Survival analysis project 2

FermaLogis Inc. Employee Turn-over Analysis

**Group 2**

* **Suvir Gupta**
* **Reena Malipatil**
* **Sriram Rajagopalan**
* **Parmod Rathee**
* **Kai-Hsun Wang**
* **Deeksha Ramesh**

Contents

[Executive Summary 1](#_Toc487226852)

[Data Preprocessing: 2](#_Toc487226853)

[Visualization: 3](#_Toc487226854)

[Modeling and Inference: 10](#_Toc487226855)

[Conclusion and Recommendations: 15](#_Toc487226856)

[Appendix: 15](#_Toc487226857)

# 

# **Executive Summary**

Employee attrition is one of the major concerns in all companies. FermaLogis is a pharmaceutics company facing this problem and we are hired to find the reasons for this cause. As per domain knowledge, there could be various reasons such as working overtime, job satisfaction level, business travel, employee working department etc. that could lead to employee turn over. FermaLogis dataset measures these variables and reports the proportions of employess that have left the company. In this survival analysis, we will profile the employees who are leaving the company and what is the maximum percentage of turn over that an attribute can cause. We decided to split the turn over types into different categories and tried to model by the combination of similar turn over types. We used different techniques in order to visualize the effect of each turn over and identified that “retirement” does not attribute to attrition and hence decided to model without considering that turn over type. Our team has used Cox’s partial likelihood estimate method to build models with censored data(considering only voluntary, involuntary resignation and termination). Cox’s regression is used to investigate the effect of variables on the time specified events to happen. The regression model suggests that Business travel (frequent travel), overtime and Job Involvement are the factors that cause increase in the Hazard rate. Similarly, other factors such as Age, Total Working Years decreases the Hazard rate. We further analyzed the variables that affect the turn over non-proportionally and identified that the variable “Years in Current Role” to be the most significant non-proportional Hazard. The Schoenfeld residual analysis suggests that Age, Total Working years, Years in current role, years with current manager and Total bonuses are the variables which have interaction factors

The content of this report describes the sampling, exploring, modifying, and modeling process that was implemented to arrive at the conclusions reported above. The report emphasizes the results that are significant and have been achieved using standard statistical method and helps identify the factors that influence voluntary resignation. Once the reasons are identified, further actions can be taken by the company based on the recommendations provided to reduce the turnover rate.

# **Data Preprocessing:**

**Variable Creation:**

On initial analysis, we observed that if the type was 0 then there was no employee turn-over and if the type was 1, then there was a turn-over due to specific reason. This reason was determined based on the turn over code ranging from 0 to 4. We created a new variable named “turnover type” in order to specify the reason behind employee attrition. The following values were assigned:

type=1 then “Retirement”

type=2 then “Voluntary Resignation”

type=3 then “Involuntary Resignation”

type=4 then “Job Termination”

type=0 then “No turnover”

**Data Clean-up:**

We observed a few columns such as “X”, “Over 18”, “Employee Count”, “Employee Number” and “Standard Hours” that were either unique identifiers or not significant to build survival models. Hence, these columns were removed from the dataset.

**Variable Encoding:**

We observed the categorical variable named Stock and assigned values “yes” and “No” based on the stock option level. For the all the values where Stock option level was greater than 1, the value “Yes” was assigned and “No” for the rest.

Finally, we created a variable “Censored” in order to take Censoring into account. If the TYPE is 0, then there is no turn over and hence “1” was assigned to the new variable “Censored”. If variable TYPE is 1, then employee turn- over has taken place and hence the value “0” was assigned to the variable “Censored”.

**Aggregation:**

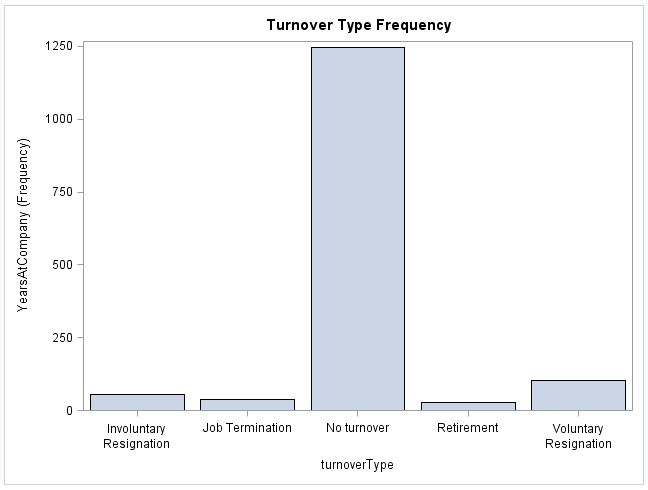
Since there were 40 columns which had details about bonuses given to the employees, we decided to aggregate them into a single variable named “Total Bonus” which gives the sum of the bonus received by each employee. To deal with the “NA” values, we decided to encode them to value 0 which means that the employee did not receive any bonus.

# **Visualization:**

To understand the distribution and the structure of the data with respect to the various parameters, procedures like FREQ, SGPLOT and LIFETEST are utilized to build visualizations and provide statistical insights.

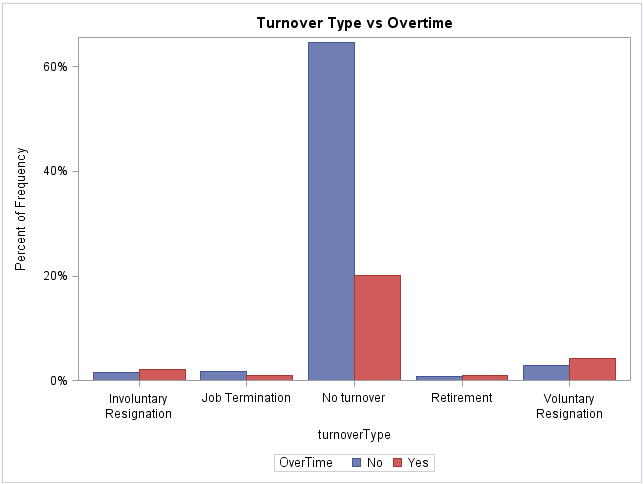
**Frequency distribution of Turnover Type**

We have plotted the frequency distribution of the variable Turnover type. As seen from the graph below, the highest reason for turnover is voluntary resignation, followed by involuntary resignation and job termination. Lot of employees leave their jobs due to family and health issues, thus necessitating better health benefits for the employees and their families.



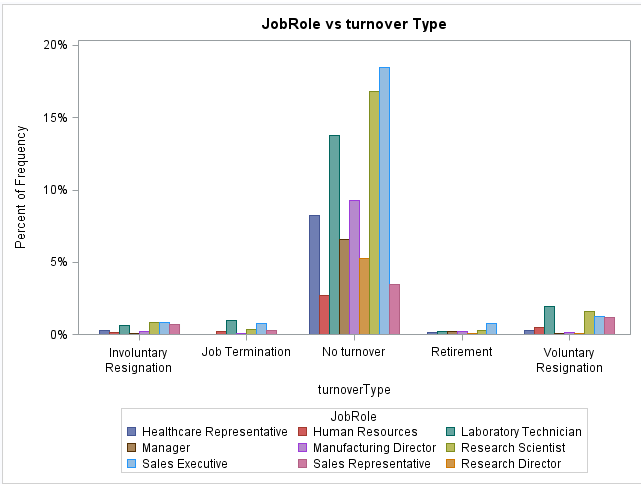
**Turnover Type and Overtime**

The Turnover type is plotted against the variable Overtime. As seen from the graph below, working overtime has been a major contributing factor in all the turnover types. The major turnover types – voluntary and involuntary resignations have a higher percent of employees who work overtime. Thus, focusing on reducing the number of overtime hours can help reduce the turnover rates.



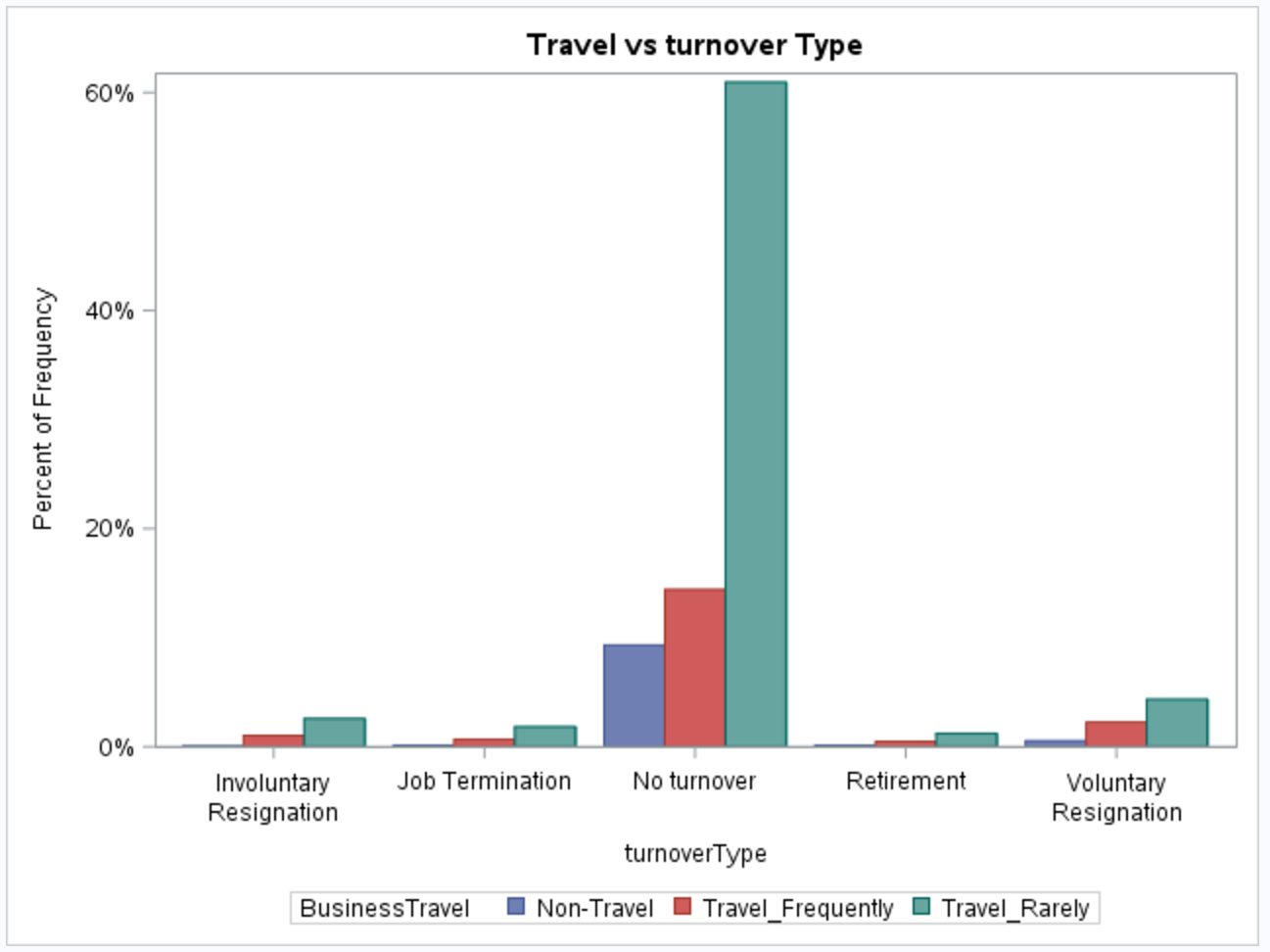
**Job Role and Turnover Type**

The SGPLOT output compares the Turnover Types with respect to various job roles. As per the graph, employees with the roles of Manager, Research Director and Human Resources have the lowest turnover. Thus, employees who are higher up the ladder are less likely to quit in comparison to young professionals. Also, employees who work as Sales Executive/Representatives and Research Scientist have the highest rate of turnover. These employees spend limited amount of time at the company and mostly resign voluntarily to pursue better opportunities.



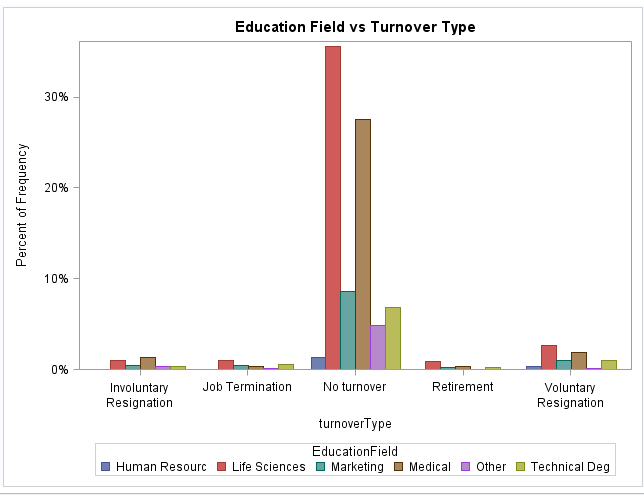
**Travel and Turnover Type**

The graph plots the frequency of travel vs. Turnover type. The values for travel are grouped into 3 clusters such as Non-Travel, Travel\_Rarely and Travel\_Frequently. As seen from the graph, a certain amount of travel also is a supportive factor for voluntary and involuntary resignation.

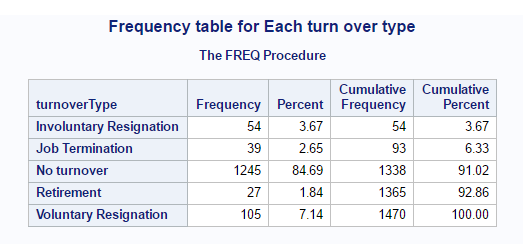


**Education Level and Turnover Type**

The graph plots the relationship between turnover type and education levels. Professionals with educational qualifications from Life Sciences, Medical and Technical Degree have the highest turnover rate. The least turnover rate being for human resources graduates.

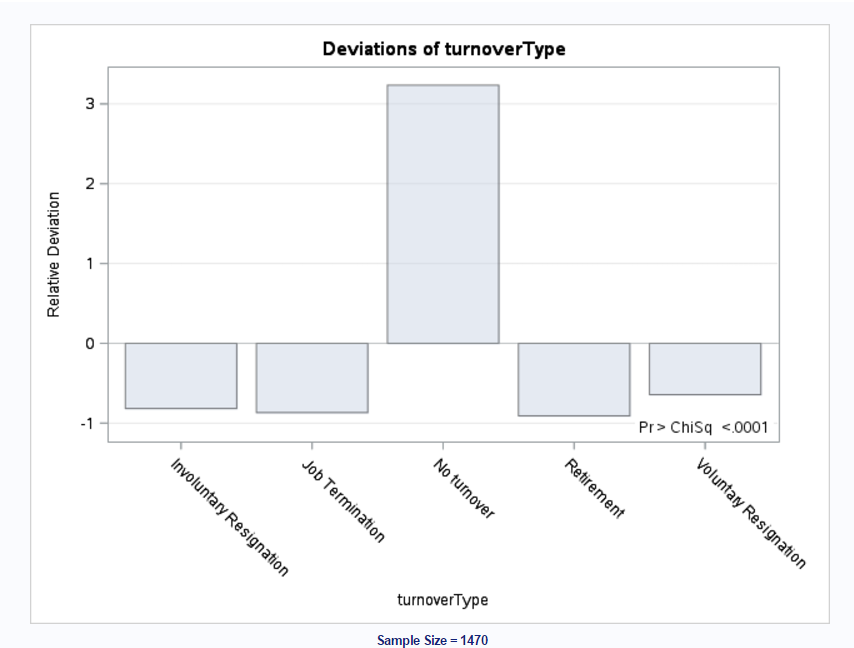


**Frequency Plot for Turn-over Type:**

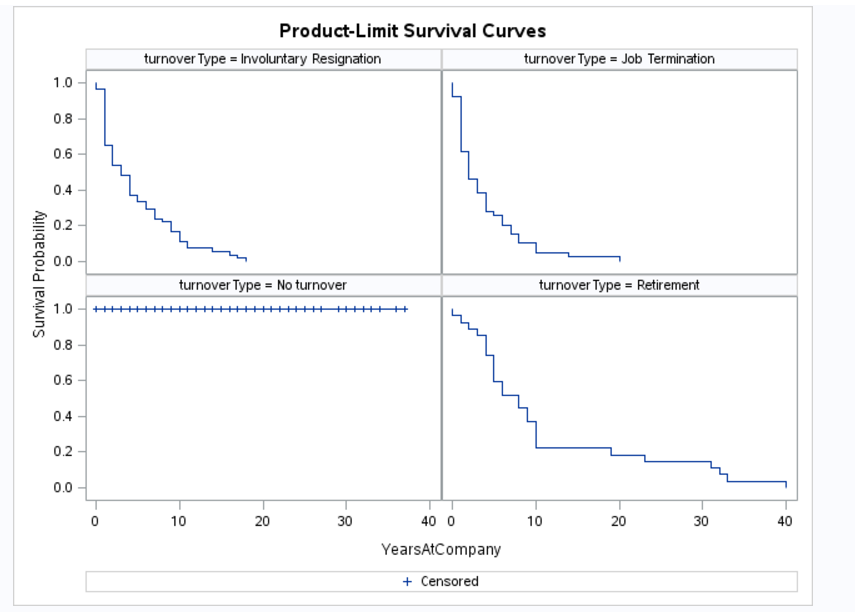


In order to statistically test the survival rates of the employees, we visualized the frequency of the different turnover types available and the percentage of employees leaving due to each type. It can be noticed that the type Voluntary Resignation has the highest frequency and percentage as well when compared to others. In order to verify if each turn-over has impact on the other with respect to attrition, we tested the statistical significance for a 95% confidence interval and observed that the test revealed that there is no relation between each type. We also observed that one turnover type reason does not impact the other. Hence the analysis can be performed by considering all the turn-over types together.

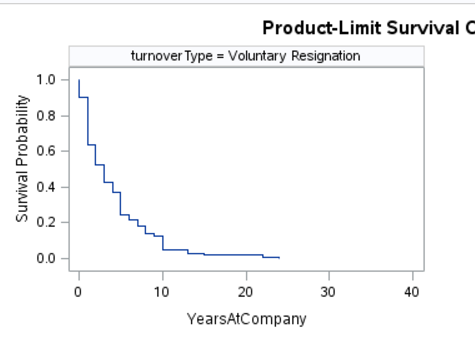
The deviation of each turn over type can be seen below with each representing a significant value thereby indicating the non-dependency.



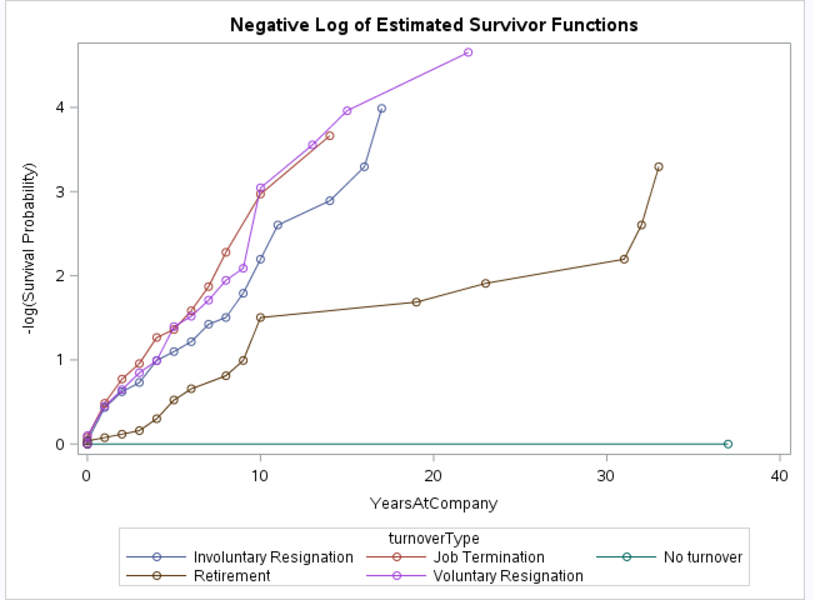
**Hazard Rates for different Attrition types:**



On plotting the Hazard rates for different turn-over types, we observed that the type “Retirement” does not have anything significant with respect to survival probability. This is self explanatory since retirement cannot be considered as Attrition. On observing the other types, it can be inferred that involuntary resignation has the highest survival probability between 0-10 years within a company and the rate slowly decreases after on. The reason may be due to personal reasons such as Health issue at a reltively higher age. Similarly, Job termination or Firing has the least survival rate when an employee stays in a company for more than 10 years which might be due to loss in work efficiency.



However, the most significant finding was when analyzing the Voluntary resignation type. There is a sharp decrease in the survival probability for employees who stay with the company between 4 and 6 years. This is the place where FermaLogis has to investigate where employees tend to resign which may account for factors such as Overtime, no bonus etc.

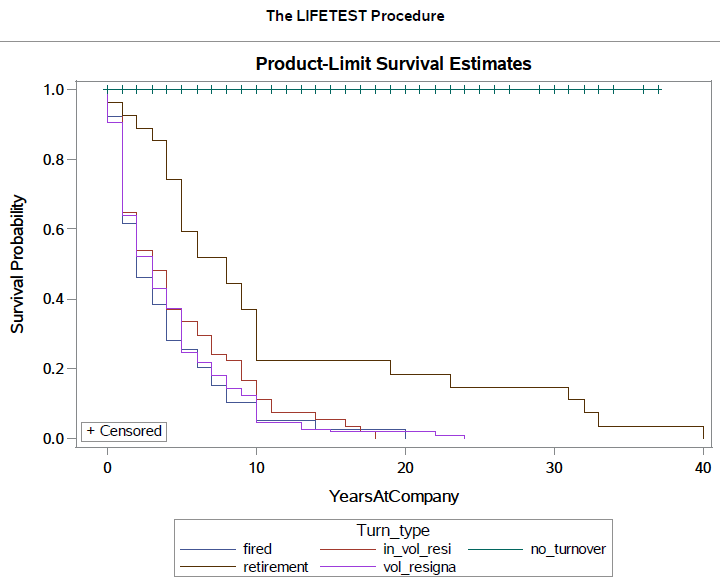


The same inference can be observed in the Negative log survival plot. As expected, the turn-over type “Retirement” does not impact the rest of the types and has a totally different probability rate. Whereas, the other types are close to one another where there might be a few common reasons which might impact the survival probabilites of an employee. This makes the need to model by considering all the turn over type together leaving behind “Retirement” and “No turnover” and also individually to study the survival rate.

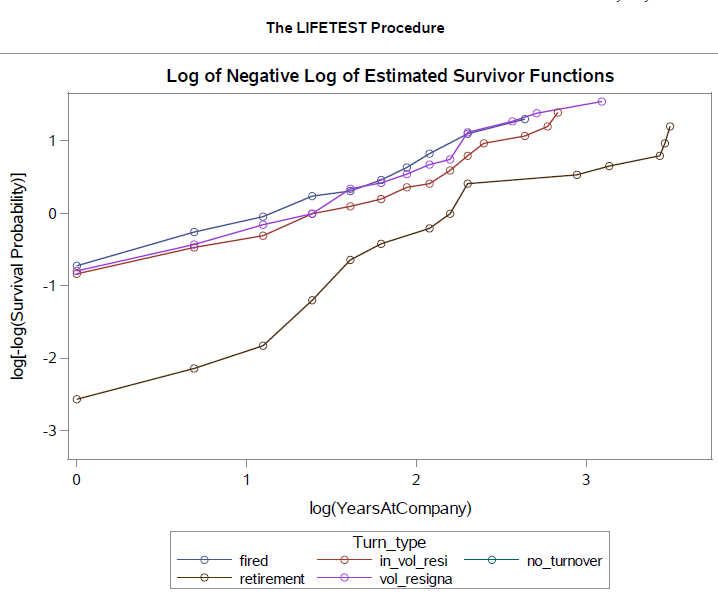
# **Modeling and Inference:**

*1)Can the events be combined or handled separately?*

We performed LIFETEST without covariates in order to determine the dependency among each turn over type so that they can be considered together.



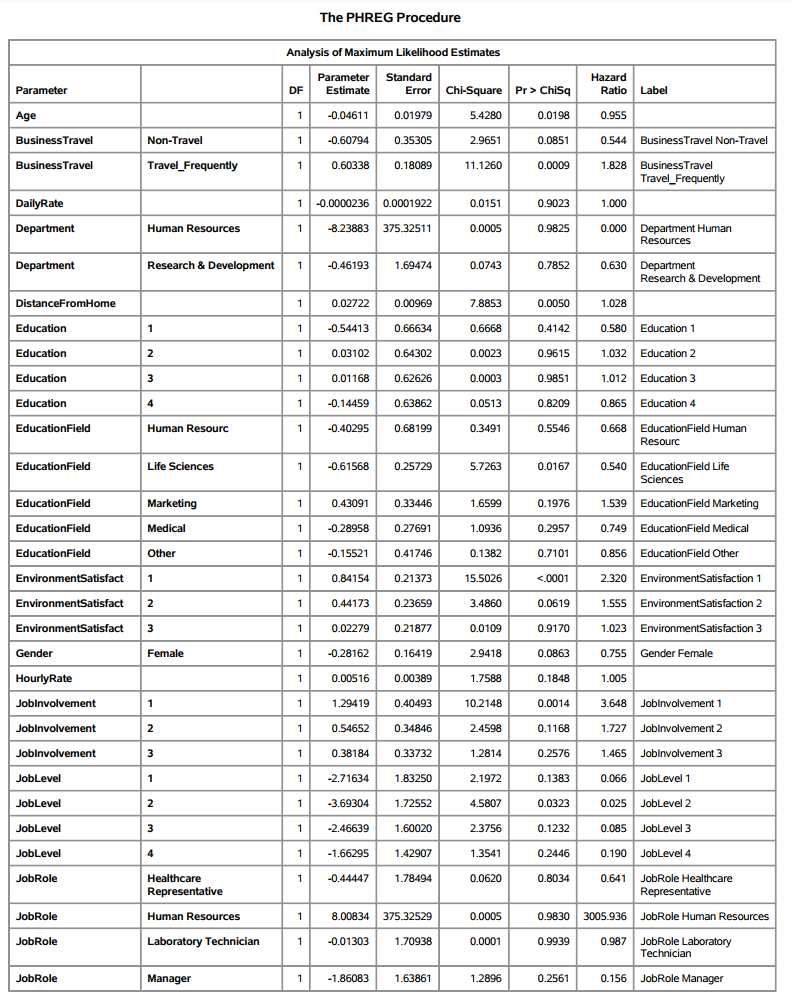
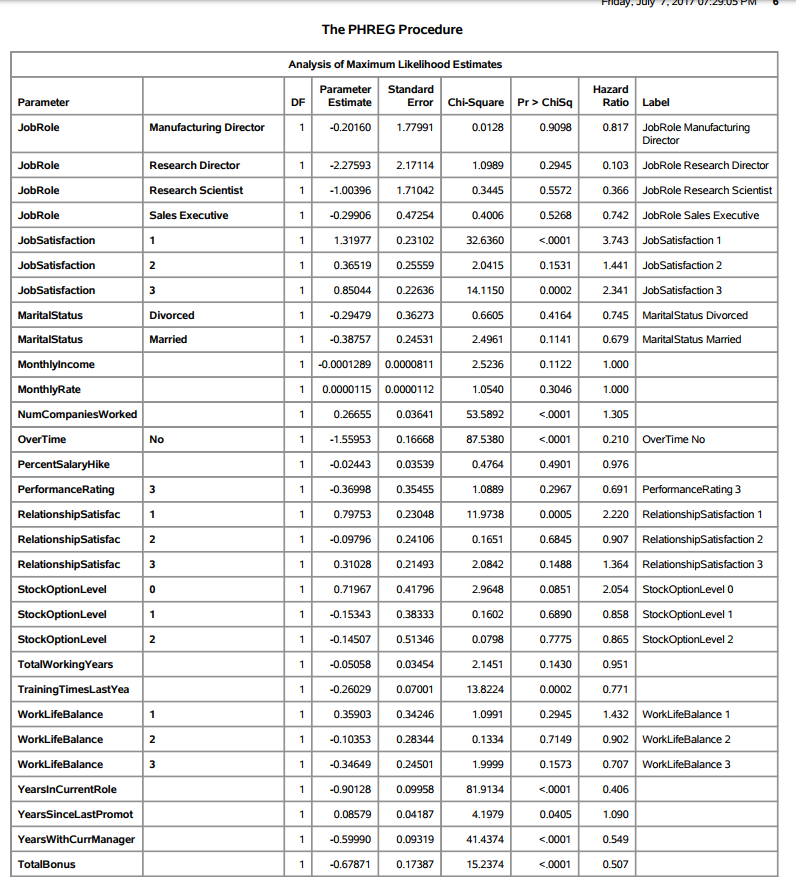
Retirement type survival curve is different from the rest of the curves. No turnover curve is parallel to the time line because they survive throughout the observation time i.e. no job attrition.



We can combine and analyze all types together except type Retirement. It is significantly different from all other types. Hence, the turn over event types can be combined except Retirement and No turn over.

*2)Attributes increasing/decreasing the Hazard rates:*

We built models using the PHREG procedure by accounting for turn over types “Voluntary Resignation”,” Involuntary Resignation” and “Termination or Fired” only. Since “Retirement” and “No-turnover” are not considered, they are censored in this approach. Following regression model was obtained:

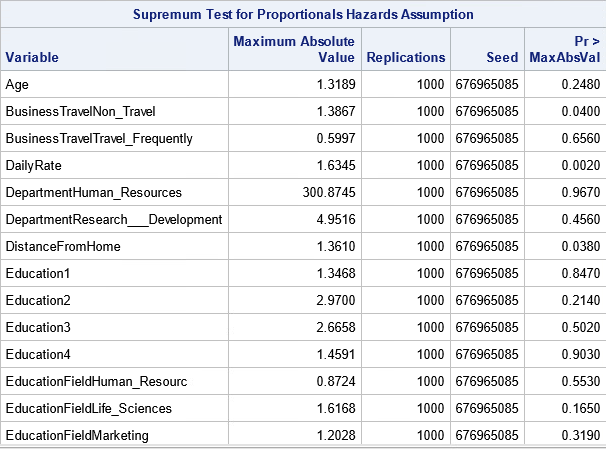


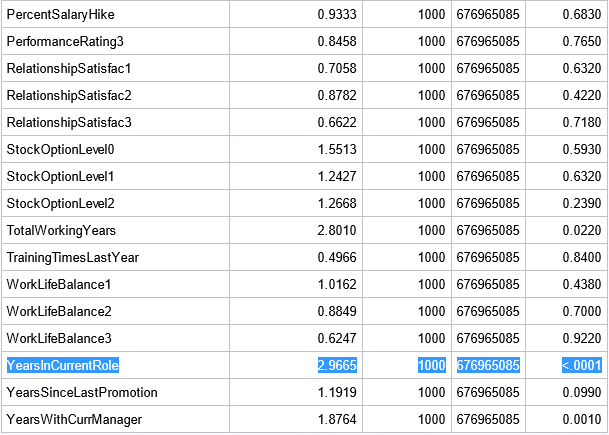
Based on the Hazard ratio from the regression model, the following are the top attributes that increases the Hazard rate are: Business travel (Travel frequent), Overtime, Job Involvement, years since last promotion, Job role – Human Resources and Stock Option.

Similarly, the top attributes that decreases the hazard rate are Age, Environment Satisfaction, total working years and Distance from Home.

*3)Are there variables which affect the hazard non-proportionally?*

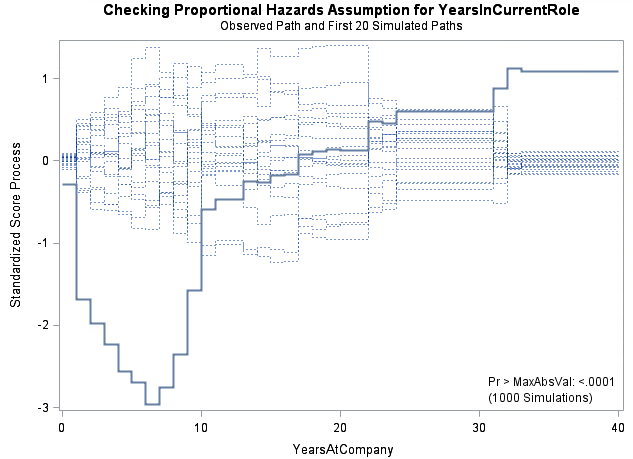
Since there were many co-variates of both categorical and numeric data type combined, we decided to look for those attributes which affect the hazard non-proportionally through supremum test for Proportional Hazards.



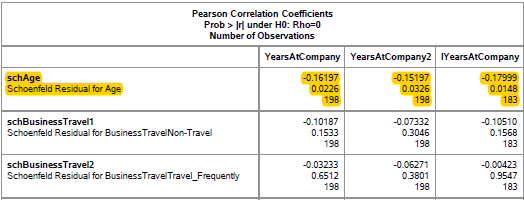


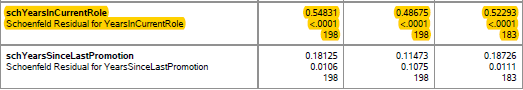
It can be inferred that “Years in Current Role” has the maximum absolute and the most important attribute which affects the Hazard non-proportionally.

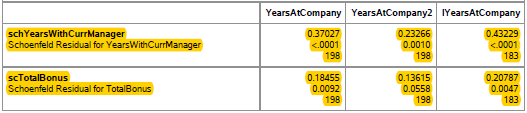
On further analyzing the attribute by simulating using the actual trend, there is a difference in actual trend and simulated one then it can be said that years in current role is having non-proportional hazard.



On further analyzing the residuals using the Schoenfeld residual analysis, it can be inferred that Age, Total Working years, Years in current role, years with current manager and Total bonuses are the variables which have interaction factors as shown below.







According to the new dataset (FermaLogis\_Event\_Type), we applied the Cox Regression Model to find out the most important factors that are affecting the employee’s retirement.

First, we modeled the all events type. The statistics are as below:



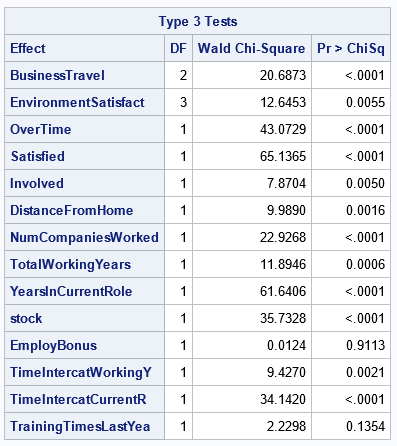
***Hazard Ratio Estimates for all events***

Base on the model of all events, we can see that the factor of employ bonus is not significant. Therefore, we tried to separate each event types (1 – Retirement, 2 - Voluntary Resignation, 3 - Involuntary Resignation (Health problems, family matters, etc.), 4 - Job Termination, Employee is Fired).

Base on the model of retirement, we know that the factor of employ bonus is significant (p-value < 0.05). In our opinion, the senior employees got higher salaries, and they got a higher cumulative bonus. Therefore, the incentive of bonus is a significant factor for employees’ turnover.



***Hazard Ratio Estimates for Retirement -*** However, employ bonus is not a significant factor for people leaving in the model of voluntary resignation, the model of involuntary resignation, and the model of job termination. The statistics results are attached below:



***Hazard Ratio Estimates for Voluntary Resignation***



***Hazard Ratio Estimates for Involuntary Resignation***



***Hazard Ratio Estimates for Job Termination, Employee is Fired***

# **Conclusion and Recommendations:**

Based on the findings and inferences above, we would like to provide the following recommendations to Fermalogis.

1. Business Travel – The Hazard rate increases exponentially for employees who travel frequently and hence fermalogis must make sure that Travel policies are made in such a way that it doesn’t affect the same employee to travel over and over again.
2. Among the turn-over types, the type “voluntary resignation” among employees who have in the company between 6 and 10 years has been the highest which can be accounted for Overtime and other factors.
3. Education Fields – Life Sciences and Medical have the highest turn-over type and hence the management should focus on employees from these fields and address the pertinent issues.
4. Job involvement – Another important factor which causes increase in hazard rate and hence the management needs to focus on encouraging employees to bring out their potentials through various forums and activities where they have a chance to showcase their skills.
5. With the modeling results, we can know that employees who retired from the company care about the bonus they have earned in their career. If the management wants to build up a sustainable business culture in this company, they should make a new bonus policy which can fully motivate their employees.

# **Appendix:**

