Question #1:

Solution 1. *For part (A),*

we have the PH model as:

$$h(t, age, sex, age * sex) = h_0(t) \exp\left(\beta_1 age + \beta_2 sex + \beta_3 sex * age\right)$$

The following SAS code fit the data into this model:

```
dm 'log; clear; output; clear;';

libname data "C:\akira\data";

/*Question 1*/
/*part A fit a cox model with age, sex, age*sex*/

data Whas_1;
    set data.Whas;
    agesex = age * sex;
run;

proc phreg data=Whas_1;
    model LENFOL*FSTAT(0) = age sex agesex;
run;
```

The output is:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
AGE	1	0.05175	0.00730	50.2241	<.0000001	1.053	Age (years)
SEX	1	1.55289	0.79231	3.8414	0.0500012	4.725	1=Male/0=Female
agesex	1	-0.02040	0.01088	3.5147	0.0608266	0.980	

We see that the interaction between age and sex is only marginally significant, and the effect of sex is also marginally significant.

To find the most parsimonious model, we tried forward, backward and stepwise methods to select explanatory variables:

```
proc phreg data=Whas_1;
    model LENFOL*FSTAT(0) = age sex agesex/selection = FORWARD;
run;

proc phreg data=Whas_1;
    model LENFOL*FSTAT(0) = age sex agesex/selection = BACKWARD;
run;

proc phreg data=Whas_1;
    model LENFOL*FSTAT(0) = age sex agesex/selection = STEPWISE;
run;
```

All three methods got the same result, that only age is kept in the model.

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
AGE	1	0.04382	0.00530	68.3133	<.0000001	1.045	Age (years)

So the most parsimonious model is:

$$h(t, age) = h_0(t) \exp\Big(\beta_1 \cdot age\Big)$$

From the output we conclude that for every one year increase in age, the risk of death increase about 4.5% *at that time.*

Part (B)*:*

Since we are using the full model, we need to consider the interaction. Thus the hazard ratio for sex, adjusting for age will be:

$$\begin{aligned} HR &= \frac{h_0(t) \exp\Big(\beta_1 \cdot age + \beta_2 \cdot 1 + \beta_3 \cdot 1 \cdot age\Big)}{h_0(t) \exp\Big(\beta_1 \cdot age + \beta_2 \cdot 0 + \beta_3 \cdot 0 \cdot age\Big)} = \exp\Big(\beta_2 + \beta_3 \cdot age\Big) \\ &= \exp\Big(1.55289 - 0.02040 \times age\Big) \end{aligned}$$

For age at 50*,*

$$HR = \exp\Big(1.55289 - 0.02040 \times 50\Big) = 1.7038$$

For age at 60*,*

$$HR = \exp\Big(1.55289 - 0.02040 \times 60\Big) = 1.3894$$

For age at 65*,*

$$HR = \exp\Big(1.55289 - 0.02040 \times 65\Big) = 1.2547$$

For age at 70*,*

$$HR = \exp\Big(1.55289 - 0.02040 \times 70\Big) = 1.1330$$

For age at 80*,*

$$HR = \exp\Big(1.55289 - 0.02040 \times 80\Big) = 0.9239$$

Now for male, the hazard ratio for a 10 *year increase is:*

$$\begin{aligned} HR &= \frac{h_0(t)\exp\Big(\beta_1(age+10)+\beta_2\cdot 1+\beta_3\cdot 1\cdot(age\ +\ 10)\Big)}{h_0(t)\exp\Big(\beta_1\cdot age+\beta_2\cdot 1+\beta_3\cdot 1\cdot age\Big)} \\ &= \exp\Big(10\beta_1+10\beta_3\Big) = \exp\Big(10(\beta_1+\beta_3)\Big) \\ &= \exp\Big(10(0.05175-0.02040)\Big) \\ &= 1.3682 \end{aligned}$$

For female, the hazard ratio for a 10 *year increase in age is:*

$$\begin{aligned} HR &= \frac{h_0(t)\exp\Big(\beta_1(age+10)+\beta_2\cdot 0+\beta_3\cdot 0\cdot(age+10)\Big)}{h_0(t)\exp\Big(\beta_1(age)+\beta_2\cdot 0+\beta_3\cdot 0\cdot age\Big)} \\ &= \exp\Big(10\beta_1\Big) = \exp\Big(10\times 0.05175\Big) \\ &= 1.6778 \end{aligned}$$

For part (C):

The estimate for survival function after adjusting for covariates is given by:

$$\hat{S}(t) = \Big[\hat{S}_0(t)\Big]^{\exp\Big(\hat{\beta}_1 age+\hat{\beta}_2 sex+\hat{\beta}_3 age*sex\Big)}$$

where

$$\hat{S}_0(t) = exp\Big[-\hat{H}_0(t)\Big]$$

and

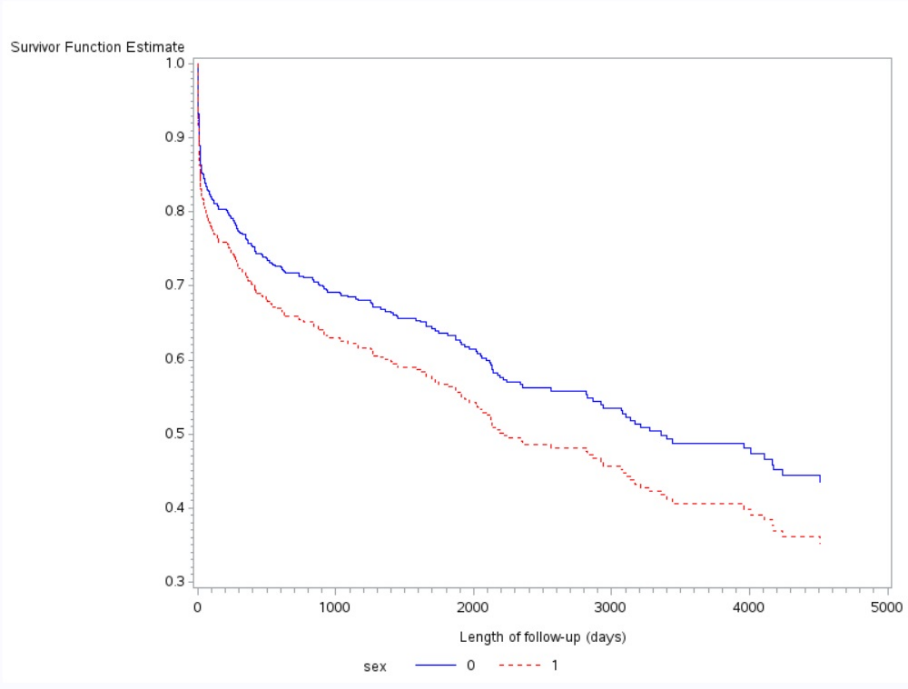
$$\hat{H}_0(t) = \sum_{t_i<t}\frac{1}{\sum_{j\in R(t_i)}\exp\Big(\hat{\beta}_1 age_j+\hat{\beta}_2 sex_j+\hat{\beta}_3 age_j*sex_j\Big)}\ \textit{(if there is no tie)}$$

$$\hat{H}_0(t) = \sum_{t_i<t}\frac{d_i}{\sum_{j\in R(t_i)}\exp\Big(\hat{\beta}_1 age_j+\hat{\beta}_2 sex_j+\hat{\beta}_3 age_j*sex_j\Big)}\ \textit{(if there are ties)}$$

The code to realize the above results is:

```
/*part C compute and graph estimated survival functions
for 65 year old males and females, and estimate median survival times*/

data groups;
    input age sex agesex;
    datalines;
    65 0 0
    65 1 65
    ;
run;

proc phreg data=Whas_1;
    model LENFOL*FSTAT(0) = age sex agesex;
    baseline covariates = groups out=cov_adj_surv
        survival = surv;
run;

proc print data=cov_adj_surv noobs;
    by sex;
run;

proc gplot data=cov_adj_surv;
    plot surv*LENFOL = sex;
    symbol1 I = STEPLJ COLOR = BLUE L = 1;
    symbol2 I = STEPLJ COLOR = Red L = 2;
run;
```

and the plot of the survival curves for male and female at age 65 is:



To estimate the median survival times, recall that by definition:

$$x_{0.5} = \inf\{t : S(t) \leq 0.5\} \text{ with estimate } \hat{x}_{0.5} = \inf\{t : \hat{S}(t) \leq 0.5\}$$

We print out the estimate for survival function at age 65 for both genders:

female:

65	0	3078	0.52631
65	0	3103	0.52173
65	0	3139	0.51712
65	0	3171	0.51252
65	0	3207	0.50795
65	0	3280	0.50335
65	0	3361	0.49850

male:

65	65	2136	0.51174
65	65	2142	0.50826
65	65	2177	0.50477
65	65	2187	0.50127
65	65	2217	0.49772
65	65	2238	0.49416
65	65	2335	0.49007

So the estimated median survival times are:

$$median_{female} = 3361, \qquad median_{male} = 2217$$

For part D:

We use AGE_{CAT} in place of AGE, and we also consider the interaction between AGE_CAT and SEX. Since there are 4 categories in AGE_CAT, in order for SAS to be able to handle it, we create our own dummy variables (thus having more control over the class statement).

We define:

$$CAT1 = \begin{cases} 1 & \text{if } AGE\_CAT = 1 \\ 0 & \text{otherwise} \end{cases} \qquad CAT2 = \begin{cases} 1 & \text{if } AGE\_CAT = 2 \\ 0 & \text{otherwise} \end{cases}$$

$$CAT3 = \begin{cases} 1 & \text{if } AGE\_CAT = 3 \\ 0 & \text{otherwise} \end{cases}$$

Hence we are using the last age category($4 = 78+$) as a reference category and our model is:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot CAT1 + \beta_2 \cdot CAT2 + \beta_3 \cdot CAT3 + \beta_4 \cdot SEX + \beta_5 \cdot CAT1 \cdot SEX + \beta_6 \cdot CAT2 \cdot SEX + \beta_7 \cdot CAT3 \cdot SEX\Big)$$

The following SAS code fit the above model to our data:

```
/*part D fit a cox model with agecat, sex, agecat*sex*/
data Whas_2;
    set data.Whas;
    IF age_cat = 1 THEN cat1 = 1;
    ELSE cat1 = 0;
    IF age_cat = 2 THEN cat2 = 1;
    ELSE cat2 = 0;
    IF age_cat = 3 THEN cat3 = 1;
    ELSE cat3 = 0;
    cat1_sex = cat1*sex;
    cat2_sex = cat2*sex;
    cat3_sex = cat3*sex;
run;

proc print data=Whas_2;
run;

proc phreg data=Whas_2;
    model LENFOL*FSTAT(0) = cat1 cat2 cat3 sex cat1_sex cat2_sex cat3_sex;
run;

proc phreg data=Whas_2;
    model LENFOL*FSTAT(0) = cat1 cat2 cat3 sex cat1_sex cat2_sex cat3_sex/selection = stepwise;
run;

proc phreg data=Whas_2;
    model LENFOL*FSTAT(0) = cat1 cat2 cat3 sex cat1_sex cat2_sex cat3_sex/selection = forward;
run;

proc phreg data=Whas_2;
    model LENFOL*FSTAT(0) = cat1 cat2 cat3 sex cat1_sex cat2_sex cat3_sex/selection = backward;
run;
```

We would like to mention that, if we do not want to manually create the dummy variables as above, equivalently we could let SAS to do it for us with the class statement. The following code will produce the exact same output:

```
proc phreg data=data.Whas;
    class age_cat;
    model LENFOL*FSTAT(0) = age_cat sex age_cat*sex;
run;
```

and we can check how SAS create the dummy variable by default:

Class Level Information				
Class	Value	Design Variables		
AGE_CAT	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	0	0	0

which is the same as our definition above.

The output is:

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
cat1	1	-1.98363	0.25424	60.8729	<.0000001	0.138	
cat2	1	-1.46588	0.24664	35.3240	<.0000001	0.231	
cat3	1	-0.99208	0.23384	17.9998	0.0000221	0.371	
SEX	1	-0.54991	0.21899	6.3056	0.0120356	0.577	1=Male/0=Female
cat1_sex	1	0.83654	0.42902	3.8020	0.0511919	2.308	
cat2_sex	1	0.75854	0.35767	4.4978	0.0339386	2.135	
cat3_sex	1	0.93861	0.32049	8.5772	0.0034039	2.556	

All main effect terms are significant, and among the interaction terms, only $cat1 * SEX$ *is marginally significant(p value* 0.05*) and the others are also significant.*

If we add option selection $= \ldots$ *after the model statement, it will only keep the three main effect* $cat1 - 3$ *and get rid of sex(regardless if we use stepwise, forward or backward). Since all covariates and interactions are significant, we decide to keep the full model, consider we are interested in the hazard ratio of SEX later.*

Now for age group $[24, 60)$*, the hazard ratio for SEX is:*

$$\begin{aligned} HR &= \frac{h_0(t)\exp\Big(\beta_1 \cdot 1 + \beta_4 \cdot 1 + \beta_5 \cdot 1 \cdot 1\Big)}{h_0(t)\exp\Big(\beta_1 \cdot 1 + \beta_4 \cdot 0 + \beta_5 \cdot 1 \cdot 0\Big)} \\ &= \exp\Big(\beta_4 + \beta_5\Big) = \exp\Big(-0.54991 + 0.83654\Big) \\ &= 1.332 \end{aligned}$$

For age group $[60, 69)$*, the hazard ratio for SEX is:*

$$\begin{aligned} HR &= \frac{h_0(t)\exp\Big(\beta_2 \cdot 1 + \beta_4 \cdot 1 + \beta_6 \cdot 1 \cdot 1\Big)}{h_0(t)\exp\Big(\beta_2 \cdot 1 + \beta_4 \cdot 0 + \beta_6 \cdot 1 \cdot 0\Big)} \\ &= \exp\Big(\beta_4 + \beta_6\Big) = \exp\Big(-0.54991 + 0.75854\Big) \\ &= 1.232 \end{aligned}$$

For age group $[69, 78)$*, the hazard ratio for SEX is:*

$$\begin{aligned} HR &= \exp\Big(\beta_4 + \beta_7\Big) = \exp\Big(-0.54991 + 0.93861\Big) \\ &= 1.475 \end{aligned}$$

For age group $78+$*, the hazard ratio for SEX is:*

$$HR = \frac{h_0(t) \cdot \exp\Big(\beta_4 \cdot 1\Big)}{h_0(t) \cdot \exp\Big(\beta_4 \cdot 0\Big)} = \exp\Big(\beta_4\Big) = \exp\Big(-0.54991\Big) = 0.5770$$

For part E:

I would prefer to report the results from model in part-D. The model in part-A says that only effect of age is significant, while the sex and interaction between sex and age are only marginally significant. The model in part-D conclude that age, sex and their interactions are all significant. From the plot in part-C we see that apparently female tends to have a longer survival time than male, given age is 65*. So adjusting for age, we are expecting to see a significant difference on survival time between different sex. This is better supported by the model in part-D.*

Since the model we choose is:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot CAT1 + \beta_2 \cdot CAT2 + \beta_3 \cdot CAT3 + \beta_4 \cdot SEX + \beta_5 \cdot CAT1 \cdot SEX + \beta_6 \cdot CAT2 \cdot SEX + \beta_7 \cdot CAT3 \cdot SEX\Big)$$

We run the following SAS code to acquire partial maximum likelihood estimate:

```
/*part E*/
proc phreg data=Whas_2;
    model LENFOL*FSTAT(0) = cat1 cat2 cat3 sex cat1_sex cat2_sex cat3_sex/risklimits;
run;
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
cat1	1	-1.98363	0.25424	60.8729	<.0000001	0.138	0.084	0.226	
cat2	1	-1.46588	0.24664	35.3240	<.0000001	0.231	0.142	0.374	
cat3	1	-0.99208	0.23384	17.9998	0.0000221	0.371	0.234	0.586	
SEX	1	-0.54991	0.21899	6.3056	0.0120356	0.577	0.376	0.886	1=Male/0=Female
cat1_sex	1	0.83654	0.42902	3.8020	0.0511919	2.308	0.996	5.352	
cat2_sex	1	0.75854	0.35767	4.4978	0.0339386	2.135	1.059	4.304	
cat3_sex	1	0.93861	0.32049	8.5772	0.0034039	2.556	1.364	4.791	

we can then compute the 95% *confidence limits for the regression coefficients:*

$$CI_{\beta_1} : \Big(\log(0.084), \log(0.226)\Big) = \Big(-2.4769, -1.4872\Big)$$
$$CI_{\beta_2} : \Big(\log(0.142), \log(0.374)\Big) = \Big(-1.9519, -0.9835\Big)$$
$$CI_{\beta_3} : \Big(\log(0.234), \log(0.586)\Big) = \Big(-1.4524, -0.5344\Big)$$
$$CI_{\beta_4} : \Big(\log(0.376), \log(0.886)\Big) = \Big(-0.9782, -0.1210\Big)$$
$$CI_{\beta_5} : \Big(\log(0.996), \log(5.352)\Big) = \Big(-0.0040, 1.6775\Big)$$
$$CI_{\beta_6} : \Big(\log(1.059), \log(4.304)\Big) = \Big(0.05733, 1.4595\Big)$$
$$CI_{\beta_7} : \Big(\log(1.364), \log(4.791)\Big) = \Big(0.3104, 1.5667\Big)$$

We also have the global test for $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	89.2000	7	<.0000001
Score	101.2329	7	<.0000001
Wald	87.7304	7	<.0000001

For part (F):

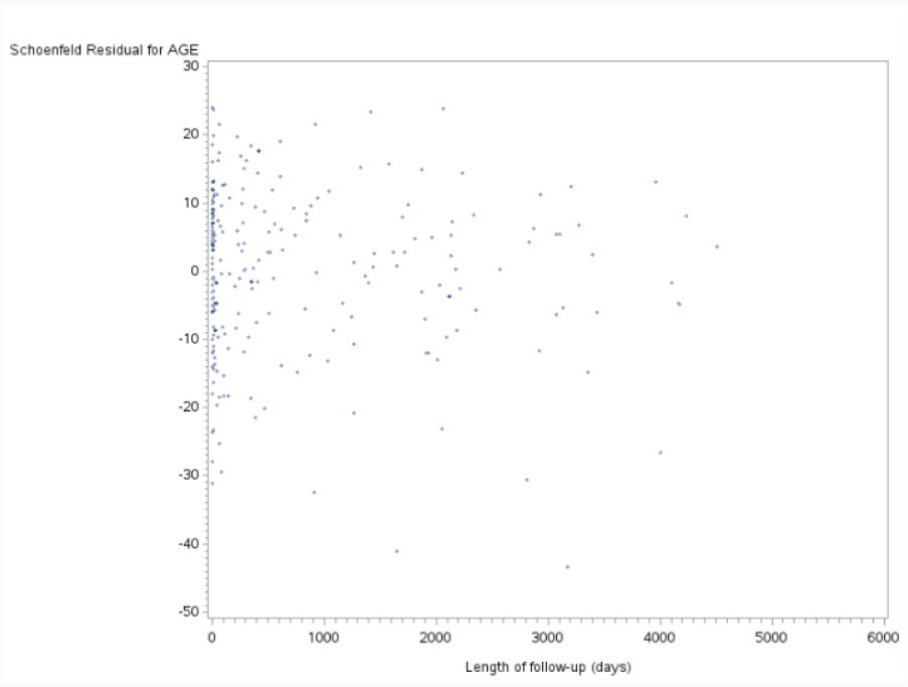
1) We plot Schoenfeld Residual for age versus the length of follow up. Since there is no obvious pattern, we find no violation of the PH model assumption for age.

2) We plot log-log survival functions versus log of length of follow up for both male and female. Since the plots are parallel, we find no violation of the PH model assumption for sex.

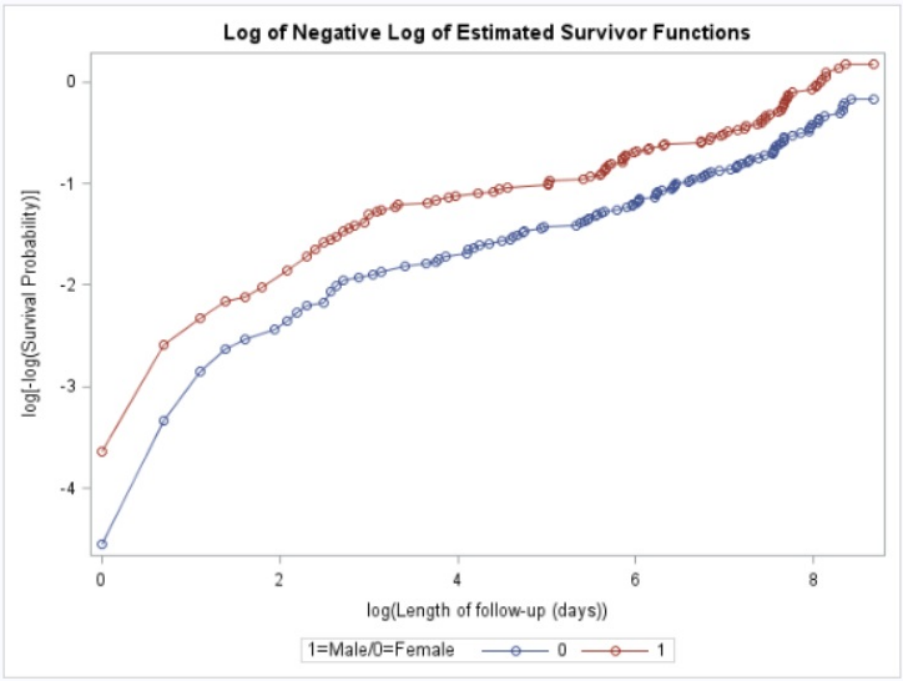
The code for 1) *and* 2) *is:*

```
/*part F validate PH model*/
/*model include sex and age*/

/*use Schoenfeld Residual test checking on age*/
/*no apparent pattern, PH model valid for age*/
proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = age sex;
    output out=resid1
        RESSCH = SCHAGE SCHSEX;
run;

proc gplot data=resid1;
    plot SCHAGE*LENFOL;
    symbol1 value = circle H=.5;
run;

/*use log-log survival functions to check on sex*/
/*two curves parallel, PH model valid for sex*/
proc lifetest data=data.Whas notable
    plots = (LLS) graphics cs=none;
    time LENFOL*FSTAT(0);
    strata sex;
run;
```

The plot for 1) *is:*



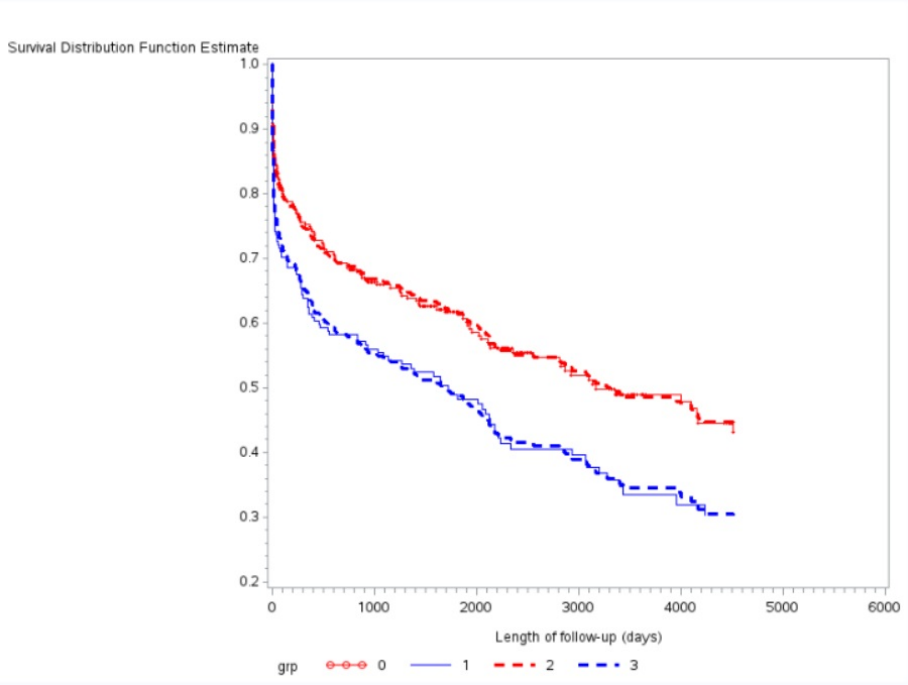
The plot for 2) *is:*



We can further check the PH model assumption with:

3) We plot the observed and expected survival function for both male and female to check on sex.

The plot is:



generated by the following code:

```
to check on sex*/
/*for both sex, the observed and expected plots are very close,
PH model is valid for sex.*/

proc lifetest data=data.Whas notable
    outsurv = observed;
    time LENFOL*FSTAT(0);
    strata sex;
run;

data covs;
    input sex @@;
    datalines;
    0 1
    ;
run;

proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = sex;
    baseline covariates = COVS out=expected
        survival = survival/nomean;
run;


data combined;
    set observed(IN = IN1) expected(IN = IN2);
    if IN1 Then grp = sex;
    else if IN2 then grp = sex + 2;
run;

proc print data=combined;
run;

proc gplot data=combined;
    plot survival*LENFOL = grp;
    symbol1 I = STEPLJ COLOR = RED L = 1;
    symbol2 I = STEPLJ COLOR = BLUE L = 1;
    symbol3 I = STEPLJ COLOR = RED L = 2 W = 2;
    symbol4 I = STEPLJ COLOR = BLUE L = 2 W = 2;
run;
```

For both male and female, the observed and expected survival functions are pretty close, so we find no violation of the PH model assumption for sex.

4) We consider the interaction with time for both age and sex. The code is:

```
/*use interaction with time to check on age and sex
we add linear interaction term here*/
/*both interaction terms are not significant, appear to be no
violation of PH assumption*/

proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = age sex aget sext;
    aget = age*LENFOL;
    sext = sex*LENFOL;
run;
```

The output is:

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
AGE	1	0.04687	0.00672	48.7004	<.0000001	1.048	Age (years)
SEX	1	0.10863	0.15995	0.4613	0.4970269	1.115	1=Male/0=Female
aget	1	-5.4977E-6	5.31812E-6	1.0687	0.3012481	1.000	
sext	1	-0.0000218	0.0001262	0.0299	0.8626044	1.000	

For both age and sex, the interactions with time are not significant, hence we find no violation of PH assumption for both age and sex.

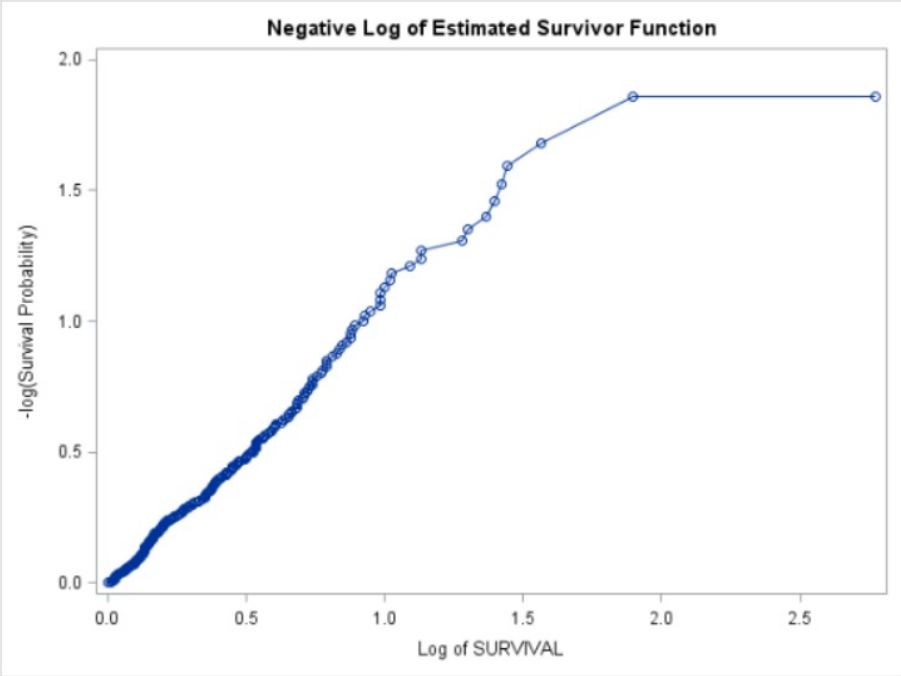
To conclude, 1)-4) all support the PH model assumption we have made for age and sex.

For part (G):

To test the goodness of fit, we first look at the log-survival plot of coxsnell residuals given by the following code:

```
/*use interaction with time to check on age and sex
we add linear interaction term here*/
/*both interaction terms are not significant, appear to be no
violation of PH assumption*/

proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = age sex aget sext;
    aget = age*LENFOL;
    sext = sex*LENFOL;
run;


/*part G testing goodness of fit, outliers,
influential observations*/

/*log survival plot of coxsnell residuals*/
proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = age sex;
    output out=resid logsurv = coxsnell;
run;

data resid2;
    set resid;
    coxsnell=-coxsnell;
run;

proc lifetest data=resid2 plots = (LS) notable;
    time coxsnell*FSTAT(0);
run;
```

The plot is:



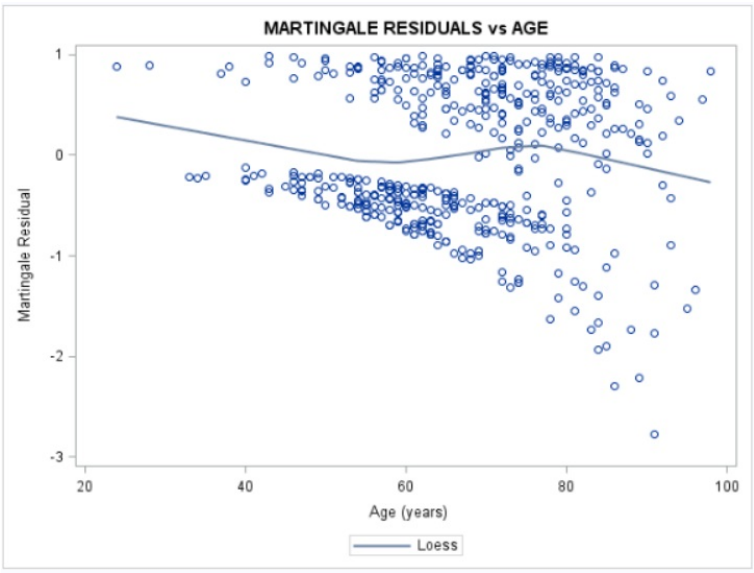
It is not deviating much from the straightline, so the coxsnell residual test tells us that the PH model from part F we fitted is adequate.

To see whether the functional form of age is appropriate (we are assuming linear form here, namely, $\beta_1 * age$), we consider using martingale residuals. We plot martingale residuals against age, along with a LOESS smoothing function. We also use ASSESS keyword in proc phreg to plot the observed cumulative martingale residual for age together with 20 simulated realizations, as well as a 4-panel display (observed cumulative martingale residual process for age together with the first 8 siulated realizations).

The code is:

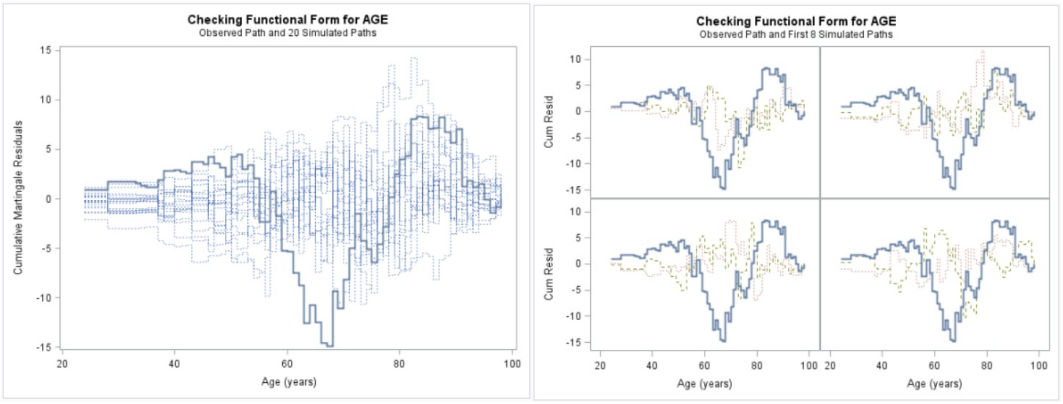
```
/*use martingale residual to check on functional form of age*/
proc phreg data=data.Whas;
    class sex;
    model LENFOL*FSTAT(0) = age sex;
    output out=MARTIN RESMART = MARTINGALE;
run;

/*no particular pattern. Current functional form of age seems
to be appropriate*/
proc sgplot data=MARTIN;
    title "MARTINGALE RESIDUALS vs AGE";
    LOESS Y = MARTINGALE X = age;
run;

/*observed cumulative martingale residuals within the spectrum
of the simulated realizations, do not see problem using age
in its current form*/
proc phreg data=data.Whas;
    class sex;
    model LENFOL*FSTAT(0) = age sex;
    ASSESS var = (age)/CRPANEL;
run;
```

The martingale residual against age plot is:



There does not seem to be an obvious pattern, so we conclude the current functional form for age is appropriate.

The cumulative martingale residual process for age is:



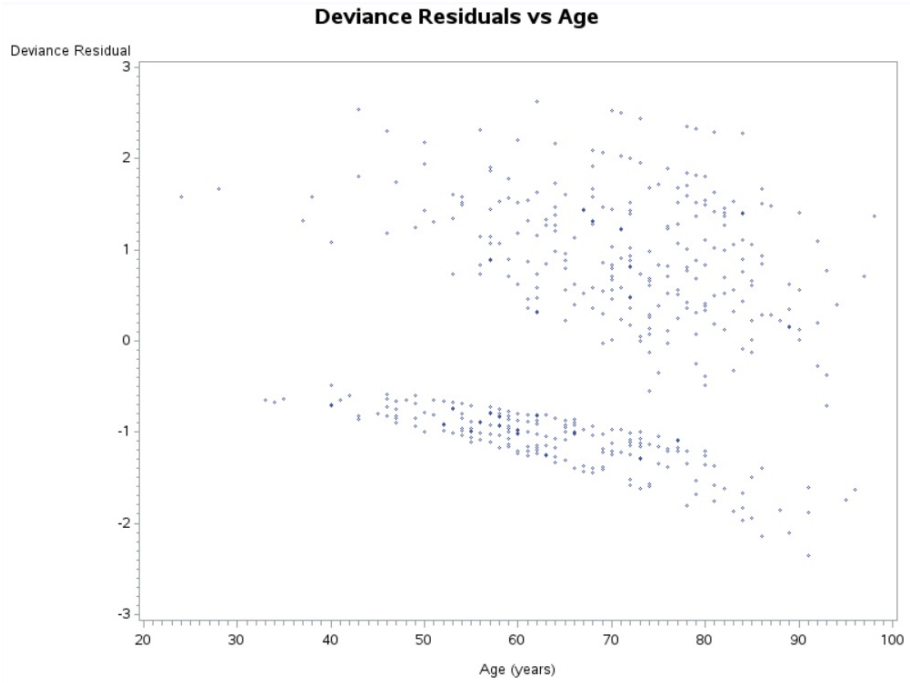
As we can see the observed cumulative martingale residual process is within the spectrum of simulated ones, again supporting the current functional form of age.

We next identify the outliers by checking deviance residuals:

We sketch the deiviance residual versus age to see the adequacy of fit for age, and use proc univariate to identity extreme observations(outliers).

The SAS code and output is:

```
/*use deviance residuals to identify outliers*/
proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = age sex;
    output out=RESID3 RESDEV = DEV;
run;

proc univariate data=RESID3;
    var dev;
run;

title "Deviance Residuals vs Age";
proc gplot data=RESID3;
    plot DEV*AGE;
    SYMBOL1 VALUE=circle H=.5;
run;
```



Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.35468	155	2.44010	336
-2.14327	62	2.50748	412
-2.10461	228	2.52423	161
-1.96894	60	2.54109	261
-1.94721	206	2.62329	1

The residual plot seem to show that there are more observations with positive values (shorter survival times than expected), but there is no obvious pattern to indicate any trend. We still think that age

is adequately fit.

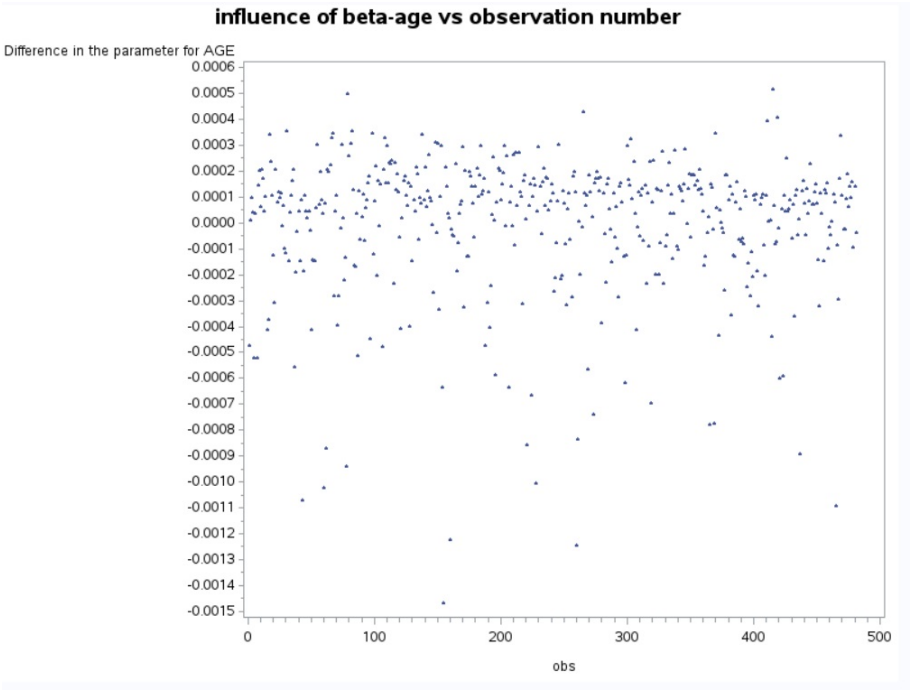
The proc univariate output finds outliers for us as shown above.

Finally we use the DFBETA keyword under the output statement of proc phreg to identify influential observations.Keep in mind that DFBETA compute $\hat{\beta} - \hat{\beta}_{(j)}$ *where* $\hat{\beta}_{(j)}$ *is the regression coefficient estimate with observation j omitted. A positive value means higher hazard rate estimate with observation j involved compared to without, hence observation j refers to an individual who has moderate values of the covariate corresponding to* β *but experience event earlier than expected. On the other hand, a negative value means lower hazard rate estimate with observation j involved compared to without, hence observation j refers to an individual who has small values of the covariate but experience the event at a time longer than expected.*

The SAS code for computing the $\hat{\beta} - \hat{\beta}_{(j)}$ *corresponding to each of the covariate age and sex is:*

```
/*use DFBETA keyword to find influential observations*/
proc phreg data=data.Whas;
    model LENFOL*FSTAT(0) = age sex;
    output out=RESID4 DFBETA=beta_age beta_sex;
run;

data resid4;
    set resid4;
    obs = _n_;
run;

proc print data=resid4;
run;

proc gplot data=resid4;
    title "influence of beta-age vs observation number";
    plot beta_age*obs;
    symbol1 value=dot H = .5;
run;
```

we also plot the difference in the parameter for age versus observation number as:



The list of $\hat{\beta} - \hat{\beta}_{(j)}$ *for all* 481 *observations are too long, so we did not print it out here in our summary. But we printed out it in SAS and find the following influential obervations(we are focusing on age because sex is not significnat in our model):*

Observation $43, 60, 160, 260$ *and* 465 *are with relatively large negative values.*

We actually did not find any influential observations with relatively large positive differences on $\hat{\beta} - \hat{\beta}_{(j)}$ *for age, at least not on the same level of those negative differences.*

*Thus completed the solution for Question #*1.

Question #2:

Solution 2. *For part* (A)*, our PH model is:*

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot age + \beta_2 \cdot drug\Big)$$

Although drug only takes values on either 50 *or* 100*, we are treating it as a continuous covariate here. The SAS code is as following:*

```
title "pneumonia cure time analysis";
proc phreg data = data.Pneumonia;
    model time*cured(0) = age drug;
run;
```

and we got the following estimate:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
AGE	1	-0.13584	0.02964	20.9981	0.0000046	0.873
DRUG	1	0.0004632	0.00635	0.0053	0.9418127	1.000

The effect of drug is **NOT** *significant at the* 0.05 *significance level(p value* 0.94*). If we ignore this fact that the drug as a continuous covariate is highly insignificant, we do have a point estimate that says* $\hat{\beta}_2 = 0.0004632$*, which leads to a hazard ratio adjusting for age as* $\exp\left(0.0004632\right) \simeq 1.000$*. This says that for each single unit increase in the drug dosage, the chance of cure(risk of event) is about the same as before the increase, namely, the drug is useless.*

For part (B)*:*

Take the amount of drug remaining in the body by the end of study as a covariate, our model is:

$$h(t) = h_0(t) \exp\left(\beta_1 \cdot age + \beta_2 \cdot remaining\ drug\right)$$

The following SAS code fit this model:

```
/*part B treating a new variable recording the
remaining drug dosage by end of study:
remain = initial dosage *exp(-0.35t)*/

data Pneumonia_b;
    set data.Pneumonia;
    remain = drug*exp(-0.35*time);
run;

proc print data=Pneumonia_b;
run;

proc phreg data=Pneumonia_b;
    model time*cured(0) = age remain;
run;
```

and the output is:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
AGE	1	-0.13893	0.03963	12.2911	0.0004551	0.870
remain	1	0.54873	0.09829	31.1693	<.0000001	1.731

under the new model the effect of drug (remaining amount in the body) is significant at the 0.05 *level. The point estimate for the drug effect here is* $\hat{\beta}_2 = 0.54873$ *which leads to a hazard ratio adjusted for age as* $\exp(0.54873) = 1.731$*. It tells that for every unit increase of remaining drug dosage in the body by the end of study, the chance of cure(risk of event) is about* 1.731 *times of the chance without increase.*

For part (C):

Our model is:

$$h(t) = h_0(t)\exp\Big(\beta_1 \cdot age + \beta_2 \cdot initial\ dosage \cdot \exp(-0.35t)\Big)$$

Thus we are looking at a model with time dependent covariate

The code is:

```
/*part C consider different effects over time for
different initial dossages*/

proc phreg data=data.Pneumonia;
    model time*cured(0) = age remain;
    remain = 0;
    remain = drug * exp(-0.35*time);
run;
```

and the output is:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
AGE	1	-0.15799	0.03413	21.4318	0.0000037	0.854
remain	1	0.17855	0.08220	4.7187	0.0298356	1.195

To compute the hazard ratio between patient with an initial dosage level of 100 *mg and patient with an initial dosage level of* 50 *mg at day* 7*, adjusting for age (assuming two patients at same age), we have:*

$$\begin{aligned} HR &= \frac{h_0(t)\exp\Big(\hat{\beta}_1 \cdot age + \hat{\beta}_2 \cdot 100 \cdot \exp\Big(-0.35 \times 7\Big)\Big)}{h_0(t)\exp\Big(\hat{\beta}_1 \cdot age + \hat{\beta}_2 \cdot 50 \cdot \exp\Big(-0.35 \times 7\Big)\Big)} = \exp\Big(\hat{\beta}_2 \cdot 50 \cdot \exp\Big(-0.35 \times 7\Big)\Big) \\ &= \exp\Big(0.17855 \cdot 50 \cdot \exp\Big(-0.35 \times 7\Big)\Big) \\ &= 2.1606 \end{aligned}$$

This is to say, at 7 *days, the patient having initial dosage of* 100 *mg has* 2.1606 *times the chance of cure of patient having initial dosage of* 50*mg, adjusting for age.*

Similarly, we can compute the hazard ratio at day 14*:*

$$\begin{aligned} HR &= \exp\Big(\hat{\beta}_2 \cdot 50 \cdot \exp\Big(-0.35 \times 14\Big)\Big) \\ &= \exp\Big(0.17855 \cdot 50 \cdot \exp\Big(-0.35 \times 14\Big)\Big) \\ &= 1.0687 \end{aligned}$$

We conclude that at 14 *days, the patient having initial dosage of* 100 *mg has* 1.0687 *times the chance of cure of patient having initial dosage of* 50 *mg.*

*Thus complete the answers for question #*2*.*

Question #3:

Solution 3. *For part (A):*

The extended Cox model is:

$$h(t) = h_0(t) \exp\Big(\beta_1 \cdot RX + \beta_2 \cdot LOGWBC + \beta_3 \cdot SEX \times TIME(t)\Big)$$

Here TIME is a new variable defined as:

$$TIME(t) = \begin{cases} 1 & if\ t < 4 \\ 3 & if\ 4 \le t < 8 \\ 5 & if\ 8 \le t < 12 \\ 7 & if\ 12 \le t < 16 \\ 9 & if\ t \ge 16 \end{cases}$$

and then SEX$\times$ TIME(t) is a time dependent covariate as given by the question.

For part (B):

Adjusted for RX and LOGWBC, the hazard ratio for the effect of SEX is:

$$HR_{SEX} = \frac{h_0(t) \exp\Big(\beta_1 \cdot RX + \beta_2 \cdot LOGWBC + \beta_3 \cdot 1 \cdot TIME(t)\Big)}{h_0(t) \exp\Big(\beta_1 \cdot RX + \beta_2 \cdot LOGWBC + \beta_3 \cdot 0 \cdot TIME(t)\Big)}$$

$$= \exp\Big(\beta_3 \cdot TIME(t)\Big) = \begin{cases} \exp\Big(\beta\Big) & if\ t < 4 \\ \exp\Big(3\beta\Big) & if\ 4 \le t < 8 \\ \exp\Big(5\beta\Big) & if\ 8 \le t < 12 \\ \exp\Big(7\beta\Big) & if\ 12 \le t < 16 \\ \exp\Big(9\beta\Big) & if\ t \ge 16 \end{cases}$$

For part (C):

To fit the model developed from part (A), we present the following SAS code:

```
/*Question 3*/

/*part C fit the model developed in part A*/
/*give estimate of the hazard ratios in part B*/
proc phreg data=data.Leukemiab;
    model weeks*relapse(0)= RX LOGWBC SEX*TIME;
    TIME = 0;
    IF weeks < 4 then TIME= 1;
    ELSE IF 4 <= weeks < 8  then TIME = 3;
    ELSE IF 8 <= weeks < 12 then TIME = 5;
    ELSE IF 12 <= weeks < 16 then TIME = 7;
    ELSE IF weeks >= 16 then TIME = 9;
run;
```

The output is:

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
RX	1	1.20367	0.47677	6.3738	0.0115815	3.332	
LOGWBC	1	1.59479	0.33007	23.3452	0.0000014	4.927	
TIME*SEX	1	-0.03549	0.08986	0.1560	0.6928970	.	TIME * SEX

We have an estimate $\hat{\beta}_3 = -0.03549$. *To estimate the hazard ratio for the effect of sex, using the formula we have developed in part B, we got:*

$$\hat{HR}_{SEX} = \exp\left(\hat{\beta}_3 \cdot TIME(t)\right) = \begin{cases} \exp\left(\hat{\beta}\right) & \text{if } t < 4 \\ \exp\left(3\hat{\beta}\right) & \text{if } 4 \le t < 8 \\ \exp\left(5\hat{\beta}\right) & \text{if } 8 \le t < 12 \\ \exp\left(7\hat{\beta}\right) & \text{if } 12 \le t < 16 \\ \exp\left(9\hat{\beta}\right) & \text{if } t \ge 16 \end{cases} = \begin{cases} 0.9651 & \text{if } t < 4 \\ 0.8990 & \text{if } 4 \le t < 8 \\ 0.8374 & \text{if } 8 \le t < 12 \\ 0.7800 & \text{if } 12 \le t < 16 \\ 0.7266 & \text{if } t \ge 16 \end{cases}$$

For part (D):

To fit the cox model stratified on sex, we present the following SAS code:

```
/*part D fit a cox model stratified on sex*/
proc sort data=data.Leukemiab out=Leukemiab_sorted;
    by SEX;
run;

proc phreg data=Leukemiab_sorted;
    by SEX;
    model weeks*relapse(0)= RX LOGWBC;
run;
```

The output for female is:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
RX	1	0.26672	0.56592	0.2221	0.6374245	1.306
LOGWBC	1	1.17012	0.49857	5.5083	0.0189269	3.222

The output for male is:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
RX	1	1.85904	0.72910	6.5013	0.0107794	6.418
LOGWBC	1	1.63890	0.51903	9.9704	0.0015908	5.150

For part (E):

The hazard ratio of RX from part C is:

$$\hat{HR}_{C,RX} = 3.332$$

while the hazard ratio of RX from part D, depends on the gender, is:

$$\hat{HR}_{D,RX,male} = 6.418$$
$$\hat{HR}_{D,RX,female} = 1.306$$

It appears that we got very different hazard ratios between part C and part D. However we would not call out either one of them as being more appropriate. The reason is that part C and part D are taking different point of views. Part C is modeling the effect of sex as a time dependent covariate, so when we are looking at the hazard ratio for RX, we are adjusting for the effect of sex as well. However for part D, the sex is not a covariate in the model, instead we stratified on the sex and only consider the hazard ratio of RX on the same gender. So the hazard ratios between C and D are not comparable.

Thus complete the answers for question #3.

Question #4.

Solution 4. *For part (A):*

For Model 1, we have all 5 covariates in the model and the cox regression model is:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot PLTLTS + \beta_2 \cdot AGE + \beta_3 \cdot SEX + \beta_4 \cdot PLTAGE + \beta_5 \cdot PLTSEX\Big)$$

For model 2, three covariates PLTLTS, AGE and SEX are included:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot PLTLTS + \beta_2 \cdot AGE + \beta_3 \cdot SEX\Big)$$

For model 3, only two covariates PLTLTS and AGE are included:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot PLTLTS + \beta_2 \cdot AGE\Big)$$

For model 4, only two covariates PLTLTS and SEX are included:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot PLTLTS + \beta_2 \cdot SEX\Big)$$

For model 5, only one covariate PLTLTS is included:

$$h(t) = h_0(t) \cdot \exp\Big(\beta_1 \cdot PLTLTS\Big)$$

For part (B):

To see the hazard ratio for the effect of platelet variable (adjusted for age and sex):

For model 1:

If SEX= 1 (male), then we have:

$$\begin{aligned} HR &= \frac{h_0(t)\cdot\exp\Big(\beta_1\cdot 1+\beta_2\cdot AGE+\beta_3\cdot 1+\beta_4\cdot 1\cdot AGE+\beta_5\cdot 1\cdot 1\Big)}{h_0(t)\cdot\exp\Big(\beta_1\cdot 0+\beta_2\cdot AGE+\beta_3\cdot 1+\beta_4\cdot 0\cdot AGE+\beta_5\cdot 0\cdot 1\Big)} \\ &= \exp\Big(\beta_1+\beta_4\cdot AGE+\beta_5\Big) \end{aligned}$$

If SEX= 0 (female), then we have:

$$\begin{aligned} HR &= \frac{h_0(t)\cdot\exp\Big(\beta_1\cdot 1+\beta_2\cdot AGE+\beta_3\cdot 0+\beta_4\cdot 1\cdot AGE+\beta_5\cdot 1\cdot 0\Big)}{h_0(t)\cdot\exp\Big(\beta_1\cdot 0+\beta_2\cdot AGE+\beta_3\cdot 0+\beta_4\cdot 0\cdot AGE+\beta_5\cdot 0\cdot 0\Big)} \\ &= \exp\Big(\beta_1+\beta_4\cdot AGE\Big) \end{aligned}$$

For model 2:

since there is no interaction, we simply have:

$$\begin{aligned} HR &= \frac{h_0(t)\cdot\exp\Big(\beta_1\cdot 1+\beta_2\cdot AGE+\beta_3\cdot SEX\Big)}{h_0(t)\cdot\exp\Big(\beta_1\cdot 0+\beta_2\cdot AGE+\beta_3\cdot SEX\Big)} \\ &= \exp\Big(\beta_1\Big) \end{aligned}$$

For model 3, 4 and 5:

since there is no interaction, we would get the same result for hazard ratio on effect of PLTLTS, which is:

$$HR = \exp\Big(\beta_1\Big)$$

For part (C):

Adjusting for AGE = 40 and SEX = 1, the hazard ratios are:

For model 1:

$$\begin{aligned} \hat{HR} &= \exp\Big(\hat{\beta}_1+\hat{\beta}_4\cdot AGE+\hat{\beta}_5\Big) \\ &= \exp\Big(\hat{\beta}_1+40\hat{\beta}_4+\hat{\beta}_5\Big) \\ &= \exp\Big(0.470+40\times(-0.008)-0.503\Big) \\ &= 0.7026 \end{aligned}$$

For model 2:

$$\hat{HR} = \exp\Big(\hat{\beta}_1\Big) = \exp\Big(-0.725\Big) = 0.4843$$

For model 3*:*

$$\hat{HR} = \exp\Big(\hat{\beta}_1\Big) = \exp\Big(-0.706\Big) = 0.4936$$

For model 4*:*

$$\hat{HR} = \exp\Big(\hat{\beta}_1\Big) = \exp\Big(-0.705\Big) = 0.4941$$

For model 5*:*

$$\hat{HR} = \exp\Big(\hat{\beta}_1\Big) = \exp\Big(-0.694\Big) = 0.4996$$

For part D:

According to the output of model 1*, the effect of both interactions are not significant since one is with p value* 0.850 *and the other is with p value* 0.532*.*

For part E:

according to model 2 *to* 5*, PLTLTS is only marginally significant regardless controlling for SEX and(or) AGE or not, and SEX and AGE effect are not significant in these models, hence we do not need to control age and(or) sex as confounders.*

For part F:

I would choose to report model 5*. First of all as we explained in part E that the other covariates do not need to be controlled for potential confounders, also from part D we see no interaction effects.*

For part G:

Now we look at the output of model 5*, it reports that the effect of platelet on survival is marginally significant (with a p value of* 0.08*).With a hazard ratio of* 0.500*, the chance of death with normal platelets at diagnosis is half of those with abnormal platelets.*

*Thus completed the answers for question #*4*.*

Question #5.

Solution 5. *For part* (*A*)*:*

I hereby confirm that I have understood both ways using statement inside proc phreg and using counting process input while dealing with time-dependent covariates.

For part (*B*)*:*

For part (*i*)*:*

we run the following code which is copied from page 165*:*

```
/*Question 5*/

/*part B*/
/*(i) run the code given on page 165*/
data RECIDCUM;
    set data.Recid;
    ARRAY emp(*) emp1-emp52;
    ARRAY cum(*) cum1-cum52;
    cum1 = emp1;
    DO i = 2 to 52;
        cum(i) = (cum(i - 1)*(i - 1) + emp(i))/i;
    END;
run;

proc phreg data=recidcum;
    where week > 1;
    model week*arrest(0) = fin age race wexp mar paro prio employ/TIEs = EFRON;
    array cumemp(*) cum1-cum52;
    employ = cumemp[week-1];
run;
```

The output is:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.36509	0.19183	3.6223	0.0570114	0.694
age	1	-0.05128	0.02217	5.3491	0.0207329	0.950
race	1	0.30021	0.30884	0.9449	0.3310173	1.350
wexp	1	-0.05113	0.21520	0.0564	0.8122104	0.950
mar	1	-0.36330	0.38366	0.8967	0.3436746	0.695
paro	1	-0.02874	0.19716	0.0213	0.8840881	0.972
prio	1	0.09120	0.02881	10.0232	0.0015458	1.095
employ	1	-0.68170	0.30917	4.8618	0.0274576	0.506

The financial aid is marginally significant (p value 0.057*), those who received financial aid has* 0.694 *as much chance of being arrested as those who did not.*

Age is significant (p value 0.02*), and for every one year older in age, there is* 95% *as much chance of being arrested as those who are one year younger.*

Race is not significant(p value 0.33*), but it does predict that those who are black has* 1.35 *as much chance of getting arrested as those who are not black.*

Working experience is not significant (p value 0.81*).*

Marital status is not significant (p value 0.34*).*

Parole status is not significant (p value 0.88*).*

Number of convictions is significant (p value 0.0015*) and people with one more convinction will have about* 9.5% *more chance of getting arrested.*

cumulative employemnt proportion is significant (p value 0027*).*

For part (ii):

The following SAS code handles the problem fully inside proc phreg, and it produce the same code as above:

```
/*(ii) code inside proc phreg to do the same
job as in (i)*/

proc phreg data=data.Recid;
    where week > 1;
    model week*arrest(0) = fin age race wexp mar paro prio employ/TIEs = EFRON;
    array emp[*] emp1-emp52;
    array cum[*] cum1-cum52;
    cum1 = emp1;
    DO i = 2 to 52;
        cum[i] = (cum[i - 1]*(i - 1) + emp[i])/i;
        IF week = i THEN employ = cum[i-1];
    END;
run;
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.36509	0.19183	3.6223	0.0570114	0.694
age	1	-0.05128	0.02217	5.3491	0.0207329	0.950
race	1	0.30021	0.30884	0.9449	0.3310173	1.350
wexp	1	-0.05113	0.21520	0.0564	0.8122104	0.950
mar	1	-0.36330	0.38366	0.8967	0.3436746	0.695
paro	1	-0.02874	0.19716	0.0213	0.8840881	0.972
prio	1	0.09120	0.02881	10.0232	0.0015458	1.095
employ	1	-0.68170	0.30917	4.8618	0.0274576	0.506

For part (C):

For part (i),

the following code using the number of switches as a time dependent covariate, adjusting for all the other covariates:

```
/*part C*/
/*(i) consider number of switches as a time dependent
covariate*/
proc phreg data=data.Recid;
    model week*arrest(0) = fin age race wexp mar paro prio num_switch/TIEs = EFRON;
    array emp[*] emp1-emp52;
    array switch[*] switch1-switch52;
    switch1 = 0;
    num_switch = 0;
    Do i = 2 to 52;
        IF emp[i] - emp[i -1] ne 0 THEN switch[i] = 1;
        ELSE switch[i] = 0;
        num_switch = num_switch + switch[i];
    END;
run;
```

The output is:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.54562	0.22920	5.6670	0.0172867	0.579
age	1	-0.00366	0.02354	0.0241	0.8765049	0.996
race	1	-0.46025	0.34361	1.7942	0.1804179	0.631
wexp	1	-0.25890	0.22181	1.3624	0.2431230	0.772
mar	1	-0.15848	0.40881	0.1503	0.6982677	0.853
paro	1	-0.22146	0.22059	1.0078	0.3154197	0.801
prio	1	0.00543	0.03045	0.0317	0.8585999	1.005
num_switch	1	0.66848	0.04471	223.5046	<.0000001	1.951

we see that the number of switches in employement status is highly significant, and every one more switch will make it almost twice as possible(hazard ratio 1.95*) of getting arrested.*

For part (ii):

The following code using the number of negative switch as a time dependent covariate, adjusting for all the other covariates:

```
/*(ii) consider number of negative switches as a time
dependent covariate*/
proc phreg data=data.Recid;
    model week*arrest(0) = fin age race wexp mar paro prio num_switch/TIEs = EFRON;
    array emp[*] emp1-emp52;
    array switch[*] switch1-switch52;
    switch1 = 0;
    num_switch = 0;
    Do i = 2 to 52;
        IF emp[i] - emp[i -1] =-1 THEN switch[i] = 1;
        ELSE switch[i] = 0;
        num_switch = num_switch + switch[i];
    END;
run;
```

The output is:

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.39002	0.19131	4.1562	0.0414831	0.677
age	1	-0.06489	0.02223	8.5212	0.0035103	0.937
race	1	0.31896	0.30848	1.0691	0.3011561	1.376
wexp	1	-0.13377	0.21220	0.3974	0.5284413	0.875
mar	1	-0.37910	0.38191	0.9853	0.3208864	0.684
paro	1	-0.07915	0.19548	0.1639	0.6855611	0.924
prio	1	0.08755	0.02879	9.2475	0.0023582	1.091
num_switch	1	-0.28849	0.11065	6.7975	0.0091286	0.749

we see that the number of negative switch is significant(p value 0.009*), and every unit increase of number of negative switch will make it* 25% *less possible to get arrested.*

we are not surprised to see this happening, in fact it is possible of having either increased or decreased chance of getting arrested when the number of negative switch increases. Because in order to have more negative switch, you also need to have more positive switch. You could not unhire a person who has alrady lost his job.

This completed the answers for question #5.

Question #6.

Solution 6. *We add two dummy covariate here for our case.*

$$switchtype\_1 = \begin{cases} 1 & if\ \#\ of\ job\ lost\ = 0 \\ 0 & otherwise \end{cases}$$

$$switchtype\_2 = \begin{cases} 1 & if\ \#\ of\ job\ lost\ = 1 \\ 0 & otherwise \end{cases}$$

Of course they are time dependent as the number of weeks progress from week 1 *to week* 52*.*

The following code fit these two time dependent covariates along with the others into the recidivism data:

```
/*Question 6*/

/*continue from 5(C)(ii), categorize number of negative switch
as a time dependent covariate*/

/*switchtype_1 is a time dependent dummy covariate indicating #of job lost = 0*/
/*switchtype_2 is a time dependent dummy covariate indicating #of job lost = 1*/

proc phreg data=data.Recid;
    model week*arrest(0) = fin age race wexp mar paro prio switchtype_1 switchtype_2/TIEs = EFRON;
    array emp[*] emp1-emp52;
    array switch[*] switch1-switch52;
    switch1 = 0;
    num_switch = 0;
    Do i = 2 to 52;
        IF emp[i] - emp[i -1] =-1 THEN switch[i] = 1;
        ELSE switch[i] = 0;
        num_switch = num_switch + switch[i];
        IF num_switch = 0 THEN switchtype_1 = 1;
        ELSE switchtype_1 = 0;
        IF num_switch = 1 THEN switchtype_2 = 1;
        ELSE switchtype_2 = 0;
    END;
run;
```

The output is:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
fin	1	-0.40104	0.19173	4.3750	0.0364703	0.670
age	1	-0.06522	0.02229	8.5581	0.0034398	0.937
race	1	0.32603	0.30839	1.1176	0.2904260	1.385
wexp	1	-0.12361	0.21232	0.3389	0.5604510	0.884
mar	1	-0.38507	0.38212	1.0155	0.3135897	0.680
paro	1	-0.07486	0.19567	0.1464	0.7020206	0.928
prio	1	0.08938	0.02879	9.6396	0.0019043	1.093
switchtype_1	1	0.65374	0.26446	6.1108	0.0134357	1.923
switchtype_2	1	0.30937	0.26551	1.3577	0.2439443	1.363

From the output we see that switchtype_1 is significant(p value 0.01). It tells that there is a significant difference of arresting time between people who have never lost the job and the people who have lost the job for at least two times. Those who have never lost a job before actually are almost twice as much possible of getting arrested as those who have lost job at least twice(hazard ratio 1.923). This may sound ridiculous, however if we think about it, those who have never lost job before may have never been hired at all. Those who have lost job at lest twice must have been hired at least twice as well. Wile they are working the chance of getting arrested should actually be smaller.

On the other hand, switchtype_2 is not significant(p value 0.24), and we do not have enough evidence showing that there is a different arresting time between those who have lost job once and those who have lost job at least twice.