Exam 2 is Wed. March 22. **You are allowed 10 sheets of notes and a calculator.** The exam covers HW1-3 and Q1-3, but emphasis is on HW4-6 and Q4-6.

18) The **Cox proportional hazards regression (PH) model** is

$$h_i(t) = h_{Y_i|\boldsymbol{x}_i}(t) = h_{Y_i|\boldsymbol{\beta}'\boldsymbol{x}_i}(t) = \exp(\boldsymbol{\beta}'\boldsymbol{x}_i)h_0(t)$$

where $h_0(t)$ is the **unknown baseline function** and $\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)$ is the **hazard ratio**.

For now, assume that the PH model is appropriate, although this assumption should be checked before performing inference.

19) The sufficient predictor $\mathbf{SP} = \boldsymbol{\beta}'\boldsymbol{x}_j = \sum_{i=1}^{p} \beta_i x_{ij}$.

variable	Estimate	Std. Error	Est/SE	or $(Est/SE)^2$	p-value
x_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	$X^2_{o,1} = z^2_{o,1}$	for H_0: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_p	$\hat{\beta}_p$	$se(\hat{\beta}_p)$	$z_{o,p} = \hat{\beta}_p/se(\hat{\beta}_p)$	$X^2_{o,p} = z^2_{o,p}$	for H_0: $\beta_p = 0$

```
SAS                              Wald          pr >
variable  df Estimate  SE        chi square    chisqu
age       1  0.1615    0.0499 10.4652       0.0012
ecog.ps   1  0.0187    0.5991 0.00097       0.9800

R         coef  exp(coef) se(coef)    z      p
age     0.1615      1.18   0.0499 3.2350 0.0012
ecog.ps 0.0187      1.02   0.5991 0.0312 0.9800

Likelihood ratio test=14.3  on 2 df, p=0.000787  n= 26
```

Shown above is output in symbols from and *SAS* and *R* . The estimated coefficient is $\hat{\beta}_j$. The Wald chi square $= X^2_{o,j}$ while p and "pr > chisqu" are both p-values.

20) The estimated sufficient predictor $\mathbf{ESP} = \hat{\boldsymbol{\beta}}'\boldsymbol{x}_j = \sum_{i=1}^{p} \hat{\beta}_i x_{ij}$. Given $\hat{\boldsymbol{\beta}}$ from output and given $\boldsymbol{x}$, be able to find ESP and $\hat{h}_i(t) = \exp(ESP)\hat{h}_0(t) = \exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x})\hat{h}_o(t)$ where $\exp(\hat{\boldsymbol{\beta}}'\boldsymbol{x})$ is the **estimated hazard ratio**. See HW4.1cd.

For tests, the p–value is an important quantity. Recall that H_0 is rejected if the p–value $< \delta$. A p–value between 0.07 and 1.0 provides little evidence that H_0 should be rejected, a p–value between 0.01 and 0.07 provides moderate evidence and a p–value less than 0.01 provides strong statistical evidence that H_0 should be rejected. Statistical evidence is not necessarily practical evidence, and reporting the p–value along with a statement of the strength of the evidence is more informative than stating that the p–value is less than some chosen value such as $\delta = 0.05$. Nevertheless, as a **homework convention**, use $\delta = 0.05$ if δ is not given.

21) See HW4 1ef, 2a. The Wald confidence interval (CI) for β_j can also be obtained from the output: the large sample 95% CI for β_j is

$$\hat{\beta}_j \pm 1.96\ se(\hat{\beta}_j).$$

22) See HW4 1gh, 2b, 3b. Investigators also sometimes test whether a predictor X_j is needed in the model given that the other $k-1$ nontrivial predictors are in the model with a **4 step Wald test of hypotheses**:
i) State the hypotheses H_0: $\beta_j = 0$ H_A: $\beta_j \neq 0$.
ii) Find the test statistic $z_{o,j} = \hat{\beta}_j / se(\hat{\beta}_j)$ or $X^2_{o,j} = z^2_{o,j}$ or obtain it from output.
iii) The p–value $= 2P(Z < -|z_{oj}|) = P(\chi^2_1 > X^2_{o,j})$. Find the p–value from output or use the standard normal table.
iv) State whether you reject Ho or fail to reject Ho and give a nontechnical sentence restating your conclusion in terms of the story problem.

If H_0 is rejected, then conclude that X_j is needed in the PH survival model given that the other $p-1$ predictors are in the model. If you fail to reject H_0, then conclude that X_j is not needed in the PH survival model given that the other $p-1$ predictors are in the model. Note that X_j could be a very useful PH survival predictor, but may not be needed if other predictors are added to the model.

For a PH, often 3 models are of interest: the **full model** that uses all p of the predictors $\boldsymbol{x}^T = (\boldsymbol{x}_R^T, \boldsymbol{x}_O^T)$, the **reduced model** that uses the r predictors $\boldsymbol{x}_R$, and the **null model** that uses none of the predictors.

The partial likelihood ratio test (**PLRT**) is used to test whether $\boldsymbol{\beta} = \boldsymbol{0}$. If this is the case, then the predictors are not needed in the PH model (so survival times $Y \perp\!\!\!\perp \boldsymbol{x}$). If $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ is not rejected, then the Kaplan Meier estimator should be used. If H_o is rejected, use the PH model.

23) See HW4 2c, 3c. The 4 step **PLRT** is
i) $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ $H_A : \boldsymbol{\beta} \neq \boldsymbol{0}$
ii) test statistic $X^2(N|F) = [-2 \log L(none)] - [-2 \log L(full)]$ is often obtained from output
iii) The p–value $= P(\chi^2_p > X^2(N|F))$ where χ^2_p has a chi–square distribution with p degrees of freedom. The p–value is often obtained from output.
iv) Reject H_0 if the p–value $< \delta$ and conclude that there is a PH survival relationship between Y and the predictors $\boldsymbol{x}$. If p–value $\geq \delta$, then fail to reject H_o and conclude that there is not a PH survival relationship between Y and the predictors $\boldsymbol{x}$.

Some *SAS* output for the PLRT is shown next. *R* output is above 20).

```
SAS Testing Global Null Hypotheses: BETA = 0
           without      with
criterion  covariates  covariates     model Chi-square
-2 LOG L   596.651      551.1888      45.463 with 3 DF (p=0.0001)
         -2 log L(None) -2 log L(Full)   X^2(N|F)
```

Let the **full model** be

$$SP = \beta_1 x_1 + \cdots + \beta_p x_p = \boldsymbol{\beta}^T \boldsymbol{x} = \alpha + \boldsymbol{\beta}_R^T \boldsymbol{x}_R + \boldsymbol{\beta}_O^T \boldsymbol{x}_O.$$

let the **reduced model**

$$SP = \beta_{R1} x_{R1} + \cdots + \beta_{Rr} x_{Rr} = \boldsymbol{\beta}_R^T \boldsymbol{x}_R$$

where the reduced model uses r of the predictors used by the full model and $\boldsymbol{x}_O$ denotes the vector of $p - r$ predictors that are in the full model but not the reduced model.

Assume that the full model is useful. Then we want to test H_0: the reduced model is good (can be used instead of the full model, so $\boldsymbol{x}_O$ is not needed in the model given $\boldsymbol{x}_R$ is in the model) versus H_A: use the full model (the full model is significantly better than the reduced model). Fit the full model and the reduced model to get $X^2(N|F)$ and $X^2(N|R)$ where $X^2(N|F)$ is used in the PLRT to test whether $\boldsymbol{\beta} = \mathbf{0}$ and $X^2(N|R)$ is used in the PLRT to test whether $\boldsymbol{\beta}_R = \mathbf{0}$ (treating the reduced model as the model in the PLRT).

variable	Estimate	Std. Error	Est/SE	or $(Est/SE)^2$	p-value
x_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	$X^2_{o,1} = z^2_{o,1}$	for Ho: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_p	$\hat{\beta}_p$	$se(\hat{\beta}_p)$	$z_{o,p} = \hat{\beta}_p/se(\hat{\beta}_p)$	$X^2_{o,p} = z^2_{o,p}$	for Ho: $\beta_p = 0$

R: Likelihood ratio test = $X^2(N|F)$ on p df

```
SAS: Testing Global Null Hypotheses: BETA = 0
Test              Chi-Square       DF     Pr > Chisq
```

Likelihood ratio $X^2(N|F)$ p pval for Ho: $\boldsymbol{\beta} = \mathbf{0}$

variable	Estimate	Std. Error	Est/SE	or $(Est/SE)^2$	p-value
x_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	$X^2_{o,1} = z^2_{o,1}$	for H_0: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_r	$\hat{\beta}_r$	$se(\hat{\beta}_r)$	$z_{o,r} = \hat{\beta}_r/se(\hat{\beta}_r)$	$X^2_{o,r} = z^2_{o,r}$	for Ho: $\beta_r = 0$

R: Likelihood ratio test = $X^2(N|R)$ on r df

```
SAS: Testing Global Null Hypotheses: BETA = 0
Test              Chi-Square       DF     Pr > Chisq
```

Likelihood ratio $X^2(N|R)$ r pval for H_0: $\boldsymbol{\beta}_R = \mathbf{0}$

The output shown on p. 3 in symbols, can be used to perform the change in PLR test. For simplicity, the reduced model used in the output is $\boldsymbol{x}_R = (x_1, ..., x_r)^T$.

Notice that $X^2(R|F) \equiv X^2(N|F) - X^2(N|R) =$

$$[-2\log L(none)] - [-2\log L(full)] - ([-2\log L(none)] - [-2\log L(red)]) =$$

$$[-2\log L(red)] - [-2\log L(full)] = -2\log\left(\frac{L(red)}{L(full)}\right).$$

24) See HW 5.1: The 4 step **change in PLR test** is

i) H_0: the reduced model is good $\quad H_A$: use the full model

ii) test statistic $X^2(R|F) = X^2(N|F) - X^2(N|R) = [-2\log L(red)] - [-2\log L(full)]$

iii) The p–value $= P(\chi^2_{p-r} > X^2(R|F))$ where χ^2_{p-r} has a chi–square distribution with $p - r$ degrees of freedom.

iv) Reject H_0 if the p–value $< \delta$ and conclude that the full model should be used. If p–value $\geq \delta$, then fail to reject H_0 and conclude that the reduced model is good.

If the reduced model leaves out a single variable x_i, then the change in PLR test becomes $H_0 : \beta_i = 0$ versus $H_A : \beta_i \neq 0$. This change in partial likelihood ratio test is a competitor of the Wald test. The change in PLRT is usually better than the Wald test if the sample size n is not large, but the Wald test is currently easier for software to produce. For large n the test statistics from the two tests tend to be very similar (asymptotically equivalent tests).

25) If the reduced model is good, then the **EE plot** of $ESP(R) = \hat{\boldsymbol{\beta}}_R^T \boldsymbol{x}_{Ri}$ versus $ESP = \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i$ should be highly correlated with the identity line with unit slope and zero intercept.

A **factor** A is a variable that takes on a categories called levels. Suppose A has a categories $c_1, ..., c_a$. Then the factor is incorporated into the PH model by using $a - 1$ indicator variables $x_{jA} = 1$ if $A = c_j$ and $x_{Aj} = 0$ otherwise, where the 1st indicator variable is omitted, eg, use $x_{2A}, ..., x_{aA}$. Each indicator has 1 degree of freedom. Hence the degrees of freedom of the $a - 1$ indicator variables associated with the factor is $a - 1$.

The x_j corresponding to variates (variables that take on numerical values) or to indicator variables from a factor are called **main effects**.

An **interaction** is a product of two or more main effects, but for a factor include products for all indicator variables of the factor.

If an interaction is in the model, also include the corresponding main effects. For example, if $x_1 x_3$ is in the model, also include the main effects x_1 and x_2.

26) Let the survival times $T_i = min(Y_i, Z_i)$, and let $\gamma_i = 1$ if $T_i = Y_i$ (uncensored) and $\gamma_i = 0$ if $T_i = Z_i$ (censored). For PH models, an **ET** plot, or ESSP, is a plot of the ESP vs T with plotting symbol 0 for censored cases and + for uncensored cases. If the ESP is a good estimator of the SP and $h_{SP}(t) = \exp(SP)h_0(t)$, then the hazard increases and survival decreases as the ESP increases. See HW5 4ab.

27) The **slice survival plot** divides the ESP into J groups of roughly the same size. For each group j, $\hat{S}_{PHj}(t)$ is computed using the $\boldsymbol{x}$ corresponding to the largest ESP in the 1st $J-1$ groups and the $\boldsymbol{x}$ corresponding to the smallest ESP in the Jth group. The Kaplan Meier estimator $\hat{S}_{KMj}(t)$ is computed from the survival times in the jth group. For each group, $\hat{S}_{PHj}(t)$ is plotted and $\hat{S}_{KMj}(t_i)$ as circles at the deaths t_i. The proportional hazards assumption is reasonable if the circles track the curve well in each of the J plots. If pointwise CI bands are added to the plot, then $\hat{S}_{KMj}$ tracks $\hat{S}_{PHj}$ well if most of the plotted circles do not fall very far outside the pointwise CI bands. See HW5 4d.

Variable selection is closely related to the change in PLR test for a reduced model. You are seeking a subset I of the variables to keep in the model. The $AIC(I)$ statistic is used as an aid in backward elimination and forward selection. The full model and the model with the smallest AIC are always of interest. Create a full model. The full model has a -2 log(L) at least as small as that of any submodel. The full model is a submodel.

Backward elimination starts with the full model with p variables and the predictor that optimizes some criterion is deleted. Then there are $p-1$ variables left and the predictor that optimizes some criterion is deleted. This process continues for models with $p-2, p-3, ..., 3$ and 2 predictors.

Forward selection starts with the model with 0 variables and the predictor that optimizes some criterion is added. Then there is p variable in the model and the predictor that optimizes some criterion is added. This process continues for models with $2, 3, ..., p-2$ and $p-1$ predictors. Both forward selection and backward elimination result in a sequence of p models $\{x_1^*\}, \{x_1^*, x_2^*\}, ..., \{x_1^*, x_2^*, ..., x_{p-1}^*\}, \{x_1^*, x_2^*, ..., x_p^*\}$ = full model.

Consider models I with r_I predictors. Often the criterion is the minimum value of $-2\log(L(\hat{\boldsymbol{\beta}}_I))$ or the minimum AIC(I) $= -2\log(L(\hat{\boldsymbol{\beta}}_I)) + 2r_I$.

Heuristically, backward elimination tries to delete the variable that will increase the -2 log(L) the least. An increase in -2 log(L) greater than 4 (if the predictor has 1 degree of freedom) may be troubling in that a good predictor may have been deleted. In practice, the backward elimination program may delete the variable such that the submodel I with k predictors has 1) the smallest AIC(I), 2) the smallest $-2\log(L(\hat{\boldsymbol{\beta}}_I))$ or 3) the biggest p–value (preferably from a change in PLR test but possibly from a Wald test) in the test $H_0 : \beta_i = 0$ versus H_A $\beta_i \neq 0$ where the current model with $k+1$ variables is treated as the full model.

Heuristically, forward selection tries to add the variable that will decrease the -2 log(L) the most. An decrease in -2 log(L) less than 4 (if the predictor has 1 degree of freedom) may be troubling in that a bad predictor may have been added. In practice, the forward selection program may add the variable such that the submodel I with k predictors has 1) the smallest AIC(I), 2) the smallest $-2\log(L(\hat{\boldsymbol{\beta}}_I))$ or 3) the smallest

p–value (preferably from a change in PLR test but possibly from a Wald test) in the test $H_0 : \beta_i = 0$ versus $H_A : \beta_i \neq 0$ where the current model with $k-1$ terms plus the predictor x_i is treated as the full model (for all variables x_i not yet in the model).

28) If an interaction (eg $x_3x_7x_9$) is in the submodel, then the main effects (x_3, x_7, and x_9) should be in the submodel.

29) If $x_{i+1}, x_{i+2}, ..., x_{i+a-1}$ are the $a-1$ indictor variables corresponding to factor A, submodel I should either contain none or all of the $a-1$ indictor variables.

30) Given a list of submodels along with the number of predictors and AIC, be able to find the "best starting submodel" I_I (the initial submodel to look at). Let I_{min} be the minimum AIC model. Then I_I is the submodel with the fewest predictors such that $AIC(I_I) \leq AIC(I_{min}) + 2$ (for a given number of predictors r_I, only consider the submodel with the smallest AIC). Also look at models I_j with fewer predictors than I_I such that $AIC(I_j) \leq AIC(I_{min}) + 7$.

31) Submodels I with more predictors than I_{min} should not be used.

32) Submodels I with $AIC(I) > AIC(I_{min}) + 7$ should not be used.

33) Assume $n > 5p$, that the full PH model is reasonable and all predictors are equally important. The following rules of thumb for a good PH submodel I are in roughly decreasing order of importance. (But for factors with $a-1$ indicators, modify ix) and x) so that the indicator with the smallest pvalue is examined.)

i) Do not use more predictors than the min AIC model I_{min}.

ii) The slice survival plots for I looks like the slice survival plot for the full model.

iii) corr(ESP,ESP(I))≥ 0.95.

iv) The plotted points in the EE plot of ESP(I) vs ESP cluster tightly about the identity line.

v) Want pvalue ≥ 0.01 for the change in PLR test that uses I as the reduced model. (So for variable selection use $\delta = 0.01$ instead of $\delta = 0.05$.)

vi) Want the number of predictors $r_I \leq n/10$.

vii) Want $-2\log(L(\hat{\boldsymbol{\beta}}_I)) \geq -2\log(L(\hat{\boldsymbol{\beta}}_{full}))$ but close.

viii) Want $AIC(I) \leq AIC(I_{min}) + 7$.

ix) Want hardly any predictors with pvalues > 0.05.

x) Want few predictors with pvalues between 0.01 and 0.05.

34) Suppose that the full model is good and is stored in M1. Let M2, M3, M4, and M5 be candidate submodels found after forward selection, backward elimination, etc. Typically one of the submodels is the min(AIC) model. Given a list of properties of each submodel, be able to pick out the "best starting submodel."

Tips: i) submodels with more predictors then the min(AIC) submodel have too many predictors.

ii) The best starting submodel I_I has $AIC(I_I) \leq AIC(I_{min}) + 2$.

iii) Submodels I with $AIC(I) > AIC(I_{min}) + 2$ are not the best starting submodel.

iv) Submodels I with a pvalue < 0.01 for the change in PLR test have too few predictors.

v) The full model may be the best starting submodel if it is the min(AIC) model and M2–M5 satisfy iii). Similarly, then min(AIC) model may be the best starting submodel.

35) In addition to the best starting submodel I_I, submodels I with fewer predictors than I_I and $AIC(I) \leq AIC(I_{min}) + 7$ are worth considering. See HW6 1i.