Exercise Sheet 6 — Survival analysis

Problem 1. German Breast Cancer Study Group 2
The data GBSG2 are available in package **TH.data**. The study investigated the effects of hormonal treatment with Tamoxifen in women suffering from node-positive breast cancer in a randomized clinical trial. The outcome of interest is recurrence-free survival (the time until recurrence of the cancer or death, whatever comes first). The data contain the following information for 686 women

- `horTh` hormonal therapy, a factor at two levels 'no' and 'yes'.
- `age` of the patients in years.
- `menostat` menopausal status, a factor at two levels 'pre' (premenopausal) and 'post' (postmenopausal).
- `tsize` tumor size (in mm).
- `tgrade` tumor grade, a ordered factor at levels 'I < II < III'.
- `pnodes` number of positive nodes.
- `progrec` progesterone receptor (in fmol).
- `estrec` estrogen receptor (in fmol).
- `time` recurrence free survival time (in days).
- `cens` censoring indicator (0- censored, 1- event).

(a) Find a good model for the data

(b) Diagnose the model. Do some variables display time-dependent effects and/or non-linear relationships? If so can you fix it?

Problem 2. Stanford heart transplant program
Consider the data set `heart.csv` that contain information about patients on the waiting list for a heart transplant. The goal is to assess whether a heart transplant increases the chances of survival. The data set contain the following variables:

- `accept.dt`: acceptance into program (waiting list)
- `tx.date`: transplant date
- `fu.date`: end of followup
- `fustat`: dead (1) or alive (0)
- `transplant`: transplant indicator

As patients have to wait before getting a transplant, `transplant` is a time-dependent variable.

(a) Transform the data in such a way that you can include `transplant` as a time-dependent variable in a Cox model.

(b) Fit the model and conclude

(c) Analyse the transplant variable as if it was known at baseline, thus ignoring that transplant is a time-dependent variable. Interpret.