OPIM 5894 - SURVIVAL ANALYSIS WITH SAS

SURVIVAL ANALYSIS FOR 'FERMALOGIS'

JULY 6, 2017

Project 2 (Group 9)

SUBMITTED BY:

XUELING CHEN
YU LI
VINIT GUPTA
RAHUL MANCHANDA
PAWAN SHIVHARE
RAJARSHI DAS

Contents

Executive Summary

This report and this part of the project is an extension to the first part of the project, where this report is tried to find better answers on some of the key factors associated to employee turnover and how those factors are contributing towards the employee attrition towards the different employee group level. The project has looked upon the different types of turnover events and reduce the types to 3, also considering affects and contribution of various parameters which are time sensitive like for example employee bonus, monthly income etc. and also dealt with non-proportionally problems. Based on the detailed analysis the project has identified that a significant proportion of the employees are leaving the organization not just for one or two reasons but due to multiple reasons, which also differ in different situation. The project found out that there are several interesting contributors for attrition for different groups and has described those findings and issues in this report in a detailed fashion based on the statistical analysis. In this part of the project, it uses very specific survival analysis concepts and procedures (as directed in the project flyer) for doing the analysis. At the end, the report provided a summary of findings as well as some carefully thought out recommendations for the CEO and management team of the 'Fermalogis', there might be some limitations associated to implementations of all the recommendations at the same time, but if these items are at least monitored by the management team, it can definitely give them better insights on how to contain the employee attrition in a long run, increasing the efficiency and productivity of the organization.

Statement of Problem

In this section of the project, the goal of this project was to employ the fundamentals of survival analysis to answer many business questions associated to employee turn overs at different employee group levels of turn over. The previous analysis was superficial at many levels, as it has not employed deeper analysis and investigation on questions as follows

- What will be the effect of the result of analysis if it combine different event types?
- What are the attributes that increase of decrease the hazard rate?
- How the bonus does affects the turn over?
- Also, if there are any other significant factors which is affecting the turnover, like for example age of the employee etc.

So, in this 2nd part of the project, as you can see in the detailed section below, the project have tried to investigate these problems by employing both the competing risk concepts and predominantly using the Cox regression model for analysis of the time-dependent covariate.

Background

All the analysis of this section of the project, is an extension of the first part of the project, which originally stared based on a request from the COO of the famous pharmaceutical company 'Fermalogis' as the COO was interested to see how the team, having the knowledge of the cutting edge 'Survival analysis' techniques, can utilize those knowledge to identify the issues associated to employee attrition in his company. The project 2 has used the data provided by the company.

Data Pre-processing and Exploration

1. Variable Type Changing

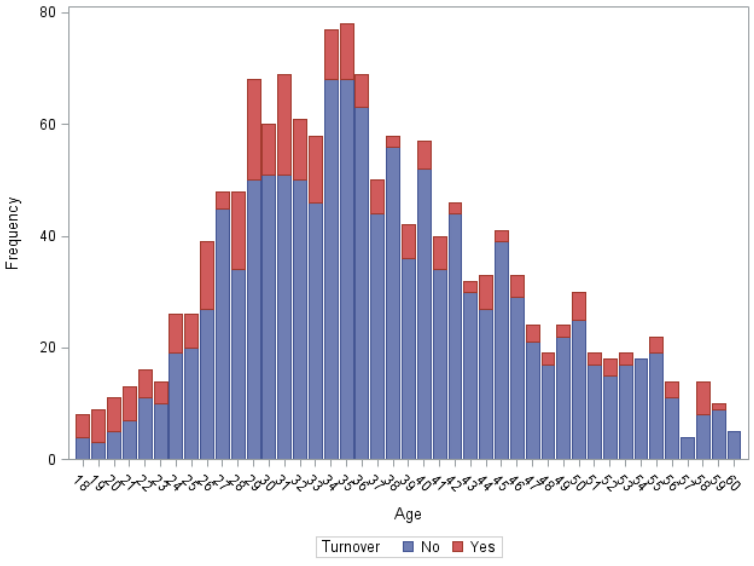
For this data set, there are some variables recoded from character to nominal type to simplify the analysis work. Here is the summary of the recoding variables.

Column Name	Explanation
Over18	1: Y 0: N
OverTime	1: Yes 0: No
JobRole	1: Healthcare Representative, Research Director, Manager Manufacturing Director 2: Laboratory Technician, Sales Representative 3: Human Resources, Manager, Research Scientist, Sales Executive
Gender	1: Male 0: Female

2. Data Exploration

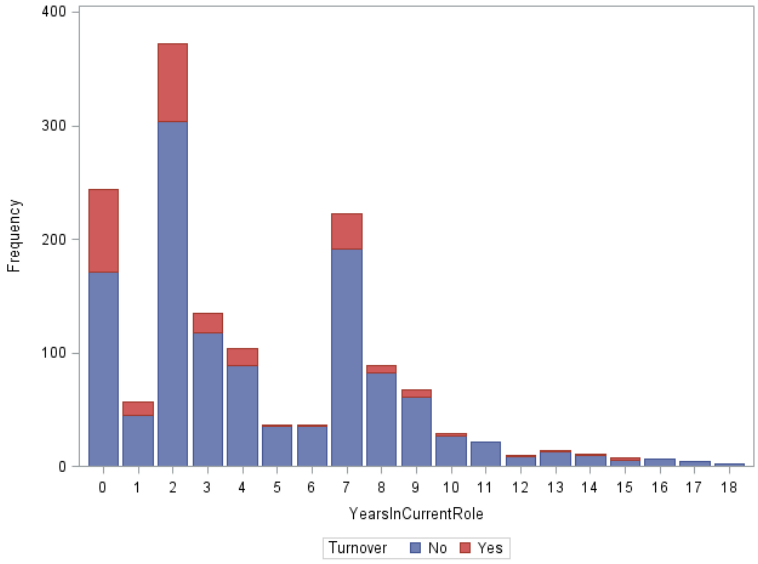
a. Potential variables that may affect the turnover frequency

- Age



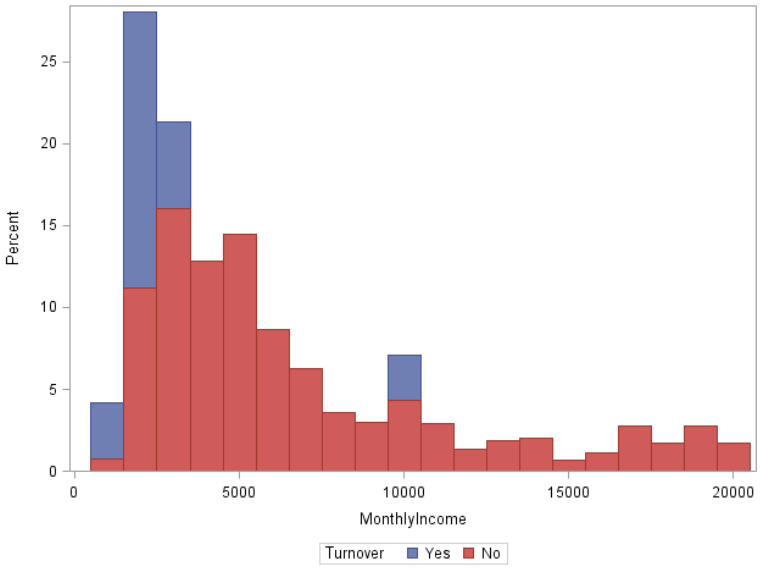
When analyzing the turnover rate of each age, it shows when employees' age is younger than 34, turnover rate is higher. By contrast, when employees' age is older than 40, turnover rate become very low and stable.

- Years at current role



It is obvious that when an employee stays in a role less than 4 years, he is more apt to leave the current position.

- Monthly Income



When monthly income increases to $5000, turnover rate decrease a lot. For the employees whose income is less than $3000, they are more willing to leave the company.

b. Variables performance based on event type

- Relationship Satisfaction

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Type by RelationshipSatisfaction

Type	RelationshipSatisfaction 1	2	3	4	Total
0	223 15.17 17.91 80.80	261 17.76 20.96 86.14	392 26.67 31.49 85.40	369 25.10 29.64 85.42	1245 84.69
1	4 0.27 14.81 1.45	3 0.20 11.11 0.99	6 0.41 22.22 1.31	14 0.95 51.85 3.24	27 1.84
2	25 1.70 23.81 9.06	26 1.77 24.76 8.58	33 2.24 31.43 7.19	21 1.43 20.00 4.86	105 7.14
3	14 0.95 25.93 5.07	8 0.54 14.81 2.64	17 1.16 31.48 3.70	15 1.02 27.78 3.47	54 3.67
4	10 0.68 25.64 3.62	5 0.34 12.82 1.65	11 0.75 28.21 2.40	13 0.88 33.33 3.01	39 2.65
Total	276 18.78	303 20.61	459 31.22	432 29.39	1470 100.00

The people who do not leave the company have a higher satisfaction for colleague relationship. So relationship may be an important factor to affect turnover rate.

3. Data Examination and Management

By looking into the data set, can find that there is a problem of data conflict (show as below). When the turnover value is Yes, there are 14 records that have the Type value of 0, they are abnormal since when Turnover value is Yes that means the specific employee is turnover so the correspond value of type should be 1, 2 , 3 or 4. So when the type value is 0, there must a problem and should modify the data set.

	X	Age	Turnove	Type	Business	DailyRat	Departn	Distance	Educatic	Educatic
22	22	36	Yes	0	Travel_Rar	1218	Sales	9	4	Life Science
125	125	31	Yes	0	Travel_Rar	249	Sales	6	4	Life Science
237	237	33	Yes	0	Travel_Rar	465	Research &	2	2	Life Science
289	289	26	Yes	0	Travel_Rar	1449	Research &	16	4	Medical
324	324	28	Yes	0	Travel_Rar	1157	Research &	2	4	Medical
369	369	40	Yes	0	Travel_Rar	575	Sales	22	2	Marketing
592	592	33	Yes	0	Travel_Rar	118	Sales	16	3	Marketing
668	668	41	Yes	0	Travel_Rar	1085	Research &	2	4	Life Science
832	832	31	Yes	0	Travel_Fre	874	Research &	15	3	Medical
839	839	42	Yes	0	Travel_Fre	481	Sales	12	3	Life Science
843	843	28	Yes	0	Travel_Rar	1485	Research &	12	1	Life Science
916	916	21	Yes	0	Travel_Fre	251	Research &	10	2	Life Science
941	941	39	Yes	0	Travel_Rar	360	Research &	23	3	Medical
1034	1034	31	Yes	0	Travel_Fre	1445	Research &	1	5	Life Science

So based on the result of previous data exploration, those error can be solved.

The people who is older than 40 or their Years in Current Role is greater than 4 years or their Monthly Income is greater than 5000 are classified as Type 0 and the value of Turnover should be No, means those people are not turn over; For the employees who have the Performance Rating equal to 4 or their Relationship Satisfaction is 4 are more likely be treated as type 3, which is Involuntary Registration, and also the Turnover should keep value of Yes; then the rest record can be treated as type 2, which is Voluntary Registration, and the Turnover value is Yes.

Analysis

1. Event Type Analysis

a. Does the different event type perform significantly differently?

There are 4 event types, when analyzing the survival and hazard rate of the observations, whether these types have different coefficient should be a very important thing to consider.

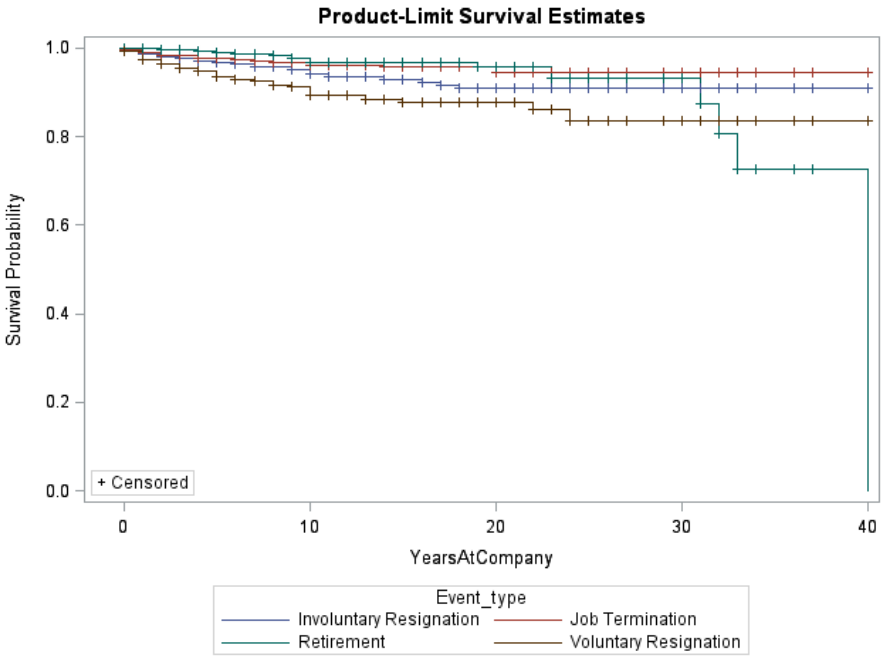
So the first thing should do here is to find out if these event types are significantly different and should be analyzed separately or not.

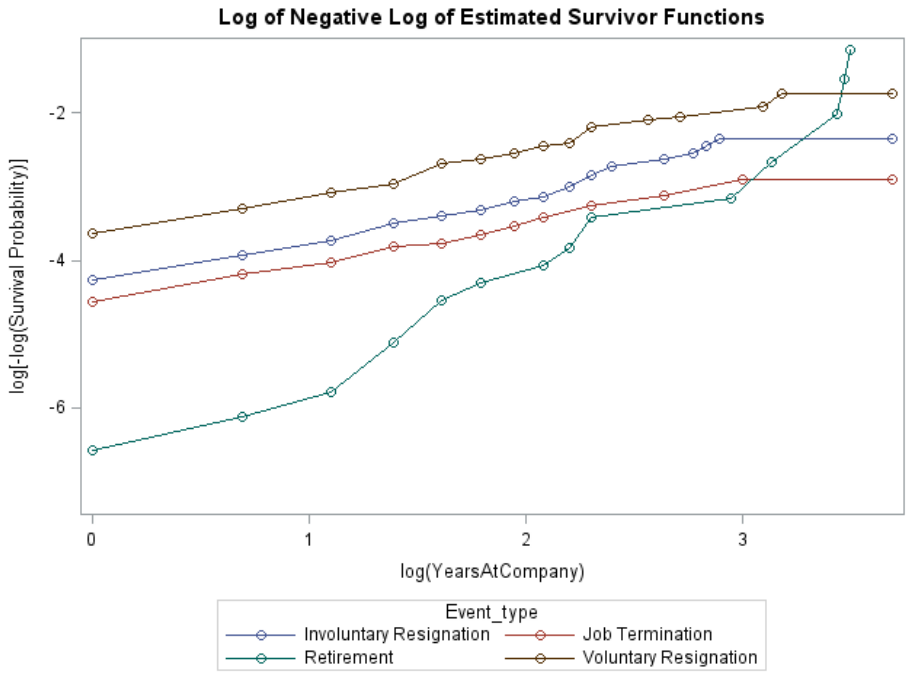
Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	64.5905	3	<.0001
Wilcoxon	71.7759	3	<.0001
-2Log(LR)	60.8222	3	<.0001

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
Event_type	Event_type	Chi-Square	Raw	Tukey-Kramer
Involuntary Resignation	Job Termination	3.4689	0.0625	0.2444
Involuntary Resignation	Retirement	8.8803	0.0029	0.0153
Involuntary Resignation	Voluntary Resignation	19.9807	<.0001	<.0001
Job Termination	Retirement	1.2488	0.2638	0.6787
Job Termination	Voluntary Resignation	40.1002	<.0001	<.0001
Retirement	Voluntary Resignation	55.5020	<.0001	<.0001

Adjustment for Multiple Comparisons for the Wilcoxon Test				
Strata Comparison			p-Values	
Event_type	Event_type	Chi-Square	Raw	Tukey-Kramer
Involuntary Resignation	Job Termination	2.1289	0.1445	0.4625
Involuntary Resignation	Retirement	13.1478	0.0003	0.0016
Involuntary Resignation	Voluntary Resignation	20.6833	<.0001	<.0001
Job Termination	Retirement	4.6955	0.0302	0.1325
Job Termination	Voluntary Resignation	36.0838	<.0001	<.0001
Retirement	Voluntary Resignation	66.8124	<.0001	<.0001

Based on p-value, voluntary resignation is significantly different from all other 3 types. Job termination is close to involuntary resignation and retirement. And all others are different.





From the survival probability estimation and log chart, it indicates that retirement has a different performance among each other. So can say these event types have different results. Therefore it is necessary to compare the different combined models with unseparated model to see whether it should be considered separately for each event type, also say whether they can use the same coefficients for building models.

Obs	Nested	Retirement	Vol_Resignation	Invol_Resignation	Job_Termination	Total	Diff	P_value
1	2300.33	78.079	918.759	544.889	320.797	1862.52	437.807	0.000

Firstly, separate each event type and use all variables to do comparison, and can find that p-value is almost zero (p-value<0.05), means the type separated models has the significant difference with the unseparated model. So it is sure that the type should be separated.

But it is not sure whether they should be all separated or not, in the next step the project is going to test and make the final decision.

b. Does it need to separate all of them?

From previous analysis, it shows voluntary resignation is significantly different from all others, yes, it is obvious that should separate it from other event type. But how to deal with job termination, involuntary resignation and retirement? Firstly, do the test to see whether Retirement and Job Termination should be separated.

Obs	Nested	Retirement	Invol_Resignation	Total	Diff	P_value
1	597.318	78.079	320.797	398.876	198.442	0.000

From p-value, can say retirement and termination are different, so should separate them as 2 different event type, they cannot use the same coefficients for the model.

Then, test the Involuntary Resignation and Retirement.

Obs	Nested	Retirement	Job_Termination	Total	Diff	P_value
1	937.37	544.889	320.797	865.686	71.684	0.107

Based on the analysis finds that the models of Involuntary Resignation and Job Termination are not significantly different, means they can use the same coefficients for modeling. So type 3 and type 4 can be combined together as type 3 in further analysis.

In conclusion, in this case it originally has 4 different event type and they have differences. Based on further analysis found that Involuntary Resignation and Job Termination could be combined together as one type, and also separate the turnover type of Retirement and Voluntary Resignation, which finally has 3 types.

2. Attributes Analysis

a. Type1 – Retirement

This project contains so many categorical variables so use one-way test first to check if these categories have very different effect on event.

In PHREG model, make the Education Field, Job Role, Job Level, Environment Satisfaction and Education test for Retirement and find those categories do not have significant difference on this event.

Then use stepwise – backward selection here to choose significant variables and 23 variables are chosen to be removed.

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Effect Label
1	PerformanceRating	1	28	0.0003	0.9865	PerformanceRating
2	MonthlyIncome	1	27	0.7432	0.3886	MonthlyIncome
3	Education	4	26	5.3338	0.2547	Education
4	JobSatisfaction	1	25	0.9432	0.3315	JobSatisfaction
5	DistanceFromHome	1	24	2.5952	0.1072	DistanceFromHome
6	YearsSinceLastPromot	1	23	2.3268	0.1272	YearsSinceLastPromotion
7	MonthlyRate	1	22	3.0621	0.0801	MonthlyRate
8	HourlyRate	1	21	2.5261	0.1120	HourlyRate
9	Department	2	20	3.8514	0.1458	Department
10	TrainingTimesLastYea	1	19	1.6933	0.1932	TrainingTimesLastYear
11	PercentSalaryHike	1	18	2.3245	0.1273	PercentSalaryHike
12	RelationshipSatisfac	3	17	5.1766	0.1593	RelationshipSatisfaction
13	EducationField	5	16	8.2037	0.1454	EducationField
14	MaritalStatus	2	15	1.9285	0.3813	MaritalStatus
15	StockOptionLevel	3	14	3.0800	0.3794	StockOptionLevel
16	WorkLifeBalance	3	13	4.5644	0.2066	WorkLifeBalance
17	EnvironmentSatisfact	3	12	4.7132	0.1940	EnvironmentSatisfaction
18	DailyRate	1	11	2.9250	0.0872	DailyRate
19	JobRole	2	10	5.1314	0.0769	
20	JobLevel	4	9	7.1013	0.1306	JobLevel
21	JobInvolvement	3	8	5.9790	0.1126	JobInvolvement
22	OverTime	1	7	2.2845	0.1307	
23	YearsWithCurrManager	1	6	3.3969	0.0653	YearsWithCurrManager

The left variables are used to do the PHREG test.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	0.29263	0.04067	51.7802	<.0001	1.340	Age
BusinessTravel	Non-Travel	1	0.18212	0.79354	0.0527	0.8185	1.200	BusinessTravel Non-Travel
BusinessTravel	Travel_Frequently	1	1.38243	0.51853	7.1080	0.0077	3.985	BusinessTravel Travel_Frequently
Gender		1	1.00950	0.50277	4.0316	0.0447	2.744	
NumCompaniesWorked		1	0.25601	0.07177	12.7259	0.0004	1.292	NumCompaniesWorked
TotalWorkingYears		1	-0.17241	0.04182	16.9948	<.0001	0.842	TotalWorkingYears
YearsInCurrentRole		1	-0.24664	0.07807	9.9809	0.0016	0.781	YearsInCurrentRole

Conclusion:

Age (34%), Travel frequently of Business Travel (298.5%), Gender (174.4%) and Number of Companies Worked (29.2%) can increase the hazard rate on retirement according to the result of hazard ratio. Also, Total Working Years (15.8%) and Years at Current Role (21.9%) can decrease the hazard rate on retirement.

b. Type 2 – Voluntary Resignation

Similar as previous analysis for type 1, analyzing the variables which have more than 2 categories first to see if each category performs differently.

For Education Field, they do not perform very differently. But for Job Role, each category is very different from each other.

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
JobRole	JobRole	Chi-Square	Raw	Tukey-Kramer
1	2	71.9251	<.0001	<.0001
1	3	16.1087	<.0001	0.0002
2	3	10.6373	0.0011	0.0032

For Job Level, level 1 is quite different from level 2-5; the second level is also different from the highest level – 5.

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
JobLevel	JobLevel	Chi-Square	Raw	Tukey-Kramer
1	2	68.0089	<.0001	<.0001
1	3	81.3351	<.0001	<.0001
1	4	98.6162	<.0001	<.0001
1	5	97.1241	<.0001	<.0001
2	3	0.7305	0.3927	0.9132
2	4	1.8480	0.1740	0.6538
2	5	5.1157	0.0237	0.1574
3	4	0.2037	0.6517	0.9914
3	5	2.0640	0.1508	0.6039
4	5	1.2215	0.2691	0.8039

For Job Satisfaction, level 1 is also different from others.

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
EnvironmentSatisfaction	EnvironmentSatisfaction	Chi-Square	Raw	Tukey-Kramer
1	2	4.0394	0.0445	0.1843
1	3	10.6043	0.0011	0.0062
1	4	6.9131	0.0086	0.0425
2	3	2.1144	0.1459	0.4655
2	4	0.6550	0.4183	0.8501
3	4	0.3492	0.5546	0.9349

For Education, level 1 is different from 3 and 4; 2 is different from 3; 3 is different from 4 and 5; 4 is different from 5. So when education level is higher, they have very different effect on turnover rate.

Adjustment for Multiple Comparisons for the Wilcoxon Test				
Strata Comparison			p-Values	
Education	Education	Chi-Square	Raw	Tukey-Kramer
1	2	0.0853	0.7702	0.9984
1	3	4.0633	0.0438	0.2583
1	4	5.6834	0.0171	0.1196
1	5	0.00985	0.9210	1.0000
2	3	4.0043	0.0454	0.2654
2	4	3.5221	0.0606	0.3299
2	5	0.0758	0.7830	0.9987
3	4	11.5659	0.0007	0.0061
3	5	6.2159	0.0127	0.0921
4	5	7.8503	0.0051	0.0407

Then use stepwise – backward method to select proper variables in this model.

Summary of Backward Elimination						
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Effect Label
1	PerformanceRating	1	27	0.0003	0.9862	PerformanceRating
2	MonthlyRate	1	26	0.0833	0.7728	MonthlyRate
3	YearsSinceLastPromot	1	25	0.1215	0.7274	YearsSinceLastPromotion
4	DailyRate	1	24	0.4251	0.5144	DailyRate
5	MaritalStatus	2	23	2.0108	0.3659	MaritalStatus
6	MonthlyIncome	1	22	0.9082	0.3406	MonthlyIncome
7	HourlyRate	1	21	1.0924	0.2959	HourlyRate
8	TotalWorkingYears	1	20	1.3466	0.2459	TotalWorkingYears
9	JobInvolvement	3	19	4.1796	0.2427	JobInvolvement
10	Department	2	18	2.9533	0.2284	Department
11	TrainingTimesLastYea	1	17	2.4110	0.1205	TrainingTimesLastYear
12	WorkLifeBalance	3	16	6.6943	0.0823	WorkLifeBalance
13	BusinessTravel	2	15	5.8637	0.0533	BusinessTravel
14	Gender	1	14	3.2195	0.0728	

From SAS analysis, 14 variables are not that significant and hence removed from model, the table below shows the variables that finally included in the model.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	-0.09801	0.01829	28.7175	<.0001	0.907	Age
DistanceFromHome		1	0.05103	0.01313	15.0986	0.0001	1.052	DistanceFromHome
Education	1	1	-1.38283	0.74072	3.4852	0.0619	0.251	Education 1
Education	2	1	-0.16918	0.69220	0.0597	0.8069	0.844	Education 2
Education	3	1	-0.01048	0.66698	0.0002	0.9875	0.990	Education 3
Education	4	1	-0.40333	0.68587	0.3456	0.5565	0.668	Education 4
EducationField	Human Resources	1	0.28673	0.61059	0.2205	0.6386	1.332	EducationField Human Resources
EducationField	Life Sciences	1	-0.54300	0.32832	2.7353	0.0982	0.581	EducationField Life Sciences
EducationField	Marketing	1	0.28069	0.42021	0.4462	0.5042	1.324	EducationField Marketing
EducationField	Medical	1	-0.73699	0.34346	4.6044	0.0319	0.479	EducationField Medical
EducationField	Other	1	-1.73466	0.78339	4.9031	0.0268	0.176	EducationField Other
EnvironmentSatisfact	1	1	0.68652	0.27631	6.1731	0.0130	1.987	EnvironmentSatisfaction 1
EnvironmentSatisfact	2	1	0.68396	0.32231	4.5032	0.0338	1.982	EnvironmentSatisfaction 2
EnvironmentSatisfact	3	1	-0.11852	0.30664	0.1494	0.6991	0.888	EnvironmentSatisfaction 3
JobLevel	1	1	0.42013	1.31122	0.1027	0.7487	1.522	JobLevel 1
JobLevel	2	1	-0.45853	1.25700	0.1331	0.7153	0.632	JobLevel 2
JobLevel	3	1	-0.88329	1.23244	0.5137	0.4736	0.413	JobLevel 3
JobLevel	4	1	-1.49176	1.33707	1.2448	0.2646	0.225	JobLevel 4
JobRole	1	1	-0.48696	0.52648	0.8555	0.3550	0.614	JobRole 1
JobRole	2	1	1.00167	0.25148	15.8657	<.0001	2.723	JobRole 2
JobSatisfaction		1	-1.03296	0.10960	88.8218	<.0001	0.356	JobSatisfaction
NumCompaniesWorked		1	0.14593	0.04877	8.9529	0.0028	1.157	NumCompaniesWorked
OverTime		1	1.80692	0.22977	61.8447	<.0001	6.092	
RelationshipSatisfac	1	1	0.86355	0.32095	7.2395	0.0071	2.372	RelationshipSatisfaction 1
RelationshipSatisfac	2	1	0.29637	0.32703	0.8213	0.3648	1.345	RelationshipSatisfaction 2
RelationshipSatisfac	3	1	0.67735	0.29941	5.1180	0.0237	1.969	RelationshipSatisfaction 3
StockOptionLevel	0	1	1.49161	0.49540	9.0656	0.0026	4.444	StockOptionLevel 0
StockOptionLevel	1	1	0.08291	0.55096	0.0226	0.8804	1.086	StockOptionLevel 1
StockOptionLevel	2	1	0.38926	0.73136	0.2833	0.5946	1.476	StockOptionLevel 2
YearsInCurrentRole		1	-0.41211	0.05690	52.4635	<.0001	0.662	YearsInCurrentRole
YearsWithCurrManager		1	-0.46327	0.05877	62.1299	<.0001	0.629	YearsWithCurrManager

Conclusion:

Age (9.3%), Education Field – Life Sciences (74.9%), Medical (15.6%) & Other (82.4%), Job Satisfaction (64.4%), Year at Current Role (33.8%) and Year with Current Manager (37.1%) have negative effect on the event – Voluntary Resignation.

Distance from Home (5.2%), Environment Satisfaction 1 (98.7%) & 2 (98.2%), Job Role – 2 (172.3%), Number of Companies Worked (15.7%), Over Time (509.2%), Relationship Satisfaction -1 (137.2%) & 2 (34.5%) & 3 (96.9%), Stock Option Level – 0 (344.4%) & 1 (8.6%) & 2 (47.6%) have positive effect on the type of Voluntary Resignation.

c. Type 3 - Involuntary & Termination

For type 3 also doing the same process of modeling and based on backward analysis, 14 variables are removed.

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Effect Label
1	MaritalStatus	2	27	0.3320	0.8470	MaritalStatus
2	Education	4	26	1.5028	0.8261	Education
3	PercentSalaryHike	1	25	0.1681	0.6818	PercentSalaryHike
4	PerformanceRating	1	24	0.2089	0.6476	PerformanceRating
5	DistanceFromHome	1	23	0.4131	0.5204	DistanceFromHome
6	HourlyRate	1	22	0.4560	0.4995	HourlyRate
7	Department	2	21	1.4024	0.4960	Department
8	DailyRate	1	20	0.6565	0.4178	DailyRate
9	Gender	1	19	1.4719	0.2250	
10	JobInvolvement	3	18	5.2956	0.1514	JobInvolvement
11	MonthlyRate	1	17	3.3927	0.0655	MonthlyRate
12	MonthlyIncome	1	16	3.7613	0.0525	MonthlyIncome
13	EnvironmentSatisfact	3	15	7.1286	0.0679	EnvironmentSatisfaction
14	EducationField	5	14	10.4410	0.0637	EducationField

The table below shows the variables that finally included in the model.

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	-0.06606	0.01730	14.5823	0.0001	0.936	Age
BusinessTravel	Non-Travel	1	-1.19460	0.60177	3.9408	0.0471	0.303	BusinessTravel Non-Travel
BusinessTravel	Travel_Frequently	1	0.34704	0.24186	2.0588	0.1513	1.415	BusinessTravel Travel_Frequently
JobLevel	1	1	12.17252	623.57347	0.0004	0.9844	193401.1	JobLevel 1
JobLevel	2	1	12.30539	623.57337	0.0004	0.9843	220882.5	JobLevel 2
JobLevel	3	1	13.66902	623.57325	0.0005	0.9825	863733.3	JobLevel 3
JobLevel	4	1	11.67258	623.57377	0.0004	0.9851	117310.9	JobLevel 4
JobRole	1	1	-0.63740	0.41173	2.3966	0.1216	0.529	JobRole 1
JobRole	2	1	0.94787	0.25587	13.7233	0.0002	2.580	JobRole 2
JobSatisfaction		1	0.39164	0.11044	12.5755	0.0004	1.479	JobSatisfaction
NumCompaniesWorked		1	0.28575	0.04281	44.5567	<.0001	1.331	NumCompaniesWorked
OverTime		1	1.02085	0.21817	21.8944	<.0001	2.776	
RelationshipSatisfac	1	1	0.17984	0.28422	0.4004	0.5269	1.197	RelationshipSatisfaction 1
RelationshipSatisfac	2	1	-0.68253	0.33180	4.2315	0.0397	0.505	RelationshipSatisfaction 2
RelationshipSatisfac	3	1	-0.56463	0.27326	4.2695	0.0388	0.569	RelationshipSatisfaction 3
StockOptionLevel	0	1	0.26361	0.44800	0.3462	0.5563	1.302	StockOptionLevel 0
StockOptionLevel	1	1	-0.45963	0.47143	0.9506	0.3296	0.632	StockOptionLevel 1
StockOptionLevel	2	1	-0.95362	0.65465	2.1219	0.1452	0.385	StockOptionLevel 2
TotalWorkingYears		1	-0.23706	0.05523	18.4213	<.0001	0.789	TotalWorkingYears
TrainingTimesLastYea		1	-0.22150	0.09099	5.9256	0.0149	0.801	TrainingTimesLastYear
WorkLifeBalance	1	1	0.29163	0.45975	0.4023	0.5259	1.339	WorkLifeBalance 1
WorkLifeBalance	2	1	0.17521	0.36965	0.2247	0.6355	1.191	WorkLifeBalance 2
WorkLifeBalance	3	1	-0.51410	0.34127	2.2694	0.1320	0.598	WorkLifeBalance 3
YearsInCurrentRole		1	-0.34480	0.05716	36.3834	<.0001	0.708	YearsInCurrentRole
YearsWithCurrManager		1	-0.28391	0.05650	25.2512	<.0001	0.753	YearsWithCurrManager

Conclusion:

Age (64%), Business Travel – Non travel (69.7%), Relationship Satisfaction – 2 (49.5%) & 3 (43.1%), Total Working Years (23.1%), Training Times Last Year (19.9%),Years in Current Role (29.2%) and Years with Current Manager (24.7%) have significantly negative effect on the turnover type of Involuntary Resignation & Termination.

Job Role -2 (158%), Job Satisfaction (47.9%), Number of Companies Worked (33.1%), Over Time (177.6%) have significantly positive effect on the turnover type of Involuntary Resignation & Termination.

3. Bonus Analysis

Bonus is a time-dependent variable in this project and whether or not an employee can gain bonus depends on whether he/she stays in the company last year. In this part, the project aims to make Bonus as a time-covariate and analyze if it has effect on different event and how it affects different event type.

In this case, only consider bonus 1 to 39, not include bonus 40 because the longest year an employee stay in the company is 40 in this dataset, and bonus depends on his last year working, so bonus 40 is relative to the 41st year which is not included here.

a. Type 1 – Retirement

At first, do the test to check if bonus in the last year and the year before last year have significant effect on retirement.

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	0.30726	0.04659	43.4920	<.0001	1.360	Age
BusinessTravel	Non-Travel	1	0.24164	0.81416	0.0881	0.7666	1.273	BusinessTravel Non-Travel
BusinessTravel	Travel_Frequently	1	1.35009	0.59107	5.2174	0.0224	3.858	BusinessTravel Travel_Frequently
Gender		1	0.98029	0.52349	3.5066	0.0611	2.665	
NumCompaniesWorked		1	0.27772	0.07773	12.7644	0.0004	1.320	NumCompaniesWorked
TotalWorkingYears		1	-0.18524	0.04653	15.8471	<.0001	0.831	TotalWorkingYears
YearsInCurrentRole		1	-0.24377	0.08099	9.0598	0.0026	0.784	YearsInCurrentRole
bonus1		1	0.17427	0.45986	0.1436	0.7047	1.190	
bonus2		1	0.27155	0.45335	0.3588	0.5492	1.312	

In this table, bonus1 is the effect of bonus in last year, bonus2 is the effect of bonus in the year before last.

Conclusion:

According to p-value, can see both of them are insignificant. So can conclude that whether or not an employee has bonus in previous 2 years cannot significantly affect retirement.

Next, use cumulative bonus to do test and analyze if it affects employee turnover.

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	0.29985	0.04427	45.8816	<.0001	1.350	Age
BusinessTravel	Non-Travel	1	0.03670	0.79756	0.0021	0.9633	1.037	BusinessTravel Non-Travel
BusinessTravel	Travel_Frequently	1	1.21540	0.58315	4.3438	0.0371	3.372	BusinessTravel Travel_Frequently
Gender		1	1.24848	0.54997	5.1533	0.0232	3.485	
NumCompaniesWorked		1	0.29355	0.07655	14.7051	0.0001	1.341	NumCompaniesWorked
TotalWorkingYears		1	-0.17275	0.04422	15.2607	<.0001	0.841	TotalWorkingYears
YearsInCurrentRole		1	-0.26456	0.08041	10.8264	0.0010	0.768	YearsInCurrentRole
bonuss		1	1.99724	0.95620	4.3627	0.0367	7.369	

Conclusion:

Bonus in this table means cumulative bonus, from p-value, find that cumulative bonus is an effective factor to affect employee turnover and it has a positive effect on event and increase hazard ratio of event, which means the more bonus an employee has before, the higher probability he chooses to retire.

b. Type 2 – Voluntary Resignation

In this table, need to focus more on last 2 parameters.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	-0.10193	0.03119	10.6793	0.0011	0.903	Age
DistanceFromHome		1	0.06937	0.01913	13.1552	0.0003	1.072	DistanceFromHome
Education	1	1	14.53167	1636	0.0001	0.9929	2046556	Education 1
Education	2	1	16.01506	1636	0.0001	0.9922	9020959	Education 2
Education	3	1	15.95980	1636	0.0001	0.9922	8535942	Education 3
Education	4	1	15.84519	1636	0.0001	0.9923	7611680	Education 4
EducationField	Human Resources	1	2.01303	0.91001	4.8934	0.0270	7.486	EducationField Human Resources
EducationField	Life Sciences	1	-0.97412	0.45108	4.6636	0.0308	0.378	EducationField Life Sciences
EducationField	Marketing	1	-0.38893	0.59903	0.4216	0.5162	0.678	EducationField Marketing
EducationField	Medical	1	-1.10206	0.47791	5.3177	0.0211	0.332	EducationField Medical
EducationField	Other	1	-16.95536	1682	0.0001	0.9920	0.000	EducationField Other
EnvironmentSatisfact	1	1	0.39725	0.36714	1.0529	0.3048	1.488	EnvironmentSatisfaction 1
EnvironmentSatisfact	2	1	0.08085	0.50590	0.0255	0.8730	1.084	EnvironmentSatisfaction 2
EnvironmentSatisfact	3	1	-0.32334	0.44951	0.5174	0.4719	0.724	EnvironmentSatisfaction 3
JobLevel	1	1	0.67691	1.75053	0.1495	0.6990	1.968	JobLevel 1
JobLevel	2	1	0.07815	1.66833	0.0022	0.9626	1.081	JobLevel 2
JobLevel	3	1	-0.55669	1.56664	0.1263	0.7223	0.573	JobLevel 3
JobLevel	4	1	-14.93287	989.36475	0.0002	0.9880	0.000	JobLevel 4
JobRole	1	1	-1.08911	0.71863	2.2969	0.1296	0.337	JobRole 1
JobRole	2	1	0.92946	0.37994	5.9645	0.0144	2.533	JobRole 2
JobSatisfaction		1	-1.09471	0.17123	40.8755	<.0001	0.335	JobSatisfaction
NumCompaniesWorked		1	0.05733	0.07002	0.6704	0.4129	1.059	NumCompaniesWorked
OverTime		1	1.90706	0.35372	29.0684	<.0001	6.733	
RelationshipSatisfac	1	1	1.03695	0.43600	5.6564	0.0174	2.821	RelationshipSatisfaction 1
RelationshipSatisfac	2	1	0.40879	0.47669	0.7354	0.3911	1.505	RelationshipSatisfaction 2
RelationshipSatisfac	3	1	0.43307	0.44752	0.9365	0.3332	1.542	RelationshipSatisfaction 3
StockOptionLevel	0	1	2.32671	1.09796	4.4906	0.0341	10.244	StockOptionLevel 0
StockOptionLevel	1	1	0.14723	1.15406	0.0163	0.8985	1.159	StockOptionLevel 1
StockOptionLevel	2	1	0.64761	1.33879	0.2340	0.6286	1.911	StockOptionLevel 2
YearsInCurrentRole		1	-0.27929	0.06289	19.7199	<.0001	0.756	YearsInCurrentRole
YearsWithCurrManager		1	-0.36634	0.06742	29.5271	<.0001	0.693	YearsWithCurrManager
bonus1		1	-0.61704	0.35888	2.9563	0.0855	0.540	
bonus2		1	-0.24402	0.32499	0.5638	0.4527	0.783	

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
bonuss		1	-0.40675	0.47636	0.7291	0.3932	0.666	

Conclusion:

According to p-value, whether an employee has bonus in previous 2 years does not have significant effect on Voluntary Resignation. Also, even consider cumulative bonus in previous year here, this still cannot have significant effect on the employees who are volunteer to leave.

c. Type 3 – Involuntary Resignation & Termination

The first table in the effect of bonus in last year and the year before last and the second table show the effect of cumulative bonus.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
bonus1		1	0.34818	0.30239	1.3258	0.2496	1.416	
bonus2		1	0.17339	0.31782	0.2976	0.5854	1.189	

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
bonuss		1	-0.03883	0.46209	0.0071	0.9330	0.962	

Conclusion:

Both the bonus in previous 2 year and cumulative year do not have significant effect on the event – Involuntary Resignation and Termination.

According to analysis of all 3 type events, can conclude that bonus, in cumulative way, do have positive effect on Retirement. In all other situations, bonus cannot work very well to decrease employee turnover rate.

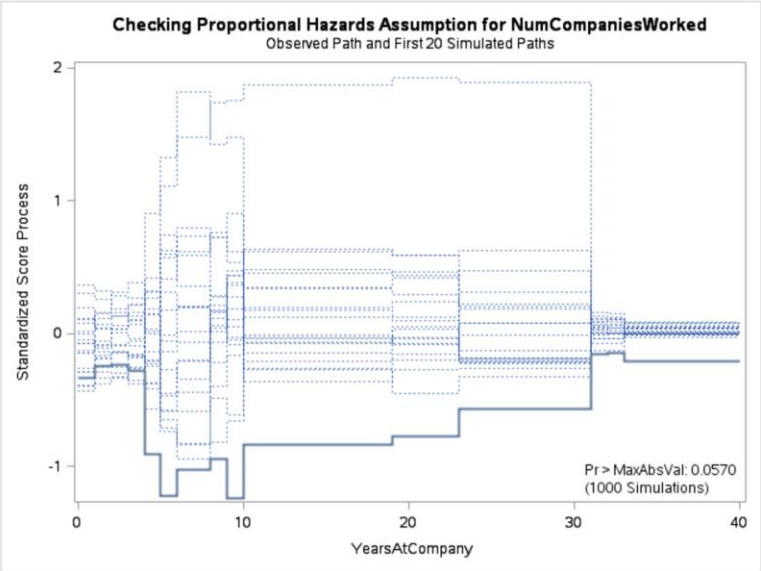
4. Non-proportionality Analysis

Some variables are time-dependent, so it may change as time change. If could define the time-covariates and deal with them properly, it will have a better understanding on employee turnover.

In this part, also do analysis based on 3 different type.

a. Type 1 – Retirement

Martingale Residual Method



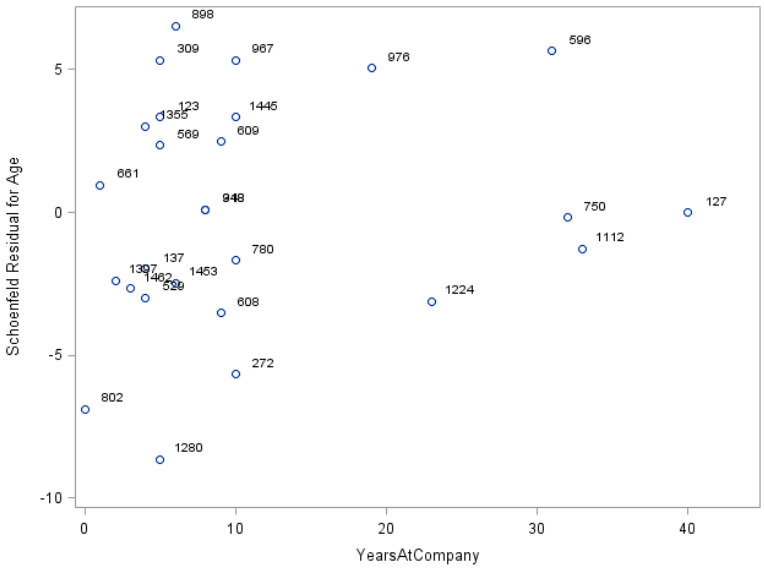
Supremum Test for Proportionals Hazards Assumption				
Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
Age	0.5325	1000	353840001	0.4380
BusinessTravelNon-Travel	0.6393	1000	353840001	0.4550
BusinessTravelTravel_Frequently	1.1953	1000	353840001	0.0560
Gender	1.3067	1000	353840001	0.1070
NumCompaniesWorked	1.2428	1000	353840001	0.0530
TotalWorkingYears	0.7171	1000	353840001	0.5050
YearsInCurrentRole	1.1557	1000	353840001	0.1750

By finding covariates deviating significantly from theoretical expectations under the proportionality assumption, Number of companies Worked may be time - dependent.

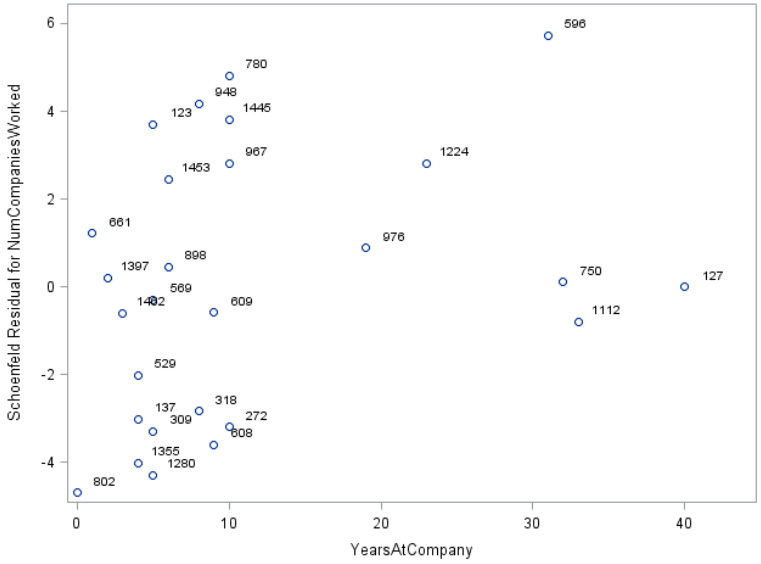
Shoenfeld Residuals Method

Firstly, try to plot residuals of some variables to see if they are independent and random distributed.

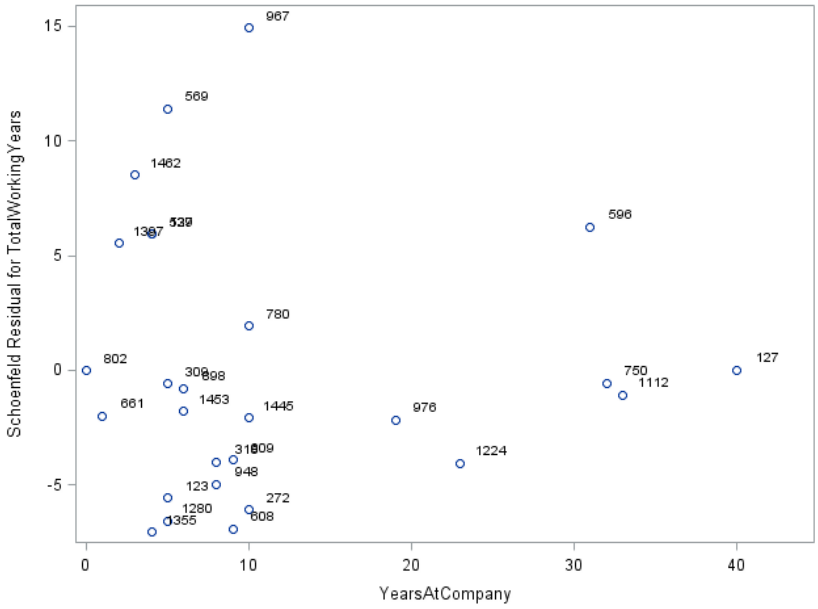
Age



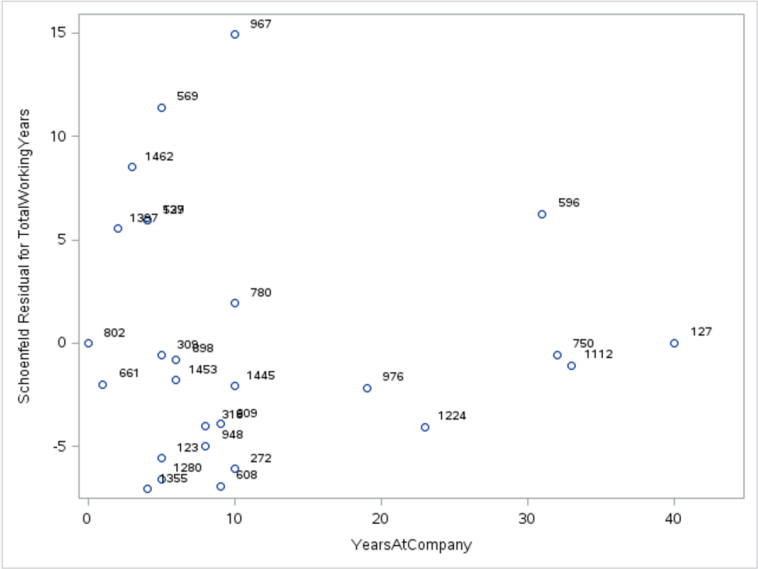
Number of Companies Worked



Total Working Years



YearsInCurrentRole



By looking at the residual above can see that the all of the residual plots are not independently distributed, so they may be time-dependent.

Then check the p-value to see which variables are non-proportional.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations			
	YearsAtCompany	lYearsAtCompany	YearsAtCompany2
schAge Schoenfeld Residual for Age	0.15216 0.4487 27	0.11544 0.5744 26	0.09631 0.6328 27
schBusinessTravel Schoenfeld Residual for BusinessTravelNon-Travel	-0.22200 0.2657 27	-0.14767 0.4716 26	-0.23483 0.2384 27
schGender Schoenfeld Residual for BusinessTravelTravel_Frequently	-0.28192 0.1543 27	-0.31536 0.1166 26	-0.21447 0.2827 27
schNumCompaniesWorked Schoenfeld Residual for Gender	-0.38569 0.0469 27	-0.38701 0.0508 26	-0.35428 0.0698 27
schTotalWorkingYears Schoenfeld Residual for NumCompaniesWorked	0.29290 0.1382 27	0.25864 0.2020 26	0.21004 0.2930 27
schYearsInCurrentRole Schoenfeld Residual for TotalWorkingYears	-0.02856 0.8875 27	-0.09938 0.6291 26	0.01192 0.9530 27

From the table can see only Number of Companies Worked interacts with Years at company is significant.

Run the model with all above decided time-dependent variables.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	0.24565	0.06516	14.2110	0.0002	1.278	Age
BusinessTravel	Non-Travel	1	0.20332	0.87276	0.0543	0.8158	1.225	BusinessTravel Non-Travel
BusinessTravel	Travel_Frequently	1	1.79147	0.55151	10.5513	0.0012	5.998	BusinessTravel Travel_Frequently
Gender		1	1.09143	0.51934	4.4166	0.0356	2.979	
NumCompaniesWorked		1	0.00416	0.13925	0.0009	0.9762	1.004	NumCompaniesWorked
TotalWorkingYears		1	-0.05143	0.06648	0.5985	0.4391	0.950	TotalWorkingYears
YearsInCurrentRole		1	-0.58169	0.13344	19.0032	<.0001	0.559	YearsInCurrentRole
yAge		1	0.01073	0.00862	1.5514	0.2129	1.011	
yNumCompaniesWorked		1	0.03678	0.01856	3.9264	0.0475	1.037	
yTotalWorkingYears		1	-0.02368	0.01117	4.4888	0.0341	0.977	
yYearsInCurrentRole		1	0.02649	0.00796	11.0841	0.0009	1.027	

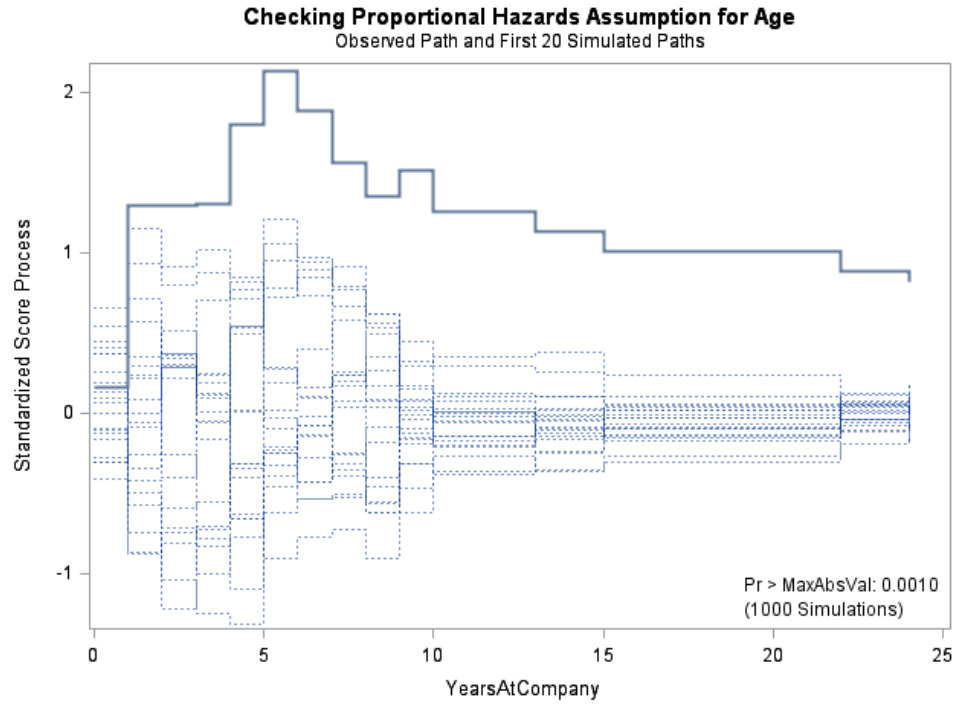
Conclusion:

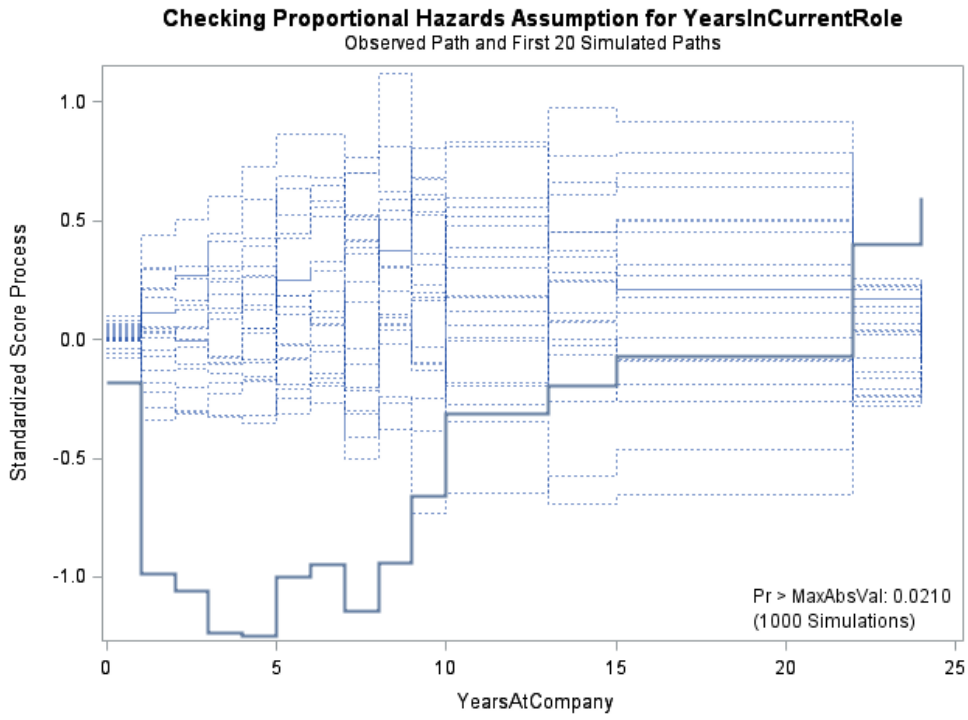
From the model, can see Number of Companies Worked (3.7%), Total Working Years (5%) and Years in Current Role (2.7%) are time-dependent covariates and only Total Working Years has a vegtive effect on hazard rate, rest are having positive relationship.

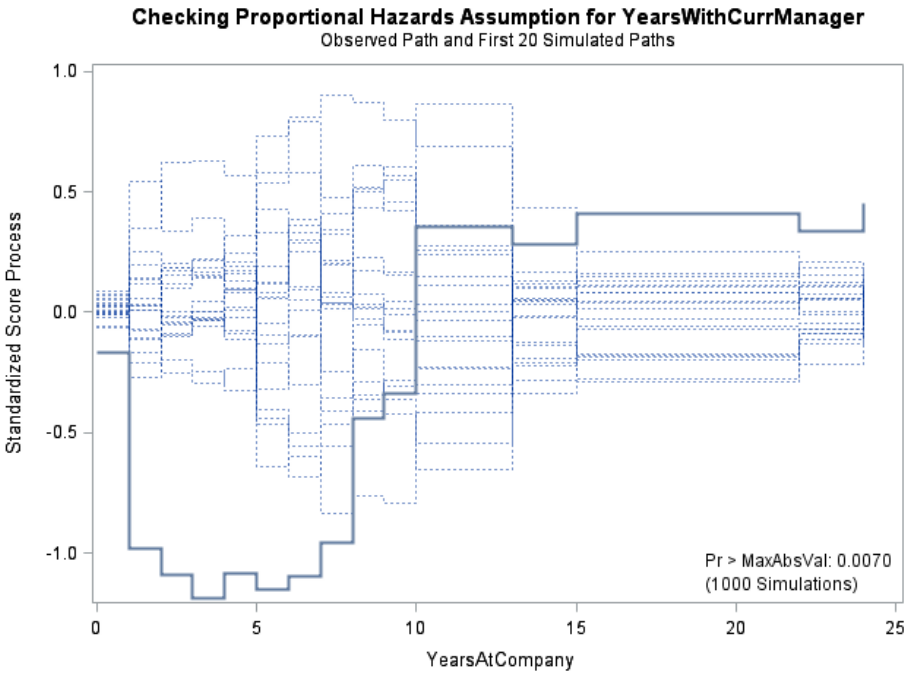
b. Type 2 – Voluntary Resignation

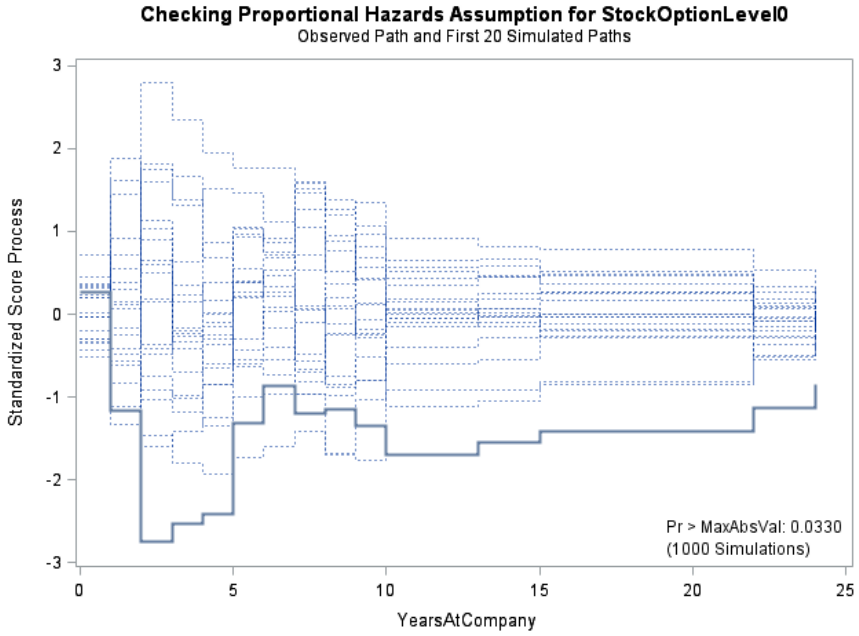
Martingale Residual Method

Based on residual deviation and p-value, find 4 time covariates.


Checking Proportional Hazards Assumption for Age
Observed Path and First 20 Simulated Paths
Standardized Score Process
2
1
0
-1
0
5
10
15
20
25
YearsAtCompany
Pr > MaxAbsVal: 0.0010
(1000 Simulations)

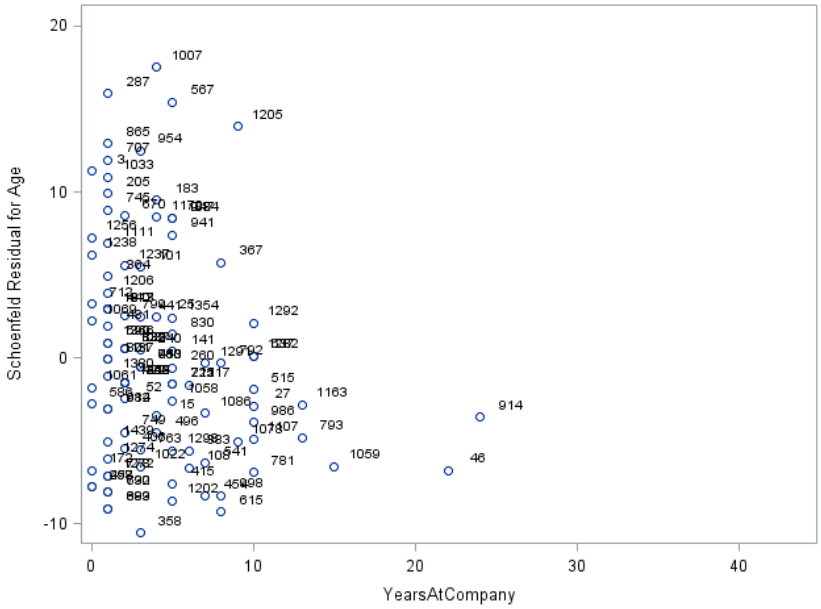

Checking Proportional Hazards Assumption for YearsInCurrentRole
Observed Path and First 20 Simulated Paths
Standardized Score Process
1.0
0.5
0.0
-0.5
-1.0
0
5
10
15
20
25
YearsAtCompany
Pr > MaxAbsVal: 0.0210
(1000 Simulations)



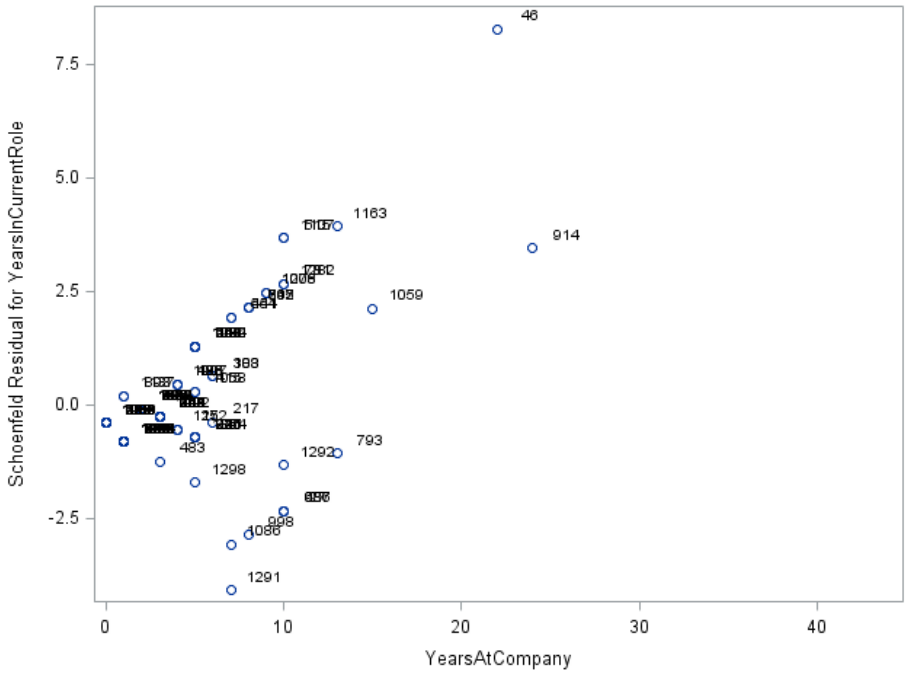


Shoenfeld Residuals Method

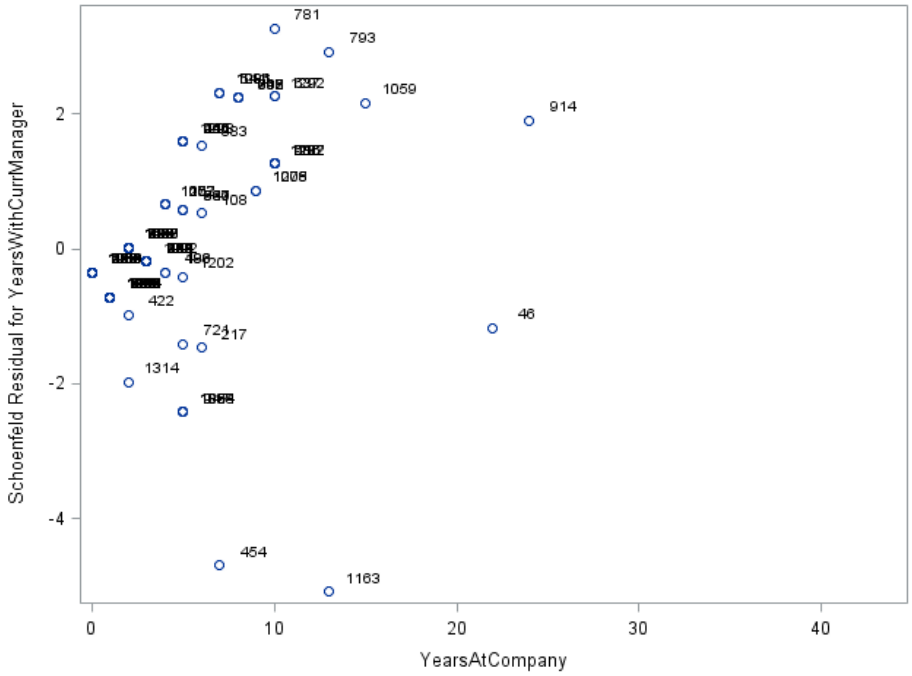
Age



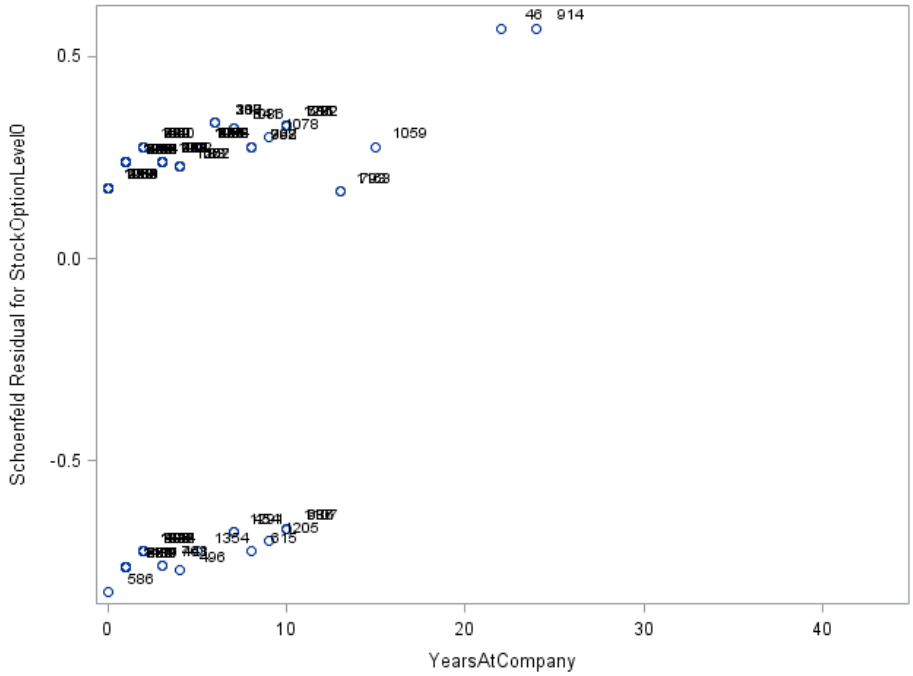
Years in Current Role



Years with Current Manager



Stock Option Level 0



By looking at the residual above can see that the residual of those four variables are not independently distributed, so they are time-dependent variables.

Then check non-proportionality based on p-value.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations			
	YearsAtCompany	lYearsAtCompany	YearsAtCompany2
schAge Schoenfeld Residual for Age	-0.21154 0.0287 107	-0.19745 0.0526 97	-0.19184 0.0478 107

From these two methods, considering Age, Years in Current Role, Years with Current Manager and Stock Option Level 0 could have non-proportionality. So put them in model with time interaction and run again to see if they are significant.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	-0.05397	0.02207	5.9614	0.0145	0.947	Age
DistanceFromHome		1	0.05003	0.01335	14.0393	0.0002	1.051	DistanceFromHome
Education	1	1	-1.37799	0.74921	3.3829	0.0659	0.252	Education 1
Education	2	1	-0.15934	0.69133	0.0531	0.8177	0.853	Education 2
Education	3	1	-0.10328	0.66013	0.0245	0.8757	0.902	Education 3
Education	4	1	-0.42702	0.68614	0.3873	0.5337	0.652	Education 4
EducationField	Human Resources	1	0.28635	0.61343	0.2179	0.6406	1.332	EducationField Human Resources
EducationField	Life Sciences	1	-0.52850	0.32925	2.5766	0.1085	0.589	EducationField Life Sciences
EducationField	Marketing	1	0.22317	0.42384	0.2773	0.5985	1.250	EducationField Marketing
EducationField	Medical	1	-0.61301	0.34661	3.1279	0.0770	0.542	EducationField Medical
EducationField	Other	1	-1.57375	0.78620	4.0068	0.0453	0.207	EducationField Other
EnvironmentSatisfact	1	1	0.66063	0.28100	5.8668	0.0154	1.975	EnvironmentSatisfaction 1
EnvironmentSatisfact	2	1	0.59989	0.32512	3.4046	0.0650	1.822	EnvironmentSatisfaction 2
EnvironmentSatisfact	3	1	-0.18260	0.31216	0.3422	0.5586	0.833	EnvironmentSatisfaction 3
JobLevel	1	1	-0.22216	1.33001	0.0279	0.8673	0.801	JobLevel 1
JobLevel	2	1	-0.96770	1.29412	0.5825	0.4453	0.372	JobLevel 2
JobLevel	3	1	-1.30272	1.31066	0.9879	0.3202	0.272	JobLevel 3
JobLevel	4	1	-1.82690	1.40423	1.6926	0.1933	0.161	JobLevel 4
JobRole	1	1	-0.75201	0.53228	1.9960	0.1577	0.471	JobRole 1
JobRole	2	1	0.97238	0.26012	13.9739	0.0002	2.644	JobRole 2
JobSatisfaction		1	-1.00711	0.10973	84.2411	<.0001	0.365	JobSatisfaction
NumCompaniesWorked		1	0.15762	0.05004	9.9224	0.0016	1.171	NumCompaniesWorked
OverTime		1	1.80561	0.23300	60.0528	<.0001	6.084	
RelationshipSatisfac	1	1	0.82328	0.32174	6.5477	0.0105	2.278	RelationshipSatisfaction 1
RelationshipSatisfac	2	1	0.25751	0.32543	0.6262	0.4288	1.294	RelationshipSatisfaction 2
RelationshipSatisfac	3	1	0.65352	0.30040	4.7329	0.0296	1.922	RelationshipSatisfaction 3
StockOptionLevel	0	1	1.11219	0.60750	3.3518	0.0671	3.041	StockOptionLevel 0
StockOptionLevel	1	1	-0.07650	0.58261	0.0172	0.8955	0.926	StockOptionLevel 1
StockOptionLevel	2	1	0.53281	0.73526	0.5251	0.4687	1.704	StockOptionLevel 2
YearsInCurrentRole		1	-0.90996	0.13730	43.9240	<.0001	0.403	YearsInCurrentRole
YearsWithCurrManager		1	-0.76151	0.12994	34.3428	<.0001	0.467	YearsWithCurrManager
yAge		1	-0.01414	0.00560	6.3811	0.0115	0.986	
yYearsInCurrentRole		1	0.06359	0.01405	20.4902	<.0001	1.066	
yYearsWithCurrManage		1	0.04433	0.01401	10.0128	0.0016	1.045	
yStockOptionLevel		1	-0.04812	0.04851	0.9842	0.3212	0.953	

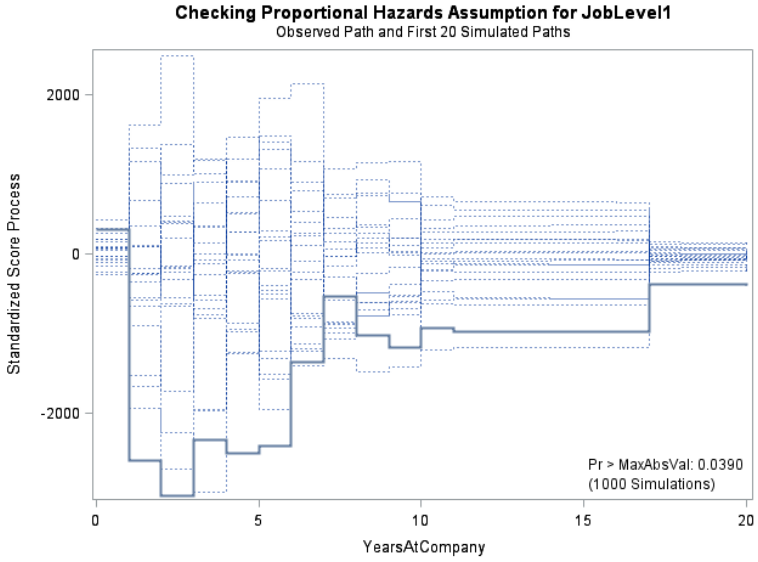
Conclusion:

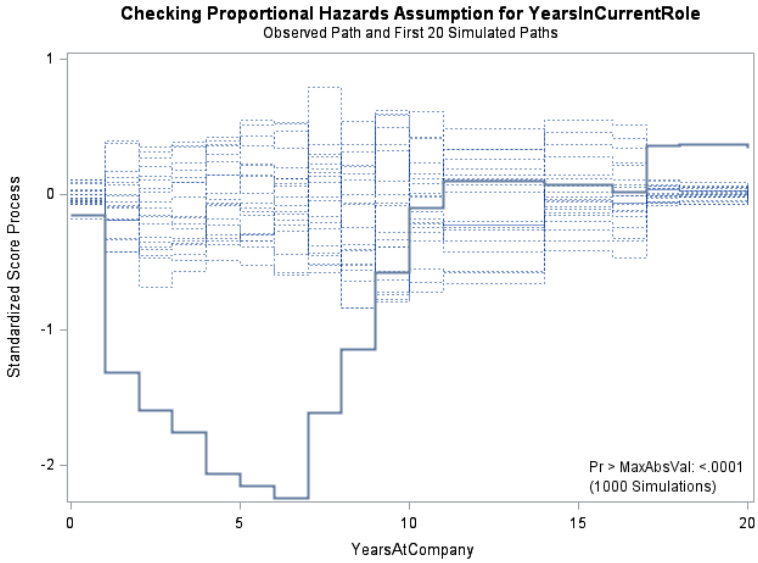
Age, Years in Current Role and Years with Current Manager are time-dependent variables (p-value <0.05). Age(1.4%) has a negative effect on the hazard rate of second type event; Years in Current Role (6.6%) and Years with Current Manager (4.5%) have positive effect.

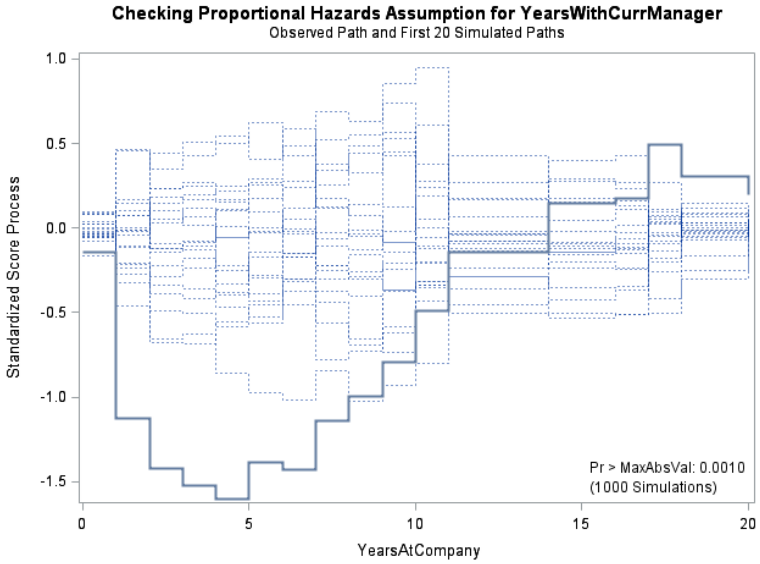
c. Type 3 – Involuntary Resignation & Termination

Martingale Residual Method

According to analysis, Job level 1, Years in Current Role, Years with Current Manager could be non-proportionality.



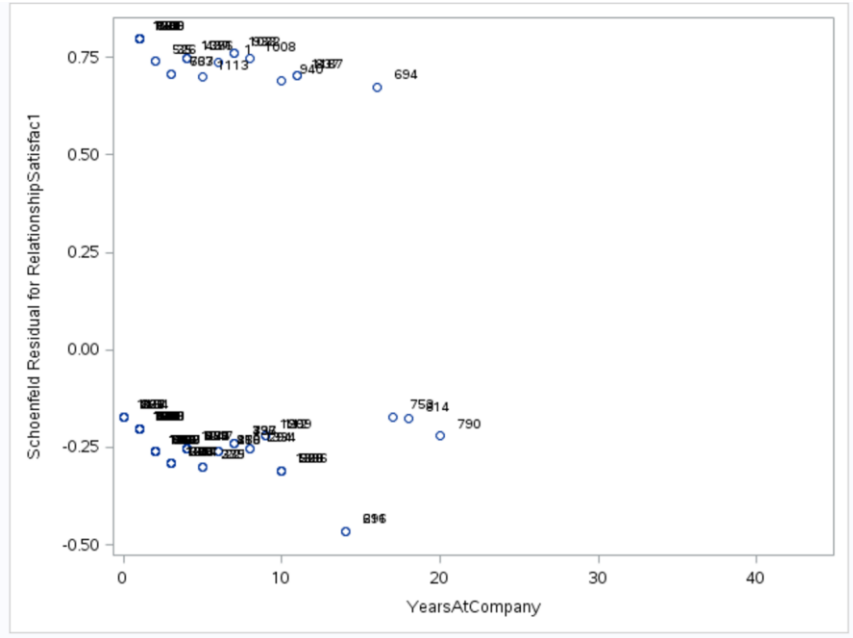




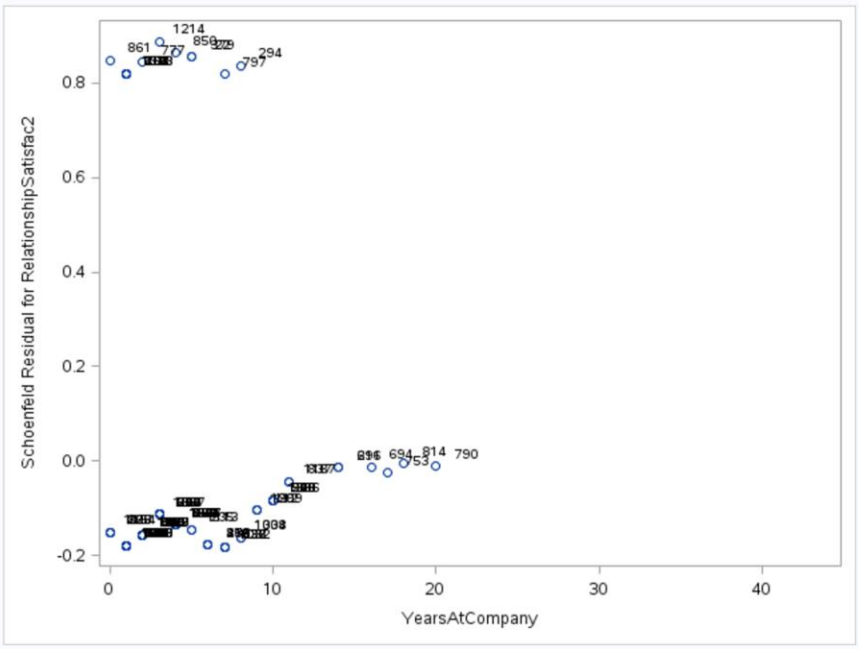
Shoenfeld Residuals Method

Residual p-value table

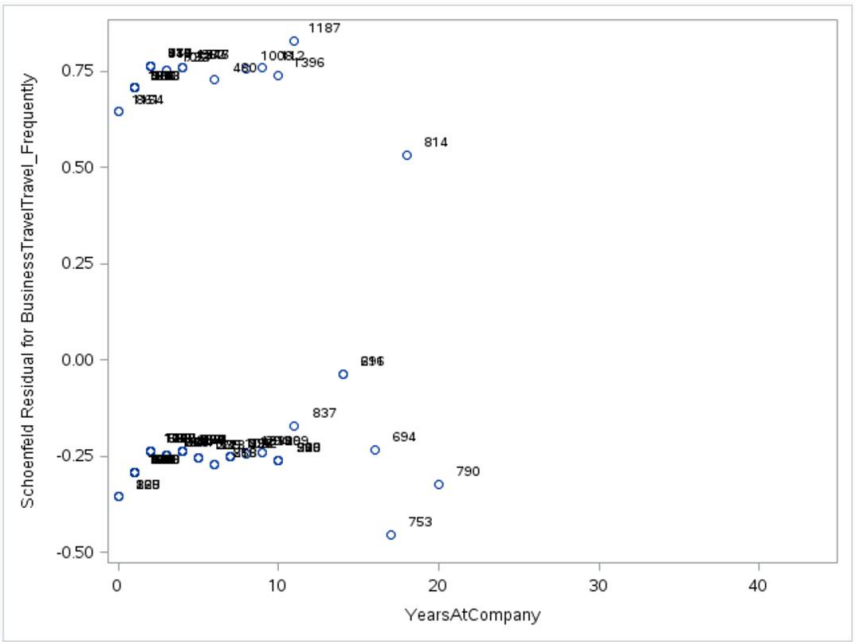
YearsInCurrentRole



YearsWithCurrManager


Schoenfeld Residual for RelationshipSatisfac2
YearsAtCompany
1214
861
294
797
694
814
790
753

JobLevel


Schoenfeld Residual for BusinessTravelTravel_Frequently
YearsAtCompany
1187
1396
814
837
694
790
753

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations			
	YearsAtCompany	lYearsAtCompany	YearsAtCompany2
schAge Schoenfeld Residual for Age	-0.02640 0.7964 98	-0.06961 0.5073 93	-0.05575 0.5856 98
schBusinessTravel Schoenfeld Residual for BusinessTravelNon-Travel	-0.09499 0.3522 98	0.01621 0.8775 93	-0.09579 0.3481 98
schJobLevel Schoenfeld Residual for BusinessTravelTravel_Frequently	-0.04999 0.6250 98	-0.03442 0.7432 93	-0.04804 0.6385 98
schJobRole Schoenfeld Residual for JobLevel1	0.14739 0.1475 98	0.20924 0.0441 93	0.11523 0.2585 98
schJobSatisfaction Schoenfeld Residual for JobLevel2	-0.12683 0.2133 98	-0.15356 0.1417 93	-0.10893 0.2856 98
schNumCompaniesWorked Schoenfeld Residual for JobLevel3	-0.04551 0.6563 98	-0.09553 0.3623 93	-0.02561 0.8024 98
schOverTime Schoenfeld Residual for JobLevel4	0.02071 0.8396 98	0.00981 0.9256 93	0.02682 0.7782 98
schRelationshipSatisfaction Schoenfeld Residual for JobRole1	-0.08039 0.4314 98	-0.06938 0.5087 93	-0.10281 0.3137 98
schStockOptionLevel Schoenfeld Residual for JobRole2	0.10922 0.2843 98	0.16374 0.1168 93	0.08715 0.3935 98
schTotalWorkingYears Schoenfeld Residual for JobSatisfaction	-0.08784 0.3897 98	-0.06832 0.3999 93	-0.07662 0.4534 98
schTrainingTimesLastYear Schoenfeld Residual for NumCompaniesWorked	0.09529 0.3506 98	0.09526 0.3637 93	0.07045 0.4906 98
schWorkLifeBalance Schoenfeld Residual for OverTime	0.02073 0.8394 98	0.10600 0.3119 93	-0.04704 0.6456 98
schYearsInCurrentRole Schoenfeld Residual for RelationshipSatisfac1	-0.01346 0.8954 98	-0.04087 0.6973 93	-0.03182 0.7557 98
schYearsWithCurrManager Schoenfeld Residual for RelationshipSatisfac2	-0.02930 0.7746 98	-0.02253 0.8303 93	-0.02552 0.8030 98

According to the scatter plot and p-value, it shows Job Role is most significant, and also Years in Current Role and Years with Current Managere are also possible to be time-dependent. So going to combine them together to consider non-proportional problem and run model again.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	-0.07213	0.01745	17.0804	<.0001	0.930	Age
BusinessTravel	Non-Travel	1	-1.14430	0.60141	3.6202	0.0571	0.318	BusinessTravel Non-Travel
BusinessTravel	Travel_Frequently	1	0.34102	0.24214	1.9835	0.1590	1.406	BusinessTravel Travel_Frequently
JobLevel	1	1	9.57560	662.51992	0.0002	0.9885	14408.83	JobLevel 1
JobLevel	2	1	9.93948	662.51935	0.0002	0.9880	20733.02	JobLevel 2
JobLevel	3	1	11.80815	662.51829	0.0003	0.9858	134343.1	JobLevel 3
JobLevel	4	1	11.53937	662.51768	0.0003	0.9861	102679.3	JobLevel 4
JobRole	1	1	-0.78337	0.41200	3.6151	0.0573	0.457	JobRole 1
JobRole	2	1	0.97245	0.25681	14.3391	0.0002	2.644	JobRole 2
JobSatisfaction		1	0.39758	0.10839	13.4552	0.0002	1.488	JobSatisfaction
NumCompaniesWorked		1	0.29439	0.04381	45.1532	<.0001	1.342	NumCompaniesWorked
OverTime		1	0.99549	0.21756	20.9372	<.0001	2.706	
RelationshipSatisfac	1	1	0.19967	0.28563	0.4886	0.4845	1.221	RelationshipSatisfaction 1
RelationshipSatisfac	2	1	-0.73278	0.33131	4.8921	0.0270	0.481	RelationshipSatisfaction 2
RelationshipSatisfac	3	1	-0.58813	0.27242	4.6607	0.0309	0.555	RelationshipSatisfaction 3
StockOptionLevel	0	1	0.29129	0.45095	0.4172	0.5183	1.338	StockOptionLevel 0
StockOptionLevel	1	1	-0.47184	0.47725	0.9774	0.3228	0.624	StockOptionLevel 1
StockOptionLevel	2	1	-0.90985	0.65208	1.9468	0.1629	0.403	StockOptionLevel 2
TotalWorkingYears		1	-0.20913	0.05218	16.0598	<.0001	0.811	TotalWorkingYears
TrainingTimesLastYea		1	-0.25077	0.09250	7.3501	0.0067	0.778	TrainingTimesLastYear
WorkLifeBalance	1	1	0.53907	0.45288	1.4168	0.2339	1.714	WorkLifeBalance 1
WorkLifeBalance	2	1	0.25556	0.36908	0.4795	0.4887	1.291	WorkLifeBalance 2
WorkLifeBalance	3	1	-0.48745	0.33945	2.0621	0.1510	0.614	WorkLifeBalance 3
YearsInCurrentRole		1	-0.74132	0.12178	37.0538	<.0001	0.476	YearsInCurrentRole
YearsWithCurrManager		1	-0.51983	0.11358	20.9476	<.0001	0.595	YearsWithCurrManager
yYearsInCurrentRole		1	0.04948	0.01174	17.7593	<.0001	1.051	
yYearsWithCurrManage		1	0.03043	0.01086	7.8595	0.0051	1.031	
yJobLevel		1	-0.05299	0.04322	1.5029	0.2202	0.948	

Conclusion:

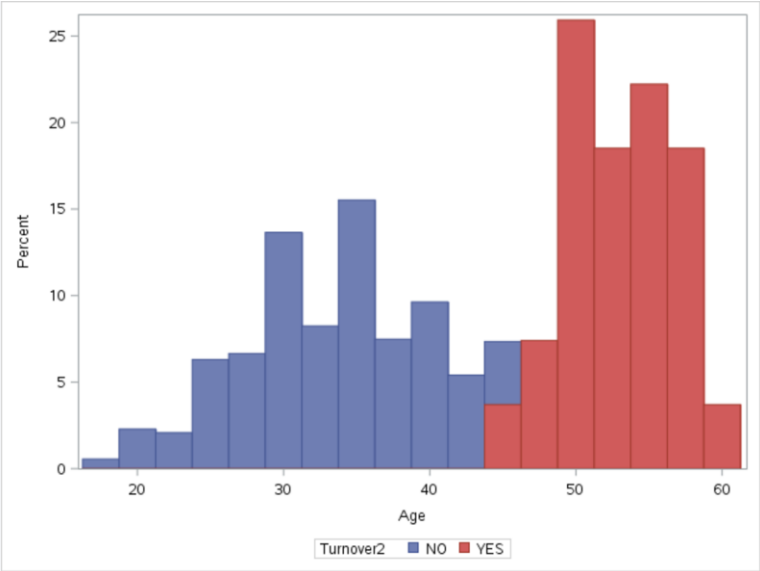
From observation of model performance, considering Years in Current Role (5.1%) and Years with Current Manager (3.1%) are 2 time-dependent variables with non-proportionality and also have positive relationship to hazard rate of type 3, whose two are significant (p-value<0.05).

5. Business Analysis

a. Type=1

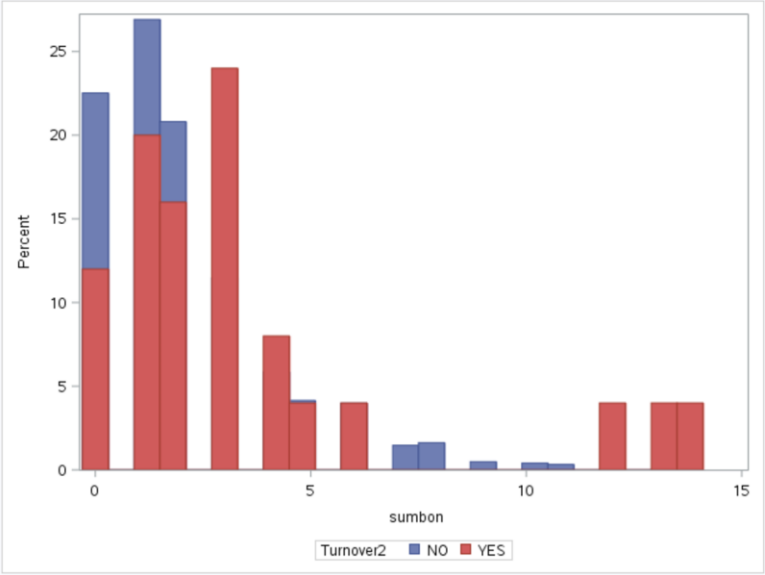
By looking at the final model, which included the cumulative bonus, it shows some business insight.

Age



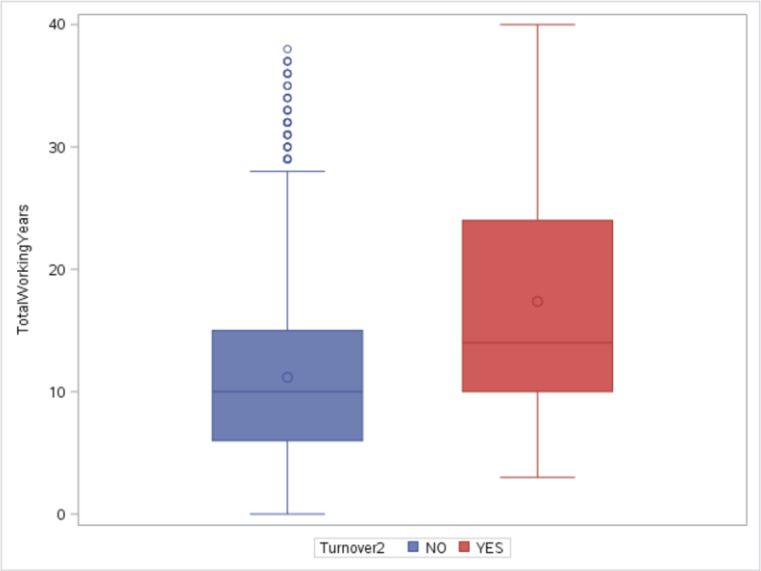
From the histogram chart, there are turnover data when age is greater than 44, it indicates that the people who are older than 44 are more likely to turnover.

Cumulative Bonus



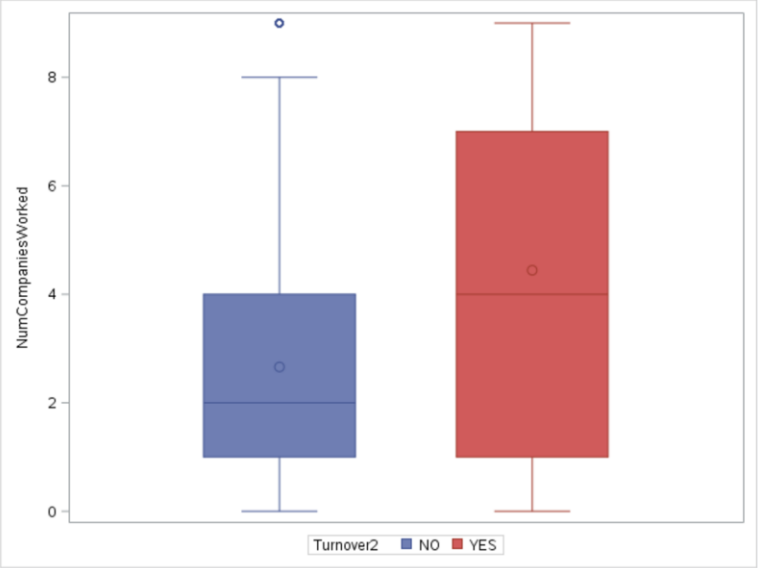
The histogram shows that employees turn over when cumulative bonus is less than 7 or is greater than 13.

TotalWorkingYears



The mean value and median value of total working years is higher in the group or turnover data, means the people with more working experience are tend to turnover.

NumCompaniesWorked



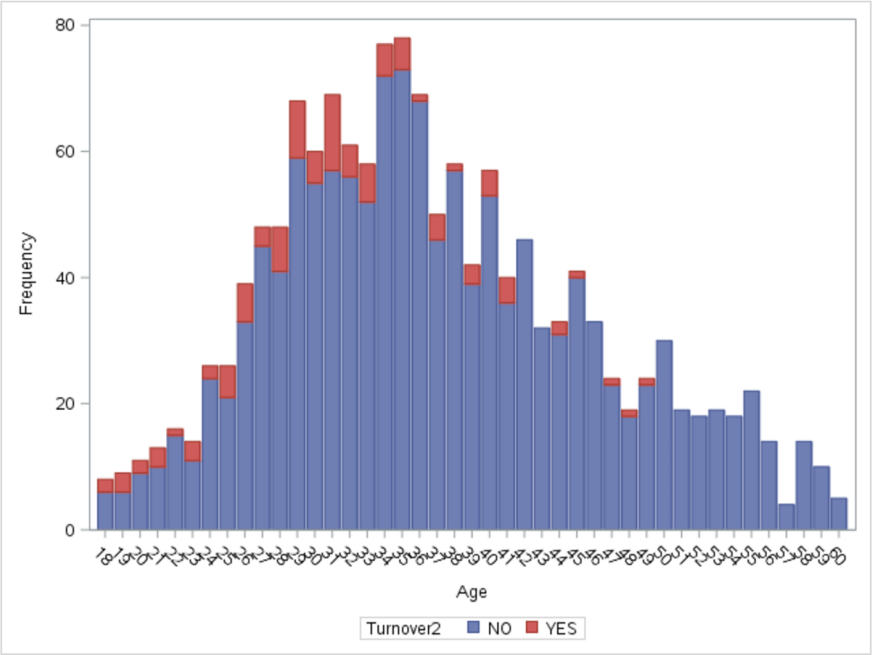
The mean value and median value of the number of companies have worked for is higher in the group or turnover data, means the people have worked on more companies are tend to turnover.

Conclusion

So conclude that the people older than 44 are more likely to turnover for type 1, which is retirement, and most of them have more working experience and have worked at more companies; also, they have higher bonus due to their long working years or have the lower bonus due to retirement, for which the observation time is started at the end of their working year.

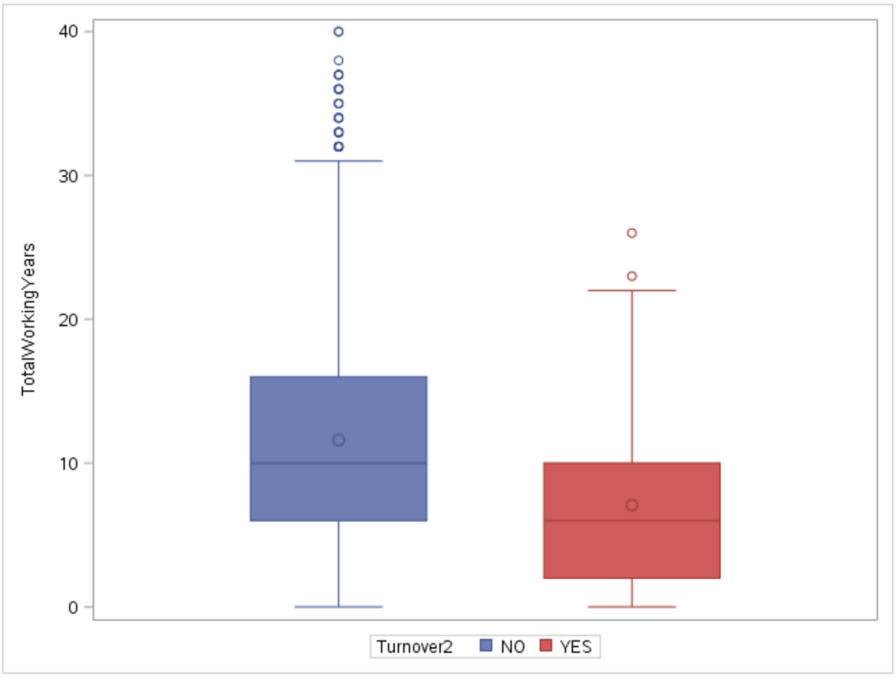
b. Type=2

By looking at the final model, which does not include any bonus variable, it shows some business insight.



From the histogram chart, there are turnover data when age is less than 49, it indicates that the people who are younger than 49 are more likely to turnover.

TotalWorkingYears



The mean value and median value of total working years is lower in the group or turnover data, means the people with less working experience are tend to turnover.

OverTime

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Turnover2 by OverTime		
	OverTime		
Turnover2	0	1	Total
NO	1008 68.57 73.95 95.64	355 24.15 26.05 85.34	1363 92.72
YES	46 3.13 42.99 4.36	61 4.15 57.01 14.66	107 7.28
Total	1054 71.70	416 28.30	1470 100.00

The frequency table shows that people who have experienced overtime work are more likely to turn over (the percentage is 14.66%, which greater than 4.36% in non-overtime group).

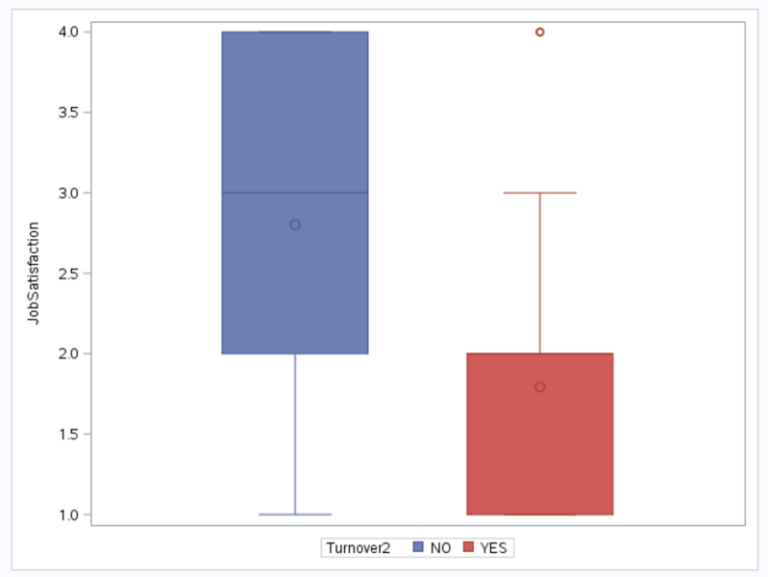
Education

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Turnover2 by Education					
	Education(Education)					
Turnover2	1	2	3	4	5	Total
NO	158 10.75 11.59 92.94	263 17.89 19.30 93.26	519 35.31 38.08 90.73	378 25.71 27.73 94.97	45 3.06 3.30 93.75	1363 92.72
YES	12 0.82 11.21 7.06	19 1.29 17.76 6.74	53 3.61 49.53 9.27	20 1.36 18.69 5.03	3 0.20 2.80 6.25	107 7.28
Total	170 11.56	282 19.18	572 38.91	398 27.07	48 3.27	1470 100.00

The frequency table shows that people who have the lower and median level of education are more likely to turn over (the percentages are greater when Education=1,2,3).

JobSatisfaction



As the box plot shows, the spread range, median value and mean value is lower in the group of turnover data that means people tend to turnover when they are not so satisfied with the company.

Conclusion

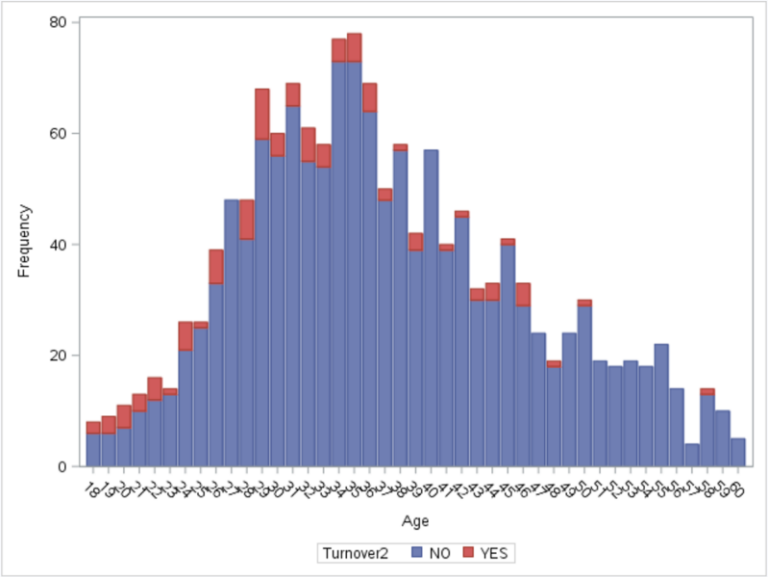
Those who are before the age of retirement and have less tolerance for the overtime work are tend to leave, they have less working experience and also their education level is median or low and not satisfied with the job.

So those young people might have thought the job is not so idea and want to find a better job and turnover voluntary, and for the older people they may want to retire early;

c. Type=3

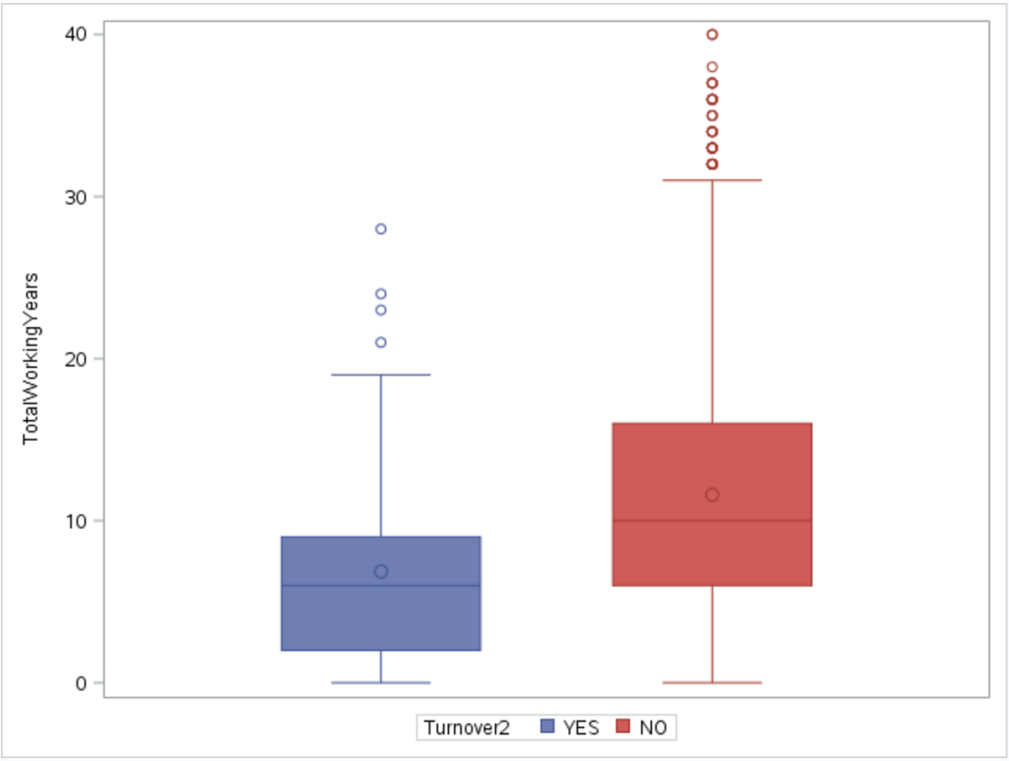
By looking at the final model, which does not include any bonus variable, it shows some business insight.

Age



From the histogram chart, there are turnover data when age is less than 50, it indicates that the people who are younger than 50 are more likely to turnover.

TotalWorkingYears



From the box plot, it shows the mean and median values of cumulative bonus are lower in the turnover group data, means people with less working experiences are tend to turnover.

OverTime

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Turnover2 by OverTime		
	OverTime		
Turnover2	0	1	Total
NO	1004 68.30 73.18 95.26	368 25.03 26.82 88.46	1372 93.33
YES	50 3.40 51.02 4.74	48 3.27 48.98 11.54	98 6.67
Total	1054 71.70	416 28.30	1470 100.00

The frequency table shows that people who have experienced overtime work are more likely to turn over (the percentage is 11.54%, which greater than 4.74% in non-overtime group).

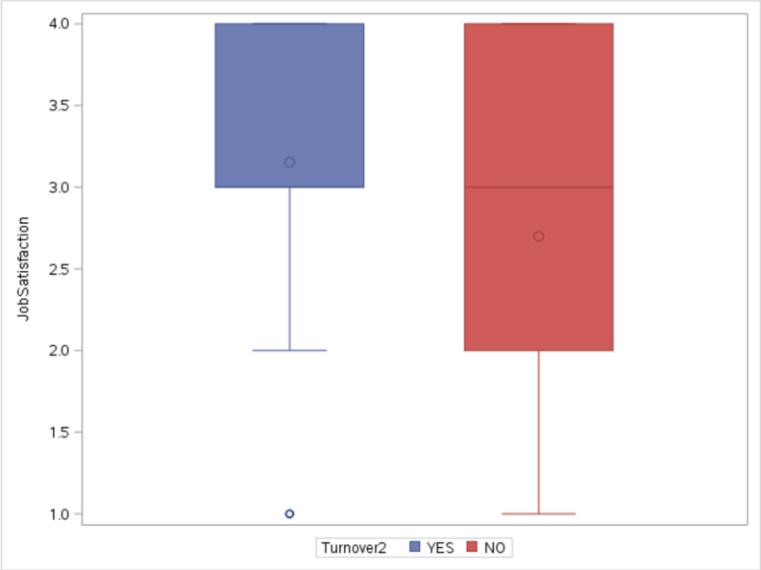
Education

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Turnover2 by Education					
	Education(Education)					
Turnover2	1	2	3	4	5	Total
NO	154 10.48 11.22 90.59	263 17.89 19.17 93.26	533 36.26 38.85 93.18	374 25.44 27.26 93.97	48 3.27 3.50 100.00	1372 93.33
YES	16 1.09 16.33 9.41	19 1.29 19.39 6.74	39 2.65 39.80 6.82	24 1.63 24.49 6.03	0 0.00 0.00 0.00	98 6.67
Total	170 11.56	282 19.18	572 38.91	398 27.07	48 3.27	1470 100.00

The frequency table shows that people who have lower level of education are more likely to turn over (the percentage is greatest when Education=1).

JobSatisfaction



The box plot shows the spread range of JobSatisfaction is smaller and greater in the group of turnover data, it indicates that people with higher job satisfaction are more likely to turnover.

JobRole

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Turnover2 by JobRole				
	JobRole			
Turnover2	1	2	3	Total
NO	449 30.54 32.73 98.03	300 20.41 21.87 87.72	623 42.38 45.41 92.99	1372 93.33
YES	9 0.61 9.18 1.97	42 2.86 42.86 12.28	47 3.20 47.96 7.01	98 6.67
Total	458 31.16	342 23.27	670 45.58	1470 100.00

The frequency table shows when job role is 2 (Laboratory Technician and Sales Representative), people are more likely to turn over (the percentage of JobRole=2 is highest 12.28%), means Laboratory Technician and Sales Representative are more likely to leave.

Conclusion

The people who is younger than 50 (before retired) and acted as Laboratory Technician or Sales Representative are tending to turn over. They generally have more working experience and satisfied about the job, but their education level is low;

That might due to those people find themselves well in the job, but they have low working efficiency so they are quit the job involuntary or be fired;

Conclusion

It is evident from the detail level analysis that have done at each granular level of data, that, there are significant permutation and combination of factors are affecting the employee turn-over at different level of the ladder of the turn over category, so, in a nutshell in this section the project tried to summarize the following key insights which are extremely significant for the company to understand the dynamics of the employee turnover in the company.

1st of all, the project found some independent contributors as follows, Like,

- Attrition rate is higher within the young population of the company on both voluntary and involuntary level. The reasons vary (As seen above in the detail analysis)
- Job role and no of years present in the job role, plays a significant role in the modulation of attrition rate, especially for the young employees.
- Obviously, the income affects the attrition, the lower the income, the more is potential for that employee group to leave the company voluntarily.

However, from the analysis that, it will be very rudimentary to analyze all the events (and the effects of other parameters on them) separately. Rather, found that, if combine the following pair of turn over types (as mentioned below), and then analyze the effect of other parameters on these types it will give more interesting business insights for explaining the dynamics of the attrition behavior for the company, so have made the following combinations.

- Combination 1: Involuntary Resignation and Job Termination
- Combination 2: Retirement and Voluntary Resignation

After the above fusion and combination of event types, there had more insights on how turn over affected in each of these groups. Some key findings are as follows, Like,

- Older people in the company with more experience and more job-related travels are more prone to retirement and voluntary resignation as these people are not that satisfied sacrificing their families and life balance with frequent travels, so they are opting for retirement like situations.
- Stagnancy in current role are also a key factor for voluntary resignation for experienced employees with higher qualifications, on the contrary people with low qualifications are preferring to hold positions as they are satisfied with the mundane work, and thus affecting the productivity of the company these low competency level, older populations are eventually getting terminated anyway.
- Low experience and lack of performance due to incompetence is also a key factor of involuntary resignation or job termination.
- Finally, have seen that in majority of the cases, bonus does not pay a significant role for voluntary resignation or termination or involuntary resignation, however, the project found an interesting business insight, that, cumulative bonus and lack of lesser bonus for many years is creating an impact for retirement population as many highly experienced people are opting for retirement as their cumulative bonus is not giving them any edge, so they are not willing to put more years in the company and retiring

Recommendations

After carefully analyzing and concluding all the different levels of impacts for turn overrate, there are also proactively constructed few recommendations for the HR and higher management of the 'Fermalogis' company, which will help the company in long run, if they monitor and implement these recommendations in a regular manner and take proactive actions to mitigate risk associated with the different level or categories of employee attrition

- More competent employees are more prone leave the company voluntarily irrespective of young or old, when there is a lack of challenging assignments, stagnancy at work. So recommend the management to engage more with these population, carefully monitor their job growth, aspirational needs, and periodically make possible arrangement to provide more assignments which are of high impact nature to these people. Reward and recognition, empowerment, autonomy of work, sound line- manger who can listen to these people's need are also key to avoid attrition for these level of people.
- From the analysis, a significant population of employee are at risk of termination, either, due to lesser contribution in productivity and value add (like older population with lesser qualification, doing mundane work and not re-trainable or capable of lateral movement in the company) or due to lack of experience and incompetence (like younger population with mediocre qualification). And recommending the higher management and HR of the company to apply the following principles for these people
    - 1. For older population with less competence, someone has been in the company for a long time and doing same mundane tasks for years, company should figure out innovative ways to re-train these populations so they can be better utilized at a higher productivity level.
    - 2. For the younger population with low competence, HR needs to be very strict in recruitment to avoid these population and filter them carefully in the selection process to avoid HIRE and FIRE scenarios.
- Finally, have seen a significant population of the employee on the verge of retirement for various reason and situation (as have already explained above), Some of these people are

highly qualified and extremely knowledgeable and experienced and might leave a void in their respective departments once they retire. To avoid such situation management associated to these people, should think of seamless transition of these people with other competent employees from the organization, way ahead of the retirement, so there is ample opportunity for proper knowledge transfer without affecting the business as usual.

- Last but not the least, can say that, although attrition is ‘inevitable’ at any point of an organization, but if management and HR team pays careful attention towards the insights and recommendation the report has provided above, and monitor them on a regular basis, these factors can be managed significantly better, and will have lesser impacts towards the overall productivity of the workforce.

Appendix

Variables explanation

Attribute-Name	Attribute Definition	Sample value(s)
Age	age of the employee when this dataset was created	41
Turnover	shows whether the employee left the company or not	'Yes'
Type	type of turnover	0: No Turnover 1: Retirement 2: Voluntary Resignation 3: Involuntary Resignation 4: Job Termination, Employee is fired
BusinessTravel	shows how much travel employee makes	Travel_Rarely
DailyRate	daily compensation of employee before any cuts/taxes	1102
Department	shows the department of the employee when this dataset was created	Sales
DistanceFromHome	commuting distance for the employee in miles	1
Education	1 - 'Below College' , 2 -'College',3 - 'Bachelor',4 - 'Master',5 - 'Doctor'	2
EducationField	shows the education field of the employee	Life Sciences
EmployeeCount	a field used for aggregation calculations	1
EmloyeeNumber	the ID of the employee	1
EnvironmentSatisfaction	a score showing how much the employee is satisfied with company's facilities (1 - 'Low',2 - 'Medium',3 - 'High', 4 - 'Very High')	2
Gender	shows the gender of the employee	Female
HourlyRate	hourly compensation of employee before any cuts/taxes	94
JobInvolvement	a score given to the employee by supervisors how much the employee is involved in company's operations (1 -'Low',2 -'Medium',3 -'High',4 -'Very High')	3
JobRole	shows the job role of the employee in the company	Sales Executive
JobLevel	shows the management level of the employee	2

JobSatisfaction	shows the last survey result of the employee about his\her job satisfaction	4
MaritalStatus	shows the marital status of the company	Single
MonthlyIncome	shows the monthly income of the employee	5993
MontlyRate	monthly compensation of employee before any cuts/taxes	19479
NumCompaniesWorked	the number of companies the employee worked before starting in the company	8
Over18	shows whether the employee is over 18 years old	Y
OverTime	shows whether employee works overtime more than 10 hours a week	Yes
PercentSalaryHike	shows the agreed yearly salary rise percent	11
PerformanceRating	a score given to the employee by supervisors how good was the performance of the employee last year (1 -'Low',2 - 'Good',3 -'Excellent',4 - 'Outstanding')	3
RelationshipSatisfaction	shows the last survey result of the employee about his\her satisfaction with other employees in the company(1- 'Low',2- 'Medium',3- 'High',4- 'Very High')	1
StandardHours	number of hours employee works for one payroll period (two weeks)	80
StockOptionLevel	shows the stock option for the employee. If your analyses give significant results for this variable, you can refer to that group as "employees having stock option level x"	0
TotalWorkingYears	shows the time the employee worked as a professional (at any company)	8
TrainingTimesLastYear	shows the number of training programs employee has attended last year	0
WorkLifeBalance	shows the employee satisfaction of the work load (4 is the highest satisfaction level) (1- 'Bad',2 - 'Good',3- 'Better',4 -'Best')	1
YearsAtCompany	Tenure at the company	6
YearsInCurrentRole	the number of years employee works in the current position	4
YearsSinceLastPromotion	shows the number of years passed since the last promotion	0
YearsWithCurrentManager	Shows the number of years with the current supervisor.	5
bonus_1-40	Shows whether the employee received bonus payments in the last 40 years. bonus_1 is last year	0……

SAS Code

```
* create library;
LIBNAME PROJ '/home/xuelingchen0/Project';
RUN;

* read data;
PROC IMPORT DATAFILE='/home/xuelingchen0/Project/FermaLogis_Event_Type3.xls'
            OUT=FermaLogis_Event_Type DBMS=xls REPLACE;
RUN;

* data exploration;
ODS GRAPHICS ON; /*to have graph trn ON*/
PROC SGPLOT DATA=FermaLogis_Event_Type;
VBAR Age/ GROUP=Turnover;
*age>=40 less turnover;

PROC SGPLOT DATA=FermaLogis_Event_Type;
VBAR YearsInCurrentRole/ GROUP=Turnover;
*YearsInCurrentRole>=4 less turnover;

PROC SGPLOT DATA=FermaLogis_Event_Type;
HISTOGRAM MonthlyIncome/ GROUP=Turnover BINWIDTH=1000;
*MonthlyIncome>5000 less turnover;

*do not consider about type 4;
PROC FREQ DATA=FermaLogis_Event_Type;
TABLE Type*PerformanceRating;
RUN;

PROC FREQ DATA=FermaLogis_Event_Type;
TABLE Type*RelationshipSatisfaction;
RUN;

PROC SGPLOT DATA=FermaLogis_Event_Type;
VBAR PerformanceRating/ GROUP=Type;
WHERE Type<>0 and Type<>4;

PROC SGPLOT DATA=FermaLogis_Event_Type;
VBAR RelationshipSatisfaction/ GROUP=Type;
WHERE Type<>0 and Type<>4;

* manage data;
DATA FermaLogis_Event_Type_Event_Type;
SET FermaLogis_Event_Type;
IF Turnover='Yes' AND Type=0 THEN DO;
   IF Age>=40 OR YearsInCurrentRole>=4 OR MonthlyIncome>=5000 THEN Turnover='No';
   ELSE IF PerformanceRating=4 OR RelationshipSatisfaction=4 THEN Type=3;
   ELSE Type=2;
END;
IF Turnover='Yes' THEN Turnover2=1;
ELSE Turnover2=0;
IF Gender='Male' THEN Gender2=1;
ELSE Gender2=0;
IF Over18='Y' THEN Over182=1;
ELSE Over182=0;
IF OverTime ='Yes' THEN OverTime2=1;
ELSE OverTime2=0;
IF JobRole='Healthcare Representative' OR JobRole='Research Director' OR JobRole='Manager' OR JobRole='Manufacturing Director' THEN JobRole2=1;
ELSE IF JobRole='Laboratory Technician' OR JobRole='Sales Representative' THEN JobRole2=2;
ELSE JobRole2=3;
DROP Turnover Gender Over18 OverTime JobRole;
RENAME Turnover2=Turnover Gender2=Gender Over182=Over18 OverTime2=OverTime JobRole2=JobRole;
RUN;

* Coding For Event Type:
0 - No turnover
1 - Retirement
2 - Voluntary Resignation
3 - Involuntary Resignation (Health problems, family matters etc.)
4 - Job Termination, Employee is Fired;

* combine data;
DATA Retirement; /*create retirement data*/
  SET FermaLogis_Event_Type_Event_Type;
  Event=(Type=1);  /*this is for censoring out other types, another way to write if statement*/
  Event_type='Retirement';
DATA Vol_Resignation;  /*create  Voluntary Resignation data*/
  SET FermaLogis_Event_Type_Event_Type;
  Event=(Type=2);
  Event_type='Voluntary Resignation';
DATA Invol_Resignation; /*create Involuntary Resignation data*/
  SET FermaLogis_Event_Type_Event_Type;
  event=(Type=3);
  Event_type='Involuntary Resignation';
DATA Job_Termination; /*create Job Termination data*/
  SET FermaLogis_Event_Type_Event_Type;
  event=(Type=4);
  Event_type='Job Termination';
DATA PROJ.Combine; /* combined the datasets and use them as strata in the graphical analysis*/
  FORMAT Event_type $23.;
  SET Retirement Vol_Resignation Invol_Resignation Job_Termination;
RUN;
```

```
92
93  * Whether ther is difference among types;
94  PROC LIFETEST DATA=PROJ.Combine PLOTS=LLS;
95    TIME YearsAtCompany*Event(0);
96    STRATA Event_type /diff=all;
97  RUN;
98  * Voluntary Resignation is different from other three types;
99  * Job Termination are close to Involuntary Resignation and Retirement;
100 * And all others are different;
101
102 * tests whether coefficients found for each event type is equal to coefficients found for the unseparated model;
103 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
104    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
105          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
106    MODEL YearsAtCompany*Type(0)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
107          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
108          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
109          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
110 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
111    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
112          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
113    MODEL YearsAtCompany*Type(0,2,3,4)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
114          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
115          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
116          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
117 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
118    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
119          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
120    MODEL YearsAtCompany*Type(0,1,3,4)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
121          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
122          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
123          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
124 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
125    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
126          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
127    MODEL YearsAtCompany*Type(0,1,2,4)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
128          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
129          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
130          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
131 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
132    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
133          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
134    MODEL YearsAtCompany*Type(0,1,2,3)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
135          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
136          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
137          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
138
139 DATA LogRatioTest0;
140     Nested = 2300.331;
141     Retirement = 78.079;
142     Vol_Resignation = 918.759;
143     Invol_Resignation = 544.889;
144     Job_Termination = 320.797;
145
146     Total = Retirement + Vol_Resignation + Invol_Resignation + Job_Termination;
147     Diff = Nested - Total;
148
149     P_value = 1 - probchi(Diff,87); *116 coef. in 4 models - 29 coef. in nested;
150 RUN;
151
152 PROC PRINT DATA = LogRatioTest0;
153     FORMAT P_Value 5.3;
154 RUN;
155 *p<0.05 means should seperate;
156
157 * test to see whether can use the same coefficients for Retirement and Job Termination;
158 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
159    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
160          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
161    MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
162          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
163          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
164          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
165 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
166    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
167          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
168    MODEL YearsAtCompany*Type(0,2,3,4)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
169          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
170          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
171          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
172 PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
173    CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
174          PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
175    MODEL YearsAtCompany*Type(0,1,2,3)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
176          EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
177          NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
178          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
179
```

```
DATA LogRatioTest1;
    Nested = 597.318;
    Retirement = 78.079;
    Invol_Resignation = 320.797;

    Total = Retirement + Invol_Resignation;
    Diff = Nested - Total;

    P_value = 1 - probchi(Diff,58); *87 coef. in 3 models - 29 coef. in nested;
RUN;

PROC PRINT DATA = LogRatioTest1;
    FORMAT P_Value 5.3;
RUN;
*p<0.05 means should seperate;


* test to see whether can use the same coefficients for Involuntary Resignation and Job Termination;
PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
   CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
         PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
         EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
         NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
   CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
         PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2,4)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
         EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
         NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;
PROC PHREG DATA=FermaLogis_Event_Type_Event_Type;
   CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
         PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2,3)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
         EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
         NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager;

DATA LogRatioTest2;
    Nested = 937.370;
    Retirement = 544.889;
    Job_Termination = 320.797;

    Total = Retirement + Job_Termination;
    Diff = Nested - Total;

    P_value = 1 - probchi(Diff,58); *87 coef. in 3 models - 29 coef. in nested;
RUN;

PROC PRINT DATA = LogRatioTest2;
    FORMAT P_Value 5.3;
RUN;
*p=0.1.7, which is greater than 0.05 means type 3 and type 4 can be merged;

*manage data;
DATA FermET; /*create retirement data*/
  SET FermaLogis_Event_Type_Event_Type;
  IF type=4 THEN type=3;
  RUN;


************************************* Type=1 ****************************************************;
* do not consider bouns_40 since no valid data included, when worked 40 years bonus will only have value in bonus_39;

********* one-way test **********;
* no big difference;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,2,3);
        STRATA EducationField/ADJUST=TUKEY;
RUN;
* no big difference;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,2,3);
        STRATA JobRole/ADJUST=TUKEY;
RUN;
* no big difference;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,2,3);
        STRATA JobLevel/ADJUST=TUKEY;
RUN;
* no big difference;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,2,3);
        STRATA EnvironmentSatisfaction/ADJUST=TUKEY;
RUN;
* no big difference;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,2,3);
        STRATA Education/ADJUST=TUKEY;
RUN;
```

```
**********Choose model**********;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
         PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
         EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
         NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
         / TIES=EFRON SELECTION=backward;
*remove 23 variables;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel;
   MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel Gender NumCompaniesWorked TotalWorkingYears YearsInCurrentRole
         / TIES=EFRON;

**********Time-dependent variable**********;

*new variable about bonus, consider 1 year before and 2 years before;
PROC PHREG DATA=FermET;
   WHERE YearsAtCompany>2;
   CLASS BusinessTravel;
   MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel Gender NumCompaniesWorked TotalWorkingYears YearsInCurrentRole
         bonus1 bonus2/ TIES=EFRON;
   ARRAY bonus_(*) bonus_1-bonus_39;
   bonus1=bonus_[YearsAtCompany-1];
   bonus2=bonus_[YearsAtCompany-2];
RUN;
*both not effective;

*the effect of bonus cumulatively-programming statement;
DATA Fermcum;
   SET FermET;
   ARRAY bon(*) bonus_1-bonus_39;
   ARRAY cum(*) cum1-cum39;
   cum1=bonus_1;
   DO i=2 TO 39;
      cum(i)=(cum(i-1)*(i-1) + bon(i))/i;
   END;
PROC PHREG DATA=Fermcum;
   WHERE YearsAtCompany>1;
   CLASS BusinessTravel;
   MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel Gender NumCompaniesWorked TotalWorkingYears YearsInCurrentRole
         bonuss/ TIES=EFRON;
   ARRAY cumbon(*) cum1-cum39;
   bonuss=cumbon[YearsAtCompany-1];
RUN;
*is effective p<0.05;

**********non-proportionally**********;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel;
   MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel Gender NumCompaniesWorked TotalWorkingYears YearsInCurrentRole
         / TIES=EFRON;
   ASSESS PH / RESAMPLE;
RUN;

*shoenfeld residuals;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel;
   MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel Gender NumCompaniesWorked TotalWorkingYears YearsInCurrentRole
         / TIES=EFRON;
    OUTPUT OUT=b RESSCH=schAge schBusinessTravel schGender schNumCompaniesWorked schTotalWorkingYears schYearsInCurrentRole;
RUN;
*Age NumCompaniesWorked TotalWorkingYears YearsInCurrentRole;

*plot the residuals and see how it's calculated;
DATA b;
    SET b;
    id= _n_;
RUN;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schNumCompaniesWorked / datalabel=id;
run;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schTotalWorkingYears / datalabel=id;
run;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schYearsInCurrentRole / datalabel=id;
run;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schAge / datalabel=id;
run;

*calculate p value;
DATA c;
  SET b;
  lYearsAtCompany=log(YearsAtCompany);
  YearsAtCompany2=YearsAtCompany**2;
PROC CORR data = c;
    VAR YearsAtCompany lYearsAtCompany YearsAtCompany2;
    WITH schAge schBusinessTravel schGender schNumCompaniesWorked schTotalWorkingYears schYearsInCurrentRole;
RUN;
```

```
** deal with non-proportional problem;
PROC PHREG DATA=FermET;
    CLASS BusinessTravel;
    MODEL YearsAtCompany*Type(0,2,3)=Age BusinessTravel Gender NumCompaniesWorked TotalWorkingYears YearsInCurrentRole
         yAge yNumCompaniesWorked yTotalWorkingYears yYearsInCurrentRole/ TIES=EFRON;
    yAge=age*YearsAtCompany;
    yNumCompaniesWorked=NumCompaniesWorked*YearsAtCompany;
    yTotalWorkingYears=TotalWorkingYears*YearsAtCompany;
    yYearsInCurrentRole=YearsInCurrentRole*YearsAtCompany;
RUN;

*********Analysis for Business problems**************;
*****Who are leaving the company? Why? ...;
*new data set;
DATA FermET1;
SET FermET;
IF Type=1 THEN Turnover2='YES';
ELSE Turnover2='NO';
sumbon=sum(OF bonus:);
yAge=age*YearsAtCompany;
yNumCompaniesWorked=NumCompaniesWorked*YearsAtCompany;
yTotalWorkingYears=TotalWorkingYears*YearsAtCompany;
yYearsInCurrentRole=YearsInCurrentRole*YearsAtCompany;
RUN;

PROC SGPLOT DATA=FermET1;
HISTOGRAM Age/ GROUP=Turnover2;
*Age>=44;
PROC SGPLOT DATA=FermET1;
HISTOGRAM sumbon/ GROUP=Turnover2;
*cummulative bonus <=6 or >=12;
PROC SGPLOT DATA=FermET1;
VBOX TotalWorkingYears/ GROUP=Turnover2;
*longer working years;
PROC SGPLOT DATA=FermET1;
VBOX NumCompaniesWorked/ GROUP=Turnover2;
*work on more company;

*conclusion: old people older than 44 are about to retired, most of them have more working experience
             and have worked at more compaies;
*Possible Reason: Those old people are close to the age of retirement;

*************************** Type=2 *********************************************;
* do not consider bouns_40 since no valid data included, when worked 40 years bonus will only have value in bonus_39;

********* one-way test **********;
* HumanResource different from Others;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,3);
        STRATA EducationField/ADJUST=TUKEY;
RUN;
* all different;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,3);
        STRATA JobRole/ADJUST=TUKEY;
RUN;
* 1 different from 2-5, 2 different from 5;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,3);
        STRATA JobLevel/ADJUST=TUKEY;
RUN;
* 1 different from 3&4;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,3);
        STRATA EnvironmentSatisfaction/ADJUST=TUKEY;
RUN;
* 1 different from 3&4, 3 different from 4;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,3);
        STRATA Education/ADJUST=TUKEY;
RUN;

*********Choose model*****************;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
         PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
         EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
         NumCompaniesWorked OverTime PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
         / TIES=EFRON SELECTION=BACKWARD;
*remove 15 variables;
PROC PHREG DATA=FermET;
   CLASS EducationField JobRole Education EnvironmentSatisfaction RelationshipSatisfaction JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age DistanceFromHome Education EducationField EnvironmentSatisfaction JobLevel JobRole
        JobSatisfaction NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel YearsInCurrentRole YearsWithCurrManager
         / TIES=EFRON;
```

```
**********Time-dependent variable************;

*new variable about bonus, consider 1 year before and 2 years before;
PROC PHREG DATA=FermET;
   WHERE YearsAtCompany>2;
   CLASS EducationField JobRole Education EnvironmentSatisfaction RelationshipSatisfaction JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age DistanceFromHome Education EducationField EnvironmentSatisfaction JobLevel JobRole
         JobSatisfaction NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel YearsInCurrentRole YearsWithCurrManager
         bonus1 bonus2/ TIES=EFRON;
   ARRAY bonus_(*) bonus_1-bonus_39;
   bonus1=bonus_[YearsAtCompany-1];
   bonus2=bonus_[YearsAtCompany-2];
RUN;
*both not effective;

*the effect of bonus cumulatively-programming statement;
DATA Fermcum;
   SET FermET;
   ARRAY bon(*) bonus_1-bonus_39;
   ARRAY cum(*) cum1-cum39;
   cum1=bonus_1;
   DO i=2 TO 39;
      cum(i)=(cum(i-1)*(i-1) + bon(i))/i;
   END;
PROC PHREG DATA=Fermcum;
   WHERE YearsAtCompany>1;
   CLASS EducationField JobRole Education EnvironmentSatisfaction RelationshipSatisfaction JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age DistanceFromHome Education EducationField EnvironmentSatisfaction JobLevel JobRole
         JobSatisfaction NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel YearsInCurrentRole YearsWithCurrManager
         bonuss/ TIES=EFRON;
   ARRAY cumbon(*) cum1-cum39;
   bonuss=cumbon[YearsAtCompany-1];
RUN;
*is ont effective;

**********non-proportionally***********;
PROC PHREG DATA=FermET;
   CLASS EducationField JobRole Education EnvironmentSatisfaction RelationshipSatisfaction JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age DistanceFromHome Education EducationField EnvironmentSatisfaction JobLevel JobRole
         JobSatisfaction NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel YearsInCurrentRole YearsWithCurrManager
         / TIES=EFRON;
   ASSESS PH / RESAMPLE;
RUN;
*Age YearsInCurrentRole YearsWithCurrManager;

*shoenfeld residuals;
PROC PHREG DATA=FermET;
   CLASS EducationField JobRole Education EnvironmentSatisfaction RelationshipSatisfaction JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age DistanceFromHome Education EducationField EnvironmentSatisfaction JobLevel JobRole
         JobSatisfaction NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel YearsInCurrentRole YearsWithCurrManager
          / TIES=EFRON;
    OUTPUT OUT=b RESSCH=schAge schDistanceFromHome schEducation schEducationField schEnvironmentSatisfaction schJobLevel schJobRole
         schJobSatisfaction schNumCompaniesWorked schOverTime schRelationshipSatisfaction schStockOptionLevel schYearsInCurrentRole schYearsWithCurrManager;
RUN;

*plot the residuals and see how it's calculated;
DATA b;
    SET b;
    id= _n_;
RUN;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schAge / datalabel=id;
run;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schYearsInCurrentRole / datalabel=id;
run;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schYearsWithCurrManager / datalabel=id;
run;

*calculate p value;
DATA c;
  SET b;
  lYearsAtCompany=log(YearsAtCompany);
  YearsAtCompany2=YearsAtCompany**2;
PROC CORR data = c;
    VAR YearsAtCompany lYearsAtCompany YearsAtCompany2;
    WITH schAge schDistanceFromHome schEducation schEducationField schEnvironmentSatisfaction schJobLevel schJobRole
         schJobSatisfaction schNumCompaniesWorked schOverTime schRelationshipSatisfaction schStockOptionLevel schYearsInCurrentRole schYearsWithCurrManager;
RUN;

** deal with non-proportional problem;
PROC PHREG DATA=FermET;
   CLASS EducationField JobRole Education EnvironmentSatisfaction RelationshipSatisfaction JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,3)=Age DistanceFromHome Education EducationField EnvironmentSatisfaction JobLevel JobRole
         JobSatisfaction NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel YearsInCurrentRole YearsWithCurrManager
          yAge yYearsInCurrentRole yYearsWithCurrManager/ TIES=EFRON;
    yAge=age*YearsAtCompany;
    yYearsInCurrentRole=YearsInCurrentRole*YearsAtCompany;
    yYearsWithCurrManager=YearsWithCurrManager*YearsAtCompany;
RUN;
```

```
**********Analysis for Business problems**************;
*****Who are leaving the company? Why? ...;
*new data set;
DATA FermET2;
SET FermET;
IF Type=2 THEN Turnover2='YES';
ELSE Turnover2='NO';
sumbon=sum(OF bonus:);
yAge=age*YearsAtCompany;
yYearsInCurrentRole=YearsInCurrentRole*YearsAtCompany;
yYearsWithCurrManager=YearsWithCurrManager*YearsAtCompany;
RUN;

PROC SGPLOT DATA=FermET2;
VBAR Age/ GROUP=Turnover2;
*Age<=41;
PROC SGPLOT DATA=FermET2;
HISTOGRAM sumbon/ GROUP=Turnover2;
*cummulative bonus <=7;
PROC SGPLOT DATA=FermET2;
VBOX TotalWorkingYears/ GROUP=Turnover2;
*less working years;
PROC FREQ DATA=FermET2;
TABLE Turnover2*OverTime;
*overtime work;
PROC FREQ DATA=FermET2;
TABLE Turnover2*Education;
*Education is median or low are more likely to leave;
PROC SGPLOT DATA=FermET2;
VBOX JobSatisfaction/ GROUP=Turnover2;
*job satisfaction is lower;

*Conclusion: young people who have less bonus in this company are tend to leave, they have less working experience
             and have less standard for the overtime work and median or low education level, not satisfied with the job;
*Possible Reason: Those young people might thought the job is not so idea and wanna find a better job so they leave volunteery;


************************************ Type=3 ************************************************;
* do not consider bouns_40 since no valid data included, when worked 40 years bonus will only have value in bonus_39;

********* one-way test **********;
* Life Sciences different from Technical Degree ;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,2);
        STRATA EducationField/ADJUST=TUKEY;
RUN;
* different from each other;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,2);
        STRATA JobRole/ADJUST=TUKEY;
RUN;
* 1 different from 2-5;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,2);
        STRATA JobLevel/ADJUST=TUKEY;
RUN;
* 1 different from 2;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,2);
        STRATA EnvironmentSatisfaction/ADJUST=TUKEY;
RUN;
* 1 different from 5;
PROC LIFETEST DATA=FermET PLOTS=S(test);
        TIME YearsAtCompany*Type(0,1,2);
        STRATA Education/ADJUST=TUKEY;
RUN;

*********Choose model*****************;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel Department EducationField JobRole MaritalStatus Education EnvironmentSatisfaction JobInvolvement
         PerformanceRating RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel DailyRate Department DistanceFromHome Education EducationField
         EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
         NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
         / TIES=EFRON SELECTION=backward;
*remove 16 variables (15 + PerformanceRating);
PROC PHREG DATA=FermET;
   CLASS BusinessTravel JobRole RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel JobLevel JobRole JobSatisfaction
         NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel TotalWorkingYears
         TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
         / TIES=EFRON;
```

```
**********Time-dependent variable************;

*new variable about bonus, consider 1 year before and 2 years before;
PROC PHREG DATA=FermET;
   WHERE YearsAtCompany>2;
   CLASS BusinessTravel JobRole RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel JobLevel JobRole JobSatisfaction
          NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel TotalWorkingYears
          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
         bonus1 bonus2/ TIES=EFRON;
   ARRAY bonus_(*) bonus_1-bonus_39;
   bonus1=bonus_[YearsAtCompany-1];
   bonus2=bonus_[YearsAtCompany-2];
RUN;
*both not effective;

*the effect of bonus cumulatively-programming statement;
DATA Fermcum;
   SET FermET;
   ARRAY bon(*) bonus_1-bonus_39;
   ARRAY cum(*) cum1-cum39;
   cum1=bonus_1;
   DO i=2 TO 39;
      cum(i)=(cum(i-1)*(i-1) + bon(i))/i;
   END;
PROC PHREG DATA=Fermcum;
   WHERE YearsAtCompany>1;
   CLASS BusinessTravel JobRole RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel JobLevel JobRole JobSatisfaction
          NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel TotalWorkingYears
          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
         bonuss/ TIES=EFRON;
   ARRAY cumbon(*) cum1-cum39;
   bonuss=cumbon[YearsAtCompany-1];
RUN;
*is not effective;

**********non-proportionally***********;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel JobRole RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel JobLevel JobRole JobSatisfaction
          NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel TotalWorkingYears
          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
         / TIES=EFRON;
   ASSESS PH / RESAMPLE;
RUN;
*YearsInCurrentRole YearsWithCurrManager;

*shoenfeld residuals;
PROC PHREG DATA=FermET;
   CLASS BusinessTravel JobRole RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
   MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel JobLevel JobRole JobSatisfaction
          NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel TotalWorkingYears
          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
          / TIES=EFRON;
    OUTPUT OUT=b RESSCH=schAge schBusinessTravel schJobLevel schJobRole schJobSatisfaction
          schNumCompaniesWorked schOverTime schRelationshipSatisfaction schStockOptionLevel schTotalWorkingYears
          schTrainingTimesLastYear schWorkLifeBalance schYearsInCurrentRole schYearsWithCurrManager;
RUN;
*Age NumCompaniesWorked TotalWorkingYears YearsInCurrentRole;

*plot the residuals and see how it's calculated;
DATA b;
    SET b;
    id= _n_;
RUN;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schYearsInCurrentRole / datalabel=id;
run;
proc sgplot data=b;
    scatter x=YearsAtCompany y=schYearsWithCurrManager / datalabel=id;
run;

*calculate p value;
DATA c;
  SET b;
  lYearsAtCompany=log(YearsAtCompany);
  YearsAtCompany2=YearsAtCompany**2;
PROC CORR data = c;
    VAR YearsAtCompany lYearsAtCompany YearsAtCompany2;
    WITH schAge schBusinessTravel schJobLevel schJobRole schJobSatisfaction
          schNumCompaniesWorked schOverTime schRelationshipSatisfaction schStockOptionLevel schTotalWorkingYears
          schTrainingTimesLastYear schWorkLifeBalance schYearsInCurrentRole schYearsWithCurrManager;
RUN;
*JobRole;
```

```
** deal with non-proportional problem;
PROC PHREG DATA=FermET;
    CLASS BusinessTravel JobRole RelationshipSatisfaction WorkLifeBalance JobLevel StockOptionLevel;
    MODEL YearsAtCompany*Type(0,1,2)=Age BusinessTravel JobLevel JobRole JobSatisfaction
          NumCompaniesWorked OverTime RelationshipSatisfaction StockOptionLevel TotalWorkingYears
          TrainingTimesLastYear WorkLifeBalance YearsInCurrentRole YearsWithCurrManager
          yYearsInCurrentRole yYearsWithCurrManager/ TIES=EFRON;
    yYearsWithCurrManager=YearsWithCurrManager*YearsAtCompany;
    yYearsInCurrentRole=YearsInCurrentRole*YearsAtCompany;
RUN;

**********Analysis for Business problems**************;
*****Who are leaving the company? Why? ...;
*new data set;
DATA FermET3;
SET FermET;
IF Type=3 THEN Turnover2='YES';
ELSE Turnover2='NO';
sumbon=sum(OF bonus:);
yYearsWithCurrManager=YearsWithCurrManager*YearsAtCompany;
yYearsInCurrentRole=YearsInCurrentRole*YearsAtCompany;
RUN;

PROC SGPLOT DATA=FermET3;
VBAR Age/ GROUP=Turnover2;
*Age<=46;
PROC SGPLOT DATA=FermET3;
VBOX sumbon/ GROUP=Turnover2;
*less cummulative bonus;
PROC SGPLOT DATA=FermET3;
VBOX TotalWorkingYears/ GROUP=Turnover2;
*more working years;
PROC FREQ DATA=FermET3;
TABLE Turnover2*OverTime;
*overtime work;
PROC FREQ DATA=FermET3;
TABLE Turnover2*Education;
*Education level is low;
PROC SGPLOT DATA=FermET3;
VBOX JobSatisfaction/ GROUP=Turnover2;
*Higher JObsatisfaction;
PROC FREQ DATA=FermET3;
TABLE Turnover2*JobRole;
*JobRole=2 'Laboratory Technician' OR 'Sales Representative';

*Conclusion: people who is younger than 46(before retired) with less bonus and worked more years, satisfied about the job and the education level is low
             also have more over time working experience;
*Possible Reason: Those people find themselves well in the job, but actually they have low working efficiency wo quit the job involunteery or be fired;
```

References

Arda Zuber, (2017). OPIM-5894 - Survival Analysis with SAS Course Materials. UConn School of Business.

Delwiche, L. D., & Slaughter, S. J. (2012). The Little SAS Book: A Primer: A Primer. SAS Institute.

Allison, P. D. (2010). Survival analysis using SAS: a practical guide. Sas Institute.