1. (Theoretical; an example of a midterm problem.) For the parametrization of the Weibull distribution “used in medical statistics”,
\[ \alpha \vartheta^\alpha t^{\alpha-1} e^{-\vartheta t^\alpha} \] for \( t \geq 0 \),
what is estimated, in terms of the parameters above, by the accelerated life model in place of the
(a) intercept  (b) scale?

2. (Theoretical.) Given a covariate \( Z \), suppose that the log survival time \( Y \) follows a linear model with a logistic error distribution, that is, \( Y = \log X = \mu + \beta Z + \sigma W \) where the density of \( W \) is given by
\[ f(w) = \frac{e^w}{(1 + e^w)^2}. \]
(a) For an individual with covariate \( Z \), find the conditional survival function \( S(x|Z) \) of the survival time \( X \) given \( Z \).
(b) The odds that an individual will die prior to time \( x \) are expressed by \( (1 - S(x|Z))/S(x|Z) \). Compute the odds of death prior to time \( x \) for this model.
(c) Consider two individuals with different covariate values. Show that, for any time \( x \), the ratio of their odds of death is independent of \( x \).

3. (Theoretical; almost an example of a midterm problem should the last part did not require some versatility to complete it under time pressure). A slightly simplified version of the marijuana smoking data is as follows: we have six respondents. The first one is 16, and recalls he started smoking already, but doesn’t recall when; the second is 14, and didn’t start smoking yet. The third recalls starting smoking sometime between the age of 13 and 17, but can’t localize it more exactly; finally, the last three started smoking at the age of 15.

For the lack of better alternatives, and also for the sake of mathematical simplicity, we assume that the survival time has uniform distribution on the interval \([12, \vartheta]\), where \( \vartheta \) is an unknown parameter.
(a) Write the likelihood for \( \vartheta \) on the basis of the data as given above.
(b) Find the maximum likelihood estimate for \( \vartheta \).

4. (Data analysis.) Following the treatment of leukaemia, patients often undergo a bone marrow transplant in order to help bring their blood cells back to a normal level. A potential fatal side effect of this is graft-versus-host disease, in which the transplanted cells attack the host cells. In a study described by Bagot et al (1988), 37 patients who were in complete remission from acute myeloid leukemia (AML) or acute lymphocytic leukaemia (ALL), or in the chronic phase of chronic myeloid leukaemia (CML), received a non-depleted allogeneic bone marrow transplant. The age of the bone marrow donor, as whether or not the donor had previously been pregnant, was recorded, together with the age of the recipient, their type of leukaemia, an index of mixed epidermal lymphocyte reactions, and whether or not the recipient developed graft-versus-host disease. The variables in this dataset are as follows:

- **patient**: Patient number (1–37)
- **time**: Survival time in days
- **status**: Status of patient (0 = alive, 1 = dead)
- **age**: Age of patient in years
- **dage**: Age of donor in years
- **yype**: Type of leukaemia (1 = AML, 2 = ALL, 3 = CML)
- **preg**: Donor pregnancy (0 = no, 1 = yes)
- **index**: Index of cell-lymphocyte reactions
- **gvhd**: Graft-versus-host disease (0 = no, 1 = yes)

Using a Weibull accelerated failure time model, investigate the dependence of the survival times on the prognostic variables. Estimate and plot the baseline survival function. Fit log-logistic and lognormal models that contain the same explanatory variables, and estimate the baseline survival function under these models. Comment on which parametric model is the most appropriate, and further examine the adequacy of this model using model-checking diagnostics.

The code is not important this time, but summarize all findings, commentaries and picture in one printed document not exceeding two pages. Feel free to be creative here. (The dataset is to be found on the course webpage.)