Survival Analysis I (CHL5209H)
Week 8 - Cox model diagnostics

Kuan Liu, PhD(c)

Feb 26th, 2020

Contact: kuan.liu@mail.utoronto.ca

Overview

1. Breslow estimator
2. COX model diagnostics
3. Modelling time-dependent covariates in COX model

Breslow estimator

UNIVERSITY OF TORONTO

- Recall from previous lecture, the Cox PH model yields the following

$$\lambda_i(t) = \lambda_0(t) exp\{\beta' x_i\},$$

  where x_i is a vector of time-independent covariates, $\lambda_0(t)$ is baseline hazard function and β is the vector of parameters (of interest) with each element interpreted as the log hazard ratio.
- If we substitute x_i with 0, what do we get?
- $\lambda_0(t)$ can be viewed as the underlying hazard for all subjects when all covariates are set to 0, hence the word "baseline".
- We can estimate β using the Cox partial likelihood which does not feature $\lambda_0(t)$.

Breslow estimator (2)

- Now, if we are interested to estimate the absolute risks of event or survival probabilities using the COX fitted model, we need an estimator for β and an estimator for $\lambda_0(t)$.
- The survival probability for individual i at time t is given as

$$
\begin{aligned}
S_i(t) &= exp\{-\int_0^t \lambda_i(u)du\} \\
&= exp\{-\int_0^t \lambda_0(u)exp\{\beta' x_i\}du\} \\
&= exp\{-\Lambda_0(t)exp\{\beta' x_i\}\}
\end{aligned}
$$

where $\Lambda_0(t)$ is the cumulative baseline hazard.

Breslow estimator (3)

UNIVERSITY OF TORONTO

- Breslow estimator is the most commonly used estimator for cumulative baseline hazard.

$$\hat{\Lambda}_0(t) = \sum_{j:t_j \leq t} \frac{d_j}{\sum_{l=1}^{n} Y_l(t_j) exp\{\hat{\beta}' x_l\}}$$

- Here, $\hat{\beta}$ are the partial likelihood estimates of the regression parameters, and the at-risk process $Y_i(t) = I_{\{T_i \geq t\}}$ is used to check whether individual i is still at risk at time t.
- t_j refers to the ordered observed event time and d_j as the number of events at each time.
- This estimator can be obtained using the profile likelihood (Good read for PhD student on derivation).
- What is the connection between Nelson-Aalen estimator and Breslow estimator? Recall $\hat{\Lambda}(t)_{NA} = \sum_{j:t_j \leq t} \frac{d_j}{n_j}$.

Breslow estimator (4)

- Demo in R.

Cox model diagnostics

UNIVERSITY OF TORONTO

- Can you name a few simple linear regression diagnostics based on residual examinations?
- It's not easy to define a counterpart of the residual in a linear regression model for censored time-to-event outcomes.
- Nevertheless, several types of residuals exist for the Cox model, and can be used for model diagnostics.
- We are going to introduce the following diagnostics today
  1. Martingale residuals
  2. Deviance residuals
  3. Score residuals and dfbeta measures
  4. Schoenfeld residuals

Martingale residuals

- Recall from lecture 1&2, the hazard function can be express with counting process notations as,

$$P(dN_i(t) = 1 \mid \mathcal{F}_{t^-}) = E[dN_i(t) \mid \mathcal{F}_{t^-}] = Y_i(t)\lambda_i(t)dt.$$

- $Y_i(t)\lambda_i(t)dt$ can be interpreted as the expected value of the counting process jump, this motivates consideration of the differences

$$dM_i(t) = dN_i(t) - Y_i(t)\lambda_i(t)dt,$$

  which have the usual kind of "observed" minus "expected" interpretation.

- Observed number of deaths (0 or 1) for subject i at time t versus the expected number of death (0 or 1) for subject i at time t based on the fitted model.

Martingale residuals (2)

- Equivalently, integrate over time (from 0 to t),

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda_i(u)du$$

or

$$N_i(t) = \int_0^t Y_i(u)\lambda_i(u)du + M_i(t)$$

- The process $M_i(t)$ has property $E[dM_i(t) \mid \mathcal{F}_{t-}] = 0$. A process with this property is called martingale process.
- The process $M_i(t)$ can thus be interpreted as 'noise', while the hazard function captures the systematic variation in the counting process $N_i(t)$.

Martingale residuals (3)

- Now incorporate Cox model, we get

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) exp\{\beta' x_i\} \lambda_0(u) du$$

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) exp\{\hat{\beta}' x_i\} d\hat{\Lambda}_0(u) \quad \text{(the estimated version)}$$

  where the cumulative baseline hazard is estimated with Breslow estimator.
- The $\hat{M}_i$ evaluated at the end of the follow-up period of individual i, is the martingale residual.
- $\hat{M}_i$ ranges between $-\infty$ and 1. Why? Hint: $\Lambda(t) = -logS(t)$.

Martingale residuals (4)

How can Martingale residuals be used for Cox model diagnostic?

- The $\hat{M}_i$ for $i = 1, \ldots, n$ can be plotted for example against continuous covariates to check for non-linearity of the covariate effects.
- This is because any systematic deviation from linearity (on the modeled log hazard ratio) will show as systematic difference in the residuals.
- If there is evidence of linearity violation - that is an additive linear combination of the covariates a not suitable, we can consider including interactions, higher order terms, splines and other function forms etc.

Deviance Residuals

- Deviance residuals are re-scaled versions of martingale residuals, to make them more symmetric around zero.

$$\hat{D}_i(t) = sign(\hat{M}_i(t))\sqrt{-2[\hat{M}_i(t) + N_i(t)log(N_i(t) - \hat{M}_i(t))]}$$

- Similar to martingale residuals, we can plot deviance residuals against continuous covariates to check for linearity.

Score residual and dfbeta measure

UNIVERSITY OF TORONTO

- Score residual and dfbeta measure the influence of the observation i on the parameter estimates when i is removed from the model fit. These are calculated separately for each regression coefficient.
- We can conduct sensitivity analysis by comparing the estimated log Hazard ratios (the Cox regression coefficients) with and without the identified influential observations.

Schoenfeld residuals

- Recall the Cox log partial likelihood is

$$l(\beta) = \sum_{i=1}^{n} e_i[\beta' x_i - log\{\sum_{l=1}^{n} Y_l(t_i) exp(\beta' x_l)\}]$$

- The score function of the above log likelihood w.r.t β is

$$\sum_{i=1}^{n} e_i[x_i - \frac{\sum_{l=1}^{n} Y_l(t_i) x_l exp(\beta' x_l)}{\sum_{l=1}^{n} Y_l(t_i) exp(\beta' x_l)}]$$

- The Schoenfeld residuals are defined from the score function as

$$\hat{r}_i(t) = x_i(t) - \frac{\sum_{l=1}^{n} Y_l(t) exp(\hat{\beta}' x_l(t)) x_l(t)}{\sum_{l=1}^{n} Y_l(t) exp(\hat{\beta}' x_l(t))},$$

and can be seen as the difference between the observed covariate and the expected covariate at each failure time. The expected covariate is a weighted-average of the covariate, weighted by each individual's likelihood of event at t.

Schoenfeld residuals (2)

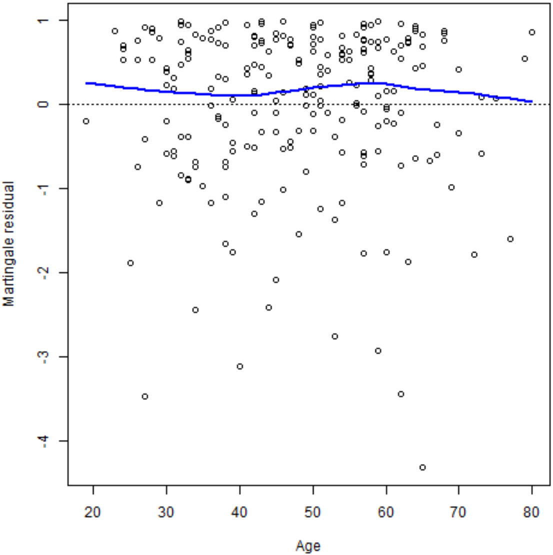
- Schoenfeld residuals are calculated for each covariates.
- Schoenfeld residuals are defined only for non-censored observations.
- Sum of the Schoenfeld residuals equals to 0. Why?
- Schoenfeld (1982) showed that these residuals are asymptotically uncorrelated and have expectation 0. Thus a plot of Schoenfeld residuals against time should be centered around zero.
- Additionally, the 'expected' covariates are obtained based on the PH assumption, thus, we can also assess PH assumption with this residual plot.
- In principle, a Schoenfeld residuals plot that shows a non-random pattern against time is evidence of violation of the PH assumption.

Demonstration in R

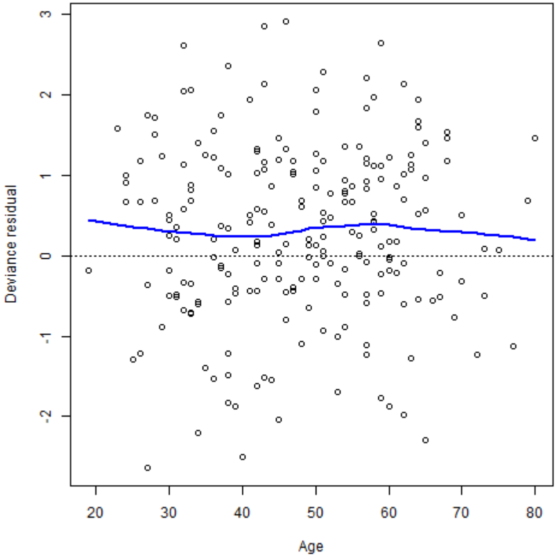
In class exercise activity. What can you conclude about the example Cox model from the following residual plots?

- Students will be divided into 5 groups.
- Each group will receive a printed copy of residual plots.
- Work in groups, write down your conclusions.
- You will be given 10 mins to work on this.

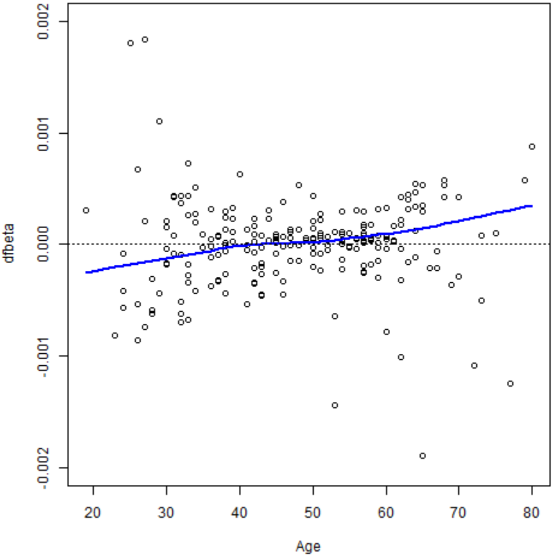
Exercise - Martingale



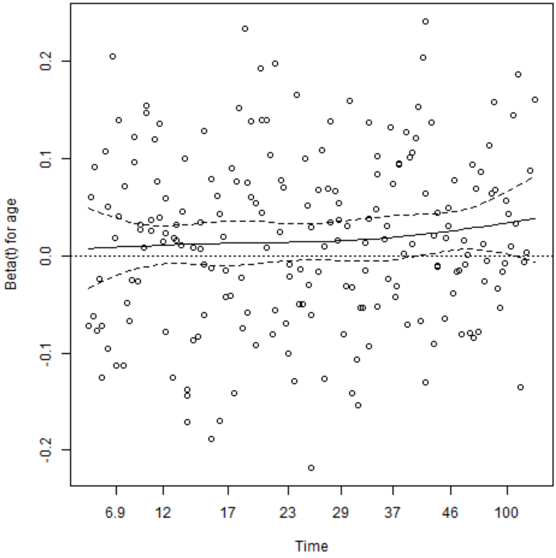
Exercise - Deviance



Exercise - dfbeta



Exercise - Schoenfeld



Testing proportionality

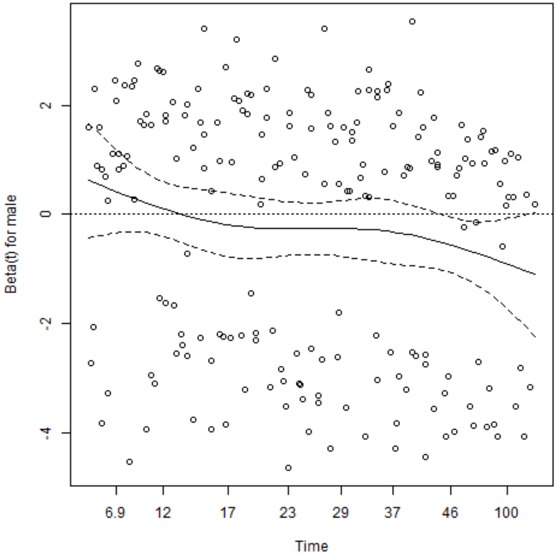
- We have just demonstrated how we can use Schoenfeld residuals to visually diagnose PH assumption for Cox model.
- In addition to plotting the Schoenfeld residuals against time, we can also test whether there is a significant correlation of the residuals with time.
- Such tests in R are calculated for each covariates by the cox.zph function.
- plot.cox.zph produces residual plots of scaled Schoenfeld residuals against time or some form of transformation.
- Another way to test PH would be to add covariate-time interaction terms into the model, and test whether these are significantly different from 0.

Tests for PH

Assessing significance of the residual time correlation

	chisq	df	p
treat	0.93	1.00	0.34
resect75	3.51	1.00	0.06
age	0.56	1.00	0.45
interval	0.20	1.00	0.66
karn	0.07	1.00	0.79
race	1.24	1.00	0.26
local	0.19	1.00	0.66
male	5.16	1.00	0.02
nitro	1.25	1.00	0.26
factor(path)	2.87	3.00	0.41
grade	0.27	1.00	0.60

Scaled Schoenfeld residual plot - male



Adding time-dependent covariates

- In the R coxph function, the covariate-time interaction terms can be added using the tt argument.
- The Schoenfeld residual plot for binary covariate male over time shows a non-random downward pattern, in addition, the correlation test is significant with a p-value$= 0.02 < 0.05$.
- We can try adding an interaction term of male and time in the model.

  coxph(Surv(weeks, event) ∼ treat + tt(treat) + resect75 + age + interval + karn + race + local + male + tt(male) + nitro + factor(path) + grade,
  *tt=function(x,t, ...) x * t, data=brain)*

tt function in R

	coef	exp(coef)	se(coef)	z	p
treat	-0.40	0.67	0.14	-2.78	0.01
resect75	-0.42	0.66	0.17	-2.55	0.01
age	0.02	1.02	0.01	2.92	0.00
interval	-0.14	0.87	0.05	-2.89	0.00
karn	-0.35	0.71	0.16	-2.17	0.03
race	0.61	1.85	0.27	2.27	0.02
local	-0.48	0.62	0.18	-2.73	0.01
male	0.15	1.16	0.23	0.66	0.51
tt(male)	-0.01	0.99	0.01	-2.41	0.02
nitro	0.51	1.67	0.16	3.27	0.00
factor(path)2	-0.69	0.50	0.22	-3.14	0.00
factor(path)3	-0.85	0.43	0.23	-3.75	0.00
factor(path)4	-0.54	0.58	0.43	-1.27	0.21
grade	-0.83	0.43	0.29	-2.85	0.00

Recap

- We learned how to obtain a Breslow estimator.
- We studied the following residual plots for Cox model diagnostics,
    1. Martinagle and Deviance residuals are use to assess linearity;
    2. Score residuals and dfbeta measures are used to identify influential observations;
    3. Schoenfeld residuals are used to assess PH assumption.
- We learned two additional ways to assess proportionality,
    1. testing significance of the correlation between Schoenfeld residuals and time;
    2. assessing the significance of the interaction term between covariates and time.