QMWS: Survival analysis workshop assignment

Instructor: Kevin McGregor **Due Date:** Tues. May 28th, 2024

Instructions: Competing this assignment is a requirement for a digital credential for this workshop. Please send your responses to me by email (`kevinmcg@yorku.ca`) by May 28th, 2024. Send written answers as well as an R script. It is pass/fail; I will return comments.

Question 1: In each case, state what kind(s) of censoring and/or truncation is present and say why.

(a) A pharmaceutical company wants to estimate survival of patients receiving treatment for a particular disease. However, they suspect that there are patients with the disease who died before study enrollment. Some individuals were observed, but did not die before the end of the study.

(b) A restaurant manager wants to estimate how long plates last before getting broken. Every Sunday, the manager counts how many plates are in stock and how many have been broken in the past week.

Question 2: Consider the following data containing time to death (in weeks) of patients with diploid tumors of the tongue:

1, 3, 4, 5, 5, 8, 12, 13, 18, 23, 26, 27, 30, 42, 56, 62, 69, 104, 104, 112, 129, 181, 8+, 67+, 76+, 104+, 176+, 231+

In R, plot the Kaplan-Meier curve for this dataset. Give a rough estimate of the median survival time.

Question 3: For this question, we will use a dataset called `gbsg`, which is from the German Breast Cancer Study Group. Load the data in R by running:

```
data(cancer, package="survival")
```

To see what the variables represent, run the command `?gbsg`.

Run a Cox proportional hazards model using hormone therapy, age, and tumor grade as predictors. Interpret each of the estimated β parameters.