Math 473 HW11 Spring 2023. **Due Monday, May 1**.

Exam 3 is Wednesday, April 26. Quiz 10 Wednesday, May 3. See old Q10 extra, problems 4) and 5) of HW 11 and problem 3) of HW 10. The final is Tuesday, May 9, 2:45 - 4:45.

1) HW6 problem 2 (2.12) discusses the recid data and how to use the infile statement.

a) Obtain the SAS program for HW 11 problem 1 from (http://parker.ad.siu.edu/Olive/survhw.txt). To execute the program, use the top menu commands "Run>Submit". An output window will appear if successful. **Warning: if you do not have the recid.txt file on e drive, then you need to change** the *infile* command in the SAS code to the drive that you are using, eg change *infile "e:redic.txt";* to *infile "f:recid.txt";* if you are using F drive. Print the 4 pages of output directly from SAS. (Put into *Word*.)

b) The generalized Cox regression model has $x_i log(time)$ interactions *finlt* to *priolt*. Which interaction was most significant (had the smallest p-value)?

c) Treat the model in b) as the full model and the model without the interactions as the reduced model. If the reduced model is good, then the Cox PH assumptions are reasonable. Test whether the reduced model is good. (The test is done exactly as for the Cox PH model.)

2) This problem produces output for the Stanford Heart Transplant data discussed in class. Obtain the SAS program for HW 11 problem 2 from (http://parker.ad.siu.edu/Olive/survhw.txt). This program was taken from Allison (1995). The time dependent variable $x_1(t) = plant = 1$ if the patient has had a transplant by time t and is 0 otherwise. The variable $x_2 = surg = 1$ if the patient has had previous heart surgery and is 0 otherwise. The variable $x_3 = ageaccpt$ is the patient's age at time of acceptance into the program. The SAS program fits a generalized Cox regression (GCR) model.

a) Print the output. (Put into *Word*.)

b) Test $\beta_1 = 0$.

c) Test $\boldsymbol{\beta} = \mathbf{0}$.

3) **2.23:** This problem produces output for the Stanford Heart Transplant data discussed in class, but *R* is used instead of SAS. Obtain the *R* program for HW 11 problem 3 from (http://parker.ad.siu.edu/Olive/survhw.txt). The time dependent variable $x_1(t) = transplant = 1$ if the patient has had a transplant by time t and is 0 otherwise. The variable $x_2 = surgery = 1$ if the patient has had previous heart surgery and is 0 otherwise. The variable $x_3 = age$ is the patient's age at time of acceptance into the program. The *R* program fits a generalized Cox regression (GCR) model. The SAS and R heart data sets seem to differ slightly and do not give the exact same answers.

a) Print the output. (Put into *Word*.)

b) Test $\beta_1 = 0$.

c) Test $\boldsymbol{\beta} = \mathbf{0}$.

4) Suppose you are estimating the mean μ of losses with $T = \overline{X}$.
actual losses 1, 2, 5, 10, 50: $\overline{X} = 13.6$,
a) Compute $T_1^*, ..., T_4^*$, where T_i^* is the sample mean of the ith bootstrap sample.
bootstrap samples:

2, 10, 1, 2, 2:

50, 10, 50, 2, 2:

10, 50, 2, 1, 1:

5, 2, 5, 1, 50:

b) Now compute the bagging estimator which is the sample mean of the T_i^*: the bagging estimator $\overline{T}^* = \frac{1}{B}\sum_{i=1}^{B} T_i^*$ where $B = 4$ is the number of bootstrap samples.

5)

```
           Estimate Std.Err 95% shorth CI
X1         -42.4846 51.2863 [-192.281, 52.492]
X2          0               [   0.000,  0.268]
X3          1.1707  0.0598  [   0.992,  1.289]
X4          0               [   0.000,  0.840]
X5          0               [   0.000,  1.916]
X6          0.1467  0.0368  [  0.0747,  0.215]
```

Given the above output, what is $\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_{min},0}$?