Cox PH Model

Deviance and Schoenfeld residuals

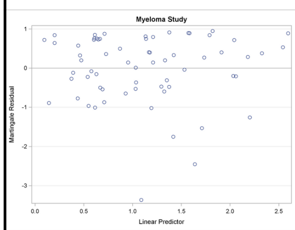
Martingale residuals

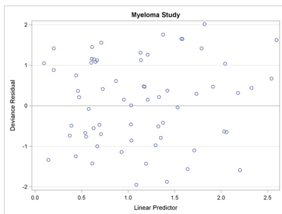
- Martingale residuals are defined as
$$r_{M,i} = \delta_i - r_{CS,i} = \delta_i - \widehat{H}(\, t_i \mid Z_i \,)$$
- The martingale residuals take values in $(-\infty, 1)$.
- The distribution of martingale residuals are quite skewed (to left).
- This skewed distribution makes it hard to identify outliers.
- A technique for creating symmetric, normalized residuals that is widely used in generalized linear modeling is to construct "Deviance Residuals".

Deviance residuals

- Deviance residual is a transformation of martingale residual to make it symmetric and normalized.
$$r_{D,i} = sign\left(r_{M,i}\right)\sqrt{-2\left[r_{M,i} + \delta_i \log\left(\delta_i - r_{M,i}\right)\right]}$$
- Deviance residuals behave like residuals from ordinary linear regression
  - Symmetric around zero, with approximated sd=1.
  - positive value ⇒ more events than expected (live shorter).
  - negative value ⇒ less events than expected (live longer).
  - Either too large positive or negative values (outside $(-3, 3)$) can be regarded as outliers.

An example





Deviance residuals

To identify outliers

- Fit a Cox PH model
- Plot deviance residuals against the linear predictors $\{Z_i^T\beta\}_{i=1}^n$
- Detect any data points that is outside of $(-3, 3)$ as outliers

Schoenfeld residuals

- Proposed by Schoenfeld (1982) to assess proportional hazard assumption in Cox PH model
  - Schoenfeld D. Residuals for the proportional hazards regresssion model. *Biometrika* 1982, 69(1):239-241
- Instead of a single residual for each individual, there is a separate residual for each event time *for each covariate*
- It is not defined for censored individuals.

Schoenfeld residuals

- The Schoenfeld residual for variable Z_l and subject $k \in D(t_i)$ who had event at t_i is defined as

$$Z_{kl} - \bar{Z}_{il}$$

  - $\bar{Z}_{il}$ i the expected value of the lth covariate at time t_i
- Then the Schoenfeld residual for variable Z_l at event time t_i is defined as

$$r_{S,li} = \sum_{k \in D(t_i)} [Z_{kl} - \bar{\boldsymbol{Z}}_{il}]$$

Schoenfeld residuals

- Schoenfeld residuals are differences between observed and expected values of each covariate at each event time.
- If the residuals exhibit a random pattern, then the effect of covariate does not change over time and PH assumption is satisfied.
- If there is a systematic pattern, then effect of covariate changes over time, and PH assumption is not satisfied.

Schoenfield residuals

- Check proportional hazard assumption for covariate Z_l.
  - plot $\{r_{S,li}\}_{i=1}^{D}$ against $\{t_i\}_{i=1}^{D}$ and examine trends
  - If the residuals exhibit a random pattern, then it indicates that the covariate effect is not changing with respect to time. That is it supports the proportional hazard assumption.
  - Any pattern in the residual plot suggests that the PH assumption is questionable.

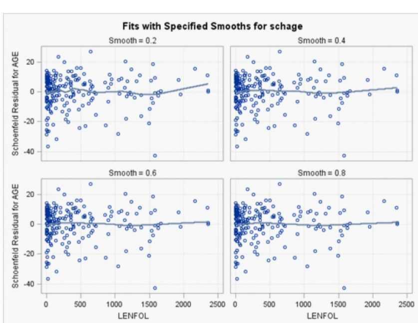
SAS Example

```
proc phreg data=whas500;
class gender;
model lenfol*fstat(0) = gender|age
bmi|bmi hr;
output out=schoen ressch=schgender
schage schgenderage
schbmi schbmibmi schhr;
run;
```

- Output Schoenfield residuals for each variable.

SAS Example

```
proc loess data = schoen;
model schage=lenfol / smooth=(0.2 0.4 0.6 0.8);
run;
```



The smooths appear mostly flat at 0, suggesting that the coefficient for age does not change over time and that proportional hazards holds for this covariate.

Summary

- Diagnosis of model assumptions
  - Functional form: Martingale residuals
  - PH assumption: Schoenfeld residuals
- Outliers: Deviance residuals.