Melanoma is a curable form of skin cancer that is the leading cause of skin cancer related deaths. Melanoma comprises 4% of all skin cancer cases and the rate of incident melanoma cases has doubled since the early 1970’s. An estimated 50,000 new cases of melanoma are expected per year with about 5,000 yearly deaths attributable to malignant melanoma.¹

Data were collected using an observational cohort study in Denmark involving 205 patients with malignant melanoma. The available covariates are sex, age, tumor thickness, and presence of ulcer. Age and tumor thickness are continuous and the other predictors are categorical. Note that I’ve also categorized the tumor thickness covariate into 3 groups and called this variable Thickness. You will need this variable to answer Question 2. The response is time to death or censoring (in days). The primary interest in the study was to evaluate whether or not these covariates are significant risk factors for failure of melanoma patients. Do the following in your analysis and incorporate your findings into your report to the medical scientists.

1. Present Kaplan-Meier estimates of the survival functions comparing males and females. Use the log-rank test to test that there is no

¹Source: http://www.melanoma.com/whatis/index.htm
difference in the distributions of survival times for the two groups.

2. Present Kaplan-Meier estimates of the survival functions for the 3 tumor thickness groups separately for males and females. Use the stratified log-rank test to test that there is no difference in the distributions of survival times for the three groups, adjusted for sex.

3. Fit the proportional hazards model to these data. In your analysis, use the continuous variable \textit{thick} and not the categorical variable \textit{ Thickness}. So the covariates under consideration are age, sex, thick, and ulcer. Start with a model that includes all main effects and decide whether or not to add interaction terms to it. Remember that the AIC criterion can add interaction terms that are not statistically significant so be sure to check the p-values of any terms it adds to see whether or not you actually want to include them in your model.

If you do not include interaction terms, then decide whether or not to remove any of the main effects. Clearly write down your final model defining all terms in it.

4. Test the proportional hazards assumption for each variable in your model. If necessary stratify on any variables in order to not violate the PH assumption. \textit{Hint: In my analysis I did not have to stratify.}

5. Present the following results.

   - The fitted model with point estimates, standard errors, p-values, and hazard ratios (point estimates and 95\% CIs for each HR).
   - Plots of estimated survival functions comparing different types of melanoma patients. You need to decide what types of patients to consider when getting the survival curves. This will of course depend on the covariates in your model. If there are 3 main effects in your model, then an appropriate $2 \times 2$ plot should be sufficient. Address the issues of statistical significance and practical importance of the predictors, where appropriate.

6. Present your R code in the appendix.