Chapter 12 Parametric Regression

PubH 7450

©Wei Pan

Email: weip@biostat.umn.edu

Http: www.biostat.umn.edu/∼weip

§12.1 Introduction

- Parametric: the baseline distribution is parametric
  $\Longrightarrow$ distribution for any subject is parametric; e.g., Weibull.

- AFT, not PHM, is discussed
  PHM can be done in an equal way; how?
  Parametric PHM: R function `phreg()` in package `eha`

- Review of AFT
  AFT 1: $S(x|Z) = S_0[\exp(Z'\theta)x]$;
  AFT 2: $Y = \log X = \mu + \gamma' Z + \sigma W$,
  (1)==(2) if i) S_0 is the survival function for r.v. $\exp(\mu + \sigma W)$;
  ii) $\theta = -\gamma$.

- Common parametric models for W (or for S_0):

S_0	W
Weibull	Extreme value distr
Log-normal	Normal
Log-logistic	Logistic
...	

§12.2 Weibull distribution for S_0

- Weibull distr.: $X \sim Weibull(\lambda, \alpha)$
  $\alpha = 1 \Longrightarrow X \sim Exp(\lambda)$;
  $E(X) = \Gamma(1 + 1/\alpha)/\lambda^{1/\alpha}$;
  $S_X(x) = \exp(-\lambda x^\alpha)$;
  $h_X(x) = \lambda\alpha x^{\alpha-1}$;

- If $Y = \log X = \mu + \sigma W$
  with $\lambda = \exp(-\mu/\sigma)$, $\sigma = 1/\alpha$, where W has a (Type-I) extreme value distr with
  $f_W(w) = \exp(w - e^w)$, $S_W(w) = \exp(-e^w)$,
  $E(W) \approx 0.57722$, $Var(W) \approx 1.64493$.

- Hence,
  $f_Y(y) = \frac{1}{\sigma}\exp\left(\frac{y-\mu}{\sigma} - e^{\frac{y-\mu}{\sigma}}\right)$,
  $S_Y(y) = \exp\left(-e^{\frac{y-\mu}{\sigma}}\right)$.

- Given right-censored data: (T_j, δ_j), $j = 1, ..., n$, again assume $X_j \sim Weibull$, then let $y_j = \log T_j$, (bad notation! why?) $L = \prod_{j-1}^{n} [f_Y(y_j)]^{\delta_j} [S_Y(y_j)]^{1-\delta_j} =$ $\prod_{j-1}^{n} \left[\frac{1}{\sigma} f_W\left(\frac{y_j-\mu}{\sigma}\right)\right]^{\delta_j} \left[S_W\left(\frac{y_j-\mu}{\sigma}\right)\right]^{1-\delta_j}$ $\Longrightarrow \hat{\mu}, \hat{\alpha}, Cov, ...$

- Now with covariates
  Data: (T_j, δ_j, Z_j), $j = 1, ..., n$.
  AFT: $\log X_j = Y_j = \mu + \gamma' Z_j + \sigma W$,
  $L = \prod_{j-1}^{n} \left[\frac{1}{\sigma} f_W\left(\frac{y_j-\mu-\gamma' Z_j}{\sigma}\right)\right]^{\delta_j} \left[S_W\left(\frac{y_j-\mu-\gamma' Z_j}{\sigma}\right)\right]^{1-\delta_j}$,
  with f_W and S_W as given before.
  $\Longrightarrow$ MLE $\hat{\mu}, \hat{\gamma}, \hat{\sigma}, Cov$.

- Inference: Wald, score, LRT, as before

- Can equally handle left-censored, interval-censored and left-truncated data!

how?
implemented in SAS Proc Lifereg.

- Example 12.2: Larynx cancer
  Covariates: $Z_1 - Z_3$, indicators of stage II-IV; Z_4, age.
  AFT model:
  $Y = \log X = \mu + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + \sigma W$
  Interpretation:
  AFT: $E(\log X|stage, age) = (\mu + \sigma EW) + \gamma' Z$.
  $E(\log X|stageI, age = z_4) = \mu^* + \gamma_4 z_4$;
  $E(\log X|stageII, age = z_4) = \mu^* + \gamma_1 + \gamma_4 z_4$;
  take a difference:
  $\gamma_1 = E(\log X|stageII, age = z_4) - E(\log X|stageI, age = z_4)$
  —-effect of stage after *adjusting for age.*
  Similarly,
  $\gamma_4 = E(\log X|stage = s, age = z_4 + 1) - E(\log X|stage = s, age = z_4)$

—-effect of age after *adjusting for stage.*

- Example 12.2: SAS
- A special case: AFT=PHM if ...

  1) AFT: $Y = \log X = \mu + \gamma' Z + \sigma W$;

  2) W has a standard extreme value distribution (i.e. S_0 is Weibull distribution).

  1)+2) $\Longrightarrow$

  $h(x|Z) = \alpha\lambda x^{\alpha-1}\exp(Z'\beta) = h_0(x)\exp(Z'\beta)$, PHM!

  with $\alpha = 1/\sigma$, $\lambda = \exp(-\mu/\sigma)$, and $\beta = -\gamma/\sigma$.

  How to prove?

$$
\begin{aligned}
AFT &\Longrightarrow S(x|Z) = S_0[\exp(Z'\theta)x] \\
&\Longrightarrow f(x|Z) = f_0([\exp(Z'\theta)x])\exp(Z'\theta) \\
&\Longrightarrow h(x|Z) = h_0([\exp(Z'\theta)x])\exp(Z'\theta) \\
&\quad = \alpha\lambda[\exp(Z'\theta)x]^{\alpha-1}\exp(Z'\theta) = \alpha\lambda x^{\alpha-1}\exp(Z'\theta\alpha) = \ldots
\end{aligned}
$$

Compare Tables 12.1 & 12.2.

- Read §12.3-12.4.

  Other parametric models; same method, interpretation,...

TABLE 12.1
Analysis of Variance Table for Stage and Age for Laryngeal Cancer Patients, Utilizing the Log Linear Model, Assuming the Weibull Distribution

Variable	*Parameter Estimate*	*Standard Error*	*Wald Chi Square*	*p-Value*
Intercept $\hat{\mu}$	3.53	0.90		
Scale $\hat{\sigma}$	0.88	0.11		
Z_1: Stage II ($\hat{\gamma}_1$)	−0.15	0.41	0.13	0.717
Z_2: Stage III ($\hat{\gamma}_2$)	−0.59	0.32	3.36	0.067
Z_3: Stage IV ($\hat{\gamma}_3$)	−1.54	0.36	18.07	<0.0001
Z_4: Age ($\hat{\gamma}_4$)	−0.02	0.01	1.87	0.172

TABLE 12.2
Parameter Estimates for the Effects of Stage and Age on Survival for Laryngeal Cancer Patients, Modeling Time Directly Assuming the Weibull Distribution

Variable	*Parameter Estimate*	*Standard Error*
Intercept $\hat{\lambda}$	0.002	0.002
Scale $\hat{\alpha}$	1.13	0.14
Z_1: Stage II ($\hat{\beta}_1$)	0.17	0.46
Z_2: Stage III ($\hat{\beta}_2$)	0.66	0.36
Z_3: Stage IV ($\hat{\beta}_3$)	1.75	0.42
Z_4: Age ($\hat{\beta}_4$)	0.02	0.01

§12.5 Diagnostics

- Class one: use properties of parametric distributions.
- First consider one-sample problem: univariate
- Model:
  $X \sim W(\lambda, \alpha) \Longrightarrow H(x) = \lambda x^{\alpha}$.
- Given data: (T_j, δ_j), $j = 1, .., n$
  $\Longrightarrow \hat{H}(x)$, a nonparametric, e.g. N-A, estimate.
- Model checking:
  Compare $\hat{H}(x)$ with assumed $H(x)$: e.g., plot $\log \hat{H}(x)$ vs $\log x$, and it should be a ...... if the model holds.
  More generally: compare a parametric estimate vs a NP estimate...
- Consider regression: (T_j, δ_j, Z_j), $j = 1, .., n$
- Model (AFT):

$\log X_j = \mu + \gamma' Z_j + \sigma W_j,$
W_j iid standard extreme value distribution
$\Longrightarrow \exp(\log X_j - \gamma' Z_j) = \exp(\mu + \sigma W_j) \overset{iid}{\sim}$ Weibull.

- X_j may not be observed because censoring/truncation, but we can use ...... $\hat{H}(x)$; then plot $\log \hat{H}(x)$ vs ......
- Fig 12.2.
- Class two: use residuals
- Cox-Snell residuals
  $r_j = \hat{H}(X_j | Z_j) \sim Exp(1)$;
  under the Weibull model, $r_j = \hat{\lambda} T_j^{\hat{\alpha}} \exp(\beta' Z_j)$.
  Apply N-A esimator to (r_j, δ_j), $j = 1, ..., n$ and see whether it is close to a straight line with slope =1 (why?).
- Fig 12.7.

- Or, look at standardized residuals,
  $s_j = \frac{\log T_j - \hat{\mu} - \hat{\gamma}' Z_j}{\hat{\sigma}}$.
  (s_j, δ_j): a censored sample from a standard extreme value distribution (under the Weibull model); equivalent to the above.

- Use other residuals
  $M_j = \delta_j - r_j$, skewed.
  $D_j = sign(M_j)\{(-2)[M_j + \delta_j \log(\delta_j - M_j)]\}^{1/2}$, more symmetric.
  check: 1) outliers by index plots
  2) overall fit: plot vs T_j; any trend suggests possible problems.

- Fig 12.10.

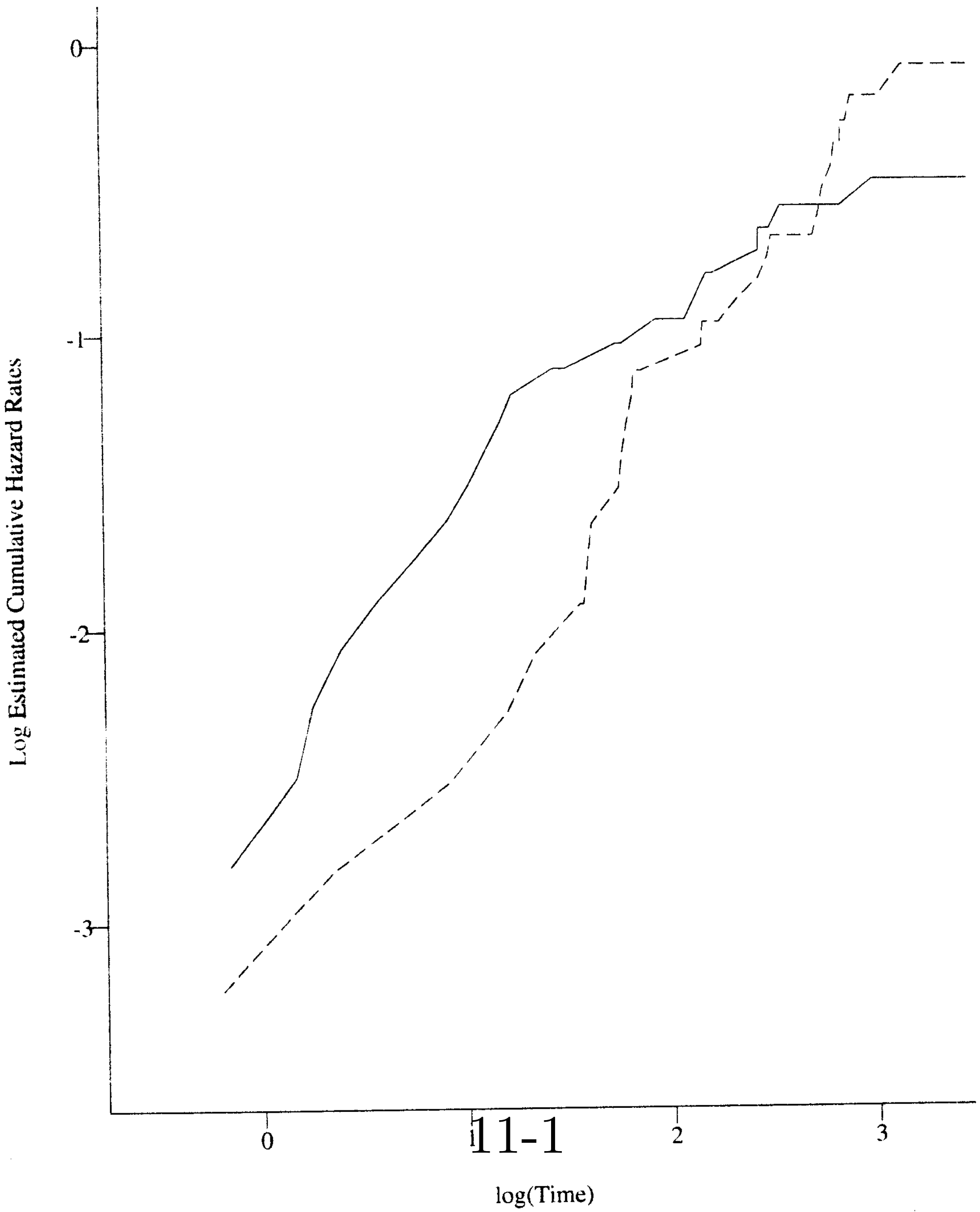
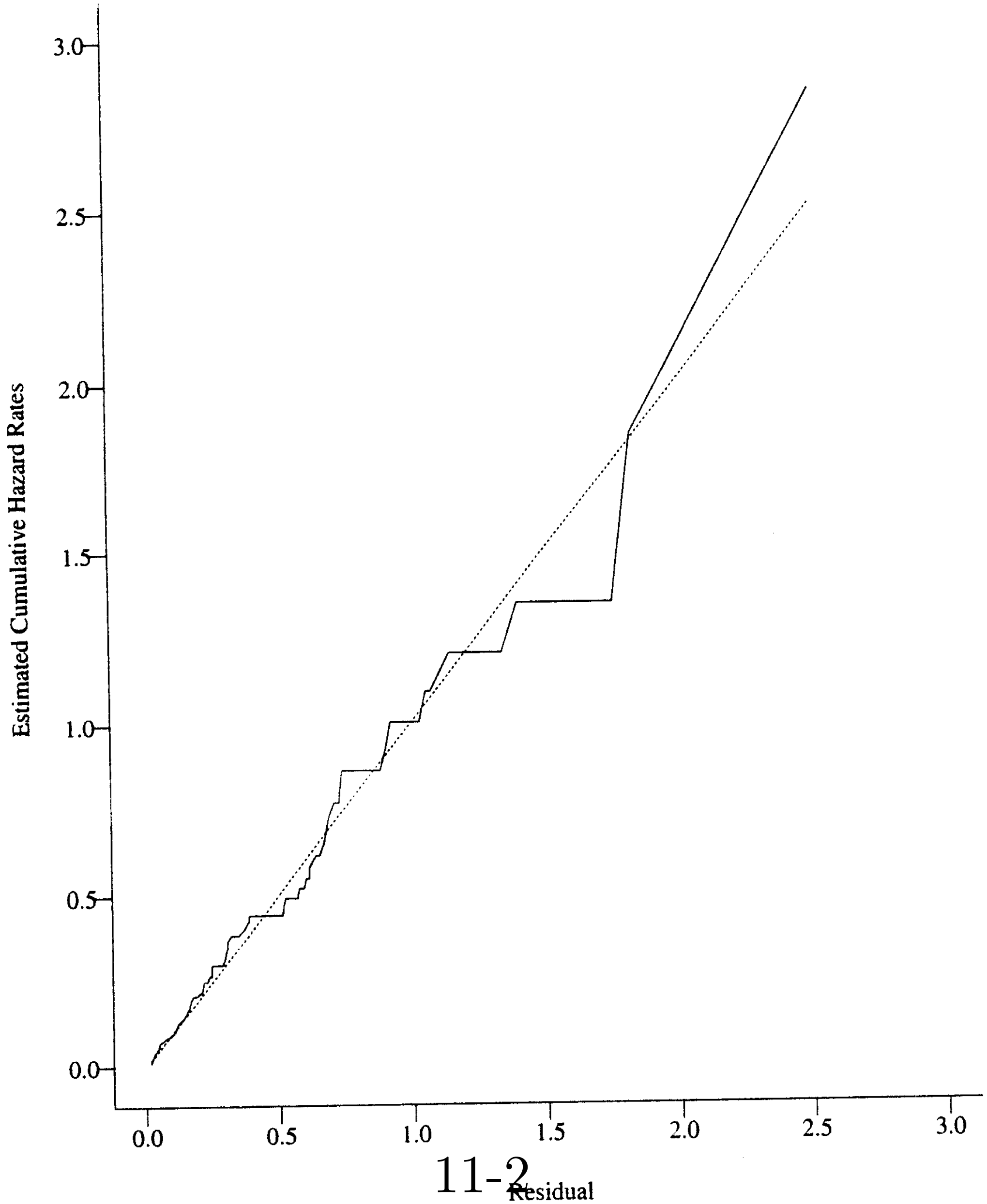


Figure 12.2 *Weibull hazard plot for the allo (solid line) and auto (dashed line) transplant groups.*



11-2

Figure 12.7 *Cox–Snell residuals to assess the fit of the Weibull regression model for the laryngeal cancer data set*

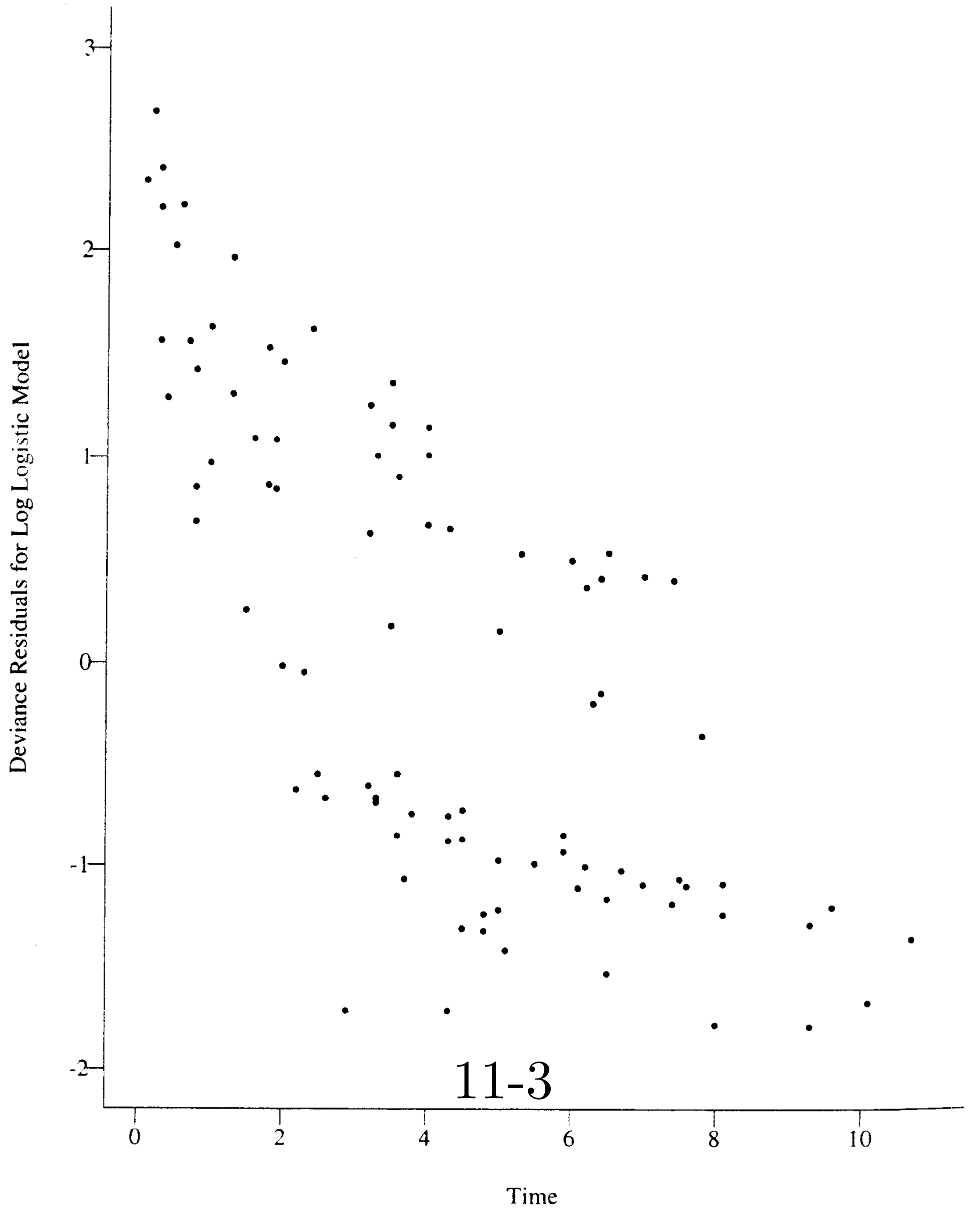


Figure 12.10 *Deviance residuals from the log logistic regression model for laryngeal cancer patients*