Week 1: Basic Concepts in Survival Analysis

Instructions (read carefully):

- Each group must submit **one** assignment as a .pdf file.
- Each member of the same group will receive the same grade.
- Please put the name of each group member on the first page.
- Use one inch margins and double spaced text.
- For each assignment, one group will be chosen in advance to present/discuss the results in class.
- All of the R code for all applied questions must be provided as a separate file (with .R as the extension) along with the homework file.
- This assignment is due **electronically** to your TA on **..** at the beginning of class.

Question 1) Using the language of "censoring" and/or "truncation" (left, right, and/or interval), explain why a prospective cohort study is often seen as higher quality than a retrospective cohort study.

Question 2) Using Figure 1, draw the line diagram for for ID = 0 that would result if this individual was left truncated.

Question 3) Please do a basic exploratory analysis of the "example_dat1.csv" dataset. No more than 1/2 page. Provide results for the exposure, the confounder, and the outcome.

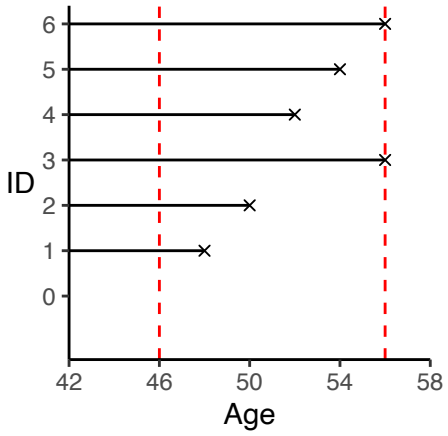
Question 4) Describe, in words, the interpretation of the CDF:

$$F(t) = P(T \leq t)$$

AND the survival function:

$$S(t) = P(T > t)$$

Figure 1: Line diagram with six hypothetical individuals who enter into a study at age 46 and who exit at age 56.



if T represents age at death from all causes, and t represents 64 years of age.

Question 5) Fit the 'survfit()' function to the "example_dat1.csv" data. Examine the R object that you get from this fit. Is there enough information in this object for you to determine the median survival time for the outcome? If so, what is the median survival time.

Question 6) Using "example_dat1.csv", plot the cumulative distribution function using the KM estimator. Interpret the curve.