The stratified Cox Procedure

1 The stratified Cox Model (SC)

Introducing the variables

- Denote by $X_1, ..., X_p$ for $p \in \mathbb{N}$ the variables satisfying the PH assumption.
- Denote by $Z_1, ..., Z_k$ for $k \in \mathbb{N}$ the variables not satisfying the PH assumption.
- We assume no interaction between the Z_j's and the X_i's for all i, j.
- Define a single new variable Z^*
  1. Categorize each Z_i
  2. Form combinations of categories (strata)
  3. The strata are the categories of Z^*
- k^* denotes the number of combinations, i.e. Z^* has k^* categories.

The general SC model

The general Cox model is given by

$$h_g(t, X) = h_{0g} \exp[\beta_1 X_1 + ... + \beta_p X_p] \tag{1}$$

with $g = 1, ..., k^*$ strata defined from Z^*

A few observations

- The same coefficients: for each stratum $\beta_1, ..., \beta_p$.
- BUT: the baseline hazard functions $h_{0g}(t)$ may be different for each stratum.
- $X_1, ..., X_p$ directly included in the model, but Z^* appears only through the different baseline hazard functions.

In R:

```
cox <- coxph(Surv(time,status) ~ X_1 + X_2 + strata(Z_1,Z_2),
       data=yourdataname, method="breslow")
```

The estimation procedure

1. We calculate the likelihood function for each strata

Strata:	$1, \ldots, k^*$
Likelihood:	$L_1, \ldots, L_{k^*}$
Hazard:	$h_1, \ldots, h_{k^*}$

2. We multiply all the likelihood function we have obtained this way.

$$L = L_1 \times \ldots \times L_{k^*}$$

3. We obtain the estimates for coefficients $\beta_1, \ldots, \beta_p$ by maximizing the log likelihood function L.

2 No interaction assumption

The interaction model is given by:

$$h_g(t) = h_{0g}(t) \exp[\beta_{1g} X_1 + ... + \beta_{pg} X_p]$$

The difference to the no-interaction model is that the coefficients β depend on the strata.
An alternative formulation:

$$h_g(t) = h_{g0}(t)[\beta_1^* X_1 + ... + \beta_p^* X_p + \sum_{j=1}^{k^*-1} \sum_{i=1}^{p} \beta_{ij}^* X_i Z_j^*]$$

where the β^* do not involve g.

In R:

```
cox_2 <- coxph(Surv(time,status) ~ X_1 * strata(Z_1,Z_2,Z_3) +
     X_2 * strata(Z_1,Z_2,Z_3), data=yourdataname, method="breslow")
```

3 Comparing: Likelihood, Wald & score test

Fisher's score vector $U(\theta)$.
The first derivative of the log-likelihood function is called *(Fisher's) score vector*,

$$U(\theta) = \frac{\partial}{\partial \theta} \ln L(\theta, X)$$

Remeber, that the derivative wrt a vector is a vector, with entires:

$$U_i(\theta) = \tfrac{\partial}{\partial \theta_i} \ln(L(\theta, X)).$$

- Still a function in θ.
- If the log-likelihood is concave we have:

$$U(\hat{\theta}) = (0, \dots, 0)$$

The Fischer information matrix $i(\theta)$.
The Fischer information matrix is a $p \times p$ matrix defined by:

$$i(\theta) = E_\theta \left[U(\theta)^T U(\theta) \right] = \left\{ -E_\theta \left[\frac{\delta^2}{\delta\theta_i\theta_j} \ln L(\theta, X) \right] \right\}_{\substack{i=1,\dots,p \\ j=1,\dots,p}}$$

Mostly this is to hard to compute, thus we estimate i with:

$$\hat{i}(\theta) =: I_{i,j}(\theta) := -\frac{\delta^2}{\delta\theta_i\delta\theta_j} \ln L(\theta, X), \quad i = 1, \dots, p, j = 1, \dots, p$$

Definition of the different tests

We already know the *likelihood ratio test* statistic:

$$LR := -2 \left(\ln L(\theta_0, X) - \ln L(\hat{\theta}, X) \right) \to \chi^2_p$$

The *Wald test* statistic is defined as:

$$W := (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \to \chi^2_p$$

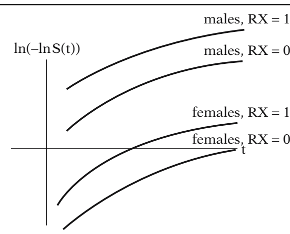
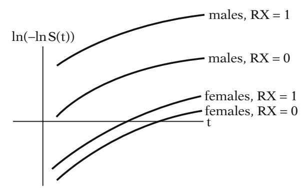
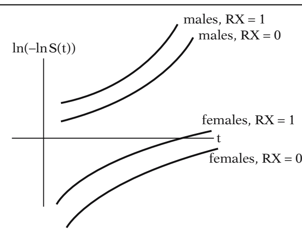
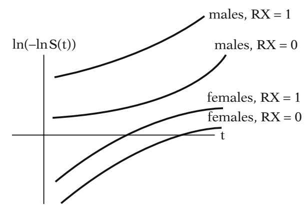
And the *score test* uses the efficient score vector:

$$S := U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \to \chi^2_p$$

Note: Score test does *not depend of* $\hat{\theta}$.

4 Graphical Comparison of the tests so far

Overview of the examples

<table>
<tr><th></th><th>No interaction</th><th>Interaction</th></tr>
<tr><td>PH Assumption</td><td>
</td><td>
</td></tr>
<tr><td>Stratified by SEX</td><td>
</td><td>
</td></tr>
</table>