Syllabus

1. Introduction
   - Survival data
   - Censoring mechanism
   - Application in medical field
2. Concepts and definitions
   - Survival function
   - Hazard function
3. Non-parametric approach
   - Life table
   - Kaplan-Meier survival estimate
   - Hazard function
   - Median and percentile survival time
4. Hypothesis testing
   - Overview – hypothesis, test statistics, p-values
   - Log-rank
   - Wilcoxon
   - Gehan test
5. Study design and sample size estimation
   - Overview
   - Survival sample size estimation
   - Accrual time and Study duration
6. Semiparametric model – proportional hazard model
   - Partial likelihood
   - Inference
   - Time varying covariates
   - Stratification
7. Model checking in the PH model
   - Model checking
   - Residuals
8. **Parametric model**
   - **Parametric proportional hazard model**
   - **Accelerate failure-time model**
9. Other topics
   - Competing risk
   - Recurrent events
   - Non-proportional hazard ratio
   - Interval censoring

Parametric Models

- Recall, we discussed the parametric distributions of one sample survival data (T, Δ)
  - Exponential
  - Piecewise exponential
  - Weibull distribution
  - Log-logistic
  - Log-normal
  - Generalized gamma distribution
  - Etc.

Likelihood Function

- Recall, the full likelihood function for right-censored data

$$L(\beta) = \prod_{i=1}^{n} h(T_i)^{\Delta_i} S(T_i)$$

$$= \prod_{i=1}^{n} f(T_i)^{\Delta_i} S(T_i)^{1-\Delta_i}$$

Fitting Exponential Distribution

- Exponential distribution is our old friend – one parameter distribution
    - Constant hazard
- The likelihood function

$$L(\beta) = \prod_{i=1}^{n} h(T_i)^{\,\Delta_i} S(T_i)$$

$$= \prod_{i=1}^{n} (\lambda)^{\Delta_i} e^{-\lambda T_i}$$

- Taking log of $L(\beta)$

$$\mathcal{L} = \log L(\beta) = \sum_{i=1}^{n} (\Delta_i \log \lambda - \lambda T_i)$$

$$= \log \lambda \sum_{i=1}^{n} \Delta_i - \lambda \sum_{i=1}^{n} T_i$$

Fitting Exponential Distribution

- Taking the first derivative of $\mathcal{L}$

$$\frac{\partial}{\partial\lambda}\mathcal{L} = \frac{d}{\hat{\lambda}} - \sum_{i=1}^{n} T_i = 0$$

$$\hat{\lambda} = \frac{d}{T}$$

where $d = \sum_{i=1}^{n} \Delta_i$, total number of events and

$T = \sum_{i=1}^{n} T_i$, total person-time

Fitting Exponential Distribution

- Taking the second derivative of $\mathcal{L}$

$$\frac{\partial^2}{\partial\lambda^2}\mathcal{L} = -\frac{d}{\lambda^2}$$

$$var(\hat{\lambda}) = \left\{-E\left(\frac{\partial^2}{\partial\lambda^2}\mathcal{L}\right)\right\}^{-1}$$

$$= \frac{d}{T^2}$$

Fitting Weibull Distribution

- The hazard function

$$h(t) = \frac{f(t)}{S(t)} = \lambda \alpha t^{\alpha - 1}$$

- The survival function

$$S(t) = e^{-\lambda t^{\alpha}}$$

- The density function

$$f(t) = \lambda \alpha t^{\alpha - 1} e^{-\lambda t^{\alpha}}$$

Fitting Weibull Distribution

- The likelihood function

$$L(\beta) = \prod_{i=1}^{n} h(T_i)^{\Delta_i} S(T_i)$$

$$= \prod_{i=1}^{n} \left(\lambda\alpha T_i{}^{\alpha-1}\right)^{\Delta_i} e^{-\lambda T_i{}^{\alpha}}$$

- Taking the first derivative of the log $L(\beta)$, the MLEs

Fitting Weibull Distribution

- Taking the first derivative of the log $L(\beta)$, the MLEs satify

$$\frac{d}{\hat{\lambda}} - \sum_{i=1}^{n} T_i^{\hat{\alpha}} = 0$$

$$\frac{d}{\hat{\lambda}} + \sum_{i=1}^{n} \Delta_i \log T_i - \hat{\lambda} \sum_{i=1}^{n} T_i^{\hat{\alpha}} \log T_i = 0$$

- No close form to solve $\hat{\lambda}$ and $\hat{\alpha}$
- Newton-Raphson numerical approach to find $\hat{\lambda}$ and $\hat{\alpha}$

One Sample Model Fitting

- Model checking usually focus on if the hazard function is constant
- Often there is not sufficient data to differentiate among many two-parameter distributions

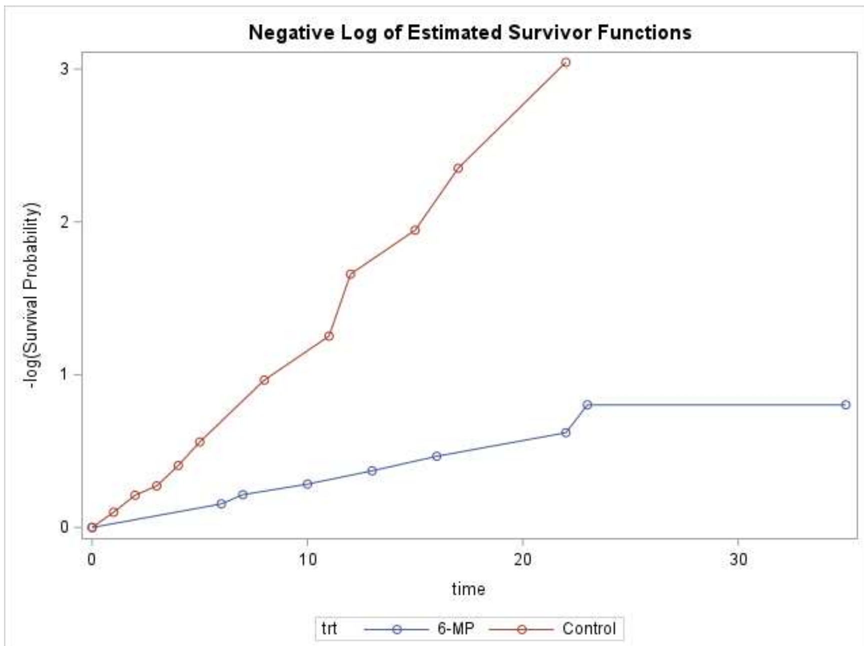
- Checking survival functions
  - Estimate survival function using the K-M estimator
  - Based on the distribution assumptions
    - Obtain MLEs for the parameters
    - Obtain the survival functions
  - Visually inspect the similarity of the survival functions

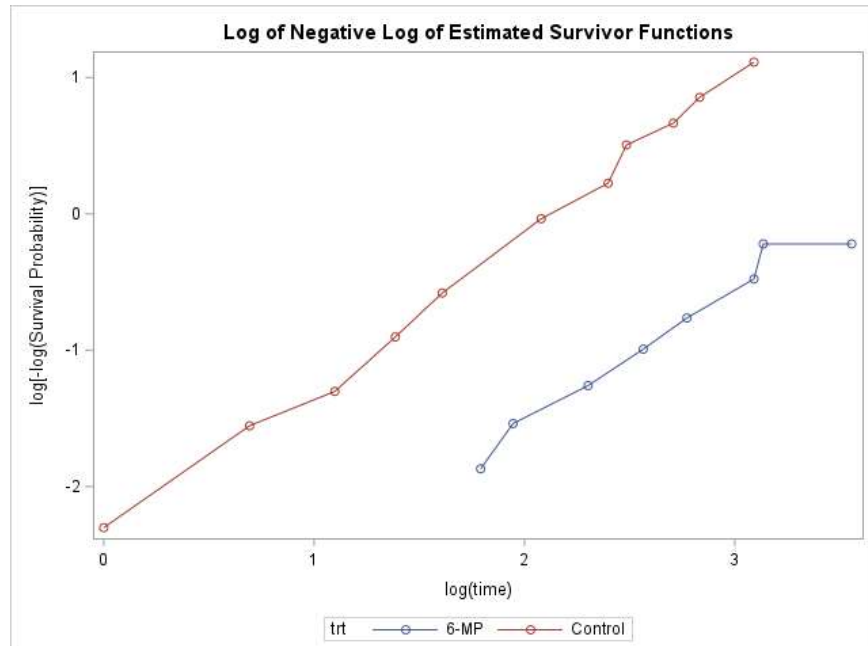
Model Fitting

- Taking $\log\left\{\hat{S}\ (t)\right\}$ and plot against t
  - Estimating cumulative hazard function
  - Exponential $\log\left\{\hat{S}\ (t)\right\} = \lambda t$
    - A straight line
  - Weibull $\log\left\{\hat{S}\ (t)\right\} = \lambda t^{\alpha}$
    - Non-linear

- Taking $\log\{-\log S(t)\}$ and plot against $\log t$
  - Exponential $\log\{-\log S(t)\} = \log \lambda t = \log \lambda + \log t$
    - The slop of a straight line is 1
  - Weibull $\log\{-\log S(t)\} = \log \lambda t^{\alpha} = \log \lambda + \alpha \log t$
    - The slop of a straight line is not 1

Example – Leukemia Data

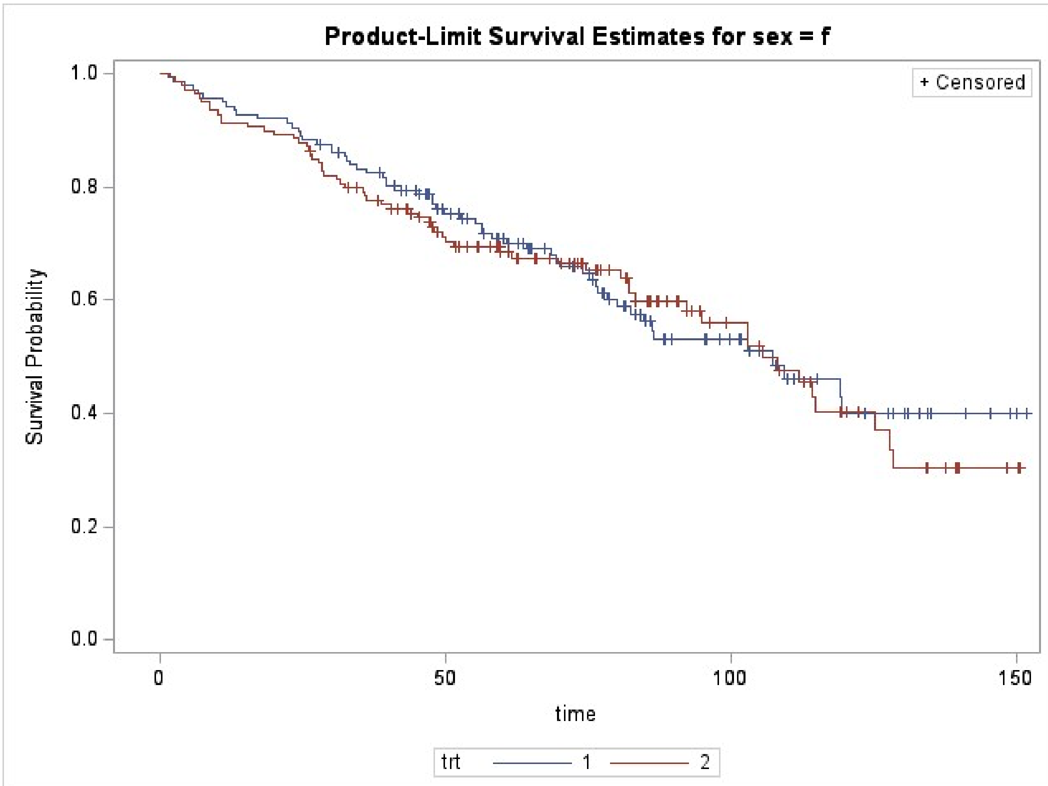




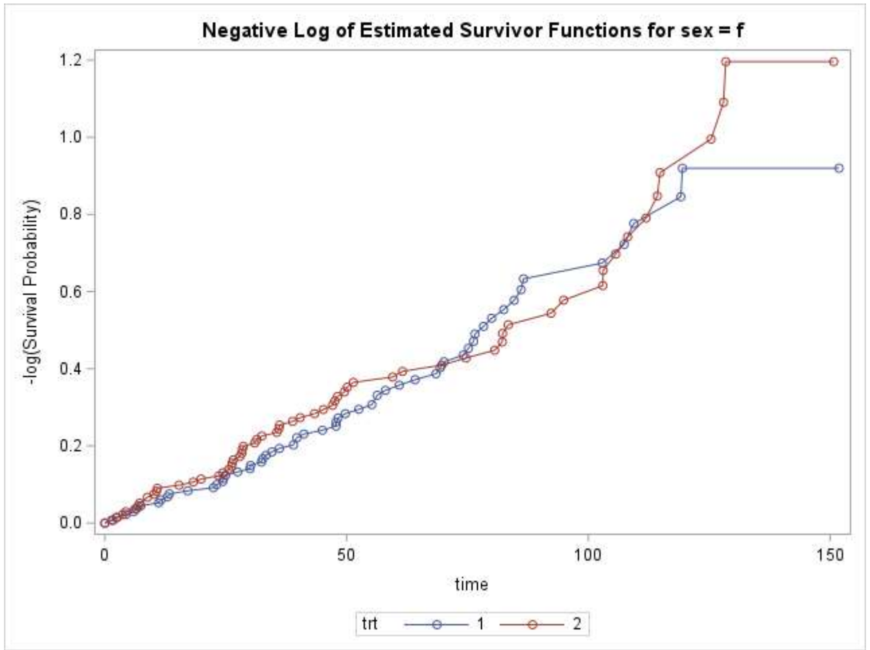
Example – PBC Data

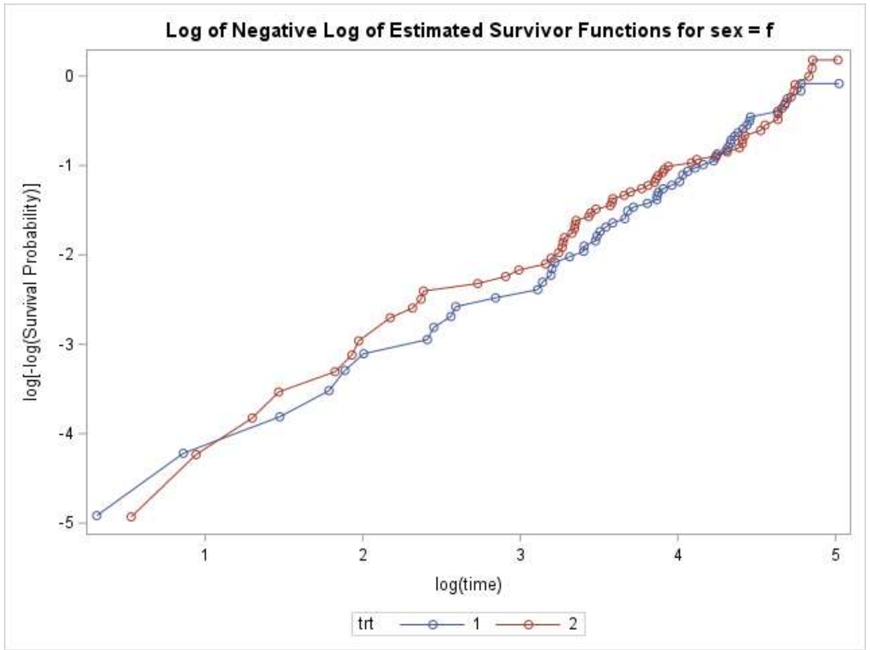
```
ods graphics on;
proc lifetest data=example method=KM  plots=(survival logsurv h loglogs)
outsurv=survival;
  time time*status(0);
  strata sex/group=trt;
  run;
ods graphics off;
```

Example – PBC Data

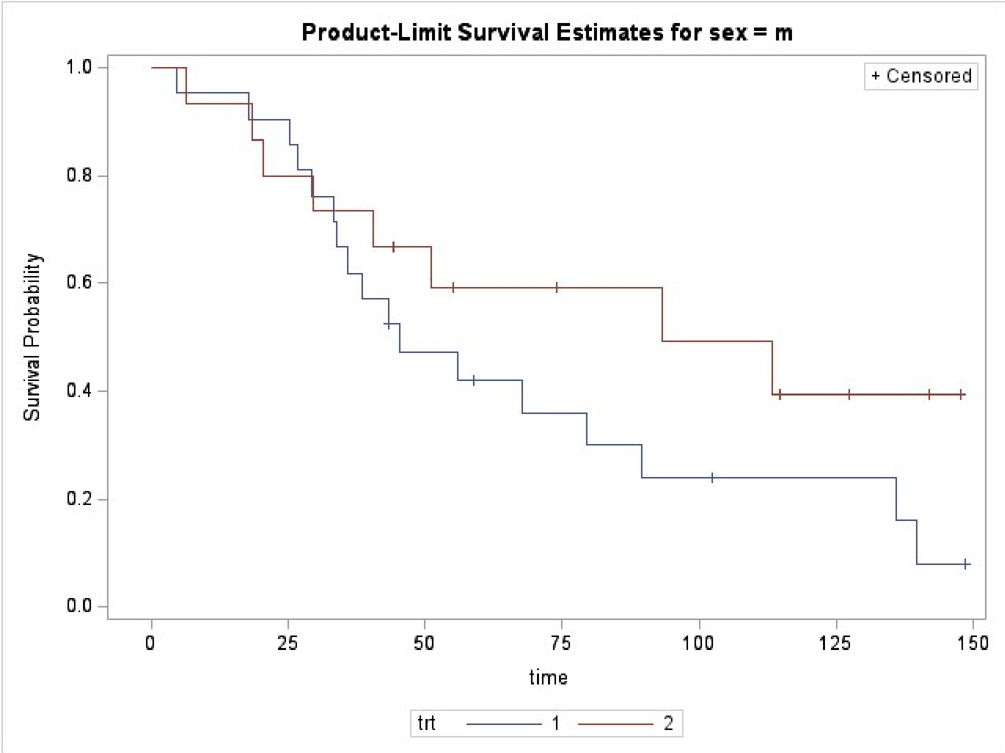


Example – PBC Data

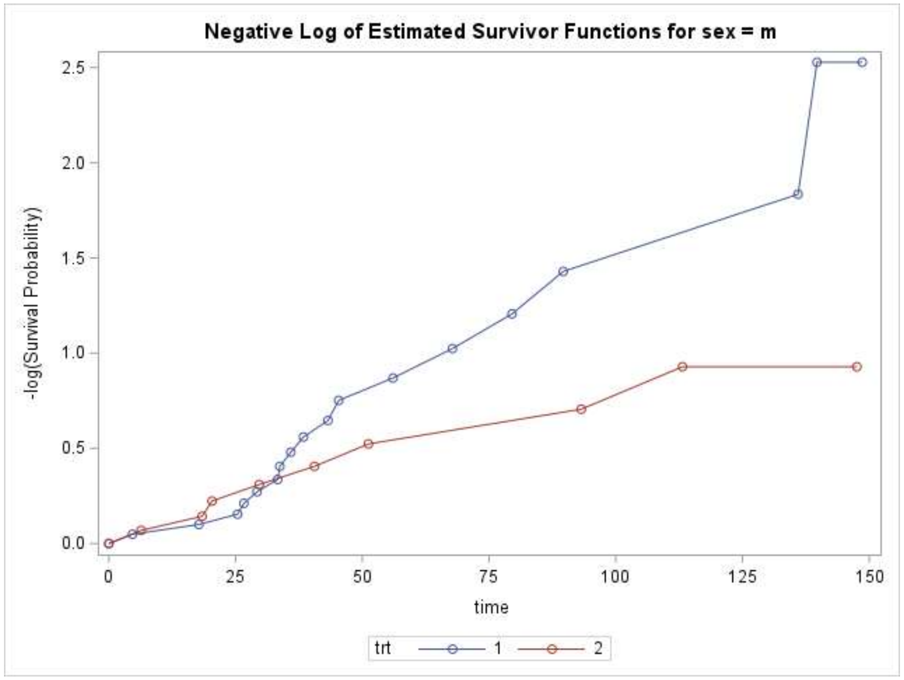


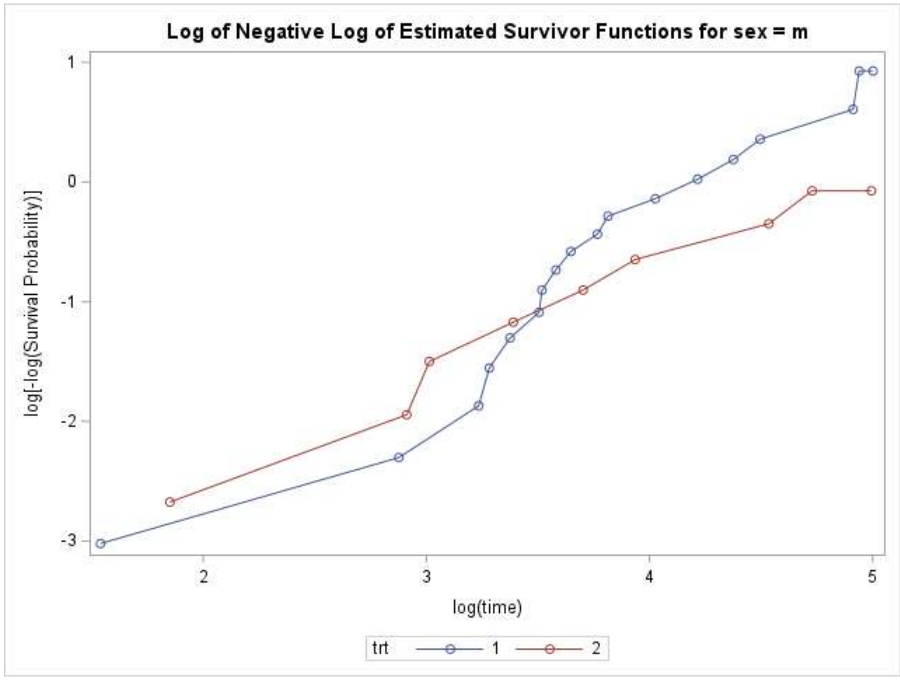


Example – PBC Data



Example – PBC Data





Parametric Regression Models

- Survival data with covariates (T, Δ, Z)
  - Parametric proportional hazard models
  - Accelerated failure-time models
    - Linear log-time models
  - Proportional odds model

- With the parametric distributions of survival data
  - Exponential
  - Piecewise exponential
  - Weibull distribution
  - Log-logistic
  - Log-normal
  - Generalized gamma distribution
  - Etc.

Parametric Proportional Hazard Models

- Recall the definition of PH model
- The Cox-regression model for survival data (T, Δ, Z)

$$h(t|Z = z) = h_0(t)e^{\beta' z}$$

where $h_0(t)$ is a baseline hazard function with known distribution

Parametric Proportional Hazard Models

- Assume baseline hazard function is Weibull

$$h_0(t) = \lambda \alpha t^{\alpha-1}$$

$$h(t|Z=z) = h_0(t)e^{\beta' z}$$
$$= \lambda \alpha t^{\alpha-1} e^{\beta' z} \quad = \lambda' \alpha t^{\alpha-1} \; (\lambda' = \lambda e^{\beta' Z})$$

- The PH model modifies the scale parameter in the Weibull distribution for subjects with covariates Z
  - The scale parameter is $\lambda e^{\beta' z}$
  - The shape parameter is still α

- Form the full likelihood to obtain the MLEs

$$\hat{\lambda}, \hat{\alpha}, \hat{\beta}$$

Parametric Proportional Hazard Models

$S(t) = e^{-H(t)}$

- The survival function is

$$S(t) = e^{-\lambda t^{\alpha} e^{\beta' z}}$$

- The estimated survival function is

$$\hat{S}(t) = e^{-\hat{\lambda} t^{\hat{\alpha}} e^{\hat{\beta}' z}}$$

- What is the advantage of parametric model?
  - Given correct model assumptions , only two parameters.
  - In comparison to partial likelihood

Breslow { assumption : piecewise , the change at events. ↓ many parameters to estimate.

Accelerated Failure-Time (AFT) Models

- AFT models for two groups

$$S_1(t) = S_0(t/\phi)$$

where ϕ indicates time scale change in Group 1 and ϕ^{-1} is called the acceleration factor

Only one covariate.

- It can be shown

$$f_1(t) = \phi^{-1} f_0(t/\phi)$$

$$h_1(t) = \phi^{-1} h_0(t/\phi)$$

AFT Models

- For survival data (T, Δ, Z)
  - Let $S_0(t)$ be the baseline hazard function
  - Let $S(t|Z)$ be the survival function with covariate Z
- AFT models can be re-parameterized

$$S(t|Z) = S_0\left(te^{\beta' Z}\right)$$

where $e^{\beta' Z}$ is called an acceleration factor as a function of Z same to ϕ^{-1}

- The hazard function

$$h(t|Z) = e^{\beta' Z} h_0(te^{\beta' Z})$$

The AFT Model in Linear Log-time Form

- The AFT

$$\log T = \mu + \gamma' Z + \sigma\epsilon$$

where $\gamma' = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ p dimensional coefficients

ϵ is the error term

$$S(t|Z) = \mathrm{P}(\mathrm{T} \geq t | Z)$$

Plug in $T = e^{\mu+\gamma' Z+\sigma\epsilon} = e^{\mu+\sigma\epsilon} e^{\gamma' Z}$, we have

$$S(t|Z) = \mathrm{P}\left(e^{\mu+\sigma\epsilon} e^{\gamma' Z} \geq t\right) = \mathrm{P}\left(e^{\mu+\sigma\epsilon} \geq t/e^{\gamma' Z}\right)$$

- The baseline survival function $S_0(t) = S(t|Z=0) = \mathrm{P}(e^{\mu+\sigma\epsilon} \geq t)$

$$S(t|Z) = S_0(te^{-\gamma' Z})$$

$-\gamma' = \beta'$

AFT Models

- The median survival time t_m with covariates Z

$$S(t_m|Z) = S_0\left(t_m e^{\beta' Z}\right) = 0.5$$

Let t_{0m} be the median time for the baseline survival function

$$t_m = t_{0m} e^{-\beta' Z} = \frac{t_{0m}}{e^{\beta' Z}}$$

- The hazard function of the AFT models

$$h(t|Z) = e^{\beta' Z} h_0(t e^{\beta' Z})$$

Weibull Distribution – AFT Models

- The AFT model

$$h(t|Z) = e^{\beta' Z} h_0(te^{\beta' Z})$$

$$h_0(t) = \lambda \alpha t^{\alpha-1}$$

$$h(t|Z) = e^{\beta' Z} \lambda \alpha (te^{\beta' Z})^{\alpha-1} \quad = e^{\beta' Z} \lambda \alpha t^{\alpha-1} e^{\beta' Z(\alpha-1)}$$

$$= e^{\alpha\beta' Z} \lambda \alpha t^{\alpha-1} \quad = e^{\alpha\beta' Z} \lambda \alpha t^{\alpha-1}$$

$$= h_0(t) e^{\alpha\beta' Z}$$

- The AFT model for Weibull distribution is a PH model

only true for Weibull and exponential

Example – Leukemia Data

Fit Statistics	
-2 Log Likelihood	94.128
AIC (smaller is better)	100.128
AICC (smaller is better)	100.760
BIC (smaller is better)	105.341

Fit Statistics (Unlogged Response)	
-2 Log Likelihood	213.159
Weibull AIC (smaller is better)	219.159
Weibull AICC (smaller is better)	219.791
Weibull BIC (smaller is better)	224.372

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
trt	1	16.6444	<.0001

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	2.2484	0.1660	1.9231	2.5737	183.51	<.0001
trt	6-MP	1	1.2673	0.3106	0.6585	1.8762	16.64	<.0001
trt	Control	0	0.0000	.	.	.	.	.
Scale		1	0.7322	0.1078	0.5486	0.9772		
Weibull Shape		1	1.3658	0.2012	1.0233	1.8228		

```
proc lifereg data=example outest=weibull
covout;
  class trt;
  model time*event (0)=trt;
run;
```

Example – PBC Data

Fit Statistics (Unlogged Response)	
-2 Log Likelihood	1712.189
Weibull AIC (smaller is better)	1720.189
Weibull AICC (smaller is better)	1720.319
Weibull BIC (smaller is better)	1735.161

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
trt	1	0.0723	0.7880
sex	1	5.0940	0.0240

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.5410	0.1888	4.1709	4.9111	578.28	<.0001
trt	1	1	-0.0386	0.1435	-0.3198	0.2427	0.07	0.7880
trt	2	0	0.0000	.	.	.	.	.
sex	f	1	0.4343	0.1924	0.0571	0.8114	5.09	0.0240
sex	m	0	0.0000	.	.	.	.	.
Scale		1	0.8600	0.0625	0.7458	0.9918		
Weibull Shape		1	1.1628	0.0846	1.0083	1.3409		

```
proc lifereg data=example
outest=weibull covout;
  class trt sex;
  model time*status(0)=trt sex
/*/dist=lnormal*/;
  output out=outw std=std quantile=0.5
p=predtime ;
  run;
```

$$L(\beta) = \prod_{i=1}^{n} h(T_i)^{\Delta_i} S(T_i)$$

$$= \prod_{i=1}^{n} \left(\lambda\alpha T_i^{\alpha-1}\right)^{\Delta_i} e^{-\lambda T_i^{\alpha}}$$

Homework 9

1. Let survival data (T_i, Δ_i) follow log-normal distribution, write the log-likelihood function. Discuss if the MLEs have closed form.
2. Use the Ovarian data to check if the survival data, by ecog status, follows
   a) the exponential distribution
   b) The Weibull distribution
3. Using the observed survival data (T_i, Δ_i, Z_i) $i = 1, 2, 3, 4, 5, 6$ to construct full likelihood function using Weibull distribution.The data are (16,1,1), (20, 0,1), (12, 1,0), (14, 0,0), (11, 1, 0), (9, 1, 1).
4. Use the Leukemia data to fit the AFT model assuming log-logistic distribution.

$(\lambda\alpha)^4 \cdot (16^{\alpha-1} \times 12^{\alpha-1} \times 11^{\alpha-1} \times 9^{\alpha-1}) \times e^{2\beta - \lambda}$