Cox Proportional Hazards Models

The purpose of Cox proportional hazards regression is to model the hazard of the event (e.g. the instantaneous hazard or rate of death at time t, conditional on surviving until time *t* or later), **which we will denote as h(t)**, as a function of a set of predictors X. We express the natural log of the hazard as a linear combination of predictors and the natural log of the baseline hazard $h_0(t)$:

$$\ln[h(t)] = \ln\,[h_0(t)] + \ \beta_1 X_1 \ldots + \ \beta_k X_k$$

Note that the intercept term we usually see in a regression model is replaced by the baseline hazard: $h_0(t)$.The baseline hazard represents the rate of the event occurring at the baseline level of the covariates (all X=0). In the Cox model the baseline hazard $h_0(t)$ is not estimated.

A ***stratified*** Cox model allows for (but does not estimate) separate baseline hazards for each level of the stratification variable while estimating a single effect for each covariate; this is useful if the assumption of proportionality of the hazards are violated by a variable that you can stratify upon. The idea is that the assumption may not hold in the pooled dataset, but the assumption *does* hold within levels of the stratification variable.

The measure of effect estimated from proportional hazards regression is the **hazard ratio**: a relative measure of how fast the event occurs in one group to how fast the event occurs in another group.

To estimate the hazard ratio (HR) for $X_1 = 1$ vs. $X_1 = 0$

$$\ln(HR) = \ln\left[\frac{h(t \mid X=1)}{h(t \mid X=0)}\right] = \ln\,[h(t\ |X = 1)] - \ \ln\,[h(t\ |X = 1)]$$

$$\ln(HR) = [\ln(\beta_0(t)) + \ \beta_1(1)] - \ [\ln(\beta_0(t)) + \ \beta_1(0)] = \ \boldsymbol{\beta_1}$$

$$HR = \exp\,(\beta_1)$$

$$95\%\ CI = \exp\,[\beta_1\ \pm (1.96 * SE(\beta_1))]$$