1. The data of a study of 90 patients diagnosed with cancer of the larynx in the 70s at a Dutch hospital is given in the file larynx.txt on www.stat.nus.edu.sg/~stachenz. The data consists of the times between first treatment and either death or the end of the study. Patients were classified into one of four stages of their disease. This data has been analyzed in Question 3, Tutorial 6.

(i) Fit a proportional hazard regression model with three covariates for stage of disease.

(ii) Determine if adding the patient’s age into the model is appropriate using a martingale residual plot based on a Cox model adjusted for disease stage. If age should not enter the model as a linear term suggest a functional form for age.

(iii) Repeat part (ii) for the covariate year of diagnostics.

(iv) Fit a model with the factor stage of disease and a linear term for age. Perform a general examination of this model using a Cox-Snell residual.

2. The data of 863 kidney transplant patients is given in the file kidtran.txt on www.stat.nus.edu.sg/~stachenz. The data contains the following variables: Time to death or on-study time, Death indicator (0=alive, 1=dead), Gender (1=male, 2=female), Race (1=white, 2=black), Age in years. This data has been analyzed in Question 2, Tutorial 6.

(i) Fit a proportional hazard regression model with covariates: gender, race, and gender by race interaction.

(ii) Check this data for possible outliers by making an appropriate plot of the deviance residuals.

(iii) For each of the three covariates in this model find the four most influential observations on the estimates of the regression coefficients. Explain why these observations are so influential.