CPH 576C, Applied Biostatistics Analysis Dr Melanie Bell

Project 1

Using the SUPPORT study data, downloadable from D2L, develop a regression model relating the set of predictors (age, sex, dzgroup, num.co, scoma, race, meanbp, hrt, temp, pafi, alb) to total cost.

Some considerations:

1. Delete any observations with zero cost from all analyses

2. For this assignment, do not impute any missing data

3. Determine whether to model costs or log costs. Justify your conclusion and use that transformation in all later steps.

4. Screen the independent variables to determine how they interrelate and whether any of the predictors are redundant

5. Fit a model to predict cost (or a transformation of it) using all predictors. For continuous predictors assume a smooth relationship but allow it to be nonlinear. Choose the complexity to allow for each predictor’s shape (i.e., degrees of freedom or knots).

6. Assess the assumptions of your fitted model.

7. Interpret the results of your final model.