

## Dataset and variables

Dataset with 48 patients suffering from multiple myeloma.

This dataset is taken from Collett, D. (2000). Modelling survival data in medical research (1st ed.). Chapman & Hall/CRC.

The dataset consists of the following variables:

PATIENT: Patient's ID

TIME: survival time

STATUS: 1: death, 0: censored

AGE: Patient's age in years

SEX: 1=males, 2=females

BUN: Levels of blood urea nitrogen

CA: serum calcium

HB: hemoglobin

PC: Percentage of plasma cells in the bone marrow

BJ: indicator variable denoting whether or not Bence-Jones protein was present in the urine

The main aim is to investigate the effect of the risk factors BUN, CA, HB, PC and BJ on the survival time of the multiple myeloma patients.

## Questions

- a) Run a univariate Cox proportional hazards regression model against BUN . Interpret the result.
- b) Run univariate Cox proportional hazards regression models for the rest of the variables in the data set.
- c) Run a multiple Cox proportional hazards regression model against all explanatory variables with  $p < 0.2$  in the univariable analysis. Interpret the results.
- d) Include AGE in #c model. compare this model to #c. Interpret the results.
- e) Run a multiple Cox model with the explanatory variables BUN, HB and SEX. Compare this model to #c. Interpret the results.
- f) Run the model diagnostics for #c. Interpret the results.
- g) Run an Exponential model with BUN, HB as explanatory variables. Interpret the results. Compare it with a respective Weibull model.