University of California, Davis
Department of Public Health Sciences

Fall 2024	**Survival Analysis**	**BST 222**
David M. Rocke		**October 15, 2024**

Homework Assignment 3

Due October 24, 2024

1. Here is some output from an analysis of the bmt data with bone marrow transplants from three types of leukemia, which are here just numbered 1, 2, and 3.

```
> summary(bmt.cox)
Call:
coxph(formula = dfsurv ~ factor(group), data = bmt)

  n= 137, number of events= 83

                   coef exp(coef) se(coef)      z Pr(>|z|)
factor(group)2 -0.5742    0.5632   0.2873 -1.999   0.0457 *
factor(group)3  0.3834    1.4673   0.2674  1.434   0.1516
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

               exp(coef) exp(-coef) lower .95 upper .95
factor(group)2    0.5632     1.7757    0.3207     0.989
factor(group)3    1.4673     0.6815    0.8688     2.478

> vcov(bmt.cox)
               factor(group)2 factor(group)3
factor(group)2     0.08254038     0.04181177
factor(group)3     0.04181177     0.07148991
```

   (a) Test the hypothesis that the coefficients of types 1 and 2 are the same. Provide a 95% confidence interval for the difference. Do this by hand and show your work.

   (b) Test the hypothesis that the coefficients of types 2 and 3 are the same. Provide a 95% confidence interval for the difference. Do this by hand and show your work.

2. The addicts data set is from a study by Caplehorn et al. ("Methadone Dosage and Retention of Patients in Maintenance Treatment," Med. J. Aust., 1991). These data comprise the times in days spent by heroin addicts from entry to departure from one of two methadone clinics. There are two further covariates, namely, prison record and methadone dose, believed to affect the survival times.

   The data set and R input code are on the website. The variables are as follows:

   - id: Subject ID
   - clinic: Clinic (1 or 2)
   - status: Survival status (0 = censored, 1 = departed from clinic)
   - time: Survival time in days
   - prison: Prison record (0 = none, 1 = any)
   - methadone: Methadone dose (mg/day)

   (a) Plot the Kaplan-Meier survival curves for the two clinics.

   (b) Test for whether the two survival curves could have come from the same process using `survdiff`.

   (c) Plot the cumulative hazards from the Nelson-Aalen estimator.

   (e) A common comparison plot for proportional hazards is the complimentary log-log survival plot which plots $\ln(-\ln[\hat{S}(x)]) = \ln \hat{H}(t)$ against $\ln(t)$. If the hazards are proportional, then so are the cumulative hazards, and after taking logs, the curves should be parallel. If the lines are straight, then the Weibull model may be appropriate. Make this plot for the two clinics using the Nelson-Aalen estimator and comment on the results. You will use the "cloglog" option in the plot command.

   (e) Construct a Cox model using only the clinic variable. Is the "survival" different at the two clinics? What is the estimated hazard (risk) ratio, a test for significance, and a 95% confidence interval?

   (f) Pick one test for the null hypothesis that the clinics do not differ. Why would you depend on this test more than the others?

   (g) Consider adding the prison and methodone variables. Which of these covariates seems to improve the model?

   (h) Plot the two survival curves from your chosen Cox model and add the two KM survival curves. What do you think?