

STAT 6390: Analysis of Survival Data

Textbook coverage: Chapter 6

Steven Chiou

Department of Mathematical Sciences,
University of Texas at Dallas

The Akaike Information Criterion

- One way to compare nested models is to use the partial likelihood ratio tests.
- One way to compare non-nested models is to use quantity like the *Akaike Information Criterion* (AIC):

$$\text{AIC} = -2 \cdot \ell(\hat{\beta}) + 2 \cdot p,$$

where $\ell(\cdot)$ can be replaced with the log-partial likelihood obtained from the Cox model.

- The AIC has two components; a “good” model should
 - fits the data well \rightarrow small value of $-2 \cdot \ell(\hat{\beta})$
 - has few parameters \rightarrow small $2 \cdot p$

The Akaike Information Criterion

- Consider the model:

```
> fit <- coxph(Surv(lenfol, fstat) ~ age + gender + bmi, data = whas100)
> fit
Call:
coxph(formula = Surv(lenfol, fstat) ~ age + gender + bmi, data = whas100)

            coef exp(coef) se(coef)      z      p
age      0.0371    1.0378   0.0127   2.92 0.0035
gender   0.1432    1.1540   0.3060   0.47 0.6397
bmi     -0.0708    0.9316   0.0361  -1.96 0.0496
```

Likelihood ratio test=21.54 on 3 df, p=8e-05
n= 100, number of events= 51

- AIC can be called with AIC function:

```
> AIC(fit)
[1] 402.6565
```

- or extractAIC (3 degrees of freedom),

```
> extractAIC(fit)
[1] 3.0000 402.6565
```

- or be computed directly with loglik.

```
> -2 * fit$loglik[2] + 2 * 3
[1] 402.6565
```

The Akaike Information Criterion

- An exhaust search can be carried out with `step` or `stepAIC` from **MASS**

```
> step(fit)
Start:  AIC=402.66
Surv(lenfol, fstat) ~ age + gender + bmi

      Df    AIC
- gender  1 400.87
<none>    402.66
- bmi     1 404.60
- age     1 410.25

Step:  AIC=400.87
Surv(lenfol, fstat) ~ age + bmi

      Df    AIC
<none>    400.87
- bmi     1 402.83
- age     1 411.78

Call:
coxph(formula = Surv(lenfol, fstat) ~ age + bmi, data = whas100)

      coef exp(coef) se(coef)      z      p
age  0.0393   1.0401   0.0119   3.31 0.00094
bmi -0.0712   0.9313   0.0361  -1.97 0.04895

Likelihood ratio test=21.32 on 2 df, p=2e-05
n= 100, number of events= 51
```

The Akaike Information Criterion

- An alternative to the AIC is the *Bayesian Information Criterion*, BIC, given by

$$\text{BIC} = -2 \cdot \ell(\hat{\beta}) + p \log(n).$$

- The key difference is in the penalty term; BIC penalizes the number of parameters by $\log(n)$, where “ n ” corresponds to the number of events.
- The BIC can be computed with

```
> extractAIC(fit, k = log(fit$nevent))
[1] 3.000 408.452
> -2 * fit$loglik[2] + log(fit$nevent) * 3
[1] 408.452
```

- The `step` function can be modified for stepwise selection with BIC:

```
> step(fit, k = log(fit$nevent))
```

- As a result, using the BIC in the model selection will tend to result in models with fewer parameters as compared to AIC.

Schoenfeld residuals

- The use of residuals for modeling checking has been well-developed in linear regression theory.
- Recall the score function,

$$S_n(\beta) = \frac{d\ell_p}{d\beta} = \sum_{i=1}^n \Delta_i \left[X_i - \left\{ \frac{\sum_{j \in R(t_i)} X_j e^{X_j \beta}}{\sum_{j \in R(t_i)} e^{X_j \beta}} \right\} \right].$$

- The estimator of the Schoenfeld residual for the i th subject is defined as

$$r_i = \Delta_i \left[X_i - \left\{ \frac{\sum_{j \in R(t_i)} X_j e^{X_j \hat{\beta}}}{\sum_{j \in R(t_i)} e^{X_j \hat{\beta}}} \right\} \right].$$

- Since $S_n(\hat{\beta}) = 0$, the sum of the Schoenfeld residuals is zero.
- Note that r_i is a p dimensional vector.

Schoenfeld residuals

- The Schoenfeld residuals are equal to zero for all censored subjects and thus contain no information about the fit of the model.
- The Schoenfeld residuals can be computed as follows:

```
> res <- resid(fit, type = "schoenfeld")
> head(res)
      age      gender      bmi
6 -11.495891 -0.4944845  6.90753182
6  -4.495891 -0.4944845  3.50526182
14 -12.564282  0.5007772 -0.01536823
44  -5.675718 -0.4947812 -1.37175453
62   9.261320  0.4997300 -9.52449179
89  10.644817  0.5204231 -6.06048648

> dim(res)
[1] 51  3

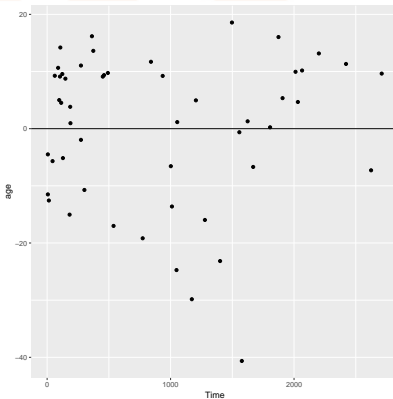
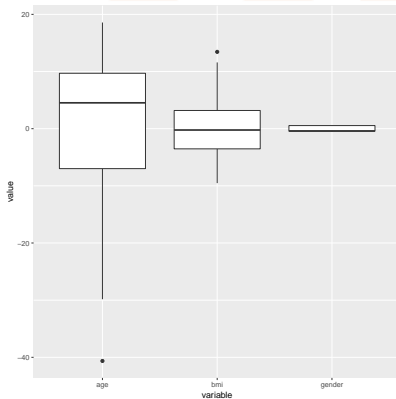
> fit$nevent
[1] 51

> colSums(res)
      age      gender      bmi
2.738574e-10  3.922418e-12 -2.151168e-12
```

Schoenfeld residuals

- Plotting these residuals versus the covariate will yield a pattern of points centered around 0.

```
> res <- as.tibble(res) %>% mutate(Time = as.numeric(rownames(res)))
> ggplot(data = reshape2::melt(res, id = "Time")) +
+   geom_boxplot(aes(x = variable, y = value))
> qplot(Time, age, data = res) + geom_hline(yintercept = 0)
```



Scaled Schoenfeld residuals

- Grambsch and Therneau (1994) proposed scaling each residual by an estimate of its variance.
- The scaled residual can be approximated as follows:

$$r_i^* = r_i \left[\widehat{\text{Var}}(r_i) \right]^{-1} \approx r_i \cdot \text{Var}(\hat{\beta}) \cdot n,$$

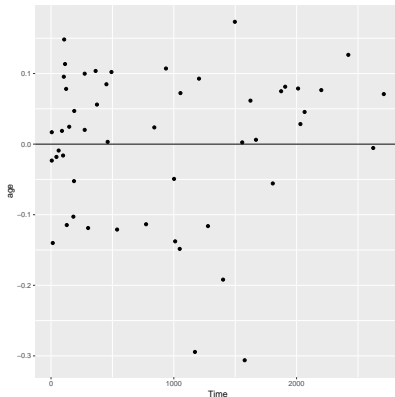
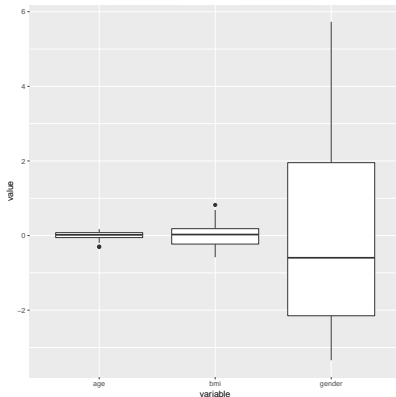
where n is the number of death.

```
> stdres <- as.tibble(resid(fit, type = "schoenfeld") %*% fit$var * fit$nevent)
> names(stdres) <- c("age", "bmi", "gender")
> stdres$Time <- res$Time
> stdres %>% print(n = 7)
# A tibble: 51 x 4
   age    bmi  gender  Time
  <dbl> <dbl>   <dbl> <dbl>
1 -0.0233 -1.46  0.394     6
2  0.0169 -2.00  0.205     6
3 -0.140   3.30 -0.0614    14
4 -0.0181 -1.97 -0.125    44
5 -0.00910 1.63 -0.579    62
6  0.0188  1.66 -0.342    89
7 -0.0161  2.17 -0.205    98
# ... with 44 more rows
```

Scaled Schoenfeld residuals

- The scaled Schoenfeld residuals can be plotted with

```
> qplot(Time, age, data = stdres) + geom_hline(yintercept = 0)
> ggplot(data = reshape2::melt(stdres, id = "Time")) +
+   geom_boxplot(aes(x = variable, y = value))
```



Scaled Schoenfeld residuals

- Recall that the proportional hazard assumption implies

$$\widehat{\text{HR}} = e^{(X_i - X_j)\hat{\beta}}, \quad (1)$$

that is free of time.

- There are number of ways to yield non-proportional hazard ratios.
- Grambsch and Therneau (1994) considered an alternative to allow covariate effect to change over time by

$$\beta(t) = \beta + \rho g(t) \approx \beta + E(r^*),$$

where $g(t)$ is a function of time.

- This suggests that plotting $\beta(t)$ against time gives a visual assessment of the proportional hazard assumption.

Scaled Schoenfeld residuals

- The $\beta(t)$ in (1) can be conveniently computed with the `cox.zph` function.

```
> zph <- cox.zph(fit)
> head(zph$y)
```

	age	gender	bmi
6	0.01383649	-1.321312	0.3232913
6	0.05399122	-1.860223	0.1338133
14	-0.10294092	3.445105	-0.1322638
44	0.01899822	-1.821987	-0.1959143
62	0.02803035	1.770130	-0.6501345
89	0.05590471	1.801303	-0.4129191

- This can be confirmed with

```
> head(stdres$age + coef(fit)["age"])
[1] 0.01383649 0.05399122 -0.10294092 0.01899822 0.02803035 0.05590471
```

Scaled Schoenfeld residuals

- Several functions of $g(t)$ have been suggested:

```
> args(cox.zph)
function (fit, transform = "km", global = TRUE)
NULL
```

Usage

```
cox.zph(fit, transform="km", global=TRUE)
```

Arguments

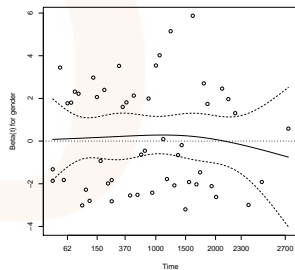
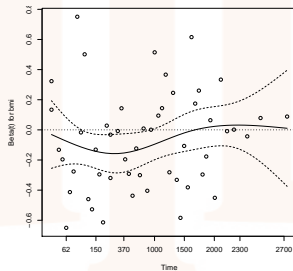
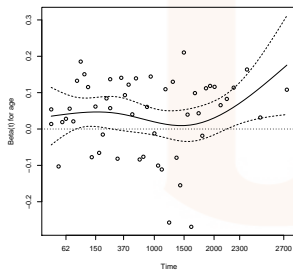
- | | |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| fit | the result of fitting a Cox regression model, using the <code>coxph</code> function. |
| transform | a character string specifying how the survival times should be transformed before the test is performed. Possible values are "km", "rank", "identity" or a function of one argument. |
| global | should a global chi-square test be done, in addition to the per-variable tests. |

- The hypothesis of $H_0 : \rho = 0$ can be tested via the partial likelihood ratio test, score test, or Wald test.

Scaled Schoenfeld residuals

- We can plot the residual with `plot`;

```
> plot(zph[1]); plot(zph[2]); plot(zph[3])
```



Scaled Schoenfeld residuals

- The plots also include a 95% confidence band for the smooth curve.
- For our dataset, there is no warning signs of violation of the proportional assumption.

```
> zph
```

	rho	chisq	p
age	0.0536	0.1986	0.656
gender	-0.0325	0.0671	0.796
bmi	0.1420	1.5793	0.209
GLOBAL	NA	1.6627	0.645

Martingale residuals

- The next collection of residuals comes from the counting process formulation of a time-to-event regression model.
- The basic formation is

$$M_i(t) = N_i(t) - \Lambda_i(t).$$

- Under our setting, we have the martingale residuals

$$m_i = \Delta_i - \hat{H}_0(t_i)e^{X_i' \hat{\beta}}.$$

- The residual is the difference between the observed value (Δ_i) of the censoring indicator and its expected value.
- Also called the Cox-Snell or modified Cox-Snell residual.

Martingale residuals

- The martingale residuals sum to 0, but range from $-\infty$ to 1.
- The sum of squares of martingale residuals can not be used as a measure of goodness of fit.
- Transformations to achieve a more symmetric distribution are helpful.
- One such transformation is the *deviance residuals*:

$$d_i = \text{sign}(m_i) \sqrt{-2 \cdot [m_i + \Delta_i \log(\Delta - m_i)]}.$$

- These deviances are symmetrically distributed with expected value 0 (if the fitted model is correct).
- The sum of squares of these residuals is the value of the likelihood ratio test in generalized linear model theory.

Reference

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.

