ACM 661 **Harsha Kankanamge** **PROJECT 2**

The data set WHAS500.SAV represents a sample of 500 subjects admitted to hospitals with acute myocardial infarction (MI) in Worcester, Massachusetts. The goal of the study is to describe factors associated with trends over time in the survival following hospital admission for acute myocardial infarction (MI).

1. In the data set WHAS500.SAV, Conduct a descriptive analysis of the variables, GENDER, BMI, LENFOL (days) and FSTAT in the data set and summarize the results.

Use length of follow-up (LENFOL) as the survival time variable, status at last follow-up (FSTAT) as the censoring variable, and 5 percent as the level of significance.

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	300	60.0	60.0	60.0
	female	200	40.0	40.0	100.0
	Total	500	100.0	100.0	

In the sample of 500 subject 60% of them are males and 40% of them are females. So in the sample male percentage is higher than female percentage.

vital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Alive	285	57.0	57.0	57.0
	Dead	215	43.0	43.0	100.0
	Total	500	100.0	100.0	

Duration of the study 57% of people were alive and 43% were died. Higher percentage of people were alive at the end of the study.

Descriptive Statistics

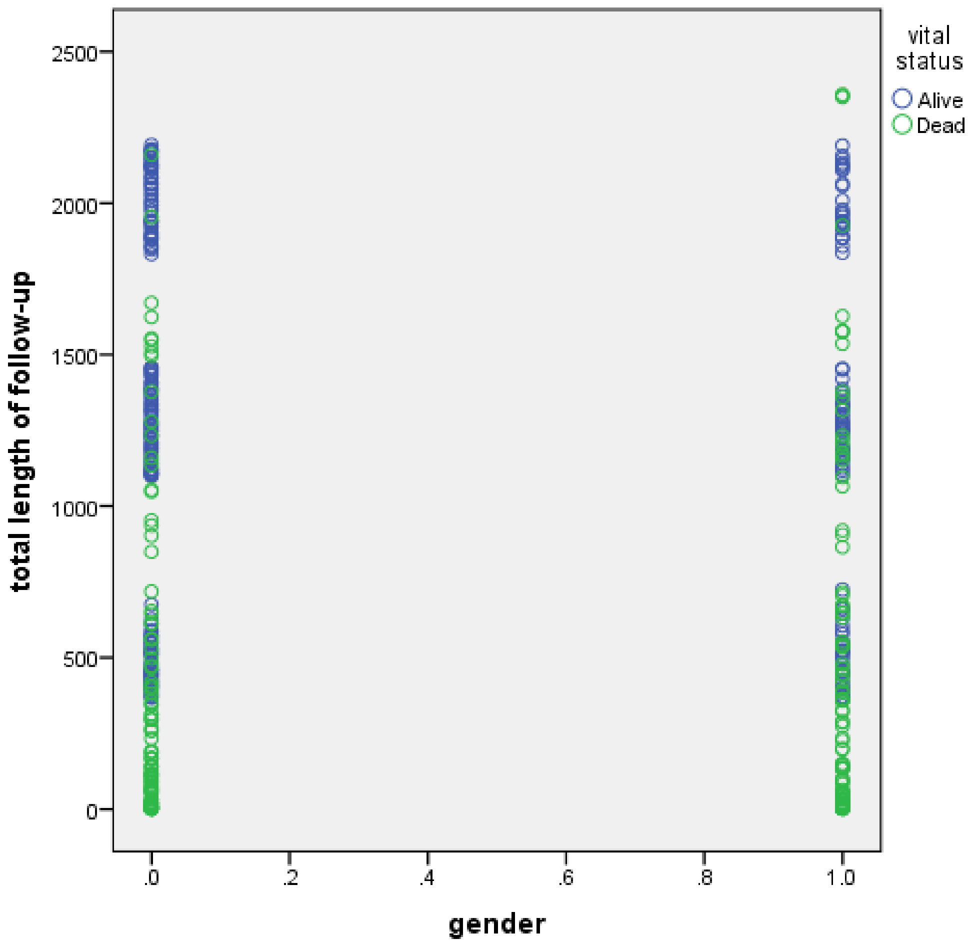
	N	Minimum	Maximum	Mean	Std. Deviation
total length of follow-up	500	1	2358	882.44	705.665
body mass index	500	13.05	44.84	26.6138	5.40566
Valid N (listwise)	500				

Mean of the total length of the follow-up time is 882.44 and std. Deviation is 705.665. Mean of the body mass index is 26.6138 and std. Deviation is 5.40566.

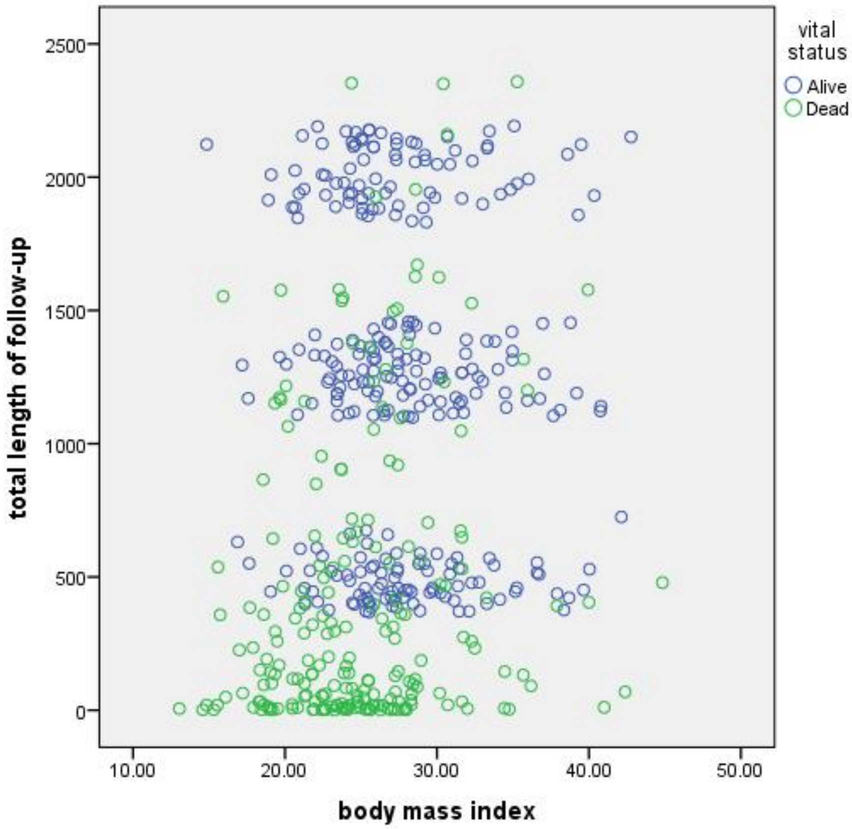
2. Plot survival time versus each covariate, setting markers by censoring status to see if there are any visual differences between survival time and the covariate by censor status.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
total length of follow-up	500	1	2358	882.44	705.665
Valid N (listwise)	500				



Most of the peoples had been died during first 600 days of the study in both gender groups. Males subjects were survived more days than female subjects. Most of male subjects and female subject were withdrawal from the study around 500 days, between 1200 to 1400 days and around 2000 days. At the end female subject was died. Closer to end of the study period a lots of male subjects were removed from the study than the female subjects.



Most of the subject ware died at the beginning of the study. As well as most of people who died had less BMI value. So people how have less BMI have high risk of having the event. Lot of subjects were removed from the study around 500 days, between 1000 days to 1500 days and around 2000 days.

3. Fit the proportional hazards model containing 'GENDER' and estimate the hazard ratio. Interpret the point estimate in words.

Categorical Variable Codings[a]

		Frequency	(1)
gender[b]	0=male	300	0
	1=female	200	1

a. Category variable: gender (gender)

b. Indicator Parameter Coding

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2447.570	7.772	1	.005	7.588	1	.006	7.588	1	.006

a. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.381	.138	7.679	1	.006	1.464	1.118	1.917

Estimated hazard ratio of gender is 1.464

A hazard ratio of 1.464 mean that at any time during the study, the per day rate of death among females is 46.4% higher than that for males.

Log hazard function $g(t,x,\beta)^\wedge = \ln [h0(t)] + 0.381*gender$

4. Is BMI a confounder of the effect of GENDER on survival? Explain the reasons for your answer.

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
body mass index	Alive	285	27.9153	5.08444	.30118
	Dead	215	24.8885	5.34633	.36462

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
body mass index	Equal variances assumed	.109	.741	6.445	498	.000	3.02674	.46960	2.10410	3.94939
	Equal variances not assumed			6.400	448.364	.000	3.02674	.47292	2.09733	3.95616

Mean ‘body mass index’ for people who have been died is 24.8885 and mean ‘body mass index’ for people who are alive is 27.9153. People who have been died has less mean body mass index compare with mean body mass index of alive people. Different is significant. Which is confirm by independent sample t test. T-statistic 6.445 and p-value = 0.000 <0.05. which means mean ‘body mass index’ is significantly different in the two groups (alive and dead). So BMI is suspected confounder of the effect of GENDER on survival.

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
body mass index	male	300	27.2689	4.82837	.27877
	female	200	25.6311	6.05204	.42794

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
body mass index	Equal variances assumed	10.491	.001	3.353	498	.001	1.63780	.48848	.67807	2.59754
	Equal variances not assumed			3.207	360.513	.001	1.63780	.51073	.63342	2.64219

Mean ‘body mass index’ for male is 27.2689 and mean ‘body mass index’ for female is 25.6311. Mean body mass index is high for males compare with females. Different is significant. Which is confirm by independent sample t test. T-statistic 3.207 and p-value = 0.001 <0.05. which means mean ‘body mass index’ is significantly different in the two gender groups (male and female). So there is a significant association between gender and body mass index.

gender * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
gender	male	Count	189	111	300
		% within gender	63.0%	37.0%	100.0%
	female	Count	96	104	200
		% within gender	48.0%	52.0%	100.0%
Total		Count	285	215	500
		% within gender	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11.016[a]	1	.001		
Continuity Correction[b]	10.412	1	.001		
Likelihood Ratio	11.003	1	.001		
Fisher's Exact Test				.001	.001
Linear-by-Linear Association	10.994	1	.001		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 86.00.

b. Computed only for a 2x2 table

37.0% of males have been died during the study. As well as 52% females have been died during the study. So higher percentage of females have been died compare with males during the study. So this deferent is significant. Which is conform by Chi-square test. Chi-square statistic is 11.016 and p-value = 0.001 < 0.05. So there is a significant association between gender and vital status.

Model with only gender.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.381	.138	7.679	1	.006	1.464	1.118	1.917

Model with only gender and BMI.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.219	.141	2.396	1	.122	1.245	.943	1.642
body mass index	-.093	.015	38.821	1	.000	.911	.885	.938

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2404.613	48.313	2	.000	50.545	2	.000	50.545	2	.000

a. Beginning Block Number 1. Method = Enter

(0.381-0.219)/0.219 = 0.7397

=73.97%

Log hazards of ‘gender’ changed from 0.381 to 0.219 when ‘BMI’ added to the model.

Log hazards of ‘gender’ decrease by about 73.97% when BMI is added to the model.
Both ‘BMI’ and ‘Gender’ are significantly associated with the ‘vital status’. (p-values <0.05)

Gender is not significant in the model when we added BMI to the model (p-value = 0.122>0.05)
So, ‘BMI’ is a major confounder in relationship between ‘gender’ and ‘vital status’. Change of Log hazards of ‘gender’ is higher than 10% when ‘BMI’ added to the model. Therefore, confounding is present.

5. Estimate the BMI-adjusted hazard ratio of GENDER and interpret in words.

Estimated BMI-adjusted hazard ratio of GENDER is 1.245

BMI-adjusted hazard ratio of GENDER is 1.245. Which means at any time during the study, the per day rate of death among females is 24.5% higher than that for males when BMI is constant.

However, Gender is not significant in the model p-value (0.122) >0.05

6. Estimate the GENDER-adjusted hazard ratio of BMI for a 5 kg/m2 increase in BMI and interpret in words.

-0.093*5 = -0.4649

The log hazard for the change of 5 kg/m2 of BMI is -0.46491 = (-0.093*5).

-0.46491 (-0.093 *5) implies that for each 5 kg/m2 increase in BMI, the log hazard of having the event decrease by 0.46491 times when gender is constant.

Estimated hazard ratio for an increase of 5 kg/m2 in BMI is exp (-0.093*5) = exp (-0.4649) = 0.6282

0.6282 Indicates that for every 5 kg/m2 increase in BMI, the hazard ratio of death decreases by 0.6282 times.

7. Is GENDER a confounder of the effect of BMI on survival? Explain the reasons for your answer.

Body mass index and vital status

Group Statistics

	vital status	N	Mean	Std. Deviation	Std. Error Mean
body mass index	Alive	285	27.9153	5.08444	.30118
	Dead	215	24.8885	5.34633	.36462

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
body mass index	Equal variances assumed	.109	.741	6.445	498	.000	3.02674	.46960	2.10410	3.94939
	Equal variances not assumed			6.400	448.364	.000	3.02674	.47292	2.09733	3.95616

Mean 'body mass index' for people who have been died is 24.8885 and mean 'body mass index' for people who are alive is 27.9153. People who have been died has less mean body mass index compare with mean body mass index of alive people. Different is significant. Which is confirm by independent sample t test. T-statistic 6.445 and p-value = 0.000 <0.05. which means mean 'body mass index' is significantly different in the two groups (alive and dead). So there is a significant association between BMI and vital status.

Body mass index and Gender

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
body mass index	male	300	27.2689	4.82837	.27877
	female	200	25.6311	6.05204	.42794

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
body mass index	Equal variances assumed	10.491	.001	3.353	498	.001	1.63780	.48848	.67807	2.59754
	Equal variances not assumed			3.207	360.513	.001	1.63780	.51073	.63342	2.64219

Mean 'body mass index' for male is 27.2689 and mean 'body mass index' for female is 25.6311. Mean body mass index is high for males compare with females. Different is significant. Which is confirm by independent sample t test. T-statistic 3.207 and p-value = 0.001 <0.05. which means mean ‘body mass index’ is significantly different in the two gender groups (male and female). So there is a significant association between gender and body mass index. Gender is suspected confounder of the effect of BMI on survival.

vital status and Gender

gender * vital status Crosstabulation

			vital status		Total
			Alive	Dead	
gender	male	Count	189	111	300
		% within gender	63.0%	37.0%	100.0%
	female	Count	96	104	200
		% within gender	48.0%	52.0%	100.0%
Total		Count	285	215	500
		% within gender	57.0%	43.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11.016[a]	1	.001		
Continuity Correction[b]	10.412	1	.001		
Likelihood Ratio	11.003	1	.001		
Fisher's Exact Test				.001	.001
Linear-by-Linear Association	10.994	1	.001		
N of Valid Cases	500				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 86.00.

b. Computed only for a 2x2 table

37.0% of males have been died during the study. As well as 52% females have been died during the study. So higher percentage of females have been died compare with males during the study. So this deferent is significant. Which is conform by Chi-square test. Chi-square statistic is 11.016 and p-value = 0.001 < 0.05. So there is a significant association between gender and vital status. Gender is suspected confounder of the effect of BMI on survival.

Model with only BMI.

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2406.994	44.165	1	.000	48.164	1	.000	48.164	1	.000

a. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
body mass index	-.098	.015	44.425	1	.000	.906	.881	.933

Model with BMI and Gender

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.219	.141	2.396	1	.122	1.245	.943	1.642
body mass index	-.093	.015	38.821	1	.000	.911	.885	.938

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2404.613	48.313	2	.000	50.545	2	.000	50.545	2	.000

a. Beginning Block Number 1. Method = Enter

(-0.098+0.093)/-0.093 = 0.05376

=5.4%

Log hazards of ‘BMI’ changed from -0.098 to -0.093 when ‘Gender’ added to the model.

(-0.098+0.093)/-0.093 = 0.05376

5.4% Log hazards of ‘BMI’ changed when ‘Gender’ added to the model

Log hazards of ‘BMI’ increase by about 5.4% when Gender is added to the model.
Both ‘BMI’ and ‘Gender’ are significantly associated with the ‘vital status’. (p-values = 0.000<0.05)

Gender is not significant in the model p-value = 0.122>0.05

However, ‘Gender’ is not a major confounder in relationship between ‘BMI’ and ‘vital status’. Since change of Log hazards of ‘BMI’ is less than 10% when ‘Gender’ added to the model. Therefore, confounding is not present.

8. Is the proportional hazards assumption reasonable in the model with BMI and GENDER? Explain.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.113	.179	.401	1	.527	1.120	.788	1.591
body mass index	-.093	.015	39.042	1	.000	.911	.885	.938
T_COV_*gender	.000	.000	.924	1	.336	1.000	1.000	1.001

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2403.690	49.265	3	.000	51.469	3	.000	51.469	3	.000

a. Beginning Block Number 1. Method = Enter

H0: β3=0

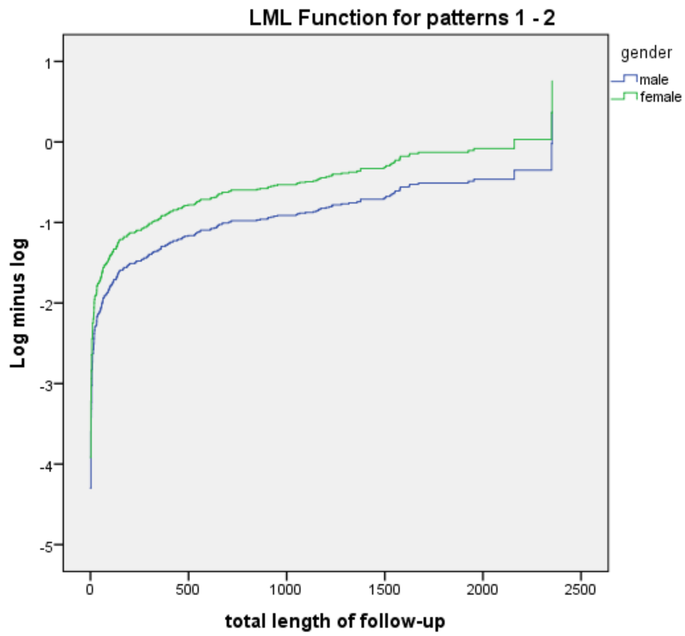
H0: β3≠0

Wald statistic = 0.924 and p-value = 0.336 >0.05 Do not reject H0.

There is not significant interaction between Tcov and gender. So T_cov*gender should not be in the model. No significant interaction between Tcov and gender indicate that proportional hazards assumption is satisfied.

Variables in the Equation

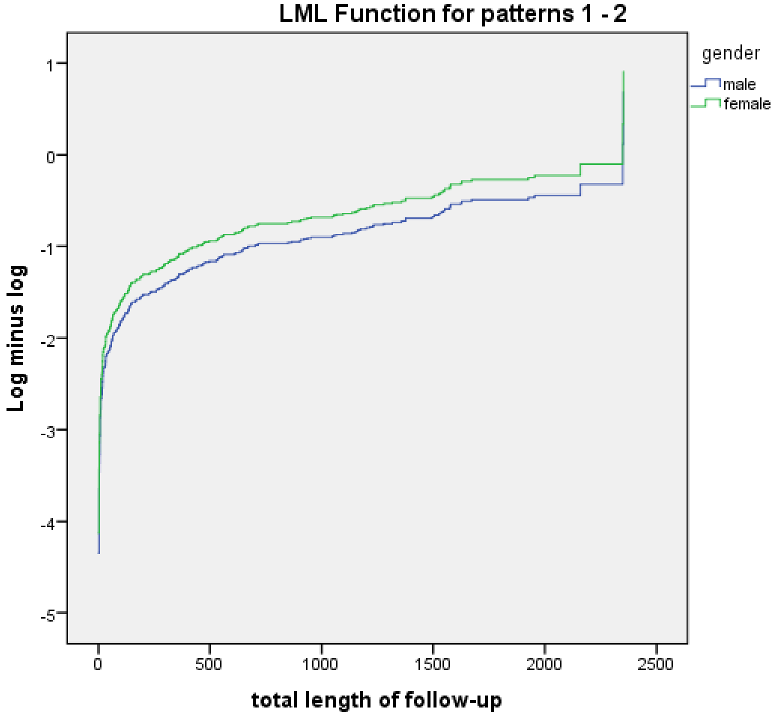
	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.381	.138	7.679	1	.006	1.464	1.118	1.917



Log minus log for female is higher than male as well as those individual strata lines are parallel. Which is appear to be difference in loghazards does not depend on the time. Parallel strata line indicate proportional hazards assumption appears to be reasonable. The different is 0.381. The difference in loghazards does not depend on time and is the proportional hazards assumption.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.219	.141	2.396	1	.122	1.245	.943	1.642
body mass index	-.093	.015	38.821	1	.000	.911	.885	.938



Log minus log for female is higher than male as well as those individual strata lines are parallel. Which is appear to be difference in loghazards does not depend on the time. Parallel strata line indicate proportional hazards assumption appears to be reasonable. The different is 0.219. When we added BMI to the model the difference between two lines is less. The difference in loghazards does not depend on time and is the proportional hazards assumption.

9. Assess the model using the Partial likelihood ratio test, Score test and Wald test.

Categorical Variable Codings[a]

		Frequency	(1)
gender[b]	0=male	300	0
	1=female	200	1

a. Category variable: gender (gender)

b. Indicator Parameter Coding

Omnibus Tests of Model Coefficients[a]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
2404.613	48.313	2	.000	50.545	2	.000	50.545	2	.000

a. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	.219	.141	2.396	1	.122	1.245	.943	1.642
body mass index	-.093	.015	38.821	1	.000	.911	.885	.938

<u>Score test</u>

H0: $\beta1=\beta2=0$

H1: At least one variable in the model is a good predictor of vital status.

Chi-square test statistic is 48.313. p-value = 0.000

p-value<0.05 reject H0

At least one variable in the model is a good predictor of vital status.

Likelihood ratio

H0: $\beta 1=\beta 2=0$

H1: At least one β_i not equal to zero.

Chi-square test statistic is 50.545.

p-value = 0.000 < 0.05 Reject H0

At least one variable is a significant predictor of vital status. Therefore, overall model is good fit.

Wald statistic

Gender

H0: $\beta 1=0$
H1: $\beta 1\neq 0$

Wald statistic = 2.396 p-value= 0.122

p-value (0.122) > 0.05 Do not reject H0

Gender is not significant. Therefore, we can say that gender is not a good predictor of FSTAT when BMI is in the model.

BMI

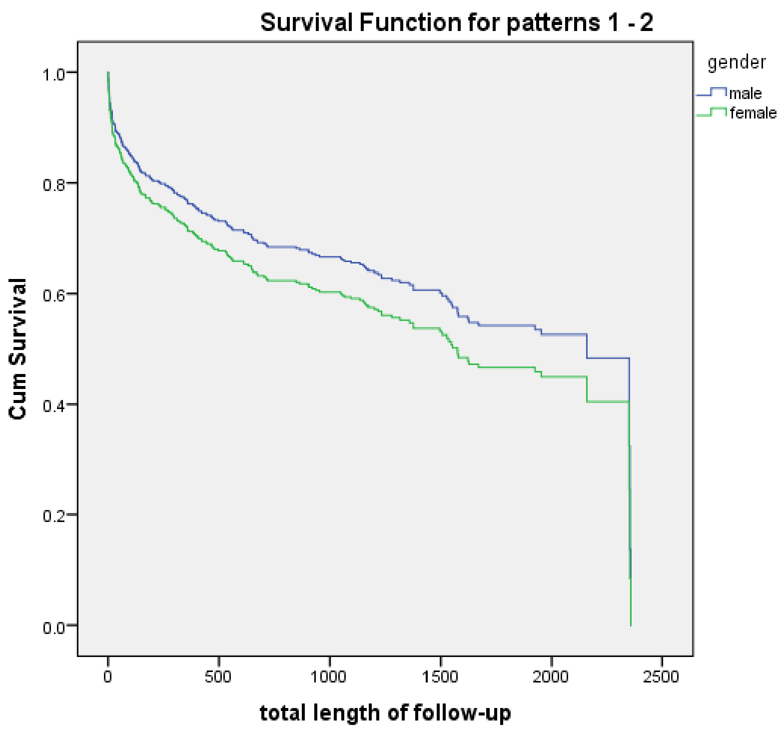
H0: $\beta 2=0$
H1: $\beta 2\neq 0$

Wald statistic = 38.821 p-value= 0.000

p-value (0.000) < 0.05 Reject H0

BMI is significant. Therefore, we can say that BMI is a good predictor of FSTAT.

10. Plot the survival functions for the gender groups and interpret them.



Within first 150 days of observation period about 20% of patients were dead in both gender group. As well as females group has lowest cumulative proportion of survival time for entire observation period compare with males. Cumulative survival proportion is keep decreasing with observation time for both gender groups.