MACQUARIE University

Faculty of Science and Engineering - Department of Mathematics and Statistics

STAT-818

(Epidemiological Methods)

ASSIGNMENT-3

(Semester 2- 2019)

Student ID: 45665761

Name: Sukhjeet Kaur

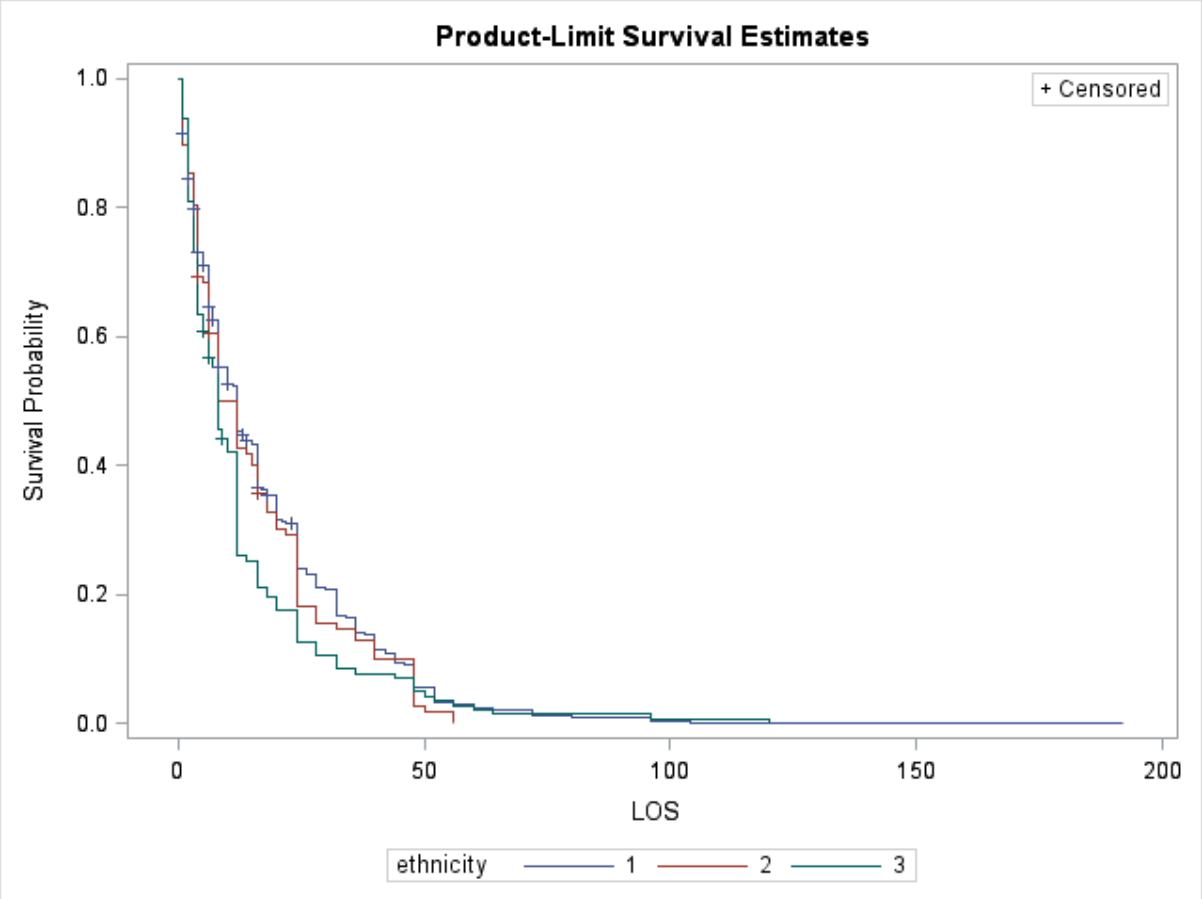
Submitted on: 30 october,2019

Master of Applied Statistics

Solution 1

a) Since the log rank test is significant as p-value is 0.0177 < 0.05 which means ethnicity is a significant factor for length of stay before adjusting for any other factors.
Kaplan Meier curve plot
Since the survival curves for ethnicity group A and B (i.e. 1 and 2 respectively) overlap each other means the length of stay could be similar for children whose mothers are from ethnicity group A or B, whereas, curve for group C (i.e.3) indicates that children of mothers with ethnicity C have significantly shorter length of stay in ICU.

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	8.0661	2	0.0177
Wilcoxon	9.5205	2	0.0086
-2Log(LR)	9.5416	2	0.0085



b) Since Wald chi square test gives p-value 0.0030 < 0.05 for factor smoking in the model. Therefore, smoking is a significant factor for length of stay before adjusting for any other factors.

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
smoking	1	8.8047	0.0030

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
smoking	0	1	-0.21884	0.07375	8.8047	0.0030	0.803	0.695 0.928	smoking 0

Hazard Ratio = 0.803

Hazard ratio 0.803 < 1, means the children whose mothers are non-smokers have reduced risk of length of stay in ICU compared to children whose mothers are smokers.

95% confidence interval is: (0.695, 0.928)

The confidence interval means the children whose mothers are non-smokers have length of stay in ICU between 69.5% to 92.8% less compared to those whose mothers are smokers. And, this result is consistent with hazard ratio as confidence interval does not contain 1.

c) The Effect of smoking and ethnicity on the association of education and length of stay is as follows:

i. Since the interaction term education*smoking in the joint test gives p-value = 0.3168 > 0.05 which is insignificant, means smoking is not an effect modifier for education.

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
education	1	0.0046	0.9457
smoking	1	0.3602	0.5484
education*smoking	1	1.0020	0.3168

To test whether smoking is a confounder two models were built one with education only and other with education after adjusting smoking and on comparing the estimates for education obtained from two models, there was no significant difference between two estimates which indicates smoking is not a confounder.

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
education	1	-0.04207	0.01730	5.9176	0.0150	0.959	0.927 0.992	education

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
education	1	3.3712	0.0663
smoking	1	6.1278	0.0133

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
education		1	-0.03262	0.01777	3.3712	0.0663	0.968	0.935	1.002	education
smoking	0	1	-0.18733	0.07568	6.1278	0.0133	0.829	0.715	0.962	smoking 0

ii. Since the interaction term education*ethnicity in the joint test gives p-value = 0.4415 > 0.05 which is insignificant, means ethnicity is not an effect modifier for education.

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
education	1	5.6218	0.0177
ethnicity	2	0.8405	0.6569
education*ethnicity	2	1.6350	0.4415

And to test for confounding, two models were built one with education only and other with education after adjusting ethnicity and on comparing the estimates for education obtained from two models, significant difference was observed between two estimates which indicates ethnicity is a confounder.

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
education	1	17.9445	<.0001
ethnicity	2	0.2450	0.8847

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
education		1	0.32763	0.07734	17.9445	<.0001	1.388	1.192	1.615	education
ethnicity	2	1	-0.07585	0.53648	0.0200	0.8876	0.927	0.324	2.653	ethnicity 2
ethnicity	3	1	-0.29762	0.61043	0.2377	0.6259	0.743	0.224	2.457	ethnicity 3

iii. The hazard ratio for education 0.968 ≈ 1 indicates that it has no effect on length of stay after adjusting for smoking. Also, the confidence interval (0.935,1.002) includes the null value 1, means there is no association between education and length of stay after adjusting for smoking.

The hazard ratio for education after adjusting for ethnicity is 1.388 > 1 means the risk in length of stay of child in ICU will increase by 1.388 times, with a year increase in schooling of mother.

And according to the confidence interval (1.192,1.695) a year increase in the schooling of mother yields an increase of 19.2% to 69.5% in the risk of length of stay (after adjusting for ethnicity).

d) Following 10 proportional hazard models were fitted to get the accurate estimate of ethnicity on length of stay.

Model Number	Parameters	-2Log L	Estimates	p-values
Model 1	Education	10498.944	-0.04207	0.0150 (significant)
Model 2	Age	10504.735	-0.00476	0.7169 (Insignificant)
Model 3	Ethnicity Group B Group C	10498.150	 0.09397 0.24063	0.0299 (significant) 0.3586 0.0092
Model 4	Smoking(non-smoker)	10496.330	-0.21884	0.0030 (significant)
Model 5	Smoking(non-smoker) Ethnicity Group B Group C	10485.971	-0.26818 0.13499 0.30343	0.0004 (significant) 0.0044 (significant) 0.1904 0.0013
Model 6	Smoking(non-smoker) Education	10492.958	-0.18733 -0.03262	0.0133 (significant) 0.0663 (Insignificant)
Model 7	Smoking(non-smoker) Age	10496.327	-0.21843 -0.00066	0.0032 (significant) 0.9600 (Insignificant)
Model 8	Smoking(non-smoker) Ethnicity Group B Group C Education	10484.564	-0.24429 0.12976 0.27767 -0.02140	0.0018 (significant) 0.0127 (significant) 0.2087 0.0042 0.2360 (Insignificant)
Model 9	Smoking(non-smoker) Ethnicity Group B Group C Age	10485.802	-0.027253 0.14054 0.30834 0.00547	0.0004 (significant) 0.0040 (significant) 0.1765 0.0012 0.6815 (Insignificant)

From above table the model with significant factors are Models 1, 3, 4 and 5. The models 1, 3, 4 are nested in model 5 so we would use likelihood ratio test for following comparisons:

Comparison of Model 1 and Model 5

χ_1^2 = 10498.944 – 10485.971 = 12.973

P-value = 0.003 < 0.05 means Model 5 is preferred.

Comparison of Model 3 and Model 5

χ_1^2 =10498.150 – 10485.971 = 12.179

P-value = 0.0005 < 0.05 means Model 5 is preferred.

Comparison of Model 4 and Model 5

χ_1^2 =10496.330 – 10485.971 = 10.359

P-value = 0.0013 < 0.05 means Model 5 is preferred.

Hence, among all models Model 5 is optimum model. (i.e. with predictors smoking and ethnicity)

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10498.944
AIC	10504.866	10500.944
SBC	10504.866	10505.737

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
education	1	-0.04207	0.01730	5.9176	0.0150	0.959	0.927 0.992	education

Figure 1- Output of model with single predictor-education

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10504.735
AIC	10504.866	10506.735
SBC	10504.866	10511.528

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
age	1	-0.00476	0.01313	0.1315	0.7169	0.995	0.970 1.021	age

Figure 2- Output of model with single predictor-age

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10498.150
AIC	10504.866	10502.150
SBC	10504.866	10511.737

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
ethnicity	2	7.0168	0.0299

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
ethnicity	2	1	0.09397	0.10237	0.8427	0.3586	1.099	0.899 1.343	ethnicity 2
ethnicity	3	1	0.24063	0.09244	6.7757	0.0092	1.272	1.061 1.525	ethnicity 3

Figure 3- Output of model with single predictor-ethnicity

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10496.330
AIC	10504.866	10498.330
SBC	10504.866	10503.123

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
smoking	1	8.8047	0.0030

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
smoking	0	1	-0.21884	0.07375	8.8047	0.0030	0.803	0.695 0.928	smoking 0

Figure 4- Output of model with single predictor-smoking

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10485.971
AIC	10504.866	10491.971
SBC	10504.866	10506.351

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
smoking	1	12.5945	0.0004
ethnicity	2	10.8621	0.0044

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
smoking	0	1	-0.26818	0.07557	12.5945	0.0004	0.765	0.659 0.887	smoking 0
ethnicity	2	1	0.13499	0.10310	1.7144	0.1904	1.145	0.935 1.401	ethnicity 2
ethnicity	3	1	0.30343	0.09438	10.3371	0.0013	1.354	1.126 1.630	ethnicity 3

Figure 5- Output of model with predictors-smoking & ethnicity

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10492.958
AIC	10504.866	10496.958
SBC	10504.866	10506.545

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
smoking	1	6.1278	0.0133
education	1	3.3712	0.0663

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
smoking	0	1	-0.18733	0.07568	6.1278	0.0133	0.829	0.715 0.962	smoking 0
education		1	-0.03262	0.01777	3.3712	0.0663	0.968	0.935 1.002	education

Figure 6- Output of model with predictors-smoking & education

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10496.327
AIC	10504.866	10500.327
SBC	10504.866	10509.914

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
age	1	0.0025	0.9600
smoking	1	8.6645	0.0032

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
age		1	-0.0006605	0.01317	0.0025	0.9600	0.999	0.974 1.025	age
smoking	0	1	-0.21843	0.07421	8.6645	0.0032	0.804	0.695 0.930	smoking 0

Figure 7- Output of model with predictors-smoking & age

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10484.564
AIC	10504.866	10492.564
SBC	10504.866	10511.738

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
smoking	1	9.7525	0.0018
ethnicity	2	8.7280	0.0127
education	1	1.4045	0.2360

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	Label
smoking	0	1	-0.24429	0.07823	9.7525	0.0018	0.783	0.672 0.913	smoking 0
ethnicity	2	1	0.12976	0.10322	1.5804	0.2087	1.139	0.930 1.394	ethnicity 2
ethnicity	3	1	0.27767	0.09695	8.2025	0.0042	1.320	1.092 1.596	ethnicity 3
education		1	-0.02140	0.01806	1.4045	0.2360	0.979	0.945 1.014	education

Figure 8- Output of model with predictors-smoking, ethnicity & education

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	10504.866	10485.802
AIC	10504.866	10493.802
SBC	10504.866	10512.976

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
age	1	0.1685	0.6815
ethnicity	2	11.0246	0.0040
smoking	1	12.7556	0.0004

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
age		1	0.00547	0.01332	0.1685	0.6815	1.005	0.980	1.032	age
ethnicity	2	1	0.14054	0.10399	1.8265	0.1765	1.151	0.939	1.411	ethnicity 2
ethnicity	3	1	0.30834	0.09513	10.5057	0.0012	1.361	1.130	1.640	ethnicity 3
smoking	0	1	-0.27253	0.07631	12.7556	0.0004	0.761	0.656	0.884	smoking 0

Figure 9- Output of model with predictors-smoking, ethnicity & age

e) The parameter estimates obtained using the model that only contains smoking and ethnicity are shown below in SAS output:

i.

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
smoking	1	12.5945	0.0004
ethnicity	2	10.8621	0.0044

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits		Label
smoking	0	1	-0.26818	0.07557	12.5945	0.0004	0.765	0.659	0.887	smoking 0
ethnicity	2	1	0.13499	0.10310	1.7144	0.1904	1.145	0.935	1.401	ethnicity 2
ethnicity	3	1	0.30343	0.09438	10.3371	0.0013	1.354	1.126	1.630	ethnicity 3

The reference category for smoking is 1=smoker and for ethnicity is 1=Group A

The estimate for smoking (status = non-smoker) is -0.26818

The estimate for ethnicity (group B) is 0.13499 and for ethnicity (group C) is 0.30343

The fitted model equation is:

$\widehat{h_i(t)} = \widehat{h_0(t)}\ \exp(-0.268 smoking_i 0 + 0.135 ethnicity_i 2 + 0.303 ethnicity_i 3)$

Where,

$\widehat{h_i(t)}$ = hazard function for *i*th individual.

$\widehat{h_0(t)}$ = baseline hazard function

$\text{Smoking}_i\ 0 = \begin{cases} 0\ if\ mother\ is\ a\ smoker \\ 1\ if\ mother\ is\ a\ non-smoker \end{cases}$

$\text{ethnicity}_i\ 2 = \begin{cases} 0\ if\ mother\ is\ from\ Group\ A \\ 1\ if\ mother\ is\ from\ Group\ B \end{cases}$

$\text{ethnicity}_i\ 3 = \begin{cases} 0\ if\ mother\ is\ from\ Group\ A \\ 1\ if\ mother\ is\ from\ Group\ C \end{cases}$

ii. The Hazards Ratio for smoking is 0.765 < 1 means there is reduced risk in length of stay of a child in ICU whose mother is a non-smoker compared to a child whose mother is a smoker.
95% confidence interval for smoking is (0.659, 0.887)
The confidence interval does not include 1 so the result is consistent with hazard ratio means the risk of length of stay of a child whose mother is a non-smoker is less by 65.9% to 88.7 % than child whose mother is a smoker.
The Hazard Ratio for ethnicity2 is 1.145 ≈ 1 means the length of stay of child in ICU is similar either the ethnicity of mother is group A or B.
Also, the confidence interval for ethnicity2 (0.935, 1.401) includes 1 indicating there is no difference in the length of stay of a child whether the mother is from ethnicity, group A or B.
The Hazard ratio for ethnicity3 is 1.354 > 1 which indicates there is an increased risk in length of stay of a child whose mother is from ethnicity-group C compared to the child whose mother is from ethnicity group A.
95% confidence interval for ethnicity3 is (1.126, 1.630) which does not include 1, so the result is consistent with hazard ratio means there is an increased risk in length of stay of a child in ICU by 12.6% to 63% whose mother is from ethnicity-group C compared to one whose mother is from ethnicity-group A

iii.

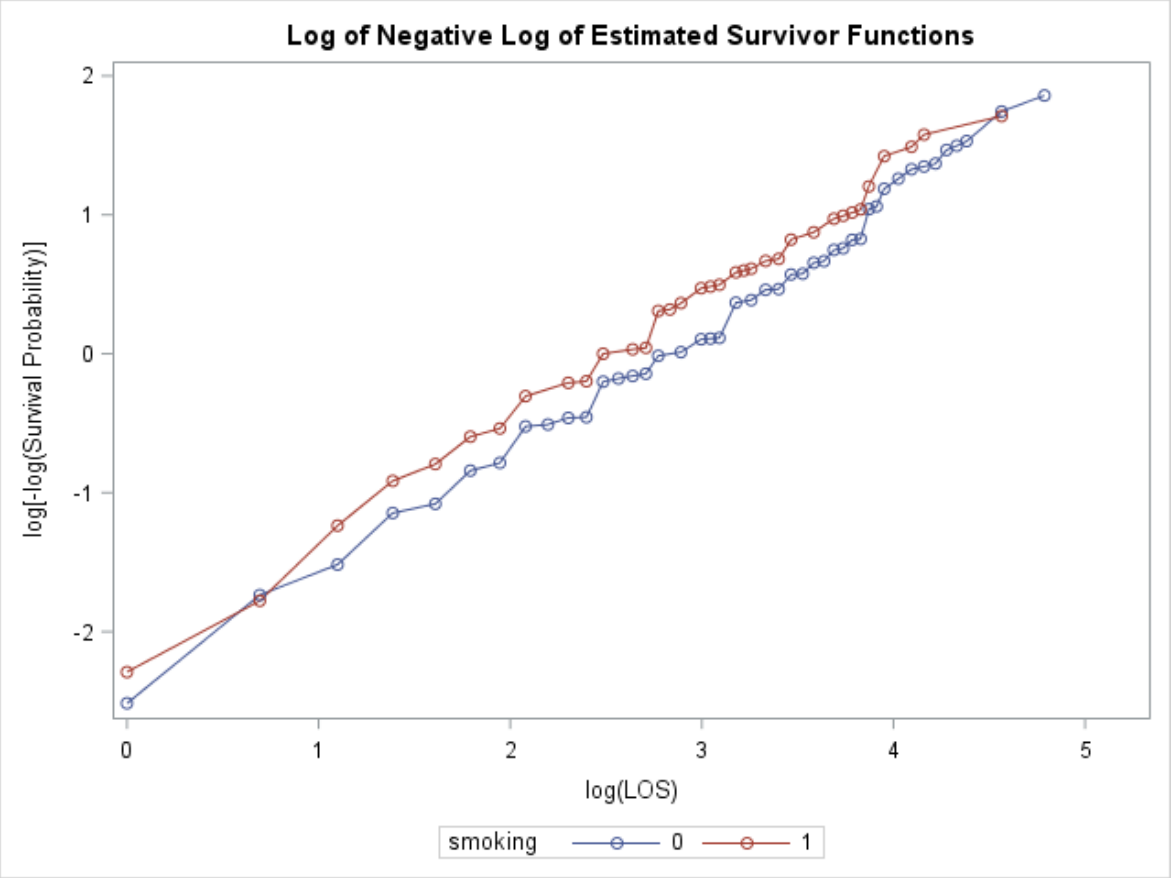


Figure 1: Ln-Ln curve for smoking

The curves for smokers and non-smokers are quite parallel but with cross over at three points (i.e. at lower and upper end), so, PH assumption does not hold. However, we can re-analyse the data using time-stratified model. To do the time

stratified analysis, we can divide the length of stay into three intervals as the curves cross at three points and then analyse the effect of smoking on length of stay for three intervals, similar to Proportional hazard model.
Or
We can fit censored linear regression model using transformation of survival time (length of stay)

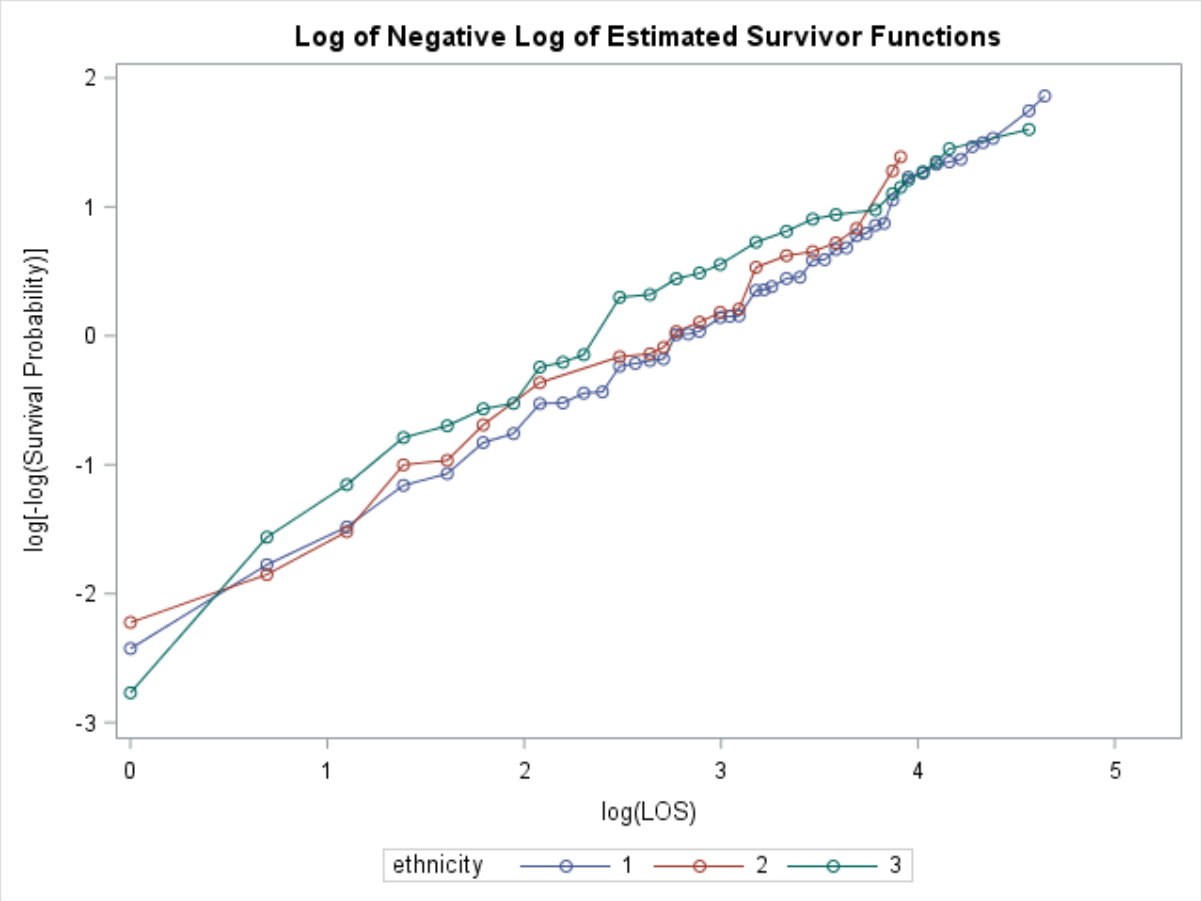


Figure 2: Ln-Ln curves for ethnicity 1=group A, 2= group B, 3= Group C

Here the curve for group C again cross over at lower and upper end with the curves for group A and B, whereas the curves for group A and group B cross over each other at multiple points. Hence, the PH assumption is violated by ethnicity.
We can re-analyse the data by dividing survival time (i.e. length of stay) into suitable number of time zones depending on time points where the curves cross and using the data created based on set of intervals, we can run PH model for each time zone and further check PH assumption for each model.
Or
We can fit censored linear regression model using transformation of survival time (length of stay)

Solution 2

REPORT

Results

In the study of survival analysis, the effect of four predictors on the length of stay of premature babies in the neonatal intensive care unit (ICU) was studied. The four predictors were age of mother at birth of a child, education level, ethnicity and smoking status of mother. The ethnicity-group A was used as reference category to compare the effect of other two groups relative to A and smoking status as smoker was used as reference to compare the effect of smoking on the survival time.

Initially, the univariate analysis was carried out in order to know the effect of each of the predictors on length of stay of child in ICU and as a result, age of mother at birth was found to have no effect on the length of stay of a child whereas education level, ethnicity and smoking status of mother have significant impact on the length of stay. Since, the ethnicity has three levels and on analysis, it was observed that infants whose mothers are from group B and A have similar effect on the length of stay as the estimate for group B is insignificant (see Table 1 below) whereas, the children whose mothers are from group C have increased risk of length of stay in ICU by 1.272 times, compared to children whose mothers are from group A. Based on smoking status, the infants whose mothers do not smoke have 0.803 times shorter length of stay in ICU compared to infants whose mothers do smoke. Then, for education it can be suggested that, it is marginally significant as the estimated hazard ratio is 0.959 which is equivalent to 1, thus it has not much impact on the length of stay.

Factors	**Estimate**	**Hazard Ratio**	**p-value**
Education	-0.042	0.959	0.0150
Age	-0.005	0.995	0.7169
Ethnicity Group B Group C	 0.094 0.241	 1.099 1.272	0.0299 0.3586 0.0092
Smoking (non-smoker)	-0.219	0.803	0.0030

Table 1: Univariate analysis output

Further, to test the joint effect of various predictors multivariate analysis was carried out for each possible combinations of the predictors, ignoring the factor - age of mother at birth of a child as it was insignificant in univariate analysis.

Before studying the association between length of stay and main effects of the predictors, the interaction effect of ethnicity with education and smoking with education was checked, which turn out to be insignificant. Moreover, the change in the effect of education on length of stay was checked by evaluating the difference in the estimates obtained from the PH models and did not change the effect of education while including ethnicity with education, a significant difference in the two estimates was observed which effects the length of stay in ICU.

There was no effect of education level of mother on length of stay as the estimated hazard ratio was 0.968 after adjusting smoking. On the other hand, ethnicity significantly changed the relationship of education with relative risk of survival time – the hazard ratio for education after adjusting for ethnicity is 1.388 > 1 means the risk in length of stay of child in ICU will increase by 1.388 times, with a year increase in schooling of mother. And according to the confidence interval (1.192,1.695) a year increase in the schooling of mother yields an increase of 19.2% to 69.5% in the risk of length of stay (after adjusting for ethnicity).

Another model with smoking and ethnicity was evaluated to study the effect on survival time, in which both smoking status of mother and ethnicity of mother gave significant results. The length of stay of a child in ICU whose mother is a non-smoker is less by 0.765 times compared to a child whose mother is a smoker and the risk of length of stay of a child is less by 65.9% to 88.7 %. While, the ethnicity groups behave differently as the Hazard Ratio for ethnicity-group B is 1.145 ≈ 1 means the length of stay of child in ICU is similar either the ethnicity of mother is group A or B, whereas the Hazard ratio for ethnicity-group C is 1.354 > 1 which indicates there is an increased risk in length of stay of a child whose mother is from ethnicity-group C compared to the child whose mother is from ethnicity group A. The risk in length of stay of a child in ICU increases by 12.6% to 63% whose mother is from ethnicity-group C compared to one whose mother is from ethnicity-group A.

Factors	**Estimate**	**Hazard Ratio**	**p-value**
Smoking(non-smoker)	-0.18733	0.829	0.0133 (significant)
Education	-0.03262	0.968	0.0663 (Insignificant)
Ethnicity			
Group B	-0.07585	0927	0.8876(Insignificant)
Group C	-0.29762	0.743	0.6259(Insignificant)
Education	0.32763	1.388	<0.0001(significant)
Smoking(non-smoker)	-0.26818	0.765	0.0004(significant)
Ethnicity			
Group B	0.13499	1.145	0.1904 (Insignificant)
Group C	0.30343	1.354	0.0013(significant)

Table 2: Multivariate analysis output

In conclusion, only two factors- smoking and ethnicity were found to have effect on the length of stay of an infant in neonatal ICU, however the validation of this analysis suggested to re-analyse the impact of these factors using different set of survival times for better understanding of the effect of these factors on length of stay.

SAS code

```
proc import datafile='ICU_2019.xlsx';
out= babies;
dbms= xlsx replace;
getnames yes;
sheet='Book1';
run;
proc print data=babies;
run;
/* part a */
proc lifetest data=babies;
time LOS*censoring(0);
strata ethnicity;
run;
/*part b */
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=smoking/risklimits;
run;
/* part c */
proc phreg data=babies;
model LOS*censoring(0)=education/risklimits;
run;
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=education smoking education*smoking; /*no effect
modification*/
run;
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=education smoking/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1");
model LOS*censoring(0)=education ethnicity education*ethnicity; /*no effect
modification*/
run;
proc phreg data=babies;
class ethnicity(ref="1");
model LOS*censoring(0)=education ethnicity/risklimits;
run;
/* for part d */
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=smoking/risklimits;
run;
proc phreg data=babies;
model LOS*censoring(0)=education/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1");
model LOS*censoring(0)=ethnicity/risklimits;
run;
proc phreg data=babies;
model LOS*censoring(0)=age/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1");
model LOS*censoring(0)=age ethnicity/risklimits;
run;
```

```
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=age smoking/risklimits;
run;
proc phreg data=babies;
model LOS*censoring(0)=age education/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1")smoking(ref="1");
model LOS*censoring(0)=age ethnicity smoking/risklimits;
run;
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=age smoking education/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1");
model LOS*censoring(0)=age ethnicity education/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1")smoking(ref="1");
model LOS*censoring(0)=age education smoking ethnicity/risklimits;
run;
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=smoking education /risklimits;
run;
proc phreg data=babies;
class smoking(ref="1");
model LOS*censoring(0)=smoking age/risklimits;
run;
proc phreg data=babies;
class ethnicity(ref="1")smoking(ref="1");
model LOS*censoring(0)= smoking ethnicity education/risklimits;
run;
/* part e */
proc phreg data=babies;
class ethnicity(ref="1")smoking(ref="1");
model LOS*censoring(0)=smoking ethnicity/risklimits;
run;
proc lifetest data=babies plots=(s,lls);
time LOS*censoring(0);
strata smoking;
run;
proc lifetest data=babies plots=(s,lls);
time LOS*censoring(0);
strata ethnicity;
run;
```