PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA

4TH OCT 2022 (VIRTUAL) & 5TH OCT 2022 (ILSM, SUNGKAI, PERAK)

9TH MALAYSIA STATISTICS CONFERENCE

Department of Statistics, Malaysia

Dealing with Uncertainties: Unearthing Measures for Recovery

Preliminary study on the survival analysis using censored lung cancer data

Siti Afiqah Muhamad Jamil[1*], Nurain Ibrahim[1]

[1]College of Computing, Informatics and Media, Universiti Teknologi MARA, Shah Alam, 40450, Selangor, Malaysia

afiqahjamil@uitm.edu.my

INTRODUCTION

- The primary goal of **Exploratory Data Analysis (EDA)** is to make data "clean," which means it should be free any redundancies.
- The aims of this study are **to perform a preliminary study prior to the application of survival method of analysis on the lung cancer censored observations.**
- This study used the **Kaplan-Meier survival curve**, proportional hazard assumption, **time-varying covariate** assumption by using **Scaled Schoenfeld residuals**, **Cox-Snell residuals** for the overall goodness of fit of the model assumption, and normality assumption using **quantile-quantile (Q-Q) plot.**
- Despite all survival analysis studies that has been employed previously, the **flow of preliminary study involving survival analysis was not comprehensively explained.**
- Thus, this study could help in **exposing appropriate assumptions of survival analysis.**

METHODOLOGY

Non-parametric distribution assumption – Kaplan-Meier (K-M) survival plot and the log-rank test for the categorical variables

Parametric distribution assumption – Check on the normality are quantile-quantile (Q-Q) plot or probability (P-P) plot

Semi-parametric distribution assumption – Cox proportional hazard model where this study had to test on the proportional hazard assumption and the time varying covariate assumption by using either Scaled Schoenfeld residuals or Unscaled Schoenfeld residuals and either Nelson-Aalen plot or log minus log graph.

RESULTS

Table 1: Descriptive statistic for continuous data

Characteristics (Continuous data)	No. (%)
Age (year)	
Mean, Standard deviation (SD)	59.85, 12.30
Median, range	61.50, (13,85)
≤ 60	26 (41.15)
> 60	28 (51.85)

Table 2: Descriptive statistic for categorical data

Characteristics (Categorical data)	No. (%)
Types of lung cancer	
Adenocarcinoma, NSCLC	19 (35.19)
Large Cell Carcinoma, NSCLC	17 (31.48)
Small Cell lung cancer, SCLC	12 (22.22)
Squamous cell carcinoma, NSCLC	6 (11.11)
Treatment of lung cancer	
Chemoradiotherapy, CCRT	32 (59.26)
Chemotherapy, Chemo	5 (9.26)
Chemotherapy and Surgery, Chemosurgery	11 (20.37)
Chemotherapy and Targeted Therapy, ChemoTarget	5 (9.26)

Figure 1: Timeline of the duration of survival

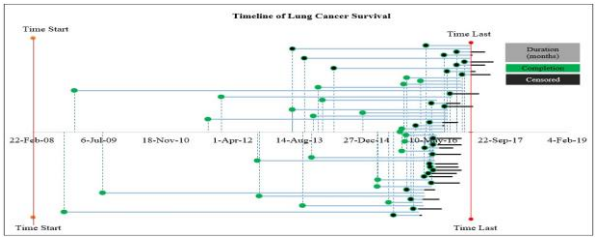


Figure 2: Kaplan Meier survival curve and log-rank test

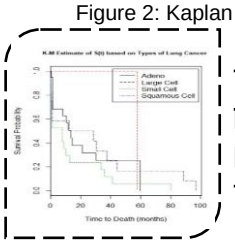


Table 3: Log-rank test's output

Log-rank (Mantel-Cox)	z-value	df	P-value
Gender	0.919	1	0.358
Races	-0.289	2	0.773
Types	0.937	3	0.349
Treatments	0.923	3	0.356

Figure 3: Proportional hazard

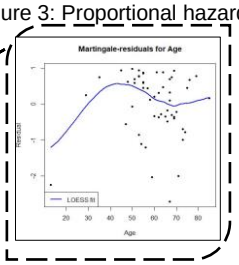


Figure 4: Schoenfeld Residuals

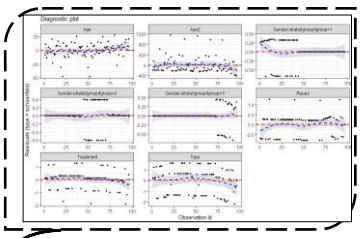


Figure 5: Q-Q plot

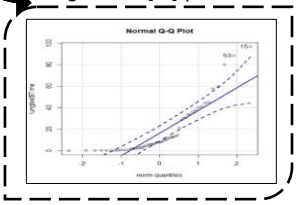
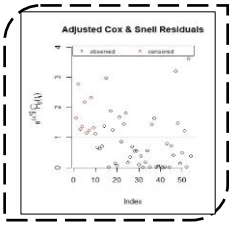


Figure 6: Adjusted Cox-Snell Residual



CONCLUSIONS

1. Kaplan-Meier survival plot with log rank test was violated.
2. Proportional hazard assumption was violated at one of the categorical variables of gender. The violation of proportional hazard proved that the covariates was not time-fixed and has been solve by using the split method.
3. The parametric assumption was satisfied as the quantile-quantile plot shows the constant pattern about 45º. Some outliers also have been detected.
4. Overall fit model using Cox-Snell residuals shows that the distribution fit the lung cancer data and cox model also quite good at handling the time-varying covariates.

Since, the assumptions for **semi-parametric cox-proportional hazard and non-parametric Kaplan-Meier of analysis have been violated** which leads to the application of **parametric methods of survival analysis**

REFERENCES

(1) Alotaibi, A., Ali, A., Brown, C., & Sherbeny, F. (2022). Exploratory Analysis of Survival and Mortality Rates among Older Lung Cancer Patients Utilizing Different Treatment Modalities. *INNOVATIONS in pharmacy*, *13*(2), 15-15. (2) Ainomugisha, S. (2022). *Comparison of semi-parametric and parametric survival analysis models for identifying predictors of virological suppression among HIV-infected adults on ART at the Joint Clinical Research Centre, Uganda* (Doctoral dissertation, Makerere University). (3) Kurniati, A. P., Wisudiawan, G. A. A., & Kusuma, G. P. (2022, July). Potentials of Clinical Pathway Analysis Using Process Mining on the Indonesia National Health Insurance Data Samples: an Exploratory Data Analysis. In *2022 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 294-299). IEEE. (4) Rossi, E., Zizzari, I. G., Di Filippo, A., Acampora, A., Pagliara, M. M., Sammarco, M. G., ... & Nuti, M. (2022). Circulating immune profile can predict survival of metastatic uveal melanoma patients: results of an exploratory study. *Human Vaccines & Immunotherapeutics*, 1-10. (5) Nieder, C., Mannsåker, B., & Yobuta, R. (2022). Dual Use of the METSSS Model Predicting Survival After Palliative Radiotherapy: An Exploratory Analysis. *Cureus*, *14*(1).