Cox PH model

Model building

Modeling goals

- Valid effect estimation of risk factors
- Make good prediction
- Parsimonious

Model Building

1. Establish the main scientific research question.
2. Check data quality
   - Univariate summaries for each variable
   - Look for
     - Errors in coding (e.g. 9999 used for missing)
     - Outliers
     - Missing values

Model building

3. Pairwise associations
   - Collinearity: Strong correlation between any two covariates
4. Bivariate analysis to exam the marginal association between each covariate and survival time.
   - For categorical covariates, plot the K-M estimators of group specific survival functions and test the group difference using Log-rank test .
   - For continuous covariates, break into quantiles or other clinically meaningful groups and do similar plot as the categorical covariates.

Model Building

5. Regression model
   - Determine variables (main effects) to be included
   - After the variables are selected, determine the functional form for the continuous covariates
     - Suggested by the clinical consideration
     - Suggested by data (martingale residues)
   - Determine any interactions
   - Assess the final model using the residual diagnosis methods.

Model Building

6. Presentation of model(s):
   - Report and interpret coefficients and SE's
   - Graphical display

Model selection

- The method for variable selection for Cox PH model are essentially the same as in linear regression models.
- In practice, model selection is conducted through a combination of
  - Prior knowledge
  - Automatic variable selection procedures
    - Forward selection
    - Backward selection
    - Stepwise selection

Backward selection

- Start with all the predictors in the model
- Remove the predictor with highest p-value greater than α_{RE}
- Refit the model and go to step 2
- Stop when all p-values are less than α_{RE}
- α_{RE} is the threshold to remove predictors from the model.
- Often a cut-off α_{RE} with value 15-20% is used.

Forward selection

- Start with no predictors in the model
  For all predictors not in the model, check their p-value if they are added to the model. Choose the one with lowest p-value less than αE
- Continue until no new predictors can be added.
- αE is the threshold to select predictors to enter the model.
- Often a cut-off αE with value 20-25% is used.

Stepwise procedure

- It is a combination of backward elimination and forward selection.
- At each step, a variable may be added or removed.
- Unlike the forward procedure, it allows one to delete variables that enter the model early.
- Unlike the backward procedure, it allows add the early removed variables back into the model.

Summary

- Model building strategies
- Variable selection in Cox PH model