HW #2 due on 10/29

**Data: KM 1.5 Times to Death for a Breast-Cancer Trial**

**Variables**

Time (months): time to death

Death: 0 alive; 1 dead

IM (immune-peroxidase status): 1 negative; 2 positive

**1. (15 pts) Exponential distribution:** Assume time to death follows an exponential distribution with hazard rate for negative IM and for positive IM.

1. (5%) write down the likelihood function based on KM 1.5 data.
2. (5%) Find the maximum likelihood estimates (MLE) for and .
3. (5%) Based on the MLE, derive the survival estimate at 3 years (i.e. 36 months) for both negative and positive IM.

**2. (45 pts) Nonparametric estimation**: Perform Kaplan-Meier and Nelson Aalen estimators on KM 1.5 by IM status using SAS and then use the output to answer the following questions.

1. Treat death as count data and fit a Poisson regression model to death with log(time) as the offset and IM (use negative as the reference) as the covariate.

(5%) How many deaths for negative and positive IM, respectively?

1. (5%) Based on results in Q4, find the survival estimate and the associated 95% CI at 3 years (i.e. 36 months) for both negative and positive IM.
2. (10%) Find the median survival time, and the associated 95% confidence interval for negative and positive IM, respectively. Explain why the median survival time for negative IM is undefined.
3. (5%) Perform Nelson-Aalen estimator on KM 1.5 by IM status using SAS/R.
4. (10%) Find a smooth (Epanechnikov kernel) hazard function estimate for negative and positive IM, respectively, using SAS/R without specifying the bandwidth (the program will choose the optimal bandwidth) and then interpret the estimate. Based on the hazard function estimates, would exponential distribution be a good choice for KM 1.5?
5. (10%) Based on b-e, write a short paragraph to summarize the survival data for negative and positive IM, respectively.

3. (40 pts) For the following small data set of survival time: 3, 4, 5+, 6, 6+, 8+, 11, 14, 15, 16+, where “+" means a right censored survival time, do the following:

* + 1. Find the Kaplan-Meier estimate of the survival function and its variance.
  1. Use the above Kaplan-Meier estimate to get an estimate and its variance of the cumulative hazard function.
  2. Find the Nelson-Aalen estimate of the cumulative hazard function and its variance.
  3. Find an estimate and its variance of the survival function using the Nelson-Aalen estimate you got in (c).