Spring, April 28, 2026

Takehome Final Exam
PQHS 435: Survival Data Analysis

Due Monday, May 4 by 5 p.m.

Open textbook and notes. I assume that you will use SAS or R. However, if you have other preference, it's ok provided it meets the techinical requirements specified.

NOTE: There should be no collaboration on the takehome final!

```
In all cases where the results of a statistical test are asked for,
    (i) EXPLAIN CLEARLY what the hypotheses H0 and H1 are.
    (ii) Find the P-value for the test indicated (and state what test you
        used).
    (iii) Interpret your results, and state whether the results are
significant (P<0.05), highly significant (P<0.01), or not significant
(P>=0.05).
```

When you use SAS, R or other software to help solve a problem, provide your programs, summarized output from your programs. Chapter 17 provides a quite comprehensive overview on surbvival analysis with R. Syntax details for various SAS procedures can be found at the website:

`https://support.sas.com/en/software/sas-stat-support.html`

1. In a clinical trial conducted at Ireland Cancer Center, 11 patients with AIDS associated Hodgkin's disease were treated with oral combination chemotherapy: 5 were treated in the HAART (highly active antiretroviral therapy) era (after Jan. 1, 1997), and 6 were treated in pre-HAART era (prior to Dec. 31, 1996). The overall survival times (months) for the 11 patients were

   ```
   Pre-HAART era (n=6):

     8  6.5  2.5  24  25+  5.5
   ```

```
HAART era (n=5):

  10  56  39.5+  60+  68+
```

where a trailing + means a right-censored value. For this question, You have to type the data by yourself (30 points).

(a) What are the probabilities of survival of all 11 patients at 6 months, 1 year, 2 years, 3 years, 4 years and 5 years? Use Kaplan-Meier method to find out. Summarize the survivals by a table as well as Kaplan-Meier plot with 95% confidence interval.

(b) Is there any significant difference of overall survival between patients treated in pre-HAART era and patients treated in HAART era? Apply the log-rank test to find out. What is the resulting two-sided P-value? What is the median survival time for patients in each era? Summarize the survival estimations for two groups using Kaplan-Meier plot.

2. **Leukaemia Remission Data** which is available at

   `http://bfox.cwru.edu/~pxf/teaching/435/final/6-MP.html`

   and was used as example in the class (see lecture notes - Lecture 23). Ignore covariate *pair* and do the following (45 points)

   (a) Using proportional hazard model to compare remission time of leukaemia patients treated by placebo or 6-MP, respectively. Provide appropriate interpretation of your finding.

   (b) Using Weibull proportional hazard model to compare remission time of leukaemia patients treated by placebo or 6-Mp. Find the medians of the estimated Weibull distribution of each treatment group. Calculate the hazard ratio between two treatments and its 95% confidence interval. Plot the estimated survival and hazard functions under each treatment.

   (c) Fitting the same data using log-logistic accelerated failure time (AFT) model, calculate the accelaration factor and interpret your finding.

3. Survival times for 50 black ducks were posted at

   `http://bfox.cwru.edu/~pxf/teaching/435/final/bducks.dat`

   For full describtion of the data, see Appendix B.3 and Table B.2 (at page 500) in our textbook (45 points).

   (a) Analyze the data using the Cox Proportional Hazards model and give appropriate interpretation of your results.

   (b) For those important covariates in your final model identified in (a), determing their functional forms by checking martingale residuals.

   (c) For those important covariates in your final model identified in (a), examining the proportaional hazard assumption for each of them.

   (d) Based on your final model, identify influential observations if any.

4. **Bone marrow transplantation for leukemia** is a standard treatment for acute leukemia. The data from a multicenter trial of patients prepared for transplantation with a radiation-free conditioning regimen is available at following URL

   `http://bfox.cwru.edu/~pxf/teaching/435/final/bmt.htm`

   In this study, several potential risk factors were measured at the time of transplantation (time-independent). Some factors are time-dependent; these are the development of acute GVHD, the development of chronic GVHD and the return of the patient's platelet count to a self-sustaining level (platelet recovery) during the transplant recovery process. For this question, you are asked to run univariate proportional hazards models (i.e. one covariate at a time) with time-dependent covariate, namely,

   (a) Estimate the effect of acute GVHD on disease free survival (DFS).

   (b) Estimate the effect of chronic GVHD on DFS.

   (c) Estimate the effect of platelet recovery on DFS.

   Please intrepret your findings in those univariate analyses (30 points).

   Note: The dataset given contains many covariates. However, for this question, you only need following variables in the dataset: T_2, δ_3, T_A, δ_A, T_C, δ_C, T_P, δ_P.