Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN 3025-0803

Vol. 3, No. 2, pp. 48-53, 2025

Applicating Kaplan-Meier Survival Analysis on Employee's Tenure by Overtime Availability Status

Ferdinand Nathaniel Widjaya[1*]

[1]*Statistics Study Program, Faculty of Science, Technology and Mathematics, Matana University, Tangerang, Indonesia*

**Corresponding author email: ferdinand.widjaya@student.matanauniversity.ac.id*

Abstract

This study aimed to analyze the impact of overtime work on the tenure of employees using the Kaplan-Meier method. The data used in this research encompassed employees' tenure records and their overtime work status in a specific company. Utilizing the Kaplan-Meier method, the analysis revealed that employees engaged in overtime work exhibited a tenure ends probability higher than those who does not. The Kaplan-Meier Survival Analysis shows a much steeper step-down probability for Overtime than those who don't. Despite the substantial disparity in tenure probabilities on the graph, furthermore statistical analysis (log-rank test) indicated that an employee engaged in overtime work did significantly influence the duration of their tenure. These findings provide crucial insights into workload during overtime and overtime work policies within organizations could affect the employee significantly.

Keywords: Employees Tenure, Kaplan-Meier Method, Overtime

1. Introduction

Employees are a crucial element in company operational and growth. The relationship between employees and the company is a mutualistic or mutually beneficial relationship; the employees give what it has in terms of thoughts and abilities, and then the company pays them with wages. Employees are responsible towards company productivity through targeted goals or responsible given. However, difficulty in the process and company's increased need of productivity could occur, forcing the employees to work extra outside their business hours. This results in excessive working hours accumulated called overtime. Overtime is when an employee has accumulated work hour more than 40 hours in a week or more than accumulated work hour per week as written before in the contract and the company is responsible for paying extra wage as a compensation (Thomas, 1992). Working overtime and get paid more sounds like an extra gain for the employees although the downsides were also increased. Employees who work overtime means that more hours spent under work pressure can cause stress and burnout, all of which leads to exhaustion. The significant effects of fatigue from overtime work could be seen from their lack of attentiveness. This certainly cause declining on focus and productivity as been studied by Tuti Surmaningsih, the productivity on an overtime shift tends to decrease by 11,8% rather than normal shift (Hanna, et al., 2005). The lack of focus and attention are also the exposing employees in a hazardous state of workplace incidents, as employees who experience fatigue from overtime are at risk of almost 3 times to experience workplace accidents compared to workers who do not experience any fatigue (Okumus, et al., 2023). Both major issues on overtime could trigger a chain reaction on the employee years of work in a company due to possibly laid off for decreased productivity, injuries from workplace accidents and any stress-related issues.

Based on the problems mentioned before, therefore become the main issues this research to find out if overtime could affect employee's years of work within a company is conducted. In this study, the method applied is the Kaplan-Meier survival analysis method to calculate the chances of employee surviving through a certain period of years working and is classified based on 2 groups, namely the accepting overtime work group and refused to working overtime group. Kaplan-Meier analysis is mainly used on studies related to chance of subjects to survive throughout period of time (Rich et al., 2010). The distinguishable main features on Kaplan-Meier are the declining staircase plot describing the probability to survive over the duration of time. The usage of Kaplan Meier method on employee years of work based on overtime availability is mainly inspired to previous study in a related field, which used Kaplan-

Meier on employee years of work based on gender and positions until their voluntary turnover (resign) (Morita et al., 1989).

Based on the background of the above problem, the problem formulation in this study is:

a) Can Kaplan-Meier method be applied to properly described the probability of overtime and non-overtime employee's years at work?
b) Does overtime significantly affects employee's years at work?

Based on the problem given above, the purpose of this study is to finding out that:

a) Applicating Kaplan-Meier survival analysis to describes the likelihood of an employee's longevity of employment based on their availability to work overtime.
b) Determining the impact of overtime towards employee's years at work

Based on the problem given above, the benefit of this study is to finding out that:

a) To contribute scientific research on human resources field studies on overtime issues
b) To contribute on socializing the effects of working overtime

The problem limitations is the dataset used in the study is a secondary data gained from a company, couldn't be generalized 100% correctly in any company or situations.

2. Literature Review

2.1. Survival Analysis

Methods include: the stages and formulas that are used in data analysis, arranged sequentially step by step. Survival Analysis is a statistical procedure to analysing the probability of time until the desired events occurred within a certain period of date, the events occurred could be death, failure or any endpoint regarding fields of study (Kleinban & Klein, 2012). To sterilize the data from other occurring events and incomplete observations, data censoring is a necessary step to reduce biased and incorrect conclusions. According to Klein & Moeschberger, there are 3 main censors in survival analysis (Klein & Moeschberger, 2003):

a) Left-Censored: An observation is considered left-censored when the observed event occurs before the beginning of the specified time observation period.
b) Right-Censored: An observation is right-censored when the subject being studied isn't experiencing the events after the specified time period or disappeared from observation before the end of the specified observation period.
c) Interval-Censored: Observations are interval-censored when the other events rather than the main occurs under the period between two specific observation time points.

The general mathematical model for given survival probability of a subject after ***t*** period of time according to Kleinban & Klein, 2012.

$$S(t) = P(T > t), t \geq 0$$
$$S(t) = 1 - F(t)$$
$$Whereas, S(0) = 1 \; and \lim_{t \to \infty} S(t) = 0$$

Where,

S(t) : Probability on surviving before *t* time
F(t) : Cumulative Probabilities on events occurring to subjects in *t* time
P (T > t) : Probability that the event occurs before *t* time

2.2. Kaplan-Meier

The Kaplan-Meier method was first introduced by Kaplan and Meier in 1958. The Kaplan-Meier method is a non-parametrical modification of the survival function used to overcome the problem of incomplete data (Zee & Xie, 2018). Kaplan-Meier step-down curve could be used on different groups of samples within the same survival plot based on their treatment or classifications which the log-rank test will be used later on to measure is there any significant difference between groups and their survival probability. The general model of a Kaplan-Meier survival probability is (Kleinban & Klein, 2012):

$$\hat{S}(t) = \prod_{t=1}^{T} \left(1 - \frac{d_{g,t}}{n_{g,t}} \right), \qquad g \in \{1, 2, \cdots, G)$$

Where,

$\hat{S}(t)$: Kaplan-Meier Survival Probability before t time
$d_{g,t}$: Number of events occurred in group g at t time
$n_{g,t}$: Number of subjects in group g at t time

2.3. Log-rank test

The Log-rank test is used after the survival probability distribution of Kaplan-Meier were created. The Log-rank test purpose it to evaluate whether the probability distribution of the Kaplan-Meier model on two or more groups has any significance differences over treatment or classifications. The log-rank test is actually a Chi-square test for larger samples or groups, as the general formula below:

$$X^2 = \frac{\left(\sum_{t=1}^{T}(O_{g,t} - E_{g,t})\right)^2}{Var\,(O_g - E_g)}$$

Where,

$O_{g,t}$: Observed event occurred in group g at t time

$E_{g,t}$: Expected events occurred group g at t time

Interpretation:

p-value < α: Reject H_0 (curves has significant difference).

p-value ≥ α: No significant difference.

3. Materials and Methods

3.1. Materials

The dataset used in this study is secondary dataset of IBM HR gained from Kaggle uploaded by Pavansubhash. The dataset key features which will be used are: Attrition (Whether the employee still work there or has resign), Overtime availability, and Years worked at the company.

3.2. Methods

This research method uses a quantitative of Kaplan-Meier survival analysis. The flow of this study can be seen as figure below. This flowchart breakdowns the process to Kaplan Meier Survival Analysis on Human Resources field, and proofing the hypothesis of group's difference using log-rank test. The process begins with the necessary data collection stage.

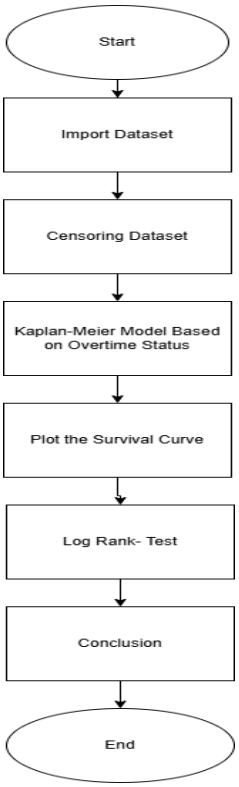

Figure 1: Research Flowchart

After the data is collected, the data is classified into two parts, namely data for overtime employee and data of employee who doesn't work overtime. Then, using attrition as an indicator of whether if the employee still occupied in the company as a survival measurement. Data censoring applies to employees left before period of observation begun (Left-censored) and employee that stays after the period of observation end (Right-censored). After the Kaplan-Meier model based on overtime groups has been created, log-rank test was performed to further solidified the hypothesis that overtime have a significant effect to employee's years at company.

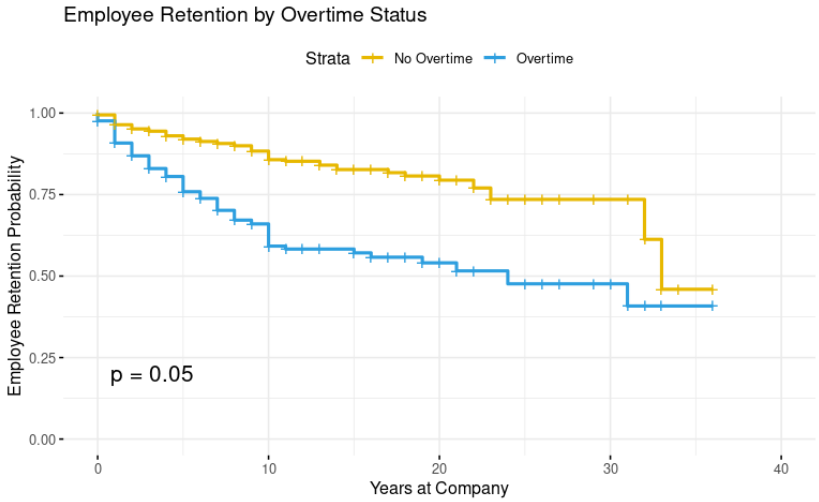

Figure 2: Kaplan Meier Survival Probability

4. Results and Discussion

The results of the Kaplan-Meier survival analysis using $\alpha = 5\%$ are summarized in the plot figure 2. The survival probability throughout years is calculated based on general Kaplan-Meier model of equation 2.2. Using the Surv() function, available in the survival() library in Rstudio software, we find out that employee who accept overtime are more likely to leave the company earlier than those who don't accept overtime as shown by the steeper drop in the curve.

OverTime=No

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	1052	6	0.994	0.00232	0.990	0.999
1	1022	31	0.964	0.00579	0.953	0.976
2	906	12	0.951	0.00679	0.938	0.965
3	814	6	0.944	0.00731	0.930	0.959
4	726	11	0.930	0.00838	0.914	0.947
5	646	7	0.920	0.00911	0.902	0.938
6	507	4	0.913	0.00974	0.894	0.932
7	455	3	0.907	0.01027	0.887	0.927
8	386	3	0.900	0.01097	0.878	0.921
9	330	6	0.883	0.01264	0.859	0.908
10	267	8	0.857	0.01534	0.827	0.887
11	179	1	0.852	0.01598	0.821	0.884
13	144	2	0.840	0.01782	0.806	0.876
14	124	2	0.827	0.01994	0.788	0.867
17	87	1	0.817	0.02186	0.775	0.861
18	80	1	0.807	0.02385	0.762	0.855
20	63	1	0.794	0.02669	0.743	0.848
22	33	1	0.770	0.03509	0.704	0.842
23	22	1	0.735	0.04787	0.647	0.835
32	6	1	0.613	0.11874	0.419	0.896
33	4	1	0.459	0.15975	0.232	0.908

Figure 3: Non-overtime Survival Distribution

OverTime=Yes

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	416	10	0.976	0.00751	0.961	0.991
1	402	28	0.908	0.01423	0.881	0.936
2	347	15	0.869	0.01684	0.836	0.902
3	312	14	0.830	0.01903	0.793	0.868
4	272	8	0.805	0.02034	0.766	0.846
5	242	14	0.759	0.02265	0.716	0.804
6	185	5	0.738	0.02383	0.693	0.786
7	161	8	0.702	0.02593	0.653	0.754
8	140	6	0.672	0.02757	0.620	0.728
9	116	2	0.660	0.02829	0.607	0.718
10	97	10	0.592	0.03254	0.531	0.659
11	65	1	0.583	0.03329	0.521	0.652
15	50	1	0.571	0.03460	0.507	0.643
16	43	1	0.558	0.03626	0.491	0.634
19	32	1	0.540	0.03909	0.469	0.623
21	22	1	0.516	0.04437	0.436	0.611
24	13	1	0.476	0.05595	0.378	0.599
31	7	1	0.408	0.07916	0.279	0.597

Figure 4: Overtime Survival Probability

As shown in figure 3 and 4 the difference between group's survival probability is highly noticeable, as overtime employees have lesser probability of surviving compared to non-overtime employee. This can be further tested using log-rank test to identified is there truly any significant difference between the group's probability of surviving.

```
Call:
survdiff(formula = surv_obj ~ OverTime, data = df)

                N Observed Expected (O-E)^2/E (O-E)^2/V
OverTime=No  1052      109    169.9      21.8      80.2
OverTime=Yes  416      127     66.1      56.2      80.2

 Chisq= 80.2  on 1 degrees of freedom, p= <2e-16
```

Figure 5:. Log Rank Test

Using α = 5%, the log rank test result p-value is smaller compared to α which means that there is significant difference between the group's survival probability. Therefore, can be concluded that overtime does significantly affect how long an employee stays at company.

5. Conclusion

Based on the study that has been conducted above, it can be concluded that Kaplan-Meier survival analysis is a perfectly suited methods in researching the probabilities of how long employees will stay within the company based on overtime status. Both of the groups turn out has their own probability of employee's survival rate. We can see this from the results from figure 4.1 which visualize the probabilities given by each group. Furthermore, the log-rank test (figure 4.4) proving significant difference in both group survival probabilities, confirming that overtime can affect how long an employee stays. It's possible due to other reasons, that non-overtime employees could have the probability of stay way higher than those who work overtime such as stress-level or work-life balance. The next suggestion for related research, is that Kaplan-Meier can be applicated to Human Resources field for deep statistical analysis and can be applicated through different dataset and groupings.

References

Hanna, A. S., Taylor, C. S., & Sullivan, K. T. (2005). Impact of extended overtime on construction labor productivity. Journal of construction engineering and management, 131(6), 734-739.

IBM HR Analytics Employee Attrition & Performance. (n.d.). Retrieved March 20, 2025, from https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset.

Klein, J. P., & Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censored and Truncated Data.

https://doi.org/10.1007/B97377

Kleinbaum, D. G., & Klein, M. (2012). Survival Analysis. https://doi.org/10.1007/978-1-4419-6646-9

Morita, J. G., Lee, T. W., & Mowday, R. T. (1989). Introducing survival analysis to organizational researchers: A selected application to turnover research. Journal of Applied Psychology, 74(2), 280–292. https://doi.org/10.1037/0021-9010.74.2.280

Okumus, D., Fariya, S., Tamer, S., Gunbeyaz, S. A., Yildiz, G., Kurt, R. E., & Barlas, B. (2023). The impact of fatigue on shipyard welding workers' occupational health and safety and performance. Ocean Engineering, 285, 115296.

Rich, J. T., Neely, J. G., Paniello, R. C., Voelker, C. C., Nussenbaum, B., & Wang, E. W. (2010). A practical guide to understanding Kaplan-Meier curves. Otolaryngology--head and neck surgery: official journal of American Academy of Otolaryngology-Head and Neck Surgery, 143(3), 331–336. https://doi.org/10.1016/j.otohns.2010.05.007

Thomas, H.R. (1992). EFFECTS OF SCHEDULES OVERTIME ON LABOR PRODUCTIVITY. Journal of Construction Engineering and Management. Vol. 118, No.1

Zee, J., & Xie, S. X. (2018). The Kaplan-Meier Method for Estimating and Comparing Proportions in a Randomized Controlled Trial with Dropouts. Biostatistics & epidemiology, 2(1), 23–33. https://doi.org/10.1080/24709360.2017.1407866