Math 473 HW 5 Spring 2023. Due Friday, March 3.

Exam 2 is Wednesday, March 22. Final is Tuesday, May 9, 2:45-4:45.

1) 2.1: Suppose that a proportional hazards model holds so that $h_{\boldsymbol{x}}(t) = \exp(\boldsymbol{\beta}'\boldsymbol{x})h_0(t)$ where $h_0(t)$ is the baseline hazard function. Let $f_0(t)$, $S_0(t)$ $F_0(t)$ and $H_0(t)$ denote the baseline pdf, survival function, distribution function and cumulative hazard function.

a) Show

$$H_{\boldsymbol{x}}(t) = \exp(\boldsymbol{\beta}'\boldsymbol{x})H_0(t).$$

b) Show

$$S_{\boldsymbol{x}}(t) = [S_0(t)]^{\exp(\boldsymbol{\beta}'\boldsymbol{x})}.$$

c) Show

$$f_{\boldsymbol{x}}(t) = f_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x})[S_0(t)]^{\exp(\boldsymbol{\beta}'\boldsymbol{x}) - 1}.$$

The *R* problems need the command library(survival) once for each time you get into *R*.

2) 2.17: The lung cancer data has the *time* until death or censoring and *status* = 0 for censored and 1 for uncensored. Then the covariates are *age*, *sex* = 1 for M and 2 for F, *ph.ecog* = Ecog performance score 0-4, *ph.karno* = a competitor to ph.ecog, *pat.karno* = patient's assessment of their karno score, *meal.cal* = calories consumed at meals excluding beverages and snacks and *wt.loss* = weight loss in last 6 months.

a) Copy and paste commands from (http://parker.ad.siu.edu/Olive/survhw.txt) for this problem into *R*. (This is actually a stratified PH model.)

Type *zfull*, then *zred1* then *zred2*. Copy and paste the resulting output into *Word*. The full model uses *age*, *ph.ecog*, *ph.karno*, *pat.karno* and *wt.loss*.

b) Test whether the reduced model that omits *age* can be used.

c) Test whether the reduced model that omits *age* and *ph.karno* can be used.

3) 2.18: Go to (http://parker.ad.siu.edu/Olive/survhw.txt) and copy and paste the source command source("http://parker.ad.siu.edu/Olive/survpack.txt") near the top of the file into *R*. This problem will use the program `bphgfit` to check the PH model with the Kaplan Meier KM estimator.

a) Copy and paste commands from (http://parker.ad.siu.edu/Olive/survhw.txt) for this problem into *R*. Copy and paste the output into *Word*. (You may need to press Enter to get the plot.)

b) Click on the plot and hold down the *Ctrl* and *c* buttons simultaneously. Then in the *Word* Edit menu, select "paste."

c) The data is remission time in weeks for leukemia patients receiving treatments A (x = 0) or B (x = 1). The indicator variable x (leuk[,3]) is the single covariate. Do a PLRT to test whether $\beta = 0$. Is there a difference in the effectiveness of the 2 treatments?

d) The solid lines in the plot correspond to the estimated PH survival function for each treatment group. The plotted points correspond to the estimated Kaplan Meier estimator for each group. If the PH model is good, then the plotted points should track the solid lines fairly well. Is the PH model good? (When $\beta = 0$, the PH model for this data is $h_0(t) = h_1(t)$, but the PH model could fail, e.g. if the survival function for treatment A is higher than that of treatment B until time t_A and then the survival

function for treatment B is higher: the survival functions cross at exactly one point $t_A > 0$.)

With some versions of R, there are three curves of circles. The center curve is the Kaplan Meier estimator while the two outer bands are pointwise CI bands.

4) 2.19: An extension of the PH model is the stratified PH model where $h_{\boldsymbol{x},j} = \exp(\boldsymbol{\beta}'\boldsymbol{x})h_{0,j}(t)$ for $j = 1, ..., K$ where $K \geq 2$ is the number of strata (groups). Testing is done in exactly the same manner as for the PH model, and the same $\boldsymbol{\beta}$ is used for each strata, only the baseline function changes. The regression in problem 1) actually used gender, male and female, as strata. If the model was good, then a PH model should hold for males and a PH model should hold for females. For the lung cancer data, females had a higher survival curve than males for $\boldsymbol{x}$ set to the average values.

A censored response plot (ESSP) is a plot of the ESP $= \hat{\boldsymbol{\beta}}'\boldsymbol{x}$ versus T, the survival times, where the symbol "0" means the time was censored and "+" uncensored. If the PH model holds, the variability of the plotted points should decrease rapidly as ESP increases.

If you got out of R after doing problem 3), you need to copy and paste the source command into R as in problem 3). The source command only needs to be done once per R session.

a) Copy and paste commands from (http://parker.ad.siu.edu/Olive/survhw.txt) for this problem into R. Click on the plot and hold down the *Ctrl* and c buttons simultaneously. Then in the *Word* Edit menu, select "paste."

b) Repeat a) except use the commands for 2.19b).

How does the variability in the plot for a narrow vertical strip at ESP = 0.5 compare to the variability for a narrow vertical strip at ESP = -1.5?

c) Copy and paste the commands for this part into R, and include the resulting plot in *Word*.

d) Copy and paste the commands for this part into R, and include the resulting plot in *Word*.

```
vlung2(2)
title("females")
```

e) The plots in c) and d) divide the ESP into 4 slices. The estimated PH survival function is evaluated at the last point in the first 3 slices and at the first point in the 4th slice. Pointwise confidence intervals are also included (dashed upper and lower lines). The plotted circles correspond to the Kaplan Meier estimator for the points in each slice. The 1st slice is in the NW corner, the 2nd slice in the NE, the 3rd slice in the SW and the 4th slice in the SE. Confidence bands that would include an entire reasonable survival function would be much wider. Hence if the plotted circles are not very far outside the pointwise CI bands, then the PH model is reasonable.

Is the PH model reasonable for males? Is the PH model reasonable for females?

With some versions of R, there are three curves of circles. The center curve is the Kaplan Meier estimator while the two outer bands are pointwise CI bands.