Math 473 HW 6 Spring 2023. Due Friday, March 10. Quiz 6 is on Wed. March 8. Exam 2 Wednesday, March 22 emphasizes HW4,5,6 and quizzes 4,5,6. **SAS forward selection, backward elimination, and stepwise selection produces too much output.** Only submit some of the produced output. The AIC line in the With Covariates column is important.

1) 2.11: Data is from a study on multiple myeloma (bone cancer) in which researchers treated 65 patients with alkylating agents. The variable *Time* is the survival time in months from diagnosis. The predictor variables are *LogBUN* (blood urea nitrogen), *HGB* (hemoglobin at diagnosis), *Platelet* (platelets at diagnosis: 0=abnormal, 1=normal), *Age* at diagnosis in years, *LogWBC*, *Frac* (fractures at diagnosis: 0=none, 1=present), *LogPBM* (log percentage of plasma cells in bone marrow), *Protein* (proteinuria at diagnosis), and *SCalc* (serum calcium at diagnosis).

a) Obtain the SAS program for this problem from (http://parker.ad.siu.edu/Olive/survhw.txt).

b) First backward elimination is considered. From the SAS output window, copy and paste the output for the full model that uses all 9 variables into *Word*. That is, scroll to the top of the output and copy and paste the following output.

```
Step  0. The model contains the following variables:

LogBUN  HGB  Platelet  Age  LogWBC  Frac  LogPBM  Protein  SCalc
.
.
.
SCalc 1 0.12595  0.10340 1.4837 0.2232 1.134
```

c) At step 7 of backward elimination, the final model considered uses LogBUN and HGB. Copy and paste the output for this model (similar to the output for b) into *Word*.

d) Backward elimination will consider 8 models. Write down the variables used for each model as well as the AIC. The first two models are shown below.

```
variables                                                  AIC
LogBUN HGB Platelet Age LogWBC Frac LogPBM Protein SCalc  310.588
LogBUN HGB          Age LogWBC Frac LogPBM Protein SCalc  308.827
```

e) Repeat d) for the 4 models considered by forward selection.

f) Repeat d) for the 4 models considered by stepwise selection.

g) For all subsets selection, complete the following table (get the 2 chisq entries).

```
variables       chisq
      2                   LogBUN HGB
      9                   full
```

h) Perform a change in PLR test if the full model uses 9 variables and the reduced model $I_{min} = I_I$ that uses LogBUN and HGB. (Use the output from b) and c).)

i) Are there any other good candidate models besides model I_I?

SAS forward selection, backward elimination, and stepwise selection produces too much output. Only submit some of the produced output. The AIC line in the With Covariates column is important.

2) 2.12: Data is from Allison (1995, p. 270). The response variable *week* is time in weeks until arrest after release from prison (right censored if week = 52). The 7 variables are *Fin* (1 for those who received financial aid, 0 else), *Age* at time of release, *Race* (1 if black, 0 else), *Wexp*(1 if inmate had full time work experience prior to conviction, 0 else), *Mar* (1 if married at time of release, 0 else), *Paro* (1 if released on parole, 0 else), *Prio* (the number of prior convictions).

a) This is a large data file. SAS needs an "end of file" marker to determine when the data ends. SAS uses a period as the end of file marker, and the period has already been added to the file. Obtain the file from (http://parker.ad.siu.edu/Olive/recid.txt) and save the file as *recid.txt* on your flash drive using the commands "File>Save Page as" with *File name recid*. (Be careful that you do not save the file as recid.txt.txt.)

b) Obtain the SAS program for this problem 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 the f drive.

c) First backward elimination is considered. Scroll to the top of the copy and paste the 1st 2 pages of output for the full model into *Word*.

d) Backward elimination will consider 5 models. Write down the variables used for each model as well as the AIC. The first two models are shown below.

variables							AIC
fin	age	race	wexp	mar	paro	prio	1332.241
fin	age	race	wexp	mar		prio	1330.429

e) Repeat d) for the 4 models considered by forward selection.

f) Repeat d) for the 5 models considered by stepwise selection.

g) For all subsets selection, complete the following table (get the 2 chisq entries).

variables	chisq	
3		fin age prio
7		full