1. The following table gives survival data from 30 patients with AML. Two possible prognostic factors are considered:

$$x_1 = \begin{cases} 1 & \text{if patient} \geq 50 \text{ years old} \\ 0 & \text{otherwise} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{if cellularity of mallow clot section is } 100\% \\ 0 & \text{otherwise} \end{cases}$$

Table. Survival Times and Data of Two Possible
Prognostic Factors and 30 AML Patients

Survival Time	x_1	x_2	Survival Time	x_1	x_2
18	0	0	8	1	0
9	0	1	2	1	1
28+	0	0	26+	1	0
31	0	1	10	1	1
39+	0	1	4	1	0
19+	0	1	3	1	0
45+	0	1	4	1	0
6	0	1	18	1	1
8	0	1	8	1	1
15	0	1	3	1	1
23	0	0	14	1	1
28+	0	0	3	1	0
7	0	1	13	1	1
12	1	0	13	1	1
9	1	0	35+	1	0

Assuming cox proportional hazard model and use the exact likelihood to handle the tie observations. Test if the prognostic factors are significant. Compare the results with the function "survdiff".

2. You are given a data set in the file "ass4.csv". It contains the following variables:

lstay	Length of stay of a resident
age	Age of a resident
trt	Nursing home assignment (1: receive treatment,0: control)
gender	Gender (1:male,0:female)
marstat	Marital status (1: married,0: not married)
hlstat	Health status (2: second best, 5: worst)
cens	Censoring indicator (1:censored, 0: discharged)

Use 123457 as the seed number to generate a sample of size 1000 and use this sample to model the home nursing home duration times, measured by days, as a function of patient characteristics.

(a) Test if the patient characteristics have no impact on the home duration times. State all three test values (Wald, likelihood ratio and score).

(b) Based on the individual z-scores, can we remove all covariates which p-values are greater than 0.05? State the null hypothesis we want to test. Give all three tests (Wald, likelihood ratio and score) and their p-values.

(c) Based on the simplified model, give a 95% confidence interval on the hazard ratio of the patient with status=4 compared with those with status=5.