The course site contains a dataset of times-to-death for laryngeal cancer patients at a single hospital. Times recorded are the interval (in years) between first treatment and either death (i.e. an event) or the end of the study (i.e. administrative censoring, which you should assume is non-informative). The variables are:

* Stage of disease at study entry (1/2/3/4)
* Time to death or censoring, in years
* Status: a death indicator (0=alive i.e. censored, 1=dead)

1. Summarize the number of *observed* deaths and the number of subjects by disease stage in a 2 × 4 table, and describe any obvious relationships.
2. Plot a Kaplan-Meier curve for all 90 subjects. Describe what the estimate of survival at year 6 *and* its confidence interval tell us about survival in the population of all laryngeal cancer patients.
3. Plot Kaplan-Meier curves for the 4 groups defined by stage of disease=1/2/3/4. Briefly summarize what this plot suggests about survival in these groups.
4. Use the log-rank test of survival in two groups; those with stage = 1 or 2 vs stage = 3 or 4. State precisely what the null hypothesis is, give the test result, and briefly interpret it.