SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

Survival Analysis of Lung Cancer Patients using PROC PHREG and PROC LIFETEST

USERS PROGRAM

Survival Analysis of Lung Cancer Patients using PROC PHREG and PROC LIFETEST

Yan Wang, Ph.D. Student in Analytics and Data Science

Advisor: Dr. Jennifer Lewis Priestley, Professor of Statistics and Data Science

College of Science and Mathematics, Kennesaw State University

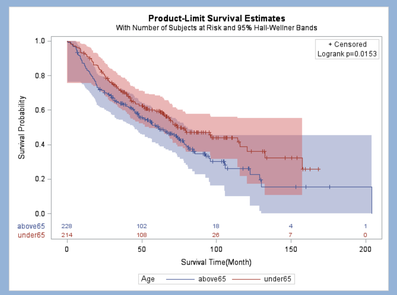
ABSTRACT

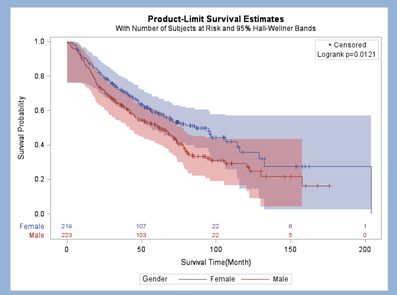
Survival analysis differs from other statistical analysis including the graphical summaries and regression modeling procedures due to that the data are almost always censored. The purpose of this project is to apply the survival analysis techniques in SAS to practical survival data, which aiming to understand the effects of Gender and Age on lung cancer patients' survival across different cancer centers (Site).

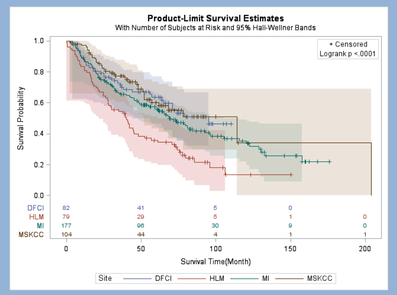
DATA DESCRIPTION

- The data set comes from National Cancer Institute's Director's Challenge Lung Study. This study was conducted on 443 lung cancer patients.
- Important variables includes Race Microarray, Test type, Gender, Age, Site, Survival Status, and Survival Time. Variables Gender, Age, and Site were selected for survival analysis.

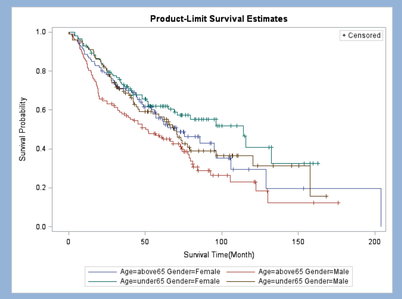
OVERALL LOG-RANK TEST

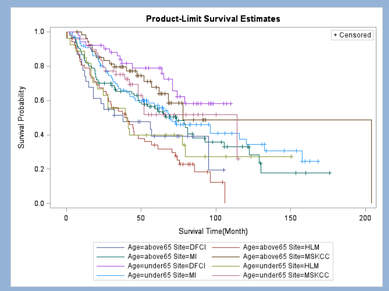


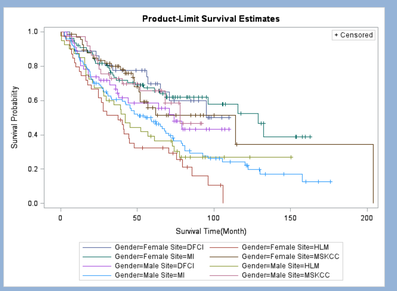




- Overall log-rank test aims to test the significance of the survival difference and can be run using PROC LIFETEST.
- Patients under-65-year-old and females have significant higher survival under lung cancer.
- Patients from HLM site have the lowest survival probability while those from MSKCC have relative higher probability, although there are some overlaps of the survival curves.







- Age effect is consistent for the two genders. Females have higher survival probability for both over-65-year-old patients and under-65-year-old patients.
- Age effect is different for patients from different sites. For DFCI site, patients under-65-year-old have better survival while for MSKCC site, patients above 65 have better survival before 80 months.
- Gender effect is also different for patients across sites. For DFCI site, females have better survival than males while for HLM site, males have better survival. Therefore, the heterogeneity of the patients across sites may increase the variation in the data and drown out the gender and age effect.

PROPORTIONAL HAZARD ASSUMPTION TEST

- Overall Kaplan Meier curves are roughly parallel, indicating that age, gender, and site satisfied the proportional hazard (PH) assumption when being considered independently.
- Goodness of Fit (GOF) testing approach provides a test statistic for assessing the PH assumption. If the PH assumption holds for certain variable, the Schoenfeld residuals for that variable will not depend on survival time.
- The p-values of GOF testing for gender, site and age are 0.8955, 0.2353, and 0.5754, respectively. This suggests that PH assumption is satisfied for these three variables.

Survival Analysis of Lung Cancer Patients using PROC PHREG and PROC LIFETEST

Yan Wang, Ph.D. Student in Analytics and Data Science

Advisor: Dr. Jennifer Lewis Priestley, Professor of Statistics and Data Science

College of Science and Mathematics, Kennesaw State University

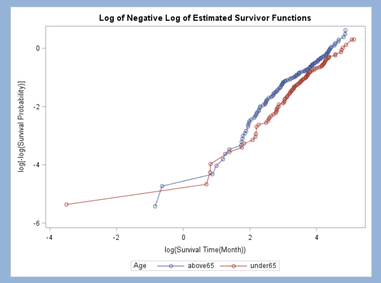
COX PROPORTIONAL HAZARD (PH) MODEL

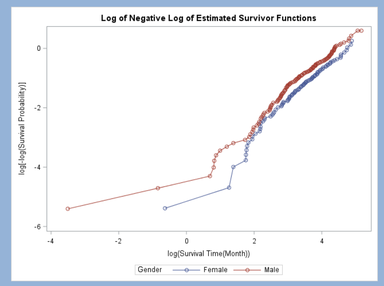
- Key assumption of a Proportional Hazard (PH) model is that hazard ratios are constant over time.
- Cox PH model can be run with PROC PHREG and it considers several explanatory variables simultaneously.
- Maximum likelihood estimations show that the hazard of death for males is 1.345 times of female when age and site is fixed (p value 0.025) while patients over-65-year-old have 1.385 higher death risk of those under-65-year-old when gender and site is fixed (p value 0.015). However, there is no significant difference in survival for different site when age and gender are fixed in the model.

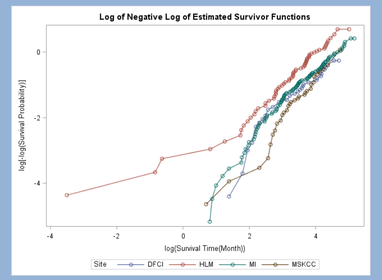
Covariate	Estimation	Chi-Square	Pr > Chisq	Hazard Ratio	95% Hazard Ratio Confidence	
Gender (Male)	0.297	4.99	0.025	1.345	1.037	1.744
Site	-0.092	1.95	0.163	0.913	0.802	1.038
Age_above65	0.326	5.97	0.015	1.385	1.067	1.800

PARAMETRIC ACCELERATED FAILURE TIME (AFT) Model

- Key assumption of a AFT model is that survival time accelerates (or decelerates) by a constant factor when comparing different levels of covariates.
- AFT model can be run with PROC LIFEREG and the most common distribution for parametric modeling of survival data is the Weibull distribution.
- Log-log of the survival function is linear with the log of time for Age, Gender and Site, which is the evident that the three covariates follows Weibull distribution.
- Results from maximum likelihood estimations are very similar with those from PH model.







Parameter	Estimation	Chi-Square	Pr > Chisq	Hazard Ratio	95% Confidence Limits	
Intercept	4.677	520.39	< 0.001	1.345	4.275	5.079
Gender (Male)	-0.286	4.72	0.023	1.335	-0.543	-0.020
Site	0.088	1.83	0.176	0.915	-0.039	0.214
Age_above65	-0.328	6.18	0.013	1.394	-0.586	-0.069
Scale	0.987				0.886	1.101
Weibull Shape	1.013				0.908	1.129

CONCLUSIONS

- The heterogeneity of the patients across sites may drown out the power of overall log-rank test on survival.
- Both semi-parametric Cox PH and parametric AFT models are good-fitting models. Females and patients under-65-year-old have better survival when compared to males (HR=1.335, $p<0.05$) and those above-65-year-old (HR=1.394, $p<0.05$).
- AFT model might act a good alternative for the Cox PH model. However, more model evaluation work should be done such as using Akaike Information Criterion to compare, and checking the robustness under small sample size or large censoring circumstances.

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL