Math 473 Exam 2 Spring 2017 Name_____

variables					AIC
rage	dage	preg	index	gvhd typ1 typ3	102.967
rage	dage		index	gvhd typ1 typ3	101.044
rage	dage			gvhd typ1 typ3	99.124
rage	dage			gvhd typ3	97.292
	dage			gvhd typ3	95.421
				gvhd typ3	95.295
				gvhd	106.130

$\leftarrow I_{min} = I_I$ $I_I = I_0$

1) Data is from Collett (2003, p. 367) regarding survival of leukemia patients after a bone marrow transplant. The predictors are *rage* = patient age, *dage* = donor age, *preg* (1 if donor pregnant, 0 else), *index* of cell-lymphocyte reactions, *gvhd* (1 if there is a graph vs host disease, 0 else), *typ1* (1 if AM leukemia, 0 else) and *typ3* (1 if CM leukemia, 0 else). The type was actually a factor with type 2 AL leukemia. Results from backward elimination are shown.

omit

a) What is the minimum AIC submodel I_{min}?

gvhd, typ3

b) What is the best starting submodel I_0?

gvhd, typ3 $= I_{min}$ in a)

do not use: more predictors than I_{min} $\leftarrow -6$ if dage, gvhd, typ3

$(AIC(I_{min}) + 2 = 97.295)$

c) Are there any other candidate submodels? Explain briefly.

no $AIC(gvhd) > AIC(I_{min}) + 7 = 102.295$

24

look at models with fewer predictors than I_0

	M1	M2	M3	M4
# of predictors	7	4	3	1
# with $0.01 \leq$ p-value ≤ 0.05	1	0	0	1
# with p-value > 0.05	5	2	1	0
$-2\log(L)$	88.967	89.292	91.185	104.130
$AIC(I)$	102.967	97.292	97.185	106.130
p-value for change in PLR test	1.0	0.955	0.696	0.019

2) The above table gives summary statistics for 4 PH regression models considered as final submodels after performing variable selection. Assume that the PH assumptions hold for all 4 models. The full model was M1. Which model should be considered as the first starting submodel I_{II}? Explain briefly why each of the other 3 submodels should not be used as the starting submodel.

$I_{II} = M3$ $(AIC(I_{min})+2 = 99.185)$

M1 and M2 have too many predictors

$AIC(M1) > AIC(I_{min})+2$ is ok

$AIC(M4) > AIC(I_{min})+2 = 99.185$

(or $AIC(M4) > AIC(I_{min})+7 = 104.185$)

10

Criterion	Without Covariates	With Covariates	Reduced Model
-2 LOG L	204.801	180.898	
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	23.9034	3	<0.0001

Parameter Variable	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
perf	1	-0.05831	0.01309	19.8512	<.0001
type	1	-0.03346	0.40642	0.0068	0.9344
trt	1	0.24725	0.35074	0.4969	0.4809

Criterion	Without Covariates	With Covariates	Full Model
-2 LOG L	204.801	177.740	
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.0615	6	0.0001

Parameter Variable	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
perf	1	-0.11415	0.03909	8.5259	0.0035
type	1	-0.49892	1.09251	0.2086	0.6479
trt	1	-0.34667	0.97596	0.1262	0.7224
perflt	1	0.01633	0.01028	2.5208	0.1124
typelt	1	0.12005	0.28977	0.1716	0.6787
trtlt	1	0.17496	0.24548	0.5080	0.4760

test is same as for PH model

e

3) The advanced lung cancer data is from Leemis (1995, p. 249). The PH model is the reduced model and has predictors *perf*, *type* and *trt*. The GCR model is the full model and adds *perflt*= perf*log(time), *typelt* = type*log(time) and *trtlt* = trt*log(time) interactions to test whether the PH assumption is reasonable. Test whether the reduced model is good.

Q5b2)

H_0 the reduced model is good H_A use the full model

$X^2(RF) = X^2(NF) - X^2(NR) = 27.0615 - 23.9034 = 3.1581$

$= 180.898 - 177.740 = 3.158$

$pval = P(X^2_3 > 3.1581) > 0.25$

df	0.25
3	4.11

fail to reject H_0, the reduced model is good

10

```
          coef exp(coef) se(coef)     z   p
histol 1.45554      4.29  0.13462 10.81 0.0
instit 0.19910      1.22  0.14389  1.38 0.17
age    0.00603      1.01  0.00125  4.83 0.0
stage  0.35003      1.42  0.04052  8.64 0.0

Likelihood ratio test= 390  on 4 df,   p=0, n = 4028
```

4) The output above is for the R `nwtco` data. Y is time until relapse. The predictors are *histol* = tumor histology from central lab, *instit* = tumor histology from local lab, *age* of patient and *stage* of disease. A PH model was used.

a) Test $\boldsymbol{\beta} = \mathbf{0}$. $H_0: \underline{\beta} = \underline{0}$ $H_A: \underline{\beta} \neq \underline{0}$

$X^2(\text{NIF}) = 390$

pval = 0

reject Ho there is a PH survival relationship between Y and the predictors histol, instit, age and stage.

b) Find a 95% CI for β_1.

$\hat{\beta}_1 \pm 1.96\, SE(\hat{\beta}_1) = 1.45554 \pm 1.96(0.13462)$

$= 1.45554 \pm 0.26386 = [1.1917, 1.7194]$

c) Do a 4 step test for $Ho: \beta_1 = 0$. $H_A: \beta_1 \neq 0$

$Z_{01} = 10.81$

pval = 0

reject Ho, histol is needed in the PH survival model given the other predictors instit, age, and stage are in the model

d) Do a 4 step test for $Ho: \beta_2 = 0$. $H_A: \beta_2 \neq 0$

$Z_{02} = 1.38$

pval = 0.17

fail to reject Ho, INSTIT IS not needed in the PH survival model given the other predictors histol, age, and stage are in the model

32

```
  coef  se(coef)    z      p
x 0.916   0.512  1.79  0.074

Likelihood ratio test= 3.38  on 1 df,   p=0.0658
```

5@) The data is for patients with acute myelogenous leukemia. Patients were in remission and Y = time until relapse. The indicator variable for treatment $x = 0$ if the patients recieved maintenance chemotherapy and was 1 otherwise.

a) From the graphs below, does it look like group 0 has a better survival rate than group 1?

Yes group 0 has a higher survival curve

a or -b

b) The graphs could differ due to chance. Do a PLR test for $\beta = 0$.

H_0 $\beta = 0$ H_1 $\beta \neq 0$

$X^2(W|F) = 3.38$

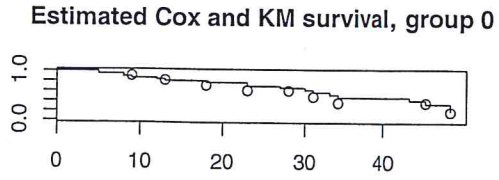
pval = 0.0658

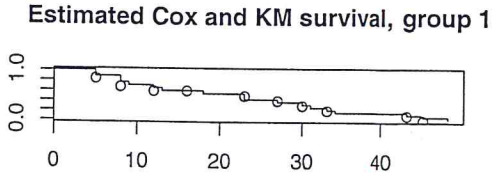
1.79 & .074 -3 but -2 if conclusion is right

fail to reject H_0 there is not a PH survival relationship between Y and X (treatment)

e c) The solid line in each plot is the estimated PH survival function while the circles correspond to the Kaplan Meier estimator. Is the proportional hazards model reasonable? Explain breifly.

Yes the KM $\hat{S}_{KM}$ circles track the cox $\hat{S}$ solid curve closely





24