Accelerated Failure Time Model

Model Diagnosis

Model Diagnosis for parametric regression

- Object: Assessing the adequacy of the fitted models to the data
- Method: Graphical diagnostic procedures (informal but very useful) using residuals.

Standardized residuals

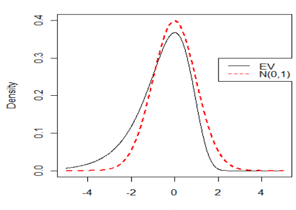
- Recall that the log-linear model is of the form $Y_i = \log\left(T_i^0\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sigma\varepsilon_i$.
  or $\varepsilon_i = \frac{\log(T_i^0) - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})}{\sigma}$.
- Define the standardized residuals as
  $$R_i = \frac{\log(T_i) - (\widehat{\beta_0} + \widehat{\beta_1} x_{i1} + \cdots + \widehat{\beta_p} x_{ip})}{\hat{\sigma}}.$$
  - $\widehat{\beta_0}, \widehat{\beta_1}, \dots, \widehat{\beta_p}$, and $\hat{\sigma}$ are the estimated parameters.
  - An approximation of the error term in the model.
  - Gives the standardized difference between observed and predicted values.

Standardized residuals

- If the log-linear model is correct, then $\{R_i\}_{i=1}^{n}$ can be viewed as censored sample from the distribution of ε.
  - For example, if a Weibull model holds, then $\{R_i\}_{i=1}^{n}$ are roughly a censored sample from a standard extreme value distribution.
  - For example, if a Loglogistic model holds, then $\{R_i\}_{i=1}^{n}$ are roughly a censored sample from a logistic distribution.
- However, a residual plot of $\{R_i\}_{i=1}^{n}$is not very useful.

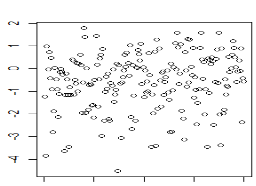
Standardized residuals

- Density of standard extreme value distribution $f(t) = \exp(t - \exp(t))$.



Standardized residuals

- Random numbers generated from standard extreme value distribution



Standardized residuals

- The traditional residual plot is not useful to assess whether error distribution assumption is satisfied.
- Instead, we will use survival plots
  - Estimate survival function using standardized residuals
  - check if the survival function follows the functional form of a given distribution.

Standardized residuals

If a Weibull model holds, then

- $\{R_i\}_{i=1}^n$ are roughly a censored sample from a standard extreme value distribution.
- $\{U_i = \exp(R_i)\}_{i=1}^n$ is roughly a censored sample from an Exponential distribution.
- Recall that for Exponential distribution

  $S(t) = \exp(-\lambda t)$ or $-\log\big(S(t)\big) = \lambda t$
- Survival plot for Weibull model
  - Obtain K-M estimator $\hat{S}(\cdot)$ based on $\{U_i\}_{i=1}^n$.
  - Plot $-\log\left(\hat{S}(U_i)\right)$ against U_i and check if straight line that go through the origin.

Standardized residuals

If a Loglogistic model holds, then

- $\{R_i\}_{i=1}^n$ are roughly a censored sample from a logistic distribution.
- $\{U_i = \exp(R_i)\}_{i=1}^n$ is roughly a censored sample from an loglogistic distribution.
- Recall that for loglogistic distribution

  $S(t) = \frac{1}{1+\lambda t^{\alpha}}$ $\quad logit\big(S(t)\big) = -\log(\lambda)$- $\alpha\log(t)$
- Survival plot for Loglogistic model
  - Obtain K-M estimator $\hat{S}(\cdot)$ based on $\{U_i\}_{i=1}^n$
  - Plot $logit\left(\hat{S}(U_i)\right)$ against $\log(U_i)$ and check if it is a straight line.

Standardized residuals

- It measures the standardized difference between the observed and the predicted log survival time.
- Definition is similar to residuals in linear regression.
- Inconvenient: for different error distribution, a different survival plot is needed.

Cox-snell residuals

- The Cox-snell residuals are defined as
  $$r_i = \hat{H}(T_i \mid \boldsymbol{x}_i)$$
  - $\hat{H}(\cdot \mid \boldsymbol{x})$ is the cumulative hazard from fitted model.
- For example, if exponential model is assumed, then
  $$H(t \mid \boldsymbol{x}) = t\exp\left(-\beta_0 - \beta_1 x_1 - \cdots - \beta_p x_p\right),$$

and $r_i = \hat{H}(T_i \mid \boldsymbol{x}_i) = T_i \exp\left(-\hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip}\right)$.

- Cox-snell residuals are very different from the usual residuals, e.g. $r_i > 0$ for all i.

Cox-snell residuals

Theorem: Let T be a random variable, and H(t) be the cumulative hazard function of T. Then
$$H(T) \sim Exp(1).$$
Proof: Let F(t) be the CDF of T. Then
$$F(T) \sim U(0,1)$$
$$\Rightarrow S(T) = 1 - F(T) \sim U(0,1)$$
$$\Rightarrow H(T) = -\log(S(T)) \sim Exp(1).$$

Cox-snell residuals

If the model fits the data, then Cox-snell residuals $\{r_i = \widehat{H}(T_i \mid \boldsymbol{x}_i)\}_{i=1}^{n}$ approximately is a censored sample from a standard exponential distribution.

- Let $\hat{S}(\cdot)$ be the K-M estimator using $\{r_i\}_{i=1}^{n}$
- a plot of $-\log[\hat{S}(r_i)]$ vs r_i should be a straight line with slope 1.

Residuals in SAS

- You can specify output of desired residuals in the OUTPUT STATEMENT
  - CRESIDUAL | CRES: Cox-Snell residuals
  - SRESIDUAL | SRES: standardized residuals

Summary

- The standardized residual is easy to understand, but not convenient.
- Different survival plots of the residuals are needed for different regression models.
- Cox-snell residual is the estimated cumulative hazard function, not traditional residuals.
- Can always been viewed as censored sample from exponential distribution.
- Always use log of survival (ls) plot to assess for goodness of fit.