STAT 645: Biostatistics - Assignment 9
Due Saturday, November 14, 11:55pm central

1. Consider the monoclonal gammopathy of undetermined significance (MGUS) dataset available under survival package. Consider age, gender, albumin, creatinine, and hemoglobin level at MGUS diagnosis as the potential covariates. Work with the dataset where these covariates are completely observed. Fit a proportional hazard (PH) model for the days-to-death from the diagnosis in terms of the covariates.

   (a) Test the association between the hazard and the age of diagnosis at the 5% level.

   (b) Estimate relative risk and its 95% CI for the death of a subject with the age of diagnosis 60 compared to the subject with the age of diagnosis 50 while all other covariates remain unchanged.

   (c) Test if there is any effect of gender, albumin, and hemoglobin at the 5% level [This is a composite hypothesis test, so do not test them separately].

   (d) Based on the initial PH model with age, gender, albumin, creatinine, and hemoglobin as the covariates, obtain the estimate and 95% CI for the 10 year survival probability for subjects with the following covariate values

Age	Gender	Albumin	Creatinine	Hemoglobin
60	Male	3	1	13.5
60	Male	3	4	13.5

   (e) Suppose now all two factor interactions among the covariates age, gender, albumin, creatinine, and hemoglobin are now included in the model along with the main effects. Apply a stepwise model selection technique to choose the best fitted model. Now based on the best fitted model obtain the estimate and 95% CI for the 10 year survival probability for subjects with covariate values mentioned in the previous question.

2. Let us consider the colon cancer data (`colon`) in the `survival` package of R. A description of the data can be found `https://stat.ethz.ch/R-manual/R-patched/library/survival/html/colon.html`. Coarsen the time to quarterly data. Replace 0 value of coarsen time by 0.1. Now, analyze the data with the number of nodes (numeric) and extent (a factor variable) as explanatory variables. Use a reasonable approach to handle the time effect on the conditional failure proability. Write a report on your analysis. Can you estimate the survival probability and 95% CI when the coarsened times are 4, 8, 12, 16, 20, 24, for subjects with 2 nodes and extent=muscle.