Evaluating and Presenting the Survival Analysis Results

OBJECTIVES

- To evaluate performance of developers in terms of time to complete tasks and success (complete correctly) using survival analysis
- To explore the ethical considerations in survival model designs

Business context and Question

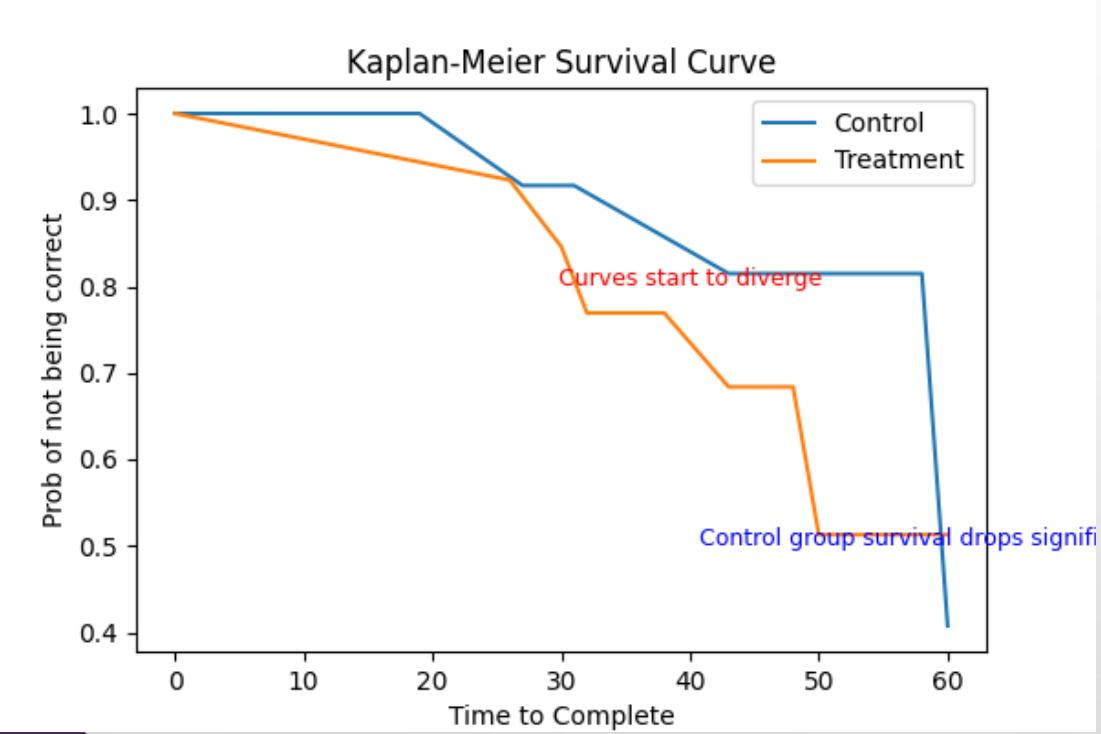
Performance measurement and prediction are key in any business undertaking.

Key question: Is there significant difference in performance between two groups of developers control and treatment in terms of terms of time to complete tasks until correct completion?

METHODOLOGY

Developers Milestones-Dataset with 26 rows and 28 columns was used to compare developers' performance over time until complete correctly and establish if significant differences using Kaplan Meier, Log rank and Cox hazard ratios.

Key Findings



Kaplan Meier shows there is difference with developers in treatment having low chance of not being completed correctly hence better performance than control as shown in the figure.

Key Findings cont.

TABLE 1:Log-Rank Test			
Statistics			**p-Value**
0.73			**0.39**
TABLE 2: Cox PH Hazard Ratios with Predictors			
Factors	Risk	95% CI	**p-Value**
Issue	0.95	0.76 - 1.2	**0.69**
Problem	0.99	0.9 - 1.1	**0.91**
Folder	1.02	0.84 - 1.24	**0.83**
File	0.97	0.48 - 1.99	**0.94**
Class	1.08	0.38 - 3.1	**0.88**
Method	0.86	0.46 - 1.6	**0.63**
Line	1.07	0.75 - 1.53	**0.71**
Solution	1	0.85 - 1.17	**0.98**
Implement	0.86	0.69 - 1.06	**0.16**
group_Treatment	0.91	0.11 - 7.64	**0.93**

- The difference noted was tested to determine is it is significant using log rank and cox proportional hazard. The results shows the difference is **not significant** this occurs when **p-value in test is greater than 0.05**.

Recommendation

- Based on Kaplan Meier I would recommend developers in treatment groups since they perform better , however, since log rank and Cox test results shows the difference were due to chance **further test** with better data quality and adequate sample size needs to be undertaken for better data driven decisions

Ethical Consideration

- **Fairness**: Data used is not representative enough, hence need of preventing discrimination against certain groups by ensuring equal representation
- **Transparency:** Interpretation should be transparent without influence of pre-existing difference like the one shown by Kaplan curves.
- **Impact:** Relying on the results to make decisions could cause unfair advantages and discriminate other developers hence need for further check with quality data.