Biostat 537: Survival Analaysis

TA Session 6

Ethan Ashby

February 20, 2024

Presentation Overview

Review of Last Time

1. The Cox model enables testing, estimation, and inference from survival data under the proportional hazards assumption.
2. The partial likelihood is the basis for estimation and inference in the Cox model – its calculation relies on the rank ordering of the event times.
3. There exist exact and approximate methods to handling tied event times.
4. (Partial) likelihood ratio tests and stepwise selection w/ AIC are model selection strategies.

Overview

Regression Diagnostics

In classical regression problems, analysts often plot the residuals ($\hat{\epsilon}_i := y_i - \hat{f}(x_i)$) against covariates/fitted values to diagnose problems and suggest remedies to functional forms of covariates.

Regression Diagnostics

In classical regression problems, analysts often plot the residuals ($\hat{\epsilon}_i := y_i - \hat{f}(x_i)$) against covariates/fitted values to diagnose problems and suggest remedies to functional forms of covariates.

In linear regression, analysts often plot residuals versus leverage, where leverage is the influence that a particular observation has on your model.

Regression Diagnostics

In classical regression problems, analysts often plot the residuals ($\hat{\epsilon}_i := y_i - \hat{f}(x_i)$) against covariates/fitted values to diagnose problems and suggest remedies to functional forms of covariates.

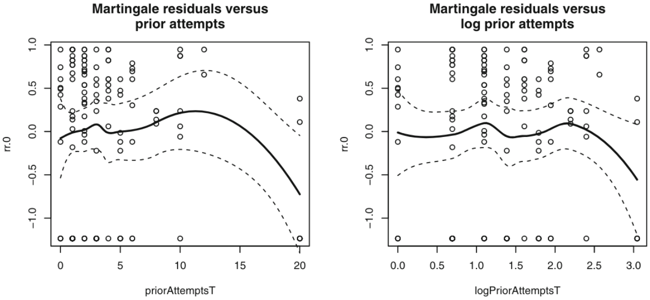
In linear regression, analysts often plot residuals versus leverage, where leverage is the influence that a particular observation has on your model.

Analogs exist for survival data!

Martingale Residuals

Martingale residuals: $m_i := \delta_i - \hat{H}_0(t_i)\exp(x_i\hat{\beta})$

1. Can reveal functional form misspecifications of covariate.



Jackknife Residuals

Some subjects may have a large influence on a Cox regression model.

Jackknife Residuals

Some subjects may have a large influence on a Cox regression model.

Jackknife residuals: difference between parameter estimate on the full data, $\hat{\beta}$, versus data where single case was deleted, $\hat{\beta}^{(-i)}$.

Jackknife Residuals

Some subjects may have a large influence on a Cox regression model.

Jackknife residuals: difference between parameter estimate on the full data, $\hat{\beta}$, versus data where single case was deleted, $\hat{\beta}^{(-i)}$.

Can be calculated using 'residuals(model, type="dfbeta")' in R.

Checking proportional hazards

The Cox model relies on the proportional hazards assumption. It is important to check its validity!

Strategy 1 for checking PH: cloglog plots

Under Cox model

$$S_1(t) = [S_0(t)]^{\exp(\beta)}$$
$$\implies \log(S_1(t)) = \exp(\beta) \cdot \log(S_0(t))$$
$$\implies \log(-\log(S_1(t))) = \beta + \log(-\log(S_0(t)))$$

This is called the *complementary log-log transformation*.

Hence, if proportional hazards holds, $\log(-\log(S_1(t)))$ and $\log(-\log(S_0(t)))$ should be parallel and separated by a constant.

Strategy 2: Schoenfeld residuals

At the *i*-th failure time, the Schoenfeld residual is:

$$\hat{r}_i := x_i - \sum_{k \in R_i} x_k \cdot \frac{\exp(x_k \beta)}{\sum_{k \in R_i} \exp(x_k \beta)} = x_i - \bar{x}(t_i)$$

If proportional hazards holds, a plot of $\hat{r}_i$ versus a covariate X should be flat line at zero.

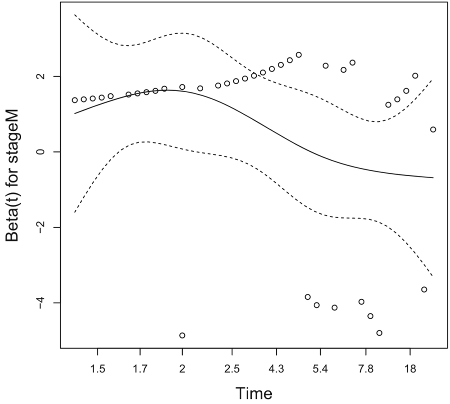
Strategy 2: Schoenfeld residuals

If proportional hazards does NOT hold and the hazard ratio varies with t, then

$$\hat{\beta}(t) \approx \hat{\beta} + \mathbb{E}[\hat{r}_i]$$

Compute in R: 'cox.zph()'. Returns a p-value for a test of whether $\hat{\beta}(t)$ is constant.

Strategy 2: Schoenfeld residuals



Overview

1. Cox Model Diagnostics
2. Time Dependent Covariates
3. AFT Regression

Stanford Heart Transplant Study

1. The 1971 study showed that patients who received heart transplant (binary fixed covariate) lived longer than patients who did not ($p=7 \times 10^{-7}$).
2. Critique the inclusion of heart transplant as a fixed covariate in the Cox model above!

Time Dependent Covariates in Cox Model

Recall the cox model assumes

$$h(t|X) = h_0(t)e^{X\beta}$$

We can incorporate time-varying covariates

$$h(t|X) = h_0(t)e^{X(t)\beta}$$

The partial likelihood becomes

$$L(\beta) = \prod_{i=1}^{D} \frac{e^{x_i(t_i)\beta}}{\sum_{k\in R_i} e^{x_k(t_i)}\beta}$$

In R

To fit a Cox model with time-dependent covariates, we must convert the time-to-event data into start-stop format.

```
> sdata<-tmerge(heart.simple, heart.simple, id=id,
+ death=event(futime,fustat), transpl=tdc(wait.time)
  )
> sdata
Row id  tstart  tstop  death  transpl
1   2        0   5     1          0
2   5        0   17    1          0
3   10       0   11    0          0
4   10       11  57    1          1
5   12       0   7     1          0
```

In R

Now we fit a Cox model to the start-stop data with Surv(start, stop, event).

```
> summary(coxph(Surv(tstart,tstop,death)~transpl, data=
    sdata))
```

Time-varying coefficient

The Cox model can readily accommodate time-varying covariates of the form.

$$h(t|X) = h_0(t)e^{X(t)\beta}$$

We can also consider time-varying coefficients

$$h(t|X) = h_0(t)e^{X\beta(t)}$$

This is a much harder task and requires determining the *time-transfer function*.

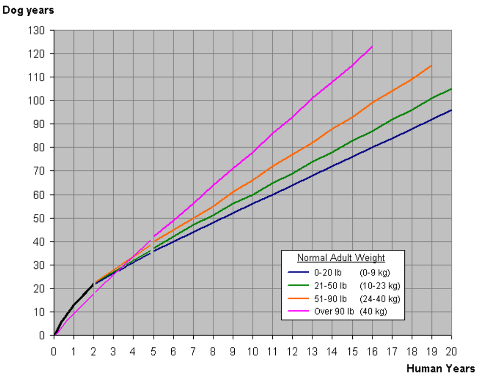
Time-transfer function

```
>result.panc2.tt<-coxph(Surv(pfs)~stage.n+tt(stage.n), +
    tt=function(x,t,...)x*log(t))
>result.panc2.tt
               coef    exp(coef)   se(coef)     z      p
stage.n        6.01    407.339     3.060      1.96     0.050
tt(stage.n)   -1.09    0.338       0.589     -1.84    0.065
```

Hence $\beta(t) = 6.01 - 1.09\log(t)$ under this tt function.

Overview

Accelerated Failure Time (AFT) models



AFT Models

Intuitively, AFT models assume that survival distributions have the same shape but are sped up or slowed down depending on covariates.

$$S(t|X) = S_0(e^{-\gamma X} t)$$
$$\iff h(t|X) = e^{-\gamma X} h_0(e^{-\gamma X} t)$$

Allows us to conclude that survival times for one-unit higher X are $e^{-\gamma}$ longer on average!

Weibull AFT

Suppose we assume survival distribution follows a Weibull shape. We can inspect the Weibull hazard and show

$$h(t|X) = e^{-\gamma/\sigma X} h_0(t)$$

Implying that a Weibull AFT is also a proportional hazards model! It is also the only model that is AFT and PH!

AFTs in R

```
>result.survreg.grp<-survreg(Surv(ttr,relapse)~grp,dist=
    "weibull")
>summary(result.survreg.grp)
```

```
              Value Std. Error     z        p
(Intercept)   5.286     0.3320 15.92 4.59e-57
grppatchOnly -1.251     0.4348 -2.88 4.00e-03
Log(scale)    0.689     0.0911  7.56 3.97e-14

Scale= 1.99
```

Summary

1. There exist residual-based diagnostics for examining functional form misspecification and PH assumptions in Cox models.
2. Start-stop and time-varying coefficients are two ways to accommodate covariates that vary with time in Cox Models.
3. AFT models are fully parametric alternatives to Cox models for regression analysis which have appealing interpretations but