Ch. 3: The Cox Proportional Hazards Model and Its Characteristics

Introduction

- Show similarity between Cox proportional hazards model and linear regression
- Cox model, properties, assumptions, etc.
- Describe how and why to use "age as the time scale" rather than "time on follow-up" as the outcome variable

1/21/2023

I. A Computer Example Using the Cox PH Model

Example of the analysis of remission time data

Leukemia Remission Data

Group 1 (n=21)		Group 2 (n=21)	
t (weeks)	log WBC	t (weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
⋮	⋮	⋮	⋮

Compare the survival experience of the two groups adjusting for the possible confounding and/or interaction effects of log WBC

explanatory ↗

T: survival time

X_1 = group status = E

X_2 = log WBC (confusing?)

If interested in confounding effect, would need a third variable X_1 & X_2.

How to use computer printout to evaluate the possible effect of treatment status on remission time adjusted for potential confounding/interaction effects of the covariate.

hazard ratio = e^{coef}

- e^{coef} gives an estimated hazard ratio for the effect of each variable adjusted for the other variables in a model.

Maximum likelihood estimation used to estimate the coefficients.

[Okay, I read this section. It's very good for telling me how to carry out an analysis of models.

II. The Formula for the Cox PH Model

$$h(t, X) = h_0(t) \times \exp\left[\sum_{i=1}^{p} \beta_i X_i\right]$$

$X = (X_1, \ldots, X_p)$

$h_0(t)$: baseline hazard — a time independent term.

time-independent X's

Time-dependent X's require the extended Cox model.

Can assume not to change once measured.

Cox model is semiparametric because the form of $h_0(t)$ is unspecified.

III. Why the Cox PH Model is Popular

GOOD

- Cox PH model is "robust" in the sense that it closely approximates the correct parametric model.
- Parametric models are preferred only if we're sure we have the correct model.
- Hence Cox is a safe choice.

- Always non-negative.

- To calculate hazard ratios, the baseline gets cancelled out, so you only need to estimate the coefficients.

- $h(t,X)$ and $S(t,X)$ can be estimated without specifying $h_0(t)$.

From j-to-i