STAT 6390: Analysis of Survival Data

Textbook coverage: Chapter 3

Steven Chiou

Department of Mathematical Sciences,
University of Texas at Dallas

Cox proportional hazards model

- The Cox model is expressed by the hazard function.
- The hazard function can be (loosely) interpreted as the risk of dying at time t.
- The Cox model has the form:

$$h(t) = h_0(t) \cdot \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots \beta_p x_p\},$$

  where
  - t is the survival time.
  - $\{x_1, \dots, x_p\}$ is a set of p covariates.
  - $\{\beta_1, \dots, \beta_p\}$ is the regression parameters; effect of covariates.
  - $h_0(t)$ is the baseline hazard. It is the value of the hazard when all x's are 0.
  - No need to specify an "intercept" term as it gets absorb to $h_0(t)$.

Cox proportional hazards model

- The quantity e^{β_i} is interpreted as the hazard ratio (HR).
  - $\beta_i > 0 \rightarrow$ HR $> 1 \rightarrow$ hazard increases $\rightarrow$ survival time decreases.
  - $\beta_i = 0 \rightarrow$ HR $= 1 \rightarrow$ no change in hazard $\rightarrow$ no change in survival time.
  - $\beta_i < 0 \rightarrow$ HR $< 1 \rightarrow$ hazard decreases $\rightarrow$ survival time increases.
- HR (and hazard) is negatively associated with the length of survival.
- The Cox model assumes the hazard curves among different patients should be proportional and cannot cross.

Fitting the Cox model in R

- We have used `coxph` to compute the Nelson–Aalen estimator.
- The usage of `coxph` is similar to that of `survreg`.

```
> fm <- Surv(lenfol, fstat) ~ (age + gender)^2 + bmi
> fit.cox <- coxph(fm, data = whas100)
> fit.aft <- survreg(fm, data = whas100)
```

- The coefficients are in opposite directions.

```
> coef(fit.cox)
        age      gender         bmi  age:gender
 0.05398825  4.16146915 -0.08737778 -0.05368624
> coef(fit.aft)
(Intercept)         age      gender         bmi  age:gender
 9.87268335 -0.06394027 -4.68945202  0.10551029  0.05915426
```

Fitting the Cox model in R

- The `summary` gives:

```
> summary(fit.cox)
Call:
coxph(formula = fm, data = whas100)

  n= 100, number of events= 51

                coef exp(coef) se(coef)      z Pr(>|z|)
age          0.05399   1.05547  0.01553  3.477 0.000507 ***
gender       4.16147  64.16572  1.82333  2.282 0.022469 *
bmi         -0.08738   0.91633  0.03724 -2.346 0.018970 *
age:gender -0.05369   0.94773  0.02438 -2.202 0.027652 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

           exp(coef) exp(-coef) lower .95 upper .95
age           1.0555    0.94744    1.0238    1.0881
gender       64.1657    0.01558    1.8000 2287.3213
bmi           0.9163    1.09131    0.8518    0.9857
age:gender    0.9477    1.05515    0.9035    0.9941

Concordance= 0.696  (se = 0.043 )
Rsquare= 0.229   (max possible= 0.985 )
Likelihood ratio test= 26.01  on 4 df,   p=3e-05
Wald test            = 21.63  on 4 df,   p=2e-04
Score (logrank) test = 23.48  on 4 df,   p=1e-04
```

Fitting the Cox model in R

- As in the result from `survreg`, the `z` column gives the *Wald statistic* computed from `coef/se(coef)`.
- The Wald statistic tests the hypothesis $H_0 : \beta = 0$.
- In this case, all covariate effects are significant at $\alpha = 0.05$.
- The $\hat{\beta}$ for `age` is positive indicating older patients have higher risk of death.
- The $\hat{\beta}$ for `bmi` is negative indicating patients with higher BMI have lower risk of death.