STAT 645: Biostatistics - Assignment 8
Due Friday, October 29, 11:55 pm central

1. Consider the dataset on 44 subjects given in the article `https://www.bmj.com/content/317/7156/468.1`. Consider prednisolone or no prednisolone as the binary treatment variable, and use it as the explanatory variable and fit lognormal, exponential, and Weibull model to the data. Then choose the best model and justify your choice.

   (a) Obtain the analytical expression of the 25th, 50th, and 75th percentile of the time-to-event of the best fitted model for the two groups.

   (b) Estimate the above three percentiles and obtain the 95% CI for the percentiles for the two groups separately.

   (c) Using a nonparametric method obtain the estimate and 95% CI for the 25th, 50th, and 75th percentiles of the time-to-event for the two groups separately. Compare and comment on the differences between these nonparametric estimates and the parametric estimates obtained in step (b).

2. Consider the colon data available in the survival package of R. Consider the subset where etype = 1 only (exclude the subjects who experienced death). You may find a descent description of the data at `https://stat.ethz.ch/R-manual/R-devel/library/survival/html/colon.html`. Consider the following seven explanatory variables, sex, age, perfor, adhere, nodes, differ, extent. Make sure to treat differ and extent as factor variables.

   (a) Build a Weibull model with the above explanatory variables and their two factor interactions. Then choose the best explanatory variables using the stepwise regression method, use both AIC and BIC.

   (b) For the best chosen models (one based on the AIC and the other based on the BIC), obtain the estimate and 95% CI for the survival probability at time 365, 730, 1095, 1460, 1825 days and for the following set of covariates. Discuss the results.

sex	age	perfor	adhere	nodes	differ	extent
Male	60	0	0	2	2	3
Male	60	0	1	2	2	3
Male	60	1	0	2	2	3
Male	60	1	1	2	2	3
Female	60	0	0	2	2	3
Female	60	0	1	2	2	3
Female	60	1	0	2	2	3
Female	60	1	1	2	2	3

   (c) Consider the Weibull model with age, sex, treatment, and nodes and their two factor interactions as the explanatory variables. Then conduct a likelihood ratio test to check age has a statistically significant effect on the time-to-event. Full points will be given only for properly writing the hypotheses, test statistics, p-value, and conclusions.