HW6- Advanced Data Analysis

1. (10pt) A random variable T is said to have a Weibull distribution is its survival function is give by
$$S(t) = e^{-(\alpha t)^\beta}$$
where $\alpha > 0$ and $\beta > 0$.

   (a) (2pt) Find the density, $f_T(t)$ of T

   (b) (2pt) Find the hazard function $\lambda(t)$ of T

   (c) (2pt) Show that
   $$\log(-\log(S(t))) = \beta\log(\alpha) + \beta\log(t)$$
   Based on this, describe a graphical method for checking whether or not the data is from a Weibull distribution.

   (d) (2pt) Consider the following data
   $$143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304$$
   and use as an estimate of $S(t_{(i)})$
   $$\hat{S}(t_{(i)}) = 1 - (i - 0.5)/n$$
   were $t_{(i)}$ is the ith ordered value and n is the sample size. Use the graphical technique in the previous question to check if a Weibull distribution is appropriate for these data

   (e) (2pt) Assume that the Weibull distribution is a good fit, use least squares approach to estimate its parameters.

2. (10pt) The data below show survival times in months of patients with Hodgkin's disease who were treated with nitrogen mustard. Group A patients received little or no prior therapy whereas Group B patients received heavy prior therapy. Starred are observations are censoring times.

$$\begin{aligned}
GroupA &: 1.25, 1.41, 4.98, 5.25, 5.38, 6.92, 8.89, 10.98, 11.18, 13.11, 13.21, 16.33, 19.77, 21.08,\\
&\quad 21.84^*, 22.07, 31.38^*, 32.61^*, 37.18^*, 42.92\\
GroupB &: 1.05, 2.92, 3.61, 4.20, 4.49, 6.72, 7.31, 9.08, 9.11, 14.49^*, 16.85, 18.82^*, 26.59^*, 30.26^*, 41.34^*
\end{aligned}$$

(a) (2.5pt) Obtain and plot the Kaplan Meier estimates of S_A and S_B, the corresponding survival functions.

(b) (2.5pt) Estimate $S_A(10)$ and $S_B(10)$ using a 95% confidence interval.

(c) (5pt) Test $H_0 : S_A = S_B$ against $H_a : S_A \neq S_B$. Use $\alpha = 0.05$.