# 1 Exercise 1

Load the survival package in R, and load the `aml` dataset. This dataset comprises survival times for patients with Acute Myelogenous Leukemia, comparing survival times for patients in the treatment arm who were maintained on chemotherapy vs. the standard of care for patients who were not maintained on chemotherapy.

1. Write a function that takes a vector of event times, and a vector of associated censoring indicators and returns the nonparametric estimators for hazard rate.

2. Use this function's output to plot the Kaplan-Meier estimator for the survival function under the null distribution for the `aml` dataset (remember, this is a step-function with steps at non-censoring event times)

3. Write a function to output the upper and lower log-log confidence bounds for $\hat{S}^{\mathrm{KM}}(t)$ for a given confidence level

4. Overlay the survival function and the 95% bounds on the same plot

5. Show that your plot matches plot made by `survival`:
   `plot(survfit(Survtime, status)~1, data=aml,conf.type="log-log"))`

6. Write a function that calculates the log-rank test-statistic $Z_1(\tau)$ where the $d_{i1}$ observations come from the maintenance arm of the trial, and $\tau$ is the maximum observed event (censored or uncensored) time in either arm. You can reuse the function you wrote in part 1 as a helper function for this part if you'd like.

7. Write a function that calculates the standard error of the log-rank test statistic, and, using the limiting normal distribution, determine the level-0.05 rejection region and determine if $Z_1(\tau)$ is in the rejection region or not (i.e. do you reject the null hypothesis of $\lambda_1(t) = \lambda_2(t) \forall t \in [0, \tau]$).

8. Verify that your code reproduces the results from `survival`:
   `survdiff(Surv(time, status) ~ x, data=aml)`

Submit all your code and plots for this exercise.

# 2 Exercise 2

Use the approach we outlined in class to derive the logit point-wise confidence interval for $\hat{S}^{\mathrm{KM}}(t)$. Specifically, construct a confidence set for $S(t)$ by first generating $C^{g(\hat{S}^{\mathrm{KM}}(t))}$ such that:

$$P(g(S(t)) \in C^{g(\hat{S}^{\mathrm{KM}}(t))}) \geq 1 - \alpha$$

and transforming back to $\hat{S}^{\mathrm{KM}}(t)$ using $g^{-1}$:

$$P(S(t) \in g^{-1}(C^{g(\hat{S}^{\mathrm{KM}}(t))})) \geq 1 - \alpha.$$

Let $g : [0, 1] \to \mathbb{R}$, and $g(x) = \log\left(\frac{x}{1-x}\right)$. It will involve expressing the logit of $\hat{S}^{\mathrm{KM}}(t)$ in terms of $\log \hat{S}^{\mathrm{KM}}(t)$ and using the framework we used to derive the log-log pointwise confidence intervals for the Kaplan-Meier estimator. You may take for granted that as $n \to \infty$

$$\frac{\hat{S}^{\mathrm{KM}}(t) - S(t)}{\sqrt{\mathrm{Var}(\hat{S}^{\mathrm{KM}}(t))}} \xrightarrow{d} \mathrm{Normal}(0, 1).$$