Survival Analysis Using Cox Proportional Hazard Modeling

for AMI(acute myocardial infarction)

Junyan Deng, MS,

Department of math and statistics, Cal State Long Beach, CA

1. Abstract

Heart Attack is a serious illness and the incidence is highly related to age. The midian survival time from KM (Kaplain Meier) estimate is approximate 2500 days. The mean survival time is 1735 days. Cox regression model and parametric distribution are used to analysis the survival rates following hospital admission for acute myocardial infarction (AMI). After checking the PH (proportional hazard) assumption, stratified Cox regression and time depending model Cox regression are to be used to adjust the model. Cpk, miord, mitype and yrgrp are not satisfied the PH assumption and need to be stratified on these variables. The long term survival time is statistically related to the period of staying in the hospital.

On the other hand, assuming the model follow a distribution and fit the model with specific distribution. Checking the FIT statistics of AIC and SBC, the Gamma distribution fits best with smallest AIC and SBC.

2. Introduction

The aim of this study is to find out the trends over time in the incidence and analysis the survival rates following hospital admission for acute myocardial infarction (AMI). The data come from The Worcester Heart Attack Study (WHAS). Data have been collected during ten 1-year periods beginning in 1975 on all AMI patients admitted to hospitals in the Worcester, Massachusetts, Metropolitan area. The data in this paper were obtained by taking 10% random sample within 6 of the cohort years from the main data set of more than 8000 admissions, and only a small subset of variables were taking into analysis.

2. Data structure

2.1 data description

The subset of predictors used are described in Table 1 alone with their codes and values. Length of follow-up (days between hospital admission and last follow-up) is used as survival time for modelling and analysis.

AGE, CPK, lenfol are continuous variables. Mitype, year and yrgrp are allocated classification variables, and the others are dummy variables.

Table 1: Description of the variables in the WHAS data set.

Data are in file WHAS.DAT

n = 481

Variable Description Codes / Units

ID Identification Code  1 - 481

AGE Age (per chart) years

SEX Gender 0 = Male

1 = Female

CPK Peak Cardiac Enzyme International Units (iu)

SHO Cardiogenic Shock Complications  0 = No

1 = Yes

CHF Left Heart Failure Complications 0 = No

 1 = Yes

MIORD  MI Order 0 = First

1 = Recurrent

MITYPE  MI Type 1 = Q-wave

2 = Not Q-wave

3 = Indeterminate

YEAR Cohort Year 1 = 1975

2 = 1978

3 = 1981

4 = 1984

5 = 1986

6 = 1988

YRGRP Grouped Cohort Year 1 = 1975 & 1978

2 = 1981 & 1984

3 = 1986 & 1988

LENSTAY Length of Hospital Stay Days between hospital discharge and Hospital Admission

DSTAT Discharge Status from Hospital 0 = Alive from 1 = Dead

LENFOL Total Length of Follow-up Days between Date of Last Follow-up and Hospital Hospital Admission Date

FSTAT Status as of Last 0 = Alive

Follow-up 1 = Dead

2.2 Correlation between variables

From Table 2, we can find out that mitype is not significant related to the survival time(lenfol), and that the other predictors contain information of the dependent variable(lenfol). Age sho and chf are highly correlated to each othter and they are highly related to the survival time. Table 3(part of the result) shows that there are ties in the data. We should take ties into consideration.

Table2

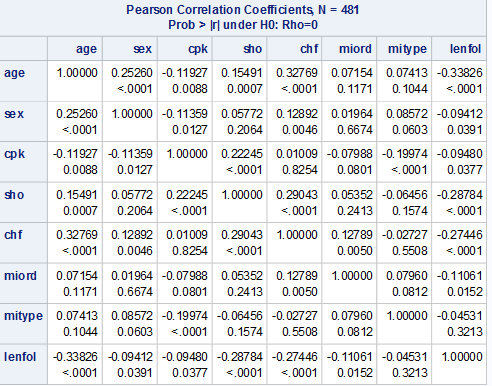
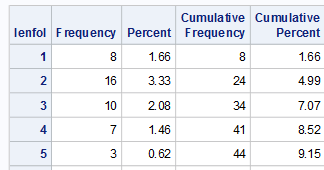


Table3



3. Method

None parameter estimate (Kaplan- Meier) can be used to get an overall look into the survival time. The disadvantage of this method is the plot would be steps. Cox regression and its extension, stratified Cox regression and time dependent Cox regression, are semi-parameter method. Cox regression are to be used to model the survival time and some influential covariates. Before deciding the final model of Cox regression, we have to check PH (proportional hazard) assumption which we assume that the hazard ratio is independent of survival time. If the some variables violate the PH assumption, stratified or time dependent method should be used to adjust the model. The advantage of cox regression is that we do not need to pre-assume the distribution of the data sample, specifically when we do not know or hard to know the distribution of the data. Cox regression model is semi-parametric and the plot is step in visual.

Parametric survival models are used to fit the data. We have to assume that the data follow a specific distribution and fit the data. Check the fitness of the model to decide which distribution would fit well. The advantage of this method is that the plot is smoother. But the data sometimes is hard to find out its distribution and need some guess and experience, and sometimes could get a wrong conclusion if assuming the wrong distribution and using a wrong model.

4. Survival analysis

4.1 Noneparameter method

4.1.1 KM curve and Log-rank test

The Kaplan\_Meier survival function estimator is calculated as:



Where ni is the number of subjects at risk and di is the number of subjects who fail, both at time ti.

KM (Kaplan-Meier) curve shows that the midian of survival time is approximate 2500 days. The survival rate is different in the some variables,such as age, sex, sho, chf, miord, using univeriate log-rank test. These variables are potentially used in Cox regression model.

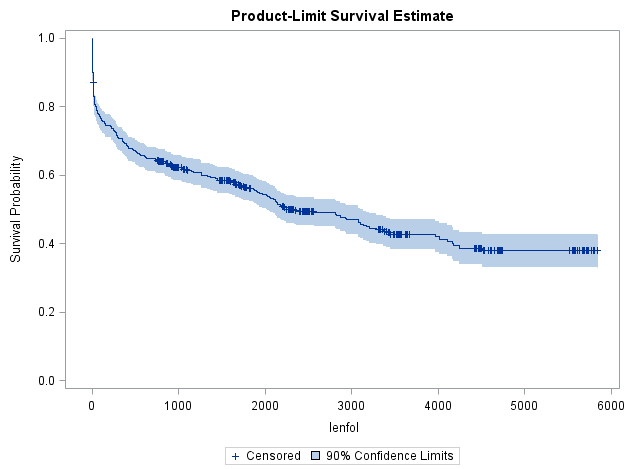
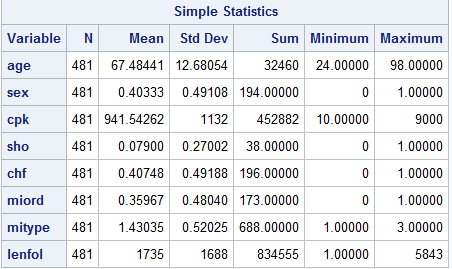
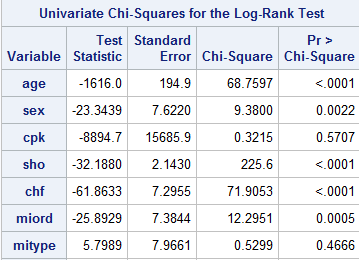
Figure 1

Table 3





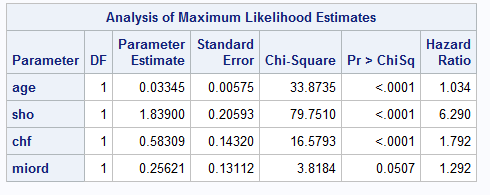
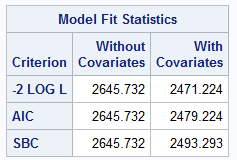
4.2 Semi-parameter method

4.2.1 Cox regression

4.2.1.1 Assumming no interaction between predicters.

Assuming no interaction between predictors and using stepwise selection to get the model, including sho, age, chf, and miord. Adding sex, cpk,mitype into the model, the statiscis -2log L increase. So this is the final model of assumming no interaction between predictors. Here we keep miord in the model for checking interaction term.

Table 5 without interaction terms



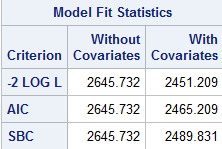
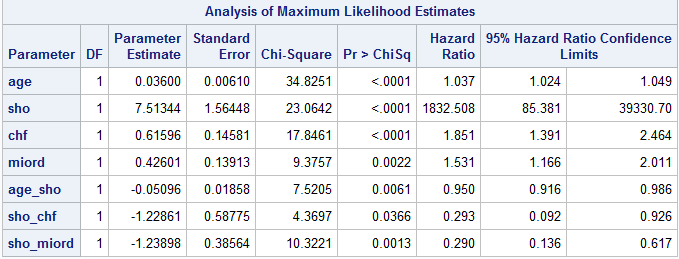
4.2.1.2 With interaction between predictors

Stepwise selection shows the interaction of age\*sho, sho\*miord and sho\*chf are significant. Compare the statistic, -2log L, we find that -2logL of the interacion model is less than the one without interaction. 2471.224-2451.209=20.015, the nested df=3, P-value is 0.002, less than 0.1. So we can conclude that the interaction model fits better. Here we can find that miord becomes significant when taking interaction into consideration.

Add cpk, sex and mitype into the model, the parameters do not change much, sex is not confounded with other variables. The P-value of sex is larger than 0.05 and it should be deleted in the model.

The final model is including age, sho, chf, miord, and the interction between age and sho, between sho and miord and between sho and chf.

Table 6 with interaction terms

5. Model adequacy checking

5.1 Check PH assumption using graph

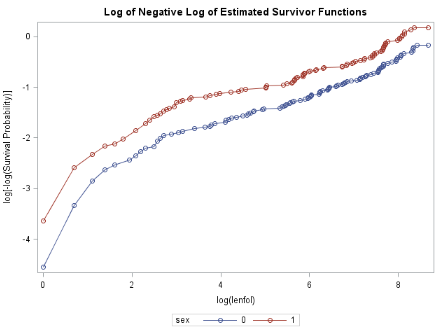
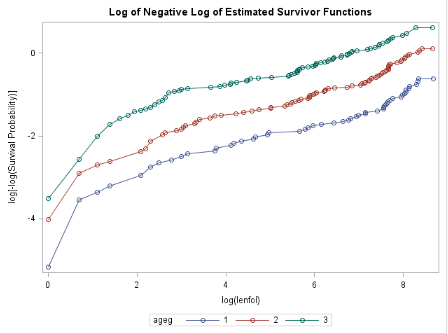
Check the PH (proportional hazard) assumption using the plot of log (-log(s (t))). If plot significant crossed, that variable violates the PH assumption, and it should be stratified later or be used time dependent variable to recode.

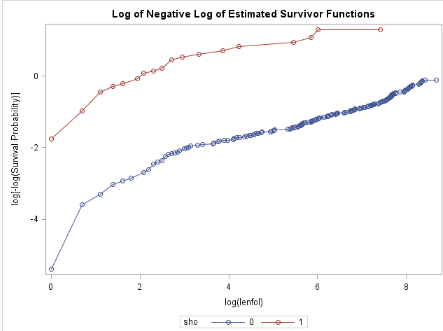
Recode age into three groups. Each group has approximate 35% percent of the total patients. Ageg=1 when age<=62. Ageg=2 when 62<age<=75. Ageg=3 when 75<age.

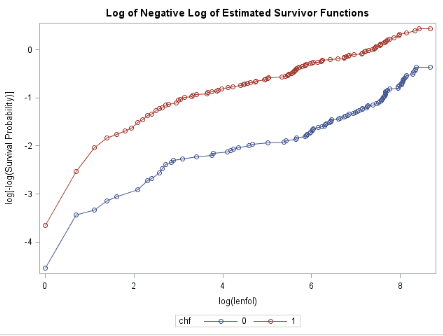
Recode cpk into three groups. Each group has approximate 35% percent of the total patients. Cpkg=1 when cpk<=98. Cpkg=2 when 98<cpk<=170. Cpkg=3 when 170<cpk.

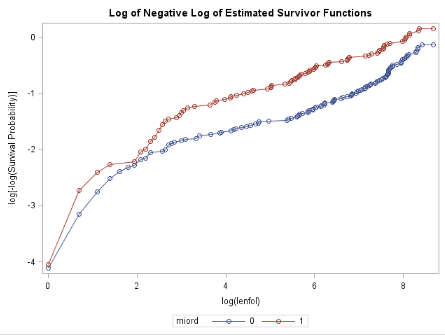
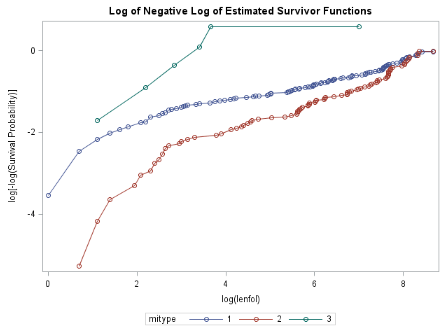
From the figures below, miord, mitype, cpkg and yrgrp violate the PH assumption, and should be stratified in Cox regression model.

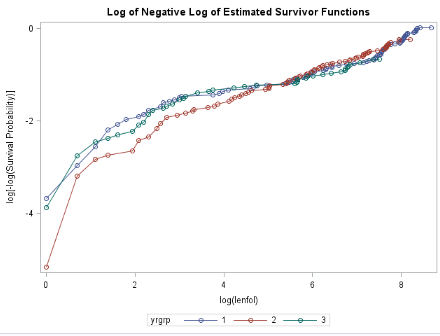
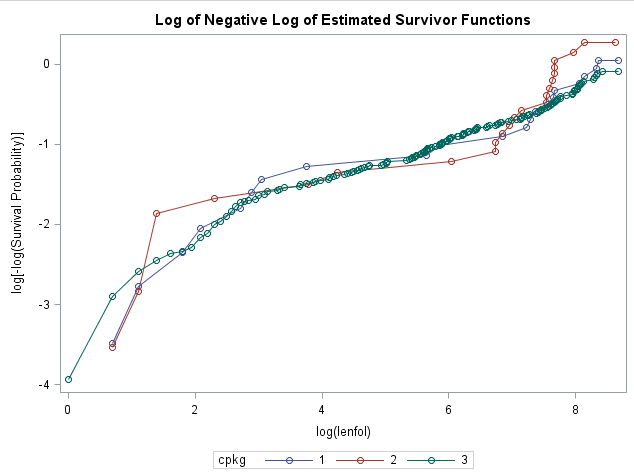
Figure 2







5.2 Check PH assumption using assess statement

Sas provide some test to check PH assumption. Assess statement can check the functional form of continuous variable and also can check PH assumption for the final model.

/\*use assess statement to check age leaner function and ph assumption.\*/

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sho chf miord age\_sho sho\_miord sho\_chf/rl ties=exact;

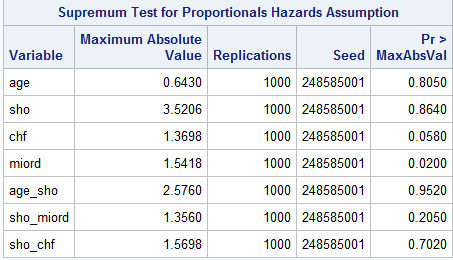
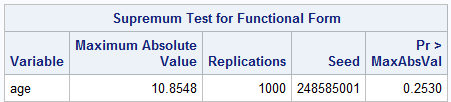
age\_sho=age\*sho;

sho\_miord=sho\*miord;

sho\_chf=sho\*chf;

assess var=(age) ph/resample;

**run**;

Table 7 assess statement

From Table 7, we can find that P-value is greater than 0.1, we can conclude the linear function of age is reasonable. The table also show that miord violate PH assumption, under the significant leva alpha=0.05.

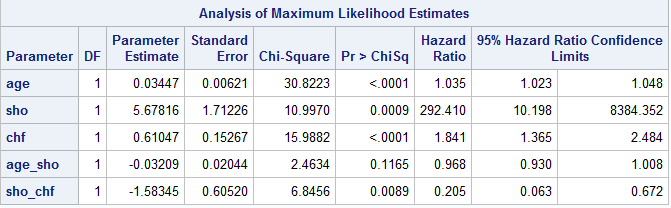
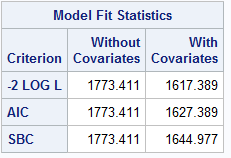
6. Stratified Cox regression

6.1 Stratified on cpkg, miord and mitype assuming they do not interact with other variables.

The interaction between age and sho becomes insignificant and the coefficient of sho and the coefficient of the interaction between sho and chf change a lot. The statistics AIC and SBC decrease much. We conclude that stratified on miord, mitype and improves the model.

Table 8

stratified on miord, mitype and assuming they are not interact with covariates



Delete age\_sho

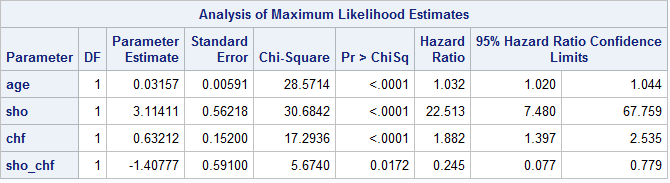
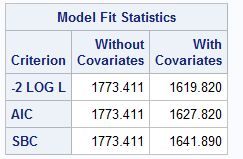
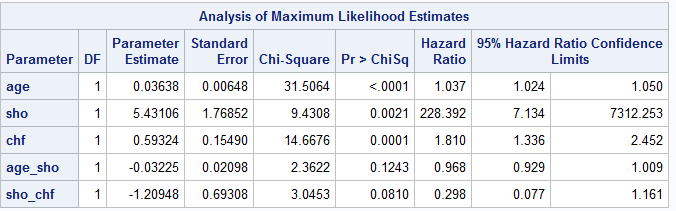
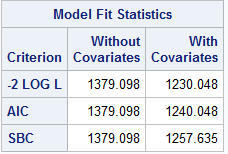
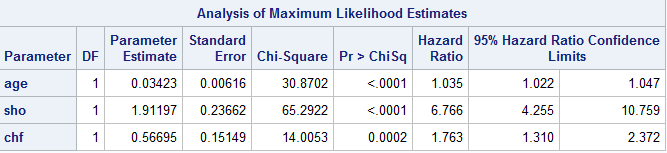
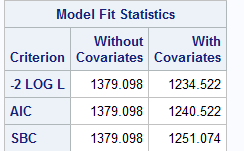


Table 9

stratified on miord, mitype, cpkg, and yrgrp assuming no interaction between stratified variables and covariates

Delete age\_sho and sho\_chf



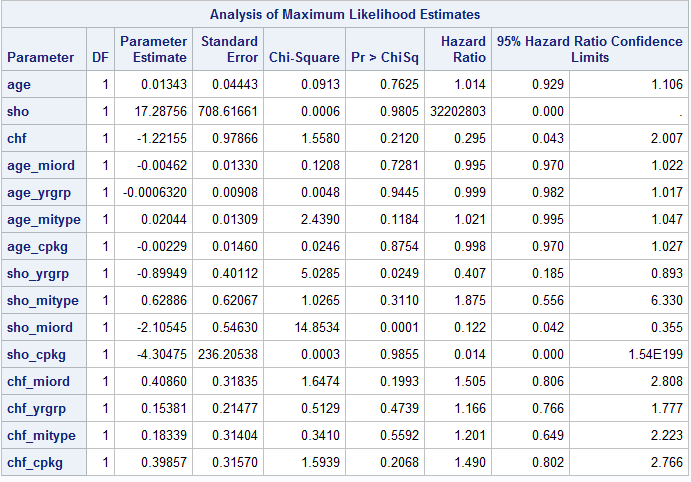
6.1.1 Stratified on miord, mitype, cpkg, and yrgrp assuming no interaction between stratified variables and covariates

Adding yrgrp into the stratified groups. The interaction between age and sho, and the interaction between sho and chf becomes insignificant. And the statistics AIC and SBC decrease much, so we can conclude that adding yrgrp into the stratified group improve the model.

6.2 stratified on miord, mitype,cpkg, and yrgrp assuming there are interaction between stratified variables and covariates

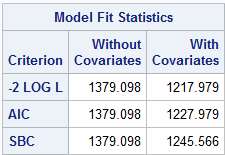
Put the interaction terms between stratified variables and covariates into the model to check whether the interaction significant or not. Table 10 shows that the interaction between yrgrp and sho is significant and the interaction between sho and miord is significant too. The main effect of sho is strongly affected.

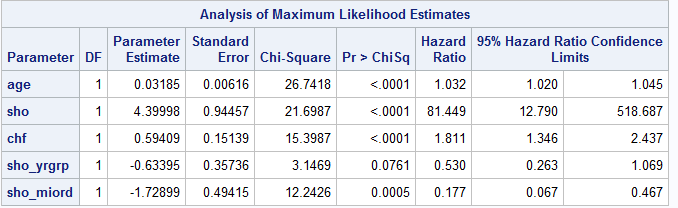
Table 10



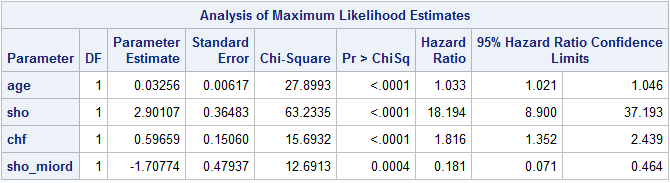
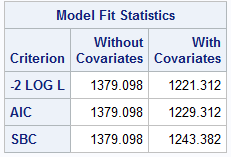
Delete insignificant terms and get the table we get that the interaction between sho and yrgrp becomes insignificant. This maybe cause by the main effect of sho. Comparing the statistics AIC and SBC with the ones of the model of assuming no interaction between stratified variables and covariates, they decrease much. And the coefficient of sho change a lot. So we conclude that the interaction between sho and miord improve the model, and conclude that the coefficient of the parameter are constant across cohort yrgrp when stratified on yrgrp, cpkg, miord and mitype.

Table 11 delete insignificant interaction terms from table 10.





Delete sho\_yrgrp

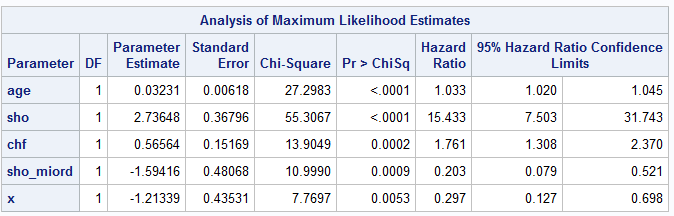
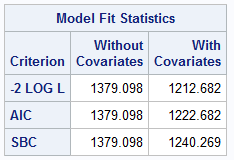


6.4 Time dependent Cox regression model

It is quite possible that the long term survival in the cohort group, yrgrp, is due to the different length of staying in the hospital. Take time dependent variable x into the model to check this suspect. Define x (t)=0 when lenfol is less than or equal to lenstay, x(t)=1 when lenfol is greater than lenstay.

Table 12 show that x is significant. We should take x time dependent variable into consideration.

Table 12 time dependent variable into the model



From now, the final Cox regression model is stratified on miord, mitype, ckpg and yrgrp, including age, sho, chf, x (time dependent variable) and the interaction between sho and miord. We can conclude that the long term survival due to the long time staying in the hospital.

The HR (age change in 1 year) is 1.033, 95% confidence interval ranging from 1.020 to 1.045. The patients one 1 year older are 3.3% more likely at risk to die.

When miord is 0, the HR (sho 1 to sho 0) is 15.443, 95% confidence interval ranging from 7.503 to 31.743. The patients with sho 1 are 15.433 times than the patient with sho 0 more likely at risk to die.

When miord is 1, the HR (sho 1 to sho 0) is 1.14.

The HR (chf 1 to chf 0) is 1.641, 95% confidence interval ranging from 1.308 to 2.370. The patients with chf 1 are 64.1% more than the patients with chf 0 to die.

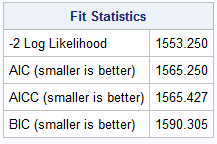
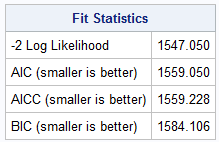
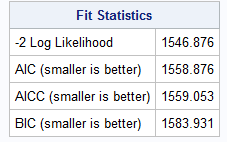
For the time dependent variable x, the HR(x=1 x=0) is 0.297, 95% confidence interval ranging from 0.127 to 0.698. The patients staying in the hospital with long time survive 3.37 (1/0.297) times longer than the staying short one.

7.1 parameter

Assuming the survival time satisfied specified distribution of exponential, Weibull, lognormal, log-logistic, and gamma, fit the model and find out that Weibull, log-logistic, lognormal, and gamma fit good, comparing AIC and SBC. These model all show that sex, cpk, and mitype are insignificant. Delete these there insignificant variables and fit again. The AIC and SBC from these models are very closed, but Gamma fits best with the smallest AIC and SBC.

Table 13 compare fit statistics

Weibull Log-logistic Log nomal

Gamma

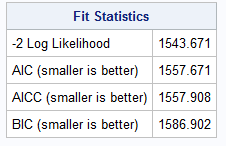
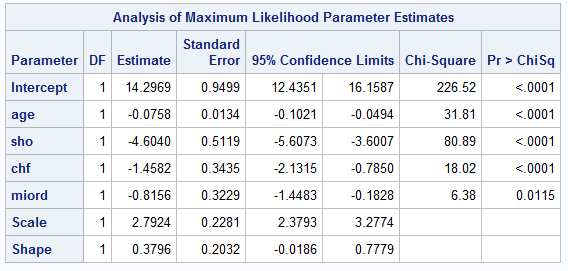
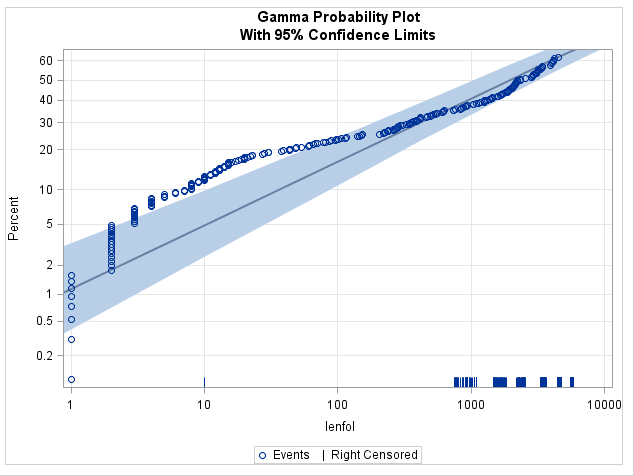


Table 13 parameter estimates of gamma model.

Gamma

All the parameters are negative, mean that they are all negative associate with the survival time.

8 conclussion

Heart attack is a serious illness and the midian survival time of paitients is approxmiate 2500 days. Age, sho, chf and mirod are highly related with the survival time. The survival time is statistically the same in gender. PH assumption was not satified on cpk, miord,mitype and yrgrp. Stratified on these variables and find out that miord interacts with sho. It need further analysis on the interaction between miord and sho. The long term survival time is related to the time staying in the hospital. This suggests patients should addmit as earlly as possible and stay as longer as possible. Age, sho (Cardiogenic Shock Complications), chf (Left Heart Failure Complications), miord (MI Order) are negative associated with the survival time based on the analysis from parameter regression model. As age increases the shorter the suvival time is. Cardiogenic shock complications decreases the survival time. Patiets who have left heart failure complications live shorter survival time. Recurrence of MI would decrease the survival time.

Appendix

Sas code

**data** whas;

infile "F:\course\590\hw\hw2\whas.dat";

input id age sex cpk sho chf miord mitype year yrgrp lenstay dstat lenfol fstat;

**run**;

**proc** **lifetest** data=whas plot=(d ls h);

time age;

test miord;

strata miord;

**run**;

**proc** **corr** data=whas plots(maxpoints=none)=matrix;

var age sex cpk sho chf miord mitype lenfol;

**run**;

/\*KM estimate\*/

**proc** **lifetest** data=whas plots=survival;

time lenfol\*fstat(**0**);

test age sex cpk sho chf miord mitype year;

**run**;

/\*check ph assumption on sex\*/

**proc** **lifetest** data=whas conftype=linear plots=(survival(cl) lls ls) outsurv=a;

time lenfol\*fstat(**0**);

strata sex;

**run**;

/\*get how many ties to see whether should take ties into consideration\*/

**proc** **freq** data=whas;

table lenfol;

**run**;

/\*stepwise selection without interaction\*/

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/selection=stepwise ties=efron sle=**0.25** sls=**0.1** ties=exact;

**run**;

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sho chf miord/ties=exact;

**run**;

/\*from stepwise we can see that cpk is not contribute much, add sex and mitype into the model to see if something confounding.\*/

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sho chf miord sex/ties=exact;

**run**;

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sho chf miord mitype/ties=exact;

**run**;

/\*stepwise selection with interaction\*/

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age | sho |chf| miord /selection=stepwise ties=efron sle=**0.5** sls=**0.05** ties=exact;

**run**;

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sho chf miord age\_sho sho\_chf sho\_miord /rl ties=exact;

age\_sho=age\*sho;

sho\_chf=sho\*chf;

sho\_miord=sho\*miord;

**run**;

/\*check ph assumption by plot\*/

**proc** **lifetest** data=whas conftype=linear plots=(lls s ls) ;

time lenfol\*fstat(**0**);

strata sho;

**run**;

**proc** **lifetest** data=whas conftype=linear plots=(lls s ls) ;

time lenfol\*fstat(**0**);

strata chf;

**run**;

**proc** **lifetest** data=whas conftype=linear plots=(lls s ls) ;

time lenfol\*fstat(**0**);

strata miord;

**run**;

**proc** **freq** data=whas;

table age;

**run**;

**proc** **freq** data=whas;

table cpk;

**run**;

**data** whas2;

set whas;

if age<=**62** then ageg=**1**;

else if **62**<age<=**75** then ageg=**2** ;

else if **75**<age then ageg=**3** ;

**run**;

**data** whas2;

set whas2;

if cpk<=**98** then cpkg=**1**;

else if **98**<cpk<=**170** then cpkg=**2**;

else if **170**<cpk then cpkg=**3**;

**run**;

**proc** **lifetest** data=whas2 conftype=linear plots=(lls s ls);

time lenfol\*fstat(**0**);

strata ageg ;

**run**;

**proc** **lifetest** data=whas2 plots=(lls s ls);

time lenfol\*fstat(**0**);

strata cpkg;

**run**;

**proc** **lifetest** data=whas plots=(lls s ls);

time lenfol\*fstat(**0**);

strata year;

**run**;

**proc** **lifetest** data=whas plots=(lls s ls);

time lenfol\*fstat(**0**);

strata yrgrp;

**run**;

/\*use assess statement to check age leaner function and ph assumption.\*/

**proc** **phreg** data=whas;

model lenfol\*fstat(**0**)=age sho chf miord age\_sho sho\_miord sho\_chf/rl ties=exact;

age\_sho=age\*sho;

sho\_miord=sho\*miord;

sho\_chf=sho\*chf;

assess var=(age ) ph/resample;

**run**;

/\*assuming miord is not interact with other variables.\*/

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf age\_sho sho\_chf/rl ties=exact;

age\_sho=age\*sho;

sho\_chf=sho\*chf;

strata miord mitype cpkg;

**run**;

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf age\_sho sho\_chf/rl ties=exact;

age\_sho=age\*sho;

sho\_chf=sho\*chf;

strata miord mitype cpkg yrgrp;

**run**;

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf/rl ties=exact;

strata miord mitype cpkg yrgrp;

**run**;

/\*check interaction with other viriables\*/

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf age\_miord age\_yrgrp age\_mitype age\_cpkg

sho\_yrgrp sho\_mitype sho\_miord sho\_cpkg

chf\_miord chf\_yrgrp chf\_mitype chf\_cpkg/rl ties=exact;

age\_miord=age\*miord;

age\_yrgrp=age\*yrgrp;

age\_mitype=age\*mitype;

age\_cpkg=age\*cpkg;

sho\_miord=sho\*miord;

sho\_yrgrp=sho\*yrgrp;

sho\_mitype=sho\*mitype;

sho\_cpkg=sho\*cpkg;

chf\_miord=chf\*miord;

chf\_yrgrp=chf\*yrgrp;

chf\_mitype=chf\*mitype;

chf\_cpkg=chf\*cpkg;

strata miord yrgrp mitype cpkg;

**run**;

/\*check interaction with other viriables\*/

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf

sho\_yrgrp sho\_miord /rl ties=exact;

sho\_miord=sho\*miord;

sho\_yrgrp=sho\*yrgrp;

strata miord yrgrp mitype cpkg;

**run**;

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf

sho\_miord /rl ties=exact;

sho\_miord=sho\*miord;

strata miord yrgrp mitype cpkg;

**run**;

**proc** **phreg** data=whas2;

model lenfol\*fstat(**0**)=age sho chf

sho\_miord x/rl ties=exact;

sho\_miord=sho\*miord;

if lenstay<lenfol then x=**0**;

else x=**1**;

strata yrgrp miord mitype cpkg;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=exponential;

probplot;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=weibull;

probplot;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=llogistic;

probplot;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=logistic;

probplot;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=lnormal;

probplot;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=normal;

probplot;

**run**;

**proc** **lifereg** data=whas;

model lenfol\*fstat(**0**)=age sex cpk sho chf miord mitype/dist=gamma;

probplot;

**run**;