Parametric models

Daniela Zugna
Dipartimento di Scienze Biomediche ed Oncologia Umana, Università di Torino
daniela.zugna@unito.it

Time functions

The probability distribution of T can be specified in many ways:

1. survivor function: $S(t)$ (Kaplan-Meier estimator/Life table)
2. probability density function: $f(t)$
3. hazard function ($\lambda(t)$)
4. cumulative hazard function ($\Lambda(t)$) (Nelson-Aalen estimator)

Hazard function

Hazard function: $\lambda(t) = lim_{\Delta t \to 0} \frac{Pr(t \leq T < t+\Delta t | T > t)}{\Delta t}$

$\lambda(t)$ is the istantaneous failure rate (>0)

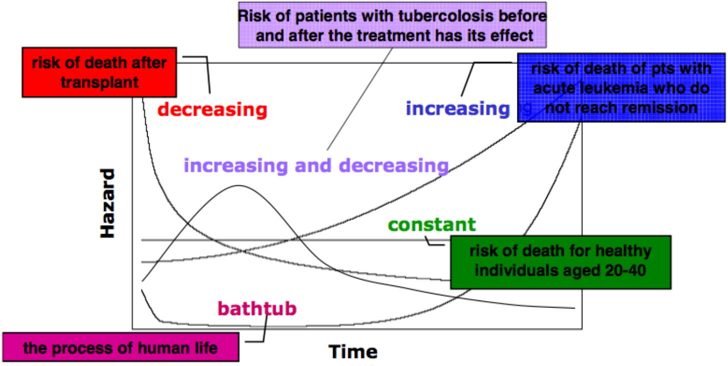
$\lambda(t)\Delta(t)$ is the probability that failure is between t and $t + \Delta(t)$ conditional on having survived until time t

$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} logS(t)$

$S(t), \lambda(t), f(t)$ are matematically equivalent

$S(t), \lambda(t)$ play a basic role in survival analysis

Hazard function



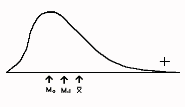
Modeling survival data

Modeling	Assumptions underlying the model	Examples
Non parametric	No assumptions	Kaplan-Meier, Nelson AAlen
Semi-parametric	Semi-specified	Cox model
Parametric	Specified functional form	Exponential, Weibull
		Log-normal, Log-logistic

Parametric models

A model assumes a well-defined functional form for $\lambda(t)$, by taking into account the main aspects of survival data:

- $T \geq 0$
- positively skewed distribution of survival time (normality is not appropriate) $\rightarrow$ normal distribution is not appropriate



- presence of censoring $\rightarrow$ least square estimation is not useful, maximum likelihood estimator is appropriate

Likelihood function

Let $i = 1, \dots, n$ be independent units

(t_i, δ_i) is observed: t_i time of occurrence of event or censoring; $\delta_i = 1$ if the event has been observed, $\delta_i = 0$ is the censoring has been observed

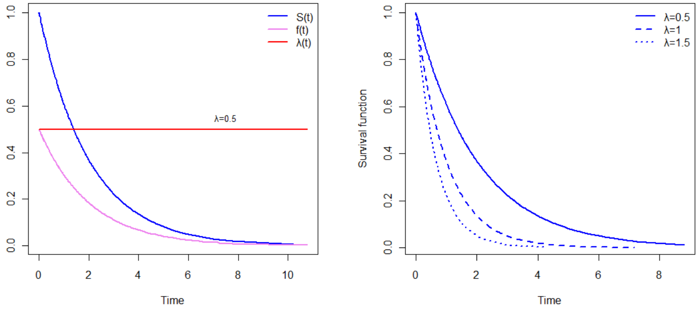
$$L(\theta) = \prod_{i=1}^{n} \left[f(t_i;\theta)^{\delta_i} S(t_i;\theta)^{1-\delta_i} \right] = \prod_{i=1}^{n} \lambda(t_i;\theta)^{\delta_i} S(t_i;\theta)$$

$$log(L(\theta)) = \sum_{i=1}^{n} log\left[\lambda(t_i;\theta)^{\delta_i} S(t_i;\theta) \right] = \sum_{i=1}^{n} \left[\delta_i log(\lambda(t_i;\theta)) + log(S(t_i;\theta)) \right]$$

Exponential model

$$T \sim exp(\lambda)$$

Hazard function	$\lambda(t) = \lambda \qquad \lambda > 0$
Cumulative hazard function	$\Lambda(t) = \int_0^t \lambda(u)du = \lambda t$
Survival function	$S(t) = exp(-\Lambda(t)) = exp(-\lambda t)$
Density function	$f(t) = -\frac{d}{dt}S(t) = \lambda exp(-\lambda t)$



Exponential model

$$log(L(\lambda)) = \sum_{i=1}^{n} [\delta_i log(\lambda(t_i;\lambda)) + log(S(t_i;\lambda))] = \sum_{i=1}^{n} [\delta_i log(\lambda) - \lambda t_i] =$$

$$= log\lambda \sum_{i=1}^{n} \delta_i - \lambda \sum_{i=1}^{n} t_i$$

$$\frac{d}{d\lambda} logL = 0 \leftrightarrow \widehat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i}$$

Estimator of failure rate when it is assumed to be constant

$$\widehat{\lambda} = \frac{numberoffailure}{totalpersontimeofobservation}$$

$\widehat{\lambda}$ corresponds to the estimator of the mortality rate heuristically introduced by epidemiologists and commonly expressed in terms of death per person times at risk

Example

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i}$$

Leukemia data: Survival in patients with cancer: is the treatment effective on the survival?

strate

failure _d: died
analysis time _t: studytime

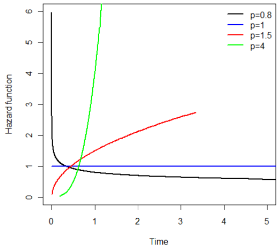
Estimated rates and lower/upper bounds of 95% confidence intervals
(48 records included in the analysis)

D	Y	Rate	Lower	Upper
31	744.0000	0.041667	0.029303	0.059247

WEIBULL MODEL

$$T \sim Weibull(\lambda, p)$$

Hazard function	$\lambda(t) = \lambda p(\lambda t)^{p-1}$	$\lambda, p > 0$
Survival function	$S(t) = exp(-(\lambda t)^p)$	
Density function	$f(t) = p\lambda^p t^{p-1} exp(-(\lambda t)^p)$	



$p < 1$ monotonically decreasing, $p = 1 \rightarrow$ exponential model, $p > 1$ monotonically increasing

Gompertz-Makeham model

Gompertz (1825) suggested that a "law of geometric progression pervades" in mortality after a certain age

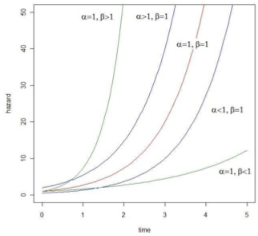
$\lambda(t) = \alpha exp(\beta t), \qquad \alpha, \beta \geq 0$

α=baseline mortality, β=age component

$S(t) = exp(-\int_0^t \alpha exp(\beta u) du) = exp(\frac{\alpha}{\beta}(1 - exp(\beta t)))$

Makeham (1860) extended the Gompertz model by adding a constant λ:

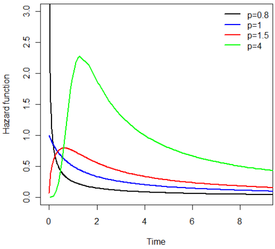
$\lambda(t) = \alpha exp(\beta t) + \lambda, \qquad S(t) = exp(-\lambda t + \frac{\alpha}{\beta}(1 - exp(\beta t)))$



Log-logistic model

$$T \sim loglogistic(\lambda, p)$$

Hazard function	$\lambda(t) = \frac{\lambda p (\lambda t)^{p-1}}{(1+(\lambda t)^p)}$	$\lambda, p > 0$
Survival function	$S(t) = \frac{1}{1+(\lambda t)^p}$	
Density function	$f(t) = \frac{p \lambda^p t^{p-1}}{(1+(\lambda t)^p)^2}$	

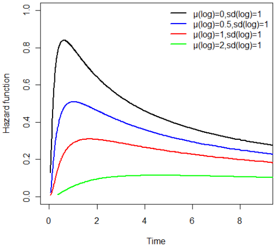


$p \leq 1$ monotonically declining hazard function, $p > 1$ the hazard function at first rises monotonically up to a maximum and then falls monotonically

Log-normal model

$$logT \sim N(\mu, \sigma^2)$$

Density function	$f(t) = \frac{1}{\sigma^2 t}\phi(\frac{logt-\mu}{\sigma^2})$	$\sigma^2 > 0$
Survival function	$S(t) = 1 - \Phi(\frac{logt-\mu}{\sigma^2})$	$\phi(t) = \frac{1}{\sqrt{2\pi}}exp(\frac{-t^2}{2})$
Hazard function	$\lambda(t) = \frac{1}{\sigma^2 t}\frac{\phi(\frac{logt-\mu}{\sigma^2})}{1-\Phi(\frac{logt-\mu}{\sigma^2})}$	$\Phi(t) = \int_0^t \phi(u)du$



The hazard function at first rises monotonically up to a maximum and then falls monotonically.

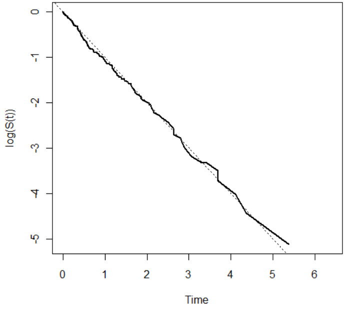
Methods to check parametric assumptions

It is important to empirically check the adeguacy of models upon which inferences are based, for example by:

- Graphical methods: they compare transformations of nonparametric estimates of survivor functions with the predictions from parametric models. Most of these approaches begin with a nonparametric estimation of a survivor function using the Kaplan-Meier estimator, and a nonparametric estimation of the cumulative hazard function using the Nelson-Aalen estimator
- Residuals and pseudo-residuals: they are calculated and used in evaluating distributional assumptions. Residuals are deviations of the observed values of the dependent variable from the values estimated under the assumptions of a specific model. When the dependent variable is not observable, as the hazard function in transition rate models, pseudo-residuals are used

Graphical methods

- Exponential model: $S(t) = exp(-\lambda t) \rightarrow log(S(t)) = -\lambda t$
  1. Consider the non parametric estimation of $S(t)$ by Kaplan-Meier estimator: $\widehat{S}_{KM}(t)$
  2. If failure time follows an exponential distribution the plot of $log(\widehat{S}_{KM}(t))$ should provide a linear graph passing through the origin vs t with slope λ figure



Graphical methods

- Exponential model: $\lambda(t) = \lambda$

  Consider the non parametric estimation of $\lambda(t)$ by life table estimator after grouping data by time intervals j:

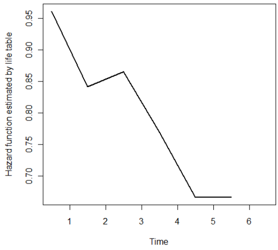
  $$\widehat{\lambda}_j = \frac{d_j}{(n_j - 1/2(d_j + c_j))(t_{j+1} - t_j)}$$

  The denominator is the product of the length of time interval by the average number of survivors at the middle point of the interval, by assuming an uniform distribution of deaths and censoring over the interval

	nsubs	nlost	nrisk	nevent	hazard	se.hazard
0-1	250	13	243.5	158	0.9604863	0.06702384
1-2	79	6	76.0	45	0.8411215	0.11375915
2-3	28	3	26.5	16	0.8648649	0.19495488
3-4	9	0	9.0	5	0.7692308	0.31754812
4-5	4	0	4.0	2	0.6666667	0.44444444
5-6	2	0	2.0	1	0.6666667	0.62853936
6-7	1	0	1.0	1	NA	NA

Graphical methods

If failure time follows an exponential distribution the plot of $\widehat{\lambda}(t))$ should provide a linear graph parallel to x-axis



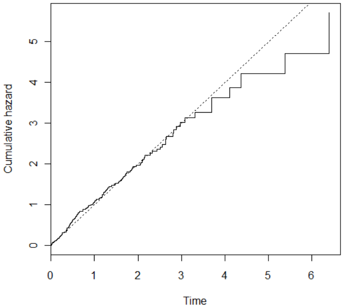
The hazard function estimate may fluctuate a lot when there are a few subjects at risk

The central death rate is plotted against time at the mid-point of the interval: smoothers are often applied, to study the shape of the hazard

Graphical methods

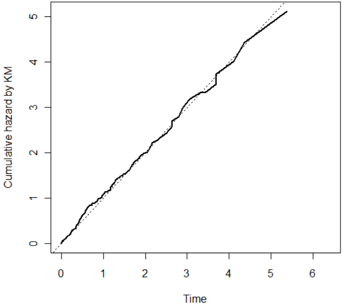
- Exponential model: $\Lambda(t) = \lambda t$ Consider the non parametric estimation of $\Lambda(t)$ by Nelson-Aalen estimator: $\widehat{\Lambda}_{NA}(t)$

  If failure time follows an exponential distribution the plot of $\widehat{\Lambda}_{NA}(t)$ should provide a linear graph passing through the origin vs t with slope λ



Graphical methods

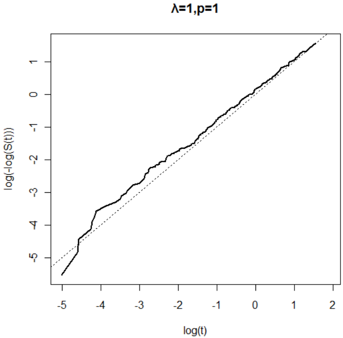
- Exponential model: $\Lambda(t) = \lambda t$
  Consider the non parametric estimation of $\Lambda(t)$ by Kaplan-Meier estimator: $\widehat{\Lambda}_{KM}(t) = -log(\widehat{S}_{KM}(t))$



Graphical methods

- Weibull model:
  $S(t) = exp(-(\lambda t)^p)$
  $log(S(t)) = -(\lambda t)^p$
  $log(-log(S(t))) = plog(\lambda) + plog(t)$

  Plot of $log(log(-log(\widehat{S}(t))))$ should be approximately linear vs $log(t)$



Graphical methods

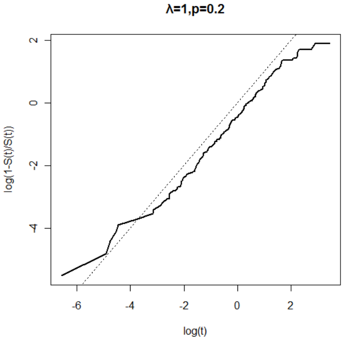
- Log-logistic model:

  $S(t) = \frac{1}{1+(\lambda t)^p}$

  $1 - (S(t)) = \frac{(\lambda t)^p}{1+(\lambda t)^p}$

  $log(\frac{1-S(t)}{S(t)}) = plog(\lambda) + plog(t)$

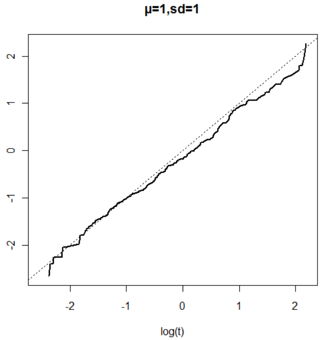
  Plot of $log(\frac{1-\widehat{S}(t)}{\widehat{S}(t)})$ should be approximately linear vs $log(t)$



Graphical methods

- Log-normal model:
  $S(t) = 1 - \Phi(\frac{log t - \mu}{\sigma^2})$
  $\Phi^{-1}(1 - S(t)) = \frac{log(t) - \mu}{\sigma^2}$

  Plot of $\Phi^{-1}(1 - \widehat{S}(t))$ should be approximately linear vs $log(t)$



Regression models

For each item, we observe $(T_i, \delta i, Z_i)$, where

- T_i is a censored failure time random variable
- δ_i is the censoring indicator
- Z_i is a set of covariates
- observations are i.i.d., independent censoring

Z_i could be

- scalar (gender, age or treatment, etc) or a vector (several covariates)
- discrete, continuous or time-varying

Regression models aim to model the relationship between survival and all of the variables of interest (explanatory variables)

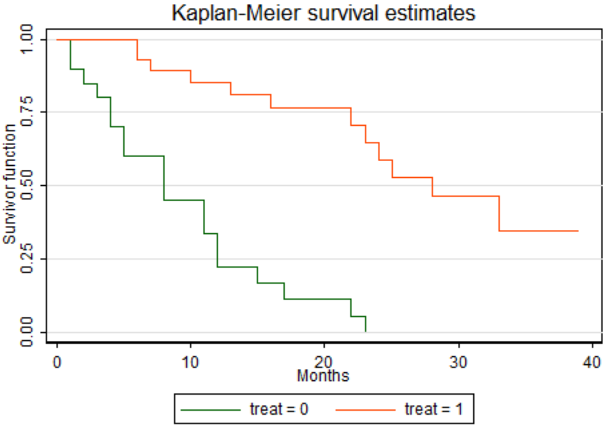
Regression models

Use of regression models to:

- adjust for imbalances in estimating and testing treatment effect
- test for and describe interactions between variables of interest
- understand prognostic factors
- model relative benefit, i.e. describe the course of a treatment effect, in terms of hazard ratio for example
- develop a model for survival probability
- identify subgroups at different prognosis, i.e. prognostic indexes based on estimated regression coefficient or parametric and semi-parametric recursive partition methods

Example

Leukemia data:



Regression models

Failure time regression models:

1. Proportional hazard models
   - Cox model
   - Exponential model
   - Weibull model
2. Accelerated failure time models
   - Exponential model
   - Weibull model
   - Log-logistic model
   - Log-normal model

Proportional hazard model

We suppose Z to be the treatment, $Z = 1$ if treated, 0 otherwise

$$\lambda(t; Z_i) = \lambda_0(t) exp(\beta' Z_i)$$

- $\lambda_0(t)$ is the same baseline hazard for every individual and it does depend only on time t
- $exp(\beta' Z_i)$ depends on covariates and causes different hazards for different individuals

$\lambda_1(t)$=hazard rate for the treated group where $\lambda_1(t) = \lambda_0(t) exp(\beta)$

$\lambda_0(t)$=hazard rate for the untreated group where $\lambda_0(t)$

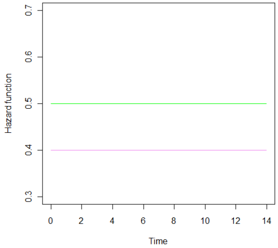
The hazard ratio, defined as the ratio between the hazard rates of treated group *vs* untreated group is given by:

$$HR = \frac{\lambda_1(t)}{\lambda_0(t)} = \frac{\lambda_0(t) exp(\beta)}{\lambda_0(t)} = exp(\beta)$$

Exponential regression model

1. One binary covariate z:

$$\lambda(t; z_i) = \lambda_0 exp(\beta' z_i) = exp(\beta_0 + \beta_1 z_i)$$



2. Several covariates $\boldsymbol{z} = (z_1, z_2, \ldots, z_k)$:

$$\lambda(t; \boldsymbol{z}_i) = \lambda_0 exp(\beta' \boldsymbol{z}_i) = exp(\beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \ldots + \beta_k z_{ki})$$

Note that the exponential function has been chosen because it assures the positivity of hazard

Leukemia data

$$\lambda(t; z_i) = exp(\beta_0 + \beta_1 z_i)$$

```
. xi:streg i.treat,dist(e)nohr

Exponential regression -- log relative-hazard form

No. of subjects =        48              Number of obs  =      48
No. of failures =        31
Time at risk    =       744
                                         LR chi2(1)     =   19.19
Log likelihood  =  -51.746927            Prob > chi2    =  0.0000

------------------------------------------------------------------------------
          _t |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_1 |  -1.60163   .3687342    -4.34   0.000    -2.324335    -.878924
       _cons | -2.248518   .2294157    -9.80   0.000    -2.698164   -1.798871
------------------------------------------------------------------------------
```

ML estimates for β_0, β_1:

$$\widehat{\beta}_0 = -2.25, 95\% CI : (-2.70, -1.80)$$
$$\widehat{\beta}_1 = -1.60, 95\% CI : (-2.32, -0.88)$$

Leukemia data

Hazard estimates:

$$\lambda(t; z = 0) = exp(-2.25) = 0.10$$
$$\lambda(t; z = 1) = exp(-2.25 - 1.60) = 0.02$$

. strate treat

Estimated rates and lower/upper bounds of 95% confidence intervals
(48 records included in the analysis)

treat	D	Y	Rate	Lower	Upper
0	19	180.0000	0.105556	0.067329	0.165486
1	12	564.0000	0.021277	0.012083	0.037465

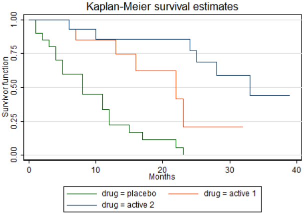
Hazard ratio estimate:

$$HR = \frac{\lambda(t;z=1)}{\lambda(t;z=0)} = \frac{exp(-2.25-1.60)}{exp(-2.25)} = exp(-1.60) = 0.20$$

Parametric test on treatment effect (based on likelihood):

$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$ $LRT : Q_{LR} = 19.19, p \leq 0.001$

Leukemia data



$$\lambda(t; z_i) = exp(\beta_0 + \beta_1 I(z_i = 1) + \beta_2 I(z_i = 2))$$

```
. xi:streg i.drug,dist(e)nohr

No. of subjects =        48              Number of obs   =      48
No. of failures =       31
Time at risk    =       744
                                    LR chi2(2)     =    20.02
Log likelihood  =  -51.330755              Prob > chi2     =   0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Idrug_2 |  -1.302057   .4682929    -2.78   0.005    -2.219894   -.3842197
    _Idrug_3 |   -1.83184   .4682929    -3.91   0.000    -2.749678   -.9140032
       _cons |  -2.248518   .2294157    -9.80   0.000    -2.698164   -1.798871
------------------------------------------------------------------------------
```

Leukemia data

ML estimates for $\beta_0, \beta_1, \beta_2$:

$$\widehat{\beta}_0 = -2.25, 95\% CI : (-2.22, -0.38)$$
$$\widehat{\beta}_1 = -1.30, 95\% CI : (-2.32, -1.80)$$
$$\widehat{\beta}_2 = -1.83, 95\% CI : (-2.75, -0.91)$$

Hazard estimates:

$$\lambda(t; z = 0) = exp(-2.25) = 0.10$$
$$\lambda(t; z = 1) = exp(-2.25 - 1.30) = 0.03$$
$$\lambda(t; z = 2) = exp(-2.25 - 1.83) = 0.02$$

Hazard ratio estimate:

$$HR_1 = \frac{\lambda(t;z=1)}{\lambda(t;z=0)} = \frac{exp(-2.25-1.30)}{exp(-2.25)} = exp(-1.30) = 0.27$$
$$HR_2 = \frac{\lambda(t;z=2)}{\lambda(t;z=0)} = \frac{exp(-2.25-1.83)}{exp(-2.25)} = exp(-1.83) = 0.16$$

Graphical check

Exponential model:

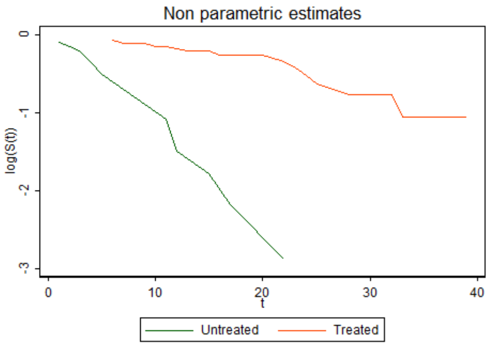
- Without covariates: $S(t) = exp(-\lambda t) \rightarrow log(S(t)) = -\lambda t$
- With covariates: $S(t;z) = exp[-exp(\beta^{'} z)t] \rightarrow log(S(t;z)) = -exp(\beta^{'} z)t$

$logS(t;z)$ is linear *vs* t

1. Calculate and plot $log\widehat{S}_{KM}(t;z)$ *vs* t
2. Exponential model is suitable if non-parametric estimates of $log\widehat{S}(t;z)$ in different groups are parallel lines

Graphical check

Leukemia data



Weibull regression model

- Without covariates: $\lambda(t) = \lambda p(\lambda t)^{p-1}$
- With covariates:

$$\lambda(t; z_i) = \lambda p(\lambda t)^{p-1} exp(\beta^{'} z_i) = \lambda_0(t) exp(\beta^{'} z_i)$$

$$\Lambda(t) = (\lambda t)^p \rightarrow \Lambda(t; z_i) = (\lambda t)^p exp(\beta^{'} z_i) = \Lambda_0(t) exp(\beta^{'} z_i)$$

$$S(t; z_i) = exp(-\Lambda(t; z_i) \rightarrow S(t; z_i) = exp(-\Lambda_0(t) exp(\beta^{'} z_i)) = exp(-\Lambda_0(t))^{exp(\beta^{'} z_i)}$$

Leukemia data

$$\lambda(t; z_i) = \lambda p(\lambda t)^{p-1} exp(\beta' z_i) = \lambda_0(t) exp(\beta' z_i)$$

```
. streg treat, dist(w) nohr

Weibull regression -- log relative-hazard form

No. of subjects =        48              Number of obs  =      48
No. of failures =        31
Time at risk    =       744
                                   LR chi2(1)     =    24.97
Log likelihood  =  -48.138622            Prob > chi2    =   0.0000

------------------------------------------------------------------------------
         _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      treat |  -1.950108   .3936052    -4.95   0.000     -2.72156   -1.178656
      _cons |  -3.610587    .628404    -5.75   0.000    -4.842236   -2.378937
-------------+----------------------------------------------------------------
      /ln_p |   .4304104   .1456647     2.95   0.003     .1449128     .715908
-------------+----------------------------------------------------------------
          p |   1.537889   .2240161                      1.155939    2.046044
        1/p |   .6502422   .0947173                      .4887481    .8650977
------------------------------------------------------------------------------
```

$\widehat{\lambda} = exp(-3.610587)$

$\widehat{\beta}_1 = -1.950108$

$p > 1 \rightarrow$ increasing hazard

Leukemia data

```
. streg treat, dist(w) hr

Weibull regression -- log relative-hazard form

No. of subjects =        48               Number of obs   =       48
No. of failures =        31
Time at risk    =       744
                                  LR chi2(1)      =     24.97
Log likelihood  =  -48.138622             Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  .1422587   .0559938    -4.95   0.000     .0657721     .307692
-------------+----------------------------------------------------------------
       /ln_p |  .4304104   .1456647     2.95   0.003     .1449128     .715908
-------------+----------------------------------------------------------------
           p |  1.537889   .2240161                      1.155939    2.046044
         1/p |  .6502422   .0947173                      .4887481    .8650977
------------------------------------------------------------------------------
```

Graphical check

- Without covariates:

$$S(t) = exp(-(\lambda t)^p) \rightarrow log(-log(S(t))) = plog(\lambda) + plog(t)$$

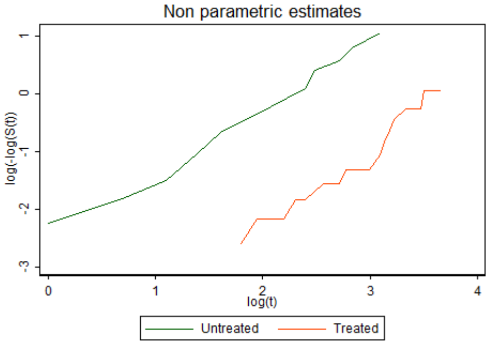
- With covariates:

$$S(t;z) = exp[(-\lambda t)^p]^{exp(\beta' z)} \rightarrow log(-log(S(t;z)) = \beta' z + plog\lambda + plogt$$

$log(-log(S(t;z))$ is linear *vs* $log(t)$ with slope p

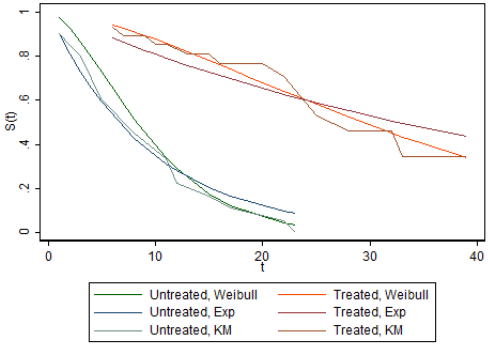
1. Calculate and plot $log(-log(\widehat{S}_{KM}(t;z))$ *vs* $logt$
2. Weibull model is suitable if non-parametric estimates of $log(-log(\widehat{S}(t;z))$ in different groups are parallel lines

Graphical check



Leukemia data

Observed vs predicted survival function



Proportional hazard model

Proportional hazard model is the most common model used for survival data, because

1. flexible choice of covariates
2. fairly easy to fit
3. standard software exists

but

"the success of Cox regression has perhaps had the unintended side-effect that practitioners too seldomly invest efforts in studying the baseline hazard ... A parametric version (of the Cox model),... if found to be adequate, would lead to more precise estimation of survival probabilities and ... concurrently contribute to a better understanding of the phenomenon under study"

Proportional odds model

Odds of failure in time: $\frac{F(t;z)}{1-F(t;z)} = w_0(t)exp(\beta^{'} z)$

Odds ratio of failure in time: $OR = \frac{F(t;z_2)/S(t;z_2)}{F(t;z_1)/S(t;z_1)} = exp(\beta^{'}(z_2 - z_1))$

- OR does not depend on $w_0(t)$
- OR is constant over time
- The effect of each covariate on OR is multiplicative (for a unit of change in the value of z_k, OR is multiplied by $exp(\beta)$)

Proportional odds model

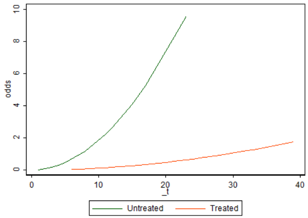
Log-logistic model is a proportional odds model:

$$\lambda(t;z) = \frac{\lambda p(\lambda t)^{p-1} exp(\beta' z)}{(1+(\lambda t)^{p}) exp(\beta' z)}$$

$$S(t;z) = \frac{1}{1+(\lambda t)^{p} exp(\beta' z)}$$

Odds of failure in time: $\frac{1-S(t;z)}{S(t;z)} = (\lambda t)^p exp(\beta' z) = w_0(t) exp(\beta' z)$

The shape of the odds function is monotone increasing for any pattern of covariates and depend on p only



Graphical check

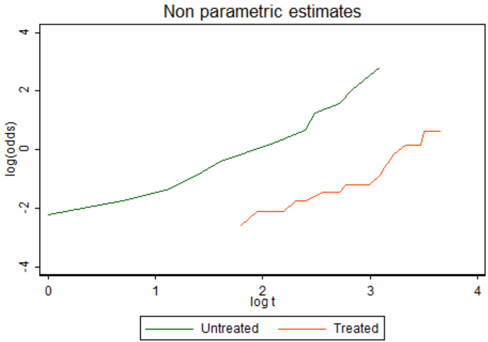
$odds = \frac{F(t;z)}{1-F(t;z)} = (\lambda t)^p exp(\beta' z)$

$log(odds) = \beta' z + plog(\lambda) + plogt$

$log(odds)$ is linear $vs\ logt$

- Calculate non parametrically and plot $log\widehat{odds}$ $vs\ logt$
- Log-logistic model is suitable if non-parametric estimates of $log\widehat{odds}$ in different groups are parallel lines

Graphical check



Log normal regression model

- Without covariates: $S(t) = 1 - \Phi(\frac{logt - \mu}{\sigma^2}) \rightarrow \Phi^{-1}(1 - S(t)) = \frac{log(t) - \mu}{\sigma^2}$
- With covariates: $S(t) = 1 - \Phi(\frac{logt - \beta^{'} z}{\sigma^2}) \rightarrow \Phi^{-1}(1 - S(t; z)) = \frac{log(t) - \beta^{'} z}{\sigma^2}$

Plot of $\Phi^{-1}(1 - \widehat{S}(t; z))$ should be approximately linear vs $log(t)$

1. Calculate non parametrically and plot $log\widehat{S}(t; z)$ vs $logt$
2. Log-normal model is suitable if non-parametric estimates of $log\widehat{S}(t; z)$ in different groups are parallel lines

Graphical methods

- Exponential model: $log(\widehat{S}(t; z))$ should provide a linear graph passing through the origin vs t

- Weibull model: $log(-log(\widehat{S}(t; z))$ should provide a linear graph *vs* $log(t)$ with slope p

- Logistic model: $log(\frac{1-\widehat{S}(t;z)}{\widehat{S}(t;z)}$ should provide a linear graph *vs* $log(t)$

- Lognormal model: $\Phi^{-1}(1 - \widehat{S}(t))$ should provide a linear graph vs $log(t)$

Accelerated failure times models

The effect of covariates is expressed directly on survival time:

$$T_Z = \frac{T_0}{exp(\theta' Z)}$$

where $T_0 = T|Z = 0$ ($Z = 0$ baseline covariates vector)

If $exp(\theta^{'} Z) > 1$ reduced failure times, if $exp(\theta^{'} Z) < 1$ accelerated failure times

$$S(t;z) = P(T > t|z) = P(\frac{T_0}{exp(\theta' z)} > t|z) = P(T_0 > texp(\theta^{'} z)|z) = S_0(texp(\theta^{'} z))$$

$$f(t;z) = lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} =$$

$$= lim_{\Delta t \to 0} \frac{1}{\Delta t} P(texp(\theta^{'} z) \leq T_0 < (t+\Delta t)exp(\theta^{'} z)) \frac{exp(\theta^{'} z)}{exp(\theta' z)} = exp(\theta^{'} z) f_0(texp(\theta^{'} z))$$

$$\lambda(t|z) = \frac{f(t;z)}{S(t;z)} = \frac{exp(\theta^{'} z) f_0(texp(\theta^{'} z))}{S_0(texp(\theta' z))} = exp(\theta^{'} z)\lambda_0(texp(\theta^{'} z))$$

$$\lambda(t;z) \propto \lambda_0(texp(\theta^{'} z))$$

Weibull model

$S(t; z) = S_0(texp(\theta^{'} z))$

$S(t; Z) = exp[(-\lambda t)^p]^{exp(\beta^{'} Z)}$

by some algebra and by a different parameterization:

$S(t; Z) = exp\Big\{ - \Big[\lambda texp(\frac{1}{p}\beta Z)\Big]^p \Big\}$

$\frac{1}{p}\beta = \theta$: parameter of an accelerated failure time

$exp(\frac{1}{p}\beta Z)$: time acceleration

Standardized residuals

In accelerated failure times models: $log(T;z) = \beta' Z + log(T_0)$

$$(t_i, \delta_i, z_i), \qquad log(t_i|z_i) = \beta' z_i + \epsilon_i, \qquad \epsilon_i = log(t_0)$$

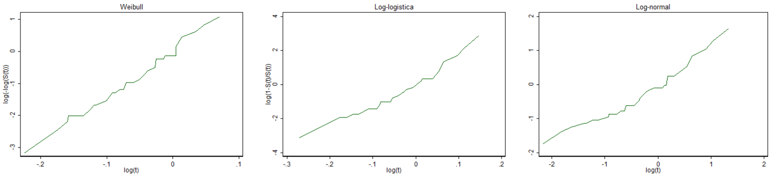
- $T_0 \sim Weibull(\lambda, p) \rightarrow$ linear model for $log(T;Z)$ with extreme value distribution of errors
- $T_0 \sim loglogistic(\lambda, p) \rightarrow$ linear model for $log(T;Z)$ with logistic distribution of errors
- $T_0 \sim log(N(\mu, \sigma)) \rightarrow$ linear model for $log(T;Z)$ with normal distribution of errors

Standardized residuals

$r_i = \frac{log(t_i) - \widehat{\beta}' z_i}{\widehat{\sigma}}$ is the standardized residual (σ the scale parameter)

The survivor function is estimated, non parametrically, on residuals $(exp(r_i), \delta_i)$, and it is plotted versus $log(t)$:

1. If $log(-log(\widehat{S}(t)))$ is linear *vs* $log(t) \rightarrow$ Weibull model
2. If $log(\frac{1-\widehat{S}(t)}{\widehat{S}(t)})$ is linear *vs* $log(t) \rightarrow$ Log-logistic model
3. If $\Phi^{-1}(1 - \widehat{S}(t))$ is linear *vs* $log(t) \rightarrow$ Log-normal model



Cox-Snell residuals

$$T \sim S(t)$$
$$S(T) \sim Unif(0,1)$$
$$-log[S(T)] \sim exp(1)$$

If the model is correct, the estimated cumulative hazard for each individual at the time of their death or censoring should be like a censored sample from a unit exponential

Pseudoresiduals: $\widehat{e}_i = \int_0^t \widehat{\lambda}(u; z_i)du \qquad i = 1, \ldots, N$

If the model is appropriate $(\widehat{e}_i, \delta_i) \sim exp(1)$

$$\Downarrow$$

a plot of $-log(\widehat{S}_{\widehat{e}}(\widehat{e}_i))$ versus $\widehat{e}_i$, calculated non-parametrically, should be approximately linear (passing through origin and with slope equal to 1)

Martingale and deviance residuals

(t_i, δ_i, z_i), $\quad \widehat{\lambda}(t_i; z_i)$, $\quad i = 1, \ldots, N$ the estimated cumulative hazard function

Martingale residuals: $\widehat{r}_{m_i} = \delta_i - \widehat{\Lambda}(t_i; z_i)$

- If the model is appropriate, a plot of $\widehat{r}_{m_i}$ versus $exp(\beta^{'} z_i)$ should be approximately linear
- Plot of $\widehat{r}_{m_i}$ versus variables not included in the model, could suggest potential relations between hazard function and those variables, by using smoothing or lowess

Deviance residuals: transformation of martingale residuals: $E(\widehat{r}_{d_i}) = 0$

- If the model is appropriate, $\widehat{r}_{d_i} \sim WN(0, 1)$ i.i.d.
- Plot of deviance residuals against covariates to look for unusual pattern

In Stata

Distribution	Metric	Survival function	Parameterization
Exponential	PH	$exp(-\lambda t)$	$\lambda = exp(z\beta)$
Exponential	AFT	$exp(-\lambda t)$	$\lambda = exp(-z\beta)$
Weibull	PH	$exp(-\lambda t^p)$	$\lambda = exp(z\beta)$
Weibull	AFT	$exp(-\lambda t^p)$	$\lambda = exp(-pz\beta)$
Loglogistic	AFT	$(1 + (\lambda t)^{1/\gamma})^{-1}$	$\lambda = exp(z\beta)$
Lognormal	AFT	$1 - \Phi(\frac{log(t)-\mu}{\sigma})$	$\mu = z\beta$