an integer K, $\Gamma(K) = (K-1)!$.

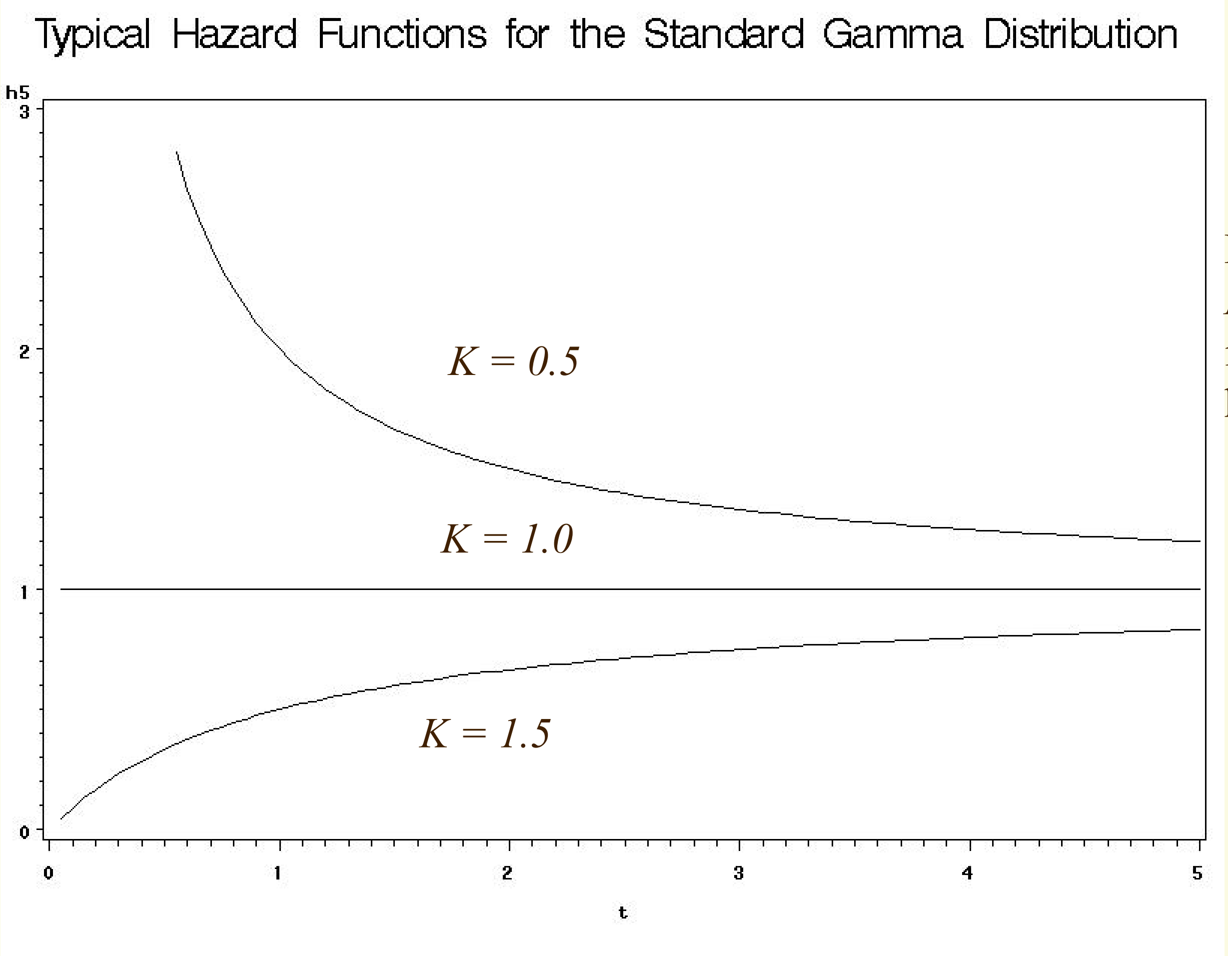
$K > 1 \Rightarrow h(0) = 0$ and increases thereafter and approaches λ as an upper limit.

$0 < K < 1 \Rightarrow h(0) = \infty$ and decreases thereafter and approaches λ as an lower limit.

$K = 1 \Rightarrow h(t)$ is a constant.

This asymptotic behaviour is different from that of the Weibull model.

Chapter 4: Estimating Parametric Regression Models with PROC LIFEREG: The Gamma model



For our example, *K=1/(0.811)^2 = 1.52* implying an increasing hazard.