Project 4

The data set (data_project_4.csv) is based on the study of Sydó et al (2014). This study aimed at knowing the relationship between age and the maximum heart rate during a 21-minute treadmill exercise known as Bruce protocol. The data set contains the following variables:

- Age – age of individuals in years
- Female – 0 (male) and 1 (female)
- Res_hr – resting heart rate
- Peak_hr – maximum heart rate during the exercise
- Finished.test – indicates the participant was able to finish the exercise completely (0 – No, 1 – Yes)

The objective of this project is to study the relationship between age and the maximum heart rate that one individual can stand. To address this objective:

1. Perform a descriptive analysis (e.g., summary statistics/plots) for each variable. Construct plots that allow the visualization of the relationship between age and the maximum heart. Calculate appropriate measures of associations for the relationship between age and the maximum heart and provide an initial interpretation of the data.
2. Now consider this data set from a perspective of survival analysis. For that, you need to interpret “peak_hr” as the “time to an event of interest” and Finished.test as indicative of right censoring (0 – completely observed; 1 – censored). Apply the Kaplan-Meier method to estimate the overall survival curve of “peak_hr”. Do the same for males and females separately.
3. Compare “peak_hr” between females and males using an appropriate statistical test. Provide your reasoning for the choice of the test.
4. Perform a Weibull regression analysis using “Peak_hr” as the outcome and the remaining variables in the data set as covariates. Is “Age” significantly influencing the outcome variable? Is the sex of the individuals significantly influencing the outcome variable?

5. Perform a residual analysis of the estimated model in 4. Use Cox-Snel residuals and perform a statistical test to validate the respective distribution.
6. Estimate a Cox's proportional hazard model using "Peak_hr" as the outcome and the remaining variables in the data set as covariates. Is "Age" significantly influencing the outcome variable? Is the sex of the individuals significantly influencing the outcome variable?
7. Discuss the idea of using the estimated model to predict the age of an individual.

Important:

Prepare a 15-minute presentation with your main findings. There will be a penalty of 0.5 points in your project grade if you exceed the time of your presentation. Upload your R script/R Markdown for code verification. Also upload your presentation as a pdf file. Failure to upload these files before classroom evaluation leads to a penalty of 0.5 points in your project grade.

Note:

The original data set was modified for the purpose of this project and, therefore, the published results should not be used as guidance.

Reference:

Sydó N, Abdelmoneim SS, Mulvagh SL, Merkely B, Gulati M, Allison TG. Relationship between exercise heart rate and age in men vs women. Mayo Clin Proc. 2014 Dec;89(12):1664-72. doi: 10.1016/j.mayocp.2014.08.018.