SURVIVAL CURVES & ESTIMATES OF EMPLOYEES IN A PHARMACEUTICAL COMPANY

MONTHLY INCOME

Since Monthly Income is a continuous variable, we created Income groups based on the quartiles. This ensures that each of the 4 group has 25% of the employees. The 4 groups are as: low income, medium income, high income, very high income. Using proc univariate, we obtained the result on the RHS and based on 25th, 50th, 75th percentiles we created the income groups as follows:

Quantiles (Definition 5)	
Level	Quantile
100% Max	19999
99%	19627
95%	17856
90%	13798
75% Q3	8380
50% Median	4919
25% Q1	2911
10%	2316
5%	2097
1%	1359
0% Min	1009

```
length income $20.;
if monthlyincome=. then income="null";else
if monthlyincome<=2911 then income="low";else
if 2911<monthlyincome<=4919 then income="medium";else
if 4919<monthlyincome<=8380 then income="high";else
income="very high";
```

income	Frequency	Percent	Cumulative Frequency	Cumulative Percent
high	368	25.03	368	25.03
low	369	25.10	737	50.14
medium	366	24.90	1103	75.03
very high	367	24.97	1470	100.00

Using Frequency procedure (proc freq) we obtain the frequency table as on the LHS. Here we see that each of the 4 groups has 25% of the employees.

- Low Income group: <=$2911
- Medium Income group: $2911 - $4919
- High Income group: $4919 - $8380
- Very High Income group: >$8380

Survival Curve: Kaplan-Meier Method

Summary of the Number of Censored and Uncensored Values					
Stratum	income	Total	Failed	Censored	Percent Censored
1	high	368	39	329	89.40
2	low	369	108	261	70.73
3	medium	366	52	314	85.79
4	very high	367	38	329	89.65
Total		1470	237	1233	83.88

The summary table on the LHS shows that attrition (as inferred from the "*failed*" column) is maximum in the **low income group** followed by **medium income group**. Attrition is comparatively very less in the **high** and **very high income groups**. The same can be inferred from the *"Percent censored"* column. We see its value is least for low income group and quite high for very high & high income groups. **Lower value of percent censored implies higher attrition.**

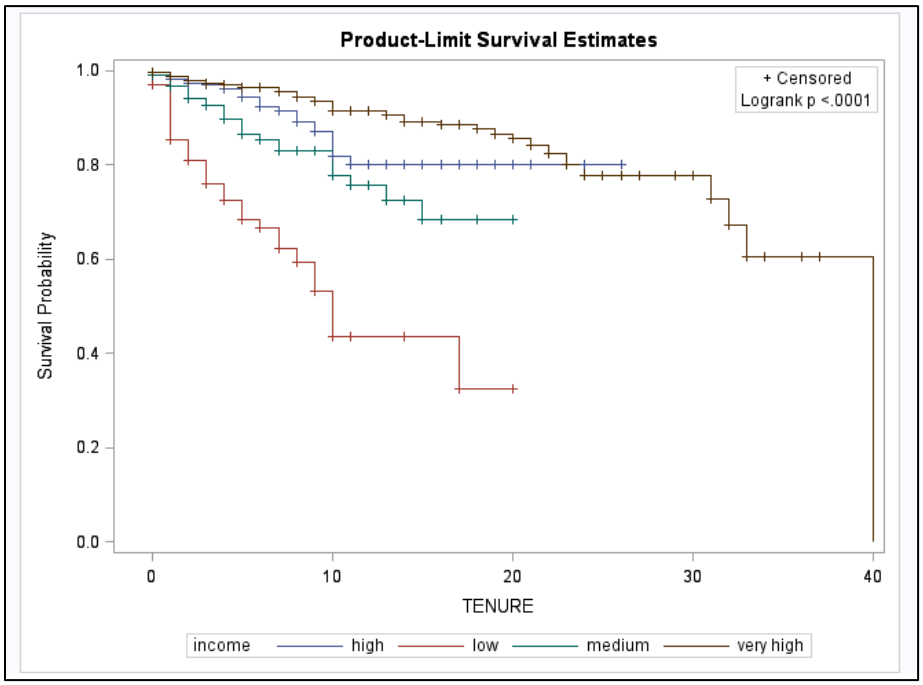
The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant i.e. <0.0001. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the 4 income groups!**

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	181.8348	3	<.0001
Wilcoxon	166.7382	3	<.0001
-2Log(LR)	172.1175	3	<.0001

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
income	income	Chi-Square	Raw	Tukey-Kramer
high	low	92.5572	<.0001	<.0001
high	medium	4.8889	0.0270	0.1202
high	very high	4.7057	0.0301	0.1318
low	medium	56.5074	<.0001	<.0001
low	very high	141.3	<.0001	<.0001
medium	very high	19.5884	<.0001	<.0001

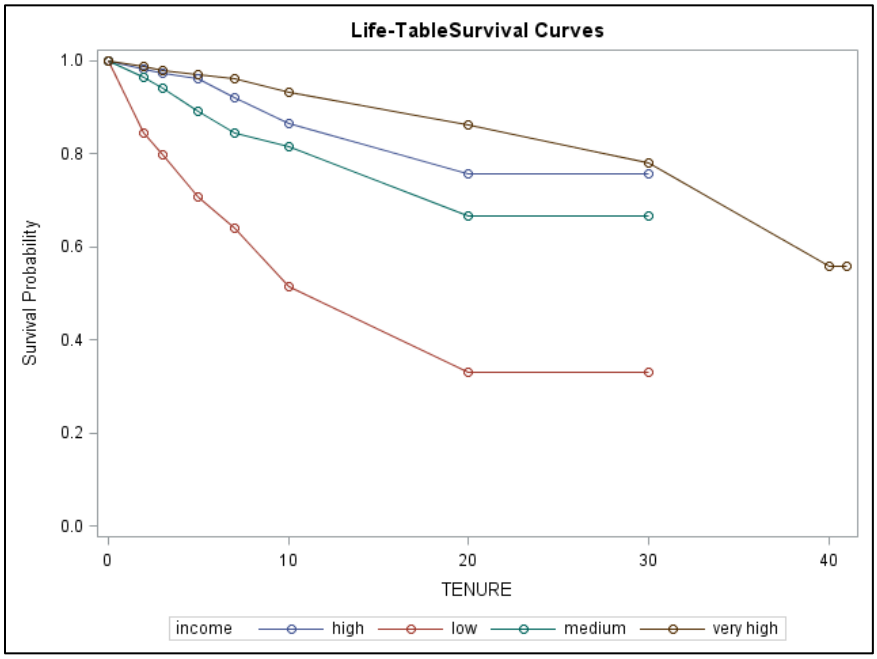
The Log rank Adjustment table on the LHS shows Strata comparison. Here the significance of Tukey-Kramer p-value suggests that the survival curve between the two strata levels is different!
Survival Curves of

- High & Low income groups is different!
- High & Medium income groups is NOT different!
- High & Very high income groups is NOT different!
- Low & Medium income groups is different!
- Low & Very high income groups is different!
- Medium & Very high income groups is different!

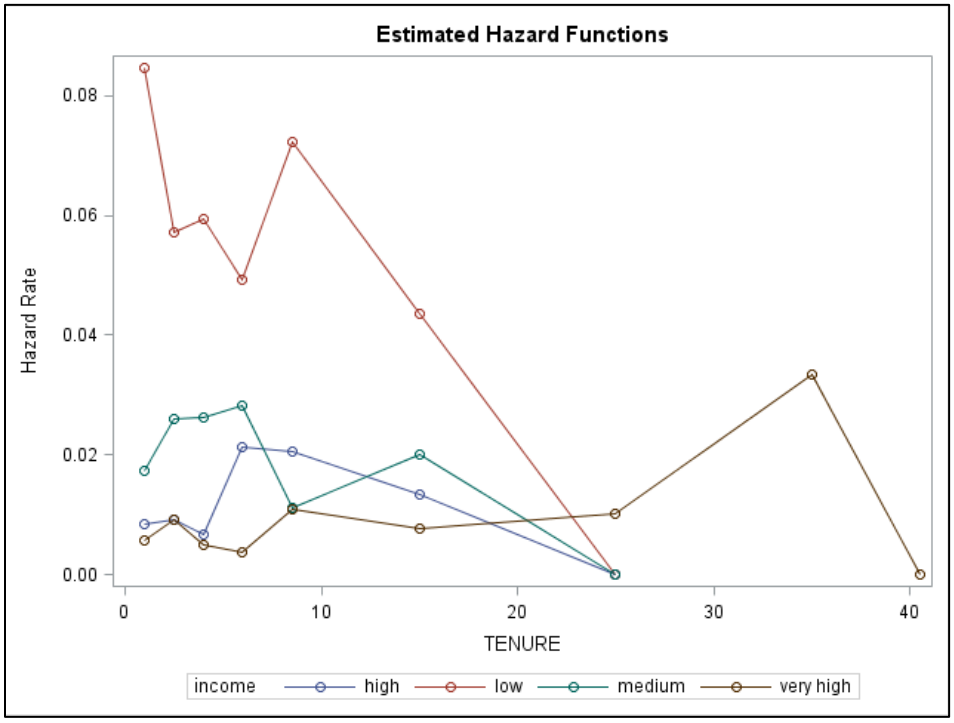


The survival plot on the LHS shows that low income employees attrite at a faster rate as compared to the other groups since its curve is the steepest. **In 1 year, around 15% of the low income employees leave the company.** The high & very high income groups attrite at a much slower rate compared to the low & middle income groups. This inference is in line with our business logic since Low & Medium income employees are mostly entry level and they are continuously on a lookout for a more stable high paying job. On the other hand, the high & very high income employees attrite less since they are already in one!

Survival Curve: Life Table Method



Estimated Hazard Plot



The hazard of attrition for the low income group is the highest from the very first year. The hazard increases sharply for high income employees after year 6, thereafter declines gradually. For medium income employees the hazard increases sharply after year 2 and then declines sharply after year 8. Again picks up after year 10 and then starts to decline at year 15. For Very high income groups, the hazard remains more or less the same till year 15. **For the initial 15 years, low income & high income employees have the highest & lowers risks of attrition respectively!**

JOB ROLE

The frequency table on the RHS shows that there are 9 job roles present in the company. **Sales Executive, Research Scientist, Laboratory Technician** are the 3 largest job roles accounting for around **60%** of the employees. With only 3.54% of the employees, Human Resources is the smallest group in the company.

JOBROLE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Healthcare Representative	131	8.91	131	8.91
Human Resources	52	3.54	183	12.45
Laboratory Technician	259	17.62	442	30.07
Manager	102	6.94	544	37.01
Manufacturing Director	145	9.86	689	46.87
Research Director	80	5.44	769	52.31
Research Scientist	292	19.86	1061	72.18
Sales Executive	326	22.18	1387	94.35
Sales Representative	83	5.65	1470	100.00

Survival Curve: Kaplan-Meier Method

Summary of the Number of Censored and Uncensored Values					
Stratum	JOBROLE	Total	Failed	Censored	Percent Censored
1	Healthcare Representative	131	9	122	93.13
2	Human Resources	52	12	40	76.92
3	Laboratory Technician	259	62	197	76.06
4	Manager	102	5	97	95.10
5	Manufacturing Director	145	10	135	93.10
6	Research Director	80	2	78	97.50
7	Research Scientist	292	47	245	83.90
8	Sales Executive	326	57	269	82.52
9	Sales Representative	83	33	50	60.24
Total		1470	237	1233	83.88

The summary table on the LHS shows that **Sales Representatives, Laboratory Technicians and Human Resources attrite the most** as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

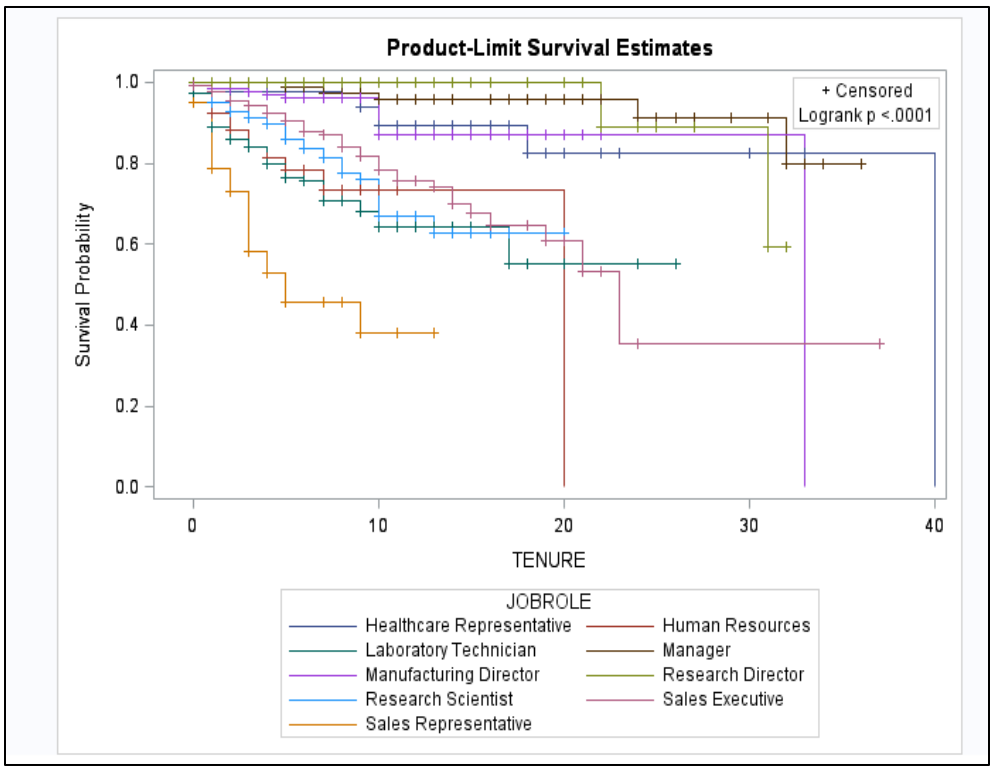
The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant i.e. <0.0001. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the Job Roles!**

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	190.4222	8	<.0001
Wilcoxon	176.6467	8	<.0001
-2Log(LR)	188.5690	8	<.0001

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
JOBROLE	JOBROLE	Chi-Square	Raw	Tukey-Kramer
Healthcare Representative	Human Resources	14.9890	0.0001	0.0035
Healthcare Representative	Laboratory Technician	36.7328	<.0001	<.0001
Healthcare Representative	Manager	0.8991	0.3430	0.9901
Healthcare Representative	Manufacturing Director	0.0116	0.9142	1.0000
Healthcare Representative	Research Director	0.00833	0.9273	1.0000
Healthcare Representative	Research Scientist	10.6273	0.0011	0.0307
Healthcare Representative	Sales Executive	3.8804	0.0489	0.5646
Healthcare Representative	Sales Representative	59.8049	<.0001	<.0001
Human Resources	Laboratory Technician	15.8925	<.0001	0.0022
Human Resources	Manager	25.5497	<.0001	<.0001
Human Resources	Manufacturing Director	13.4415	0.0002	0.0076
Human Resources	Research Director	17.7646	<.0001	0.0008
Human Resources	Research Scientist	0.4300	0.5120	0.9993
Human Resources	Sales Executive	0.3091	0.5782	0.9998
Human Resources	Sales Representative	31.1410	<.0001	<.0001
Laboratory Technician	Manager	48.1013	<.0001	<.0001
Laboratory Technician	Manufacturing Director	34.5771	<.0001	<.0001
Laboratory Technician	Research Director	41.4478	<.0001	<.0001
Laboratory Technician	Research Scientist	5.7080	0.0169	0.2897
Laboratory Technician	Sales Executive	9.2811	0.0023	0.0588
Laboratory Technician	Sales Representative	0.2704	0.6031	0.9999
Manager	Manufacturing Director	1.1022	0.2938	0.9809
Manager	Research Director	1.2208	0.2692	0.9737
Manager	Research Scientist	16.9732	<.0001	0.0013
Manager	Sales Executive	7.3888	0.0066	0.1412
Manager	Sales Representative	79.2338	<.0001	<.0001
Manufacturing Director	Research Director	0.000690	0.9790	1.0000
Manufacturing Director	Research Scientist	9.7366	0.0018	0.0473
Manufacturing Director	Sales Executive	3.4639	0.0627	0.6407
Manufacturing Director	Sales Representative	55.6142	<.0001	<.0001
Research Director	Research Scientist	11.6980	0.0006	0.0181
Research Director	Sales Executive	4.1009	0.0429	0.5256
Research Director	Sales Representative	73.0058	<.0001	<.0001
Research Scientist	Sales Executive	0.7340	0.3916	0.9950
Research Scientist	Sales Representative	7.0049	0.0081	0.1675
Sales Executive	Sales Representative	11.9818	0.0005	0.0157

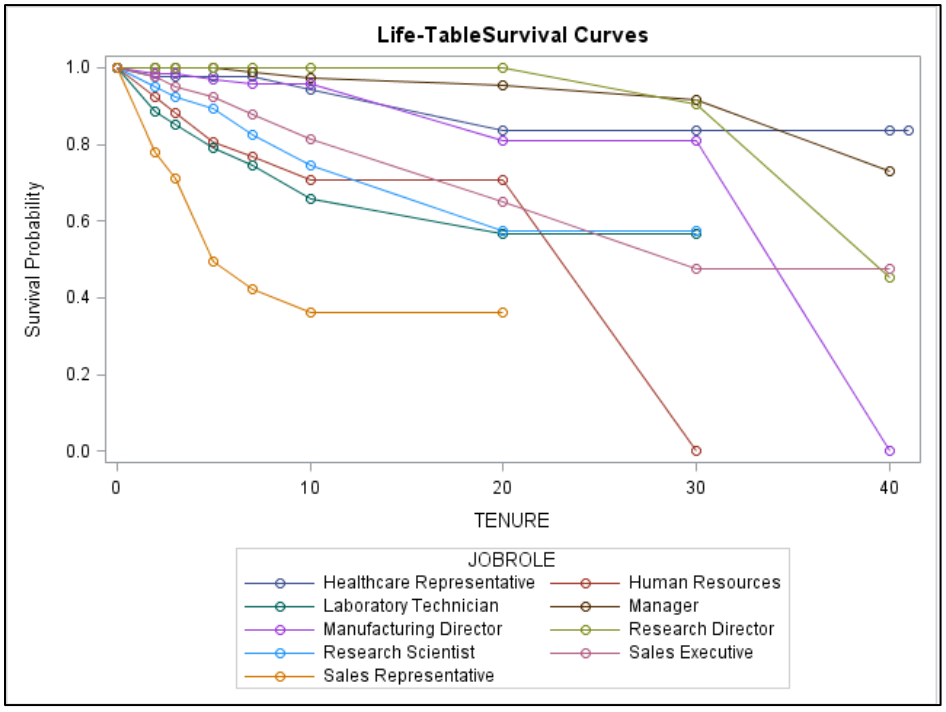
The Log rank Adjustment table on the LHS shows Strata comparison. Here the significance of Tukey-Kramer p-value suggests that the survival curve between the two strata levels is different! The survival curves of between the following Job Roles is different:

- Laboratory Technician & Healthcare Representative
- Laboratory Technician & Manager
- Laboratory Technician & Manufacturing Director
- Laboratory Technician & Research Director
- Sales Representative & Manager
- Sales Representative & Manufacturing Director
- Sales Representative & Research Director
- Sales Representative & Healthcare Representative
- Human Resources & Manager

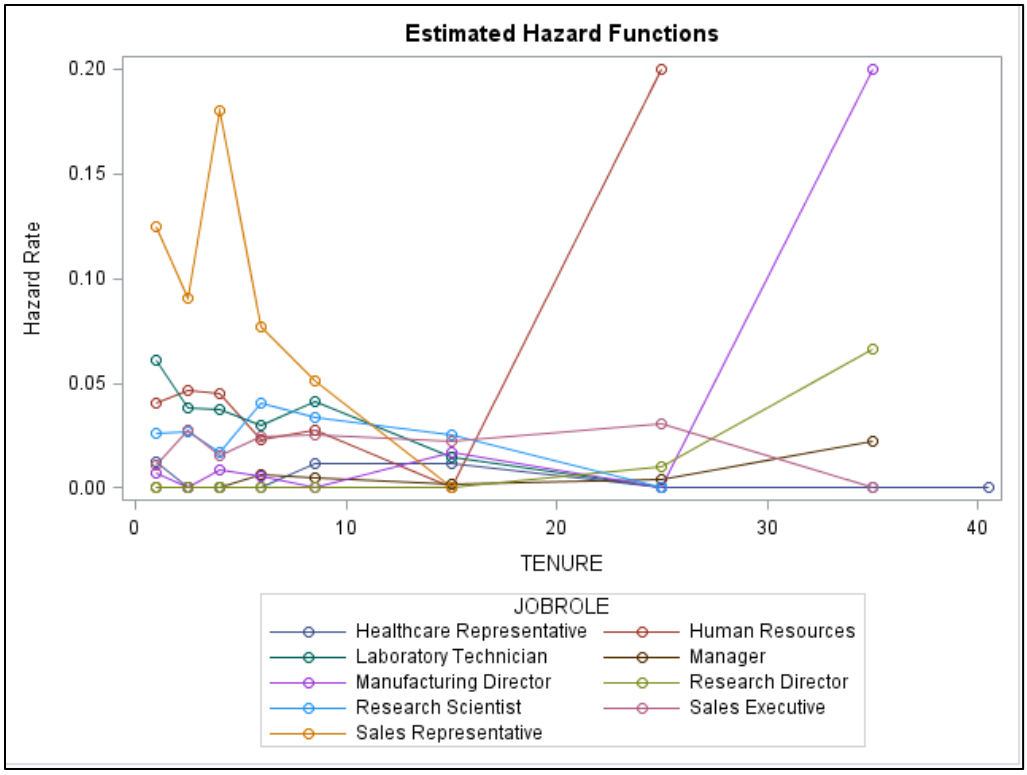


The survival plot on the LHS shows that **Sales representative has the steepest curve followed by Laboratory Technician and Human Resources which implies that they attrite the most**. **In 1 year, 22% of the Sales representatives and 10% of the Laboratory Technicians leave the company.** On the other hand, Research Director, Manager have the most stable survival curves followed by Manufacturing Director & Healthcare representative. The steeper the survival curve is, the faster is the attrition over tenure. **The senior job roles of Research Director, Manager, Manufacturing Director have lesser attrition over time. It is interesting to see that though Healthcare representative is not a senior role, it has less attrition over time.** Obviously, employees having senior job roles don't frequently switch companies as compared to their younger counterparts. They are normally well-settled in their job.

Survival Curve: Life Table Method



Estimated Hazard Plot



The hazard for Sales Representative is the quite high from the very start. Attrition hazard for Human Resources & Manufacturing Director increases sharply after year 15 and 25 respectively. The hazard for Sales executive increases after year 2 and then declines after year 4. For Laboratory Technician, the hazard decreases sharply after year 2 and then gradually decreases till year 8 after which it again picks up. **For the initial 10 years, Sales representative has the highest risk of attrition while Research Director has a constant 0 risk of attrition!**

JOB LEVEL

JOBLEVEL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	543	36.94	543	36.94
2	534	36.33	1077	73.27
3	218	14.83	1295	88.10
4	106	7.21	1401	95.31
5	69	4.69	1470	100.00

Job Level shows the management level of the employee. From the frequency table on the LHS, we can say that 73.27% of the employees belong to low management levels with Job Level 1,2. In other words, they have junior positions in the company.

Survival Curve: Kaplan-Meier Method

The summary table on the RHS shows that employees with Job Level 1 attrite the most while employees with job Level 4,5 attrite the least as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.** In other words, **senior management attrite the least while junior employees attrite the most**. This has already been established in the previous analyses.

Summary of the Number of Censored and Uncensored Values

Stratum	JOBLEVEL	Total	Failed	Censored	Percent Censored
1	1	543	143	400	73.66
2	2	534	52	482	90.26
3	3	218	32	186	85.32
4	4	106	5	101	95.28
5	5	69	5	64	92.75
Total		1470	237	1233	83.88

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	189.2851	4	<.0001
Wilcoxon	159.1823	4	<.0001
-2Log(LR)	197.5725	4	<.0001

The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant i.e. <0.0001. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the Job Levels!**

Adjustment for Multiple Comparisons for the Logrank Test				
Strata Comparison			p-Values	
JOBLEVEL	JOBLEVEL	Chi-Square	Raw	Tukey-Kramer
1	2	102.9	<.0001	<.0001
1	3	100.6	<.0001	<.0001
1	4	156.5	<.0001	<.0001
1	5	155.1	<.0001	<.0001
2	3	3.5903	0.0581	0.3200
2	4	1.4318	0.2315	0.7534
2	5	5.0605	0.0245	0.1617
3	4	1.2343	0.2666	0.8008
3	5	0.00830	0.9274	1.0000
4	5	1.5813	0.2086	0.7173

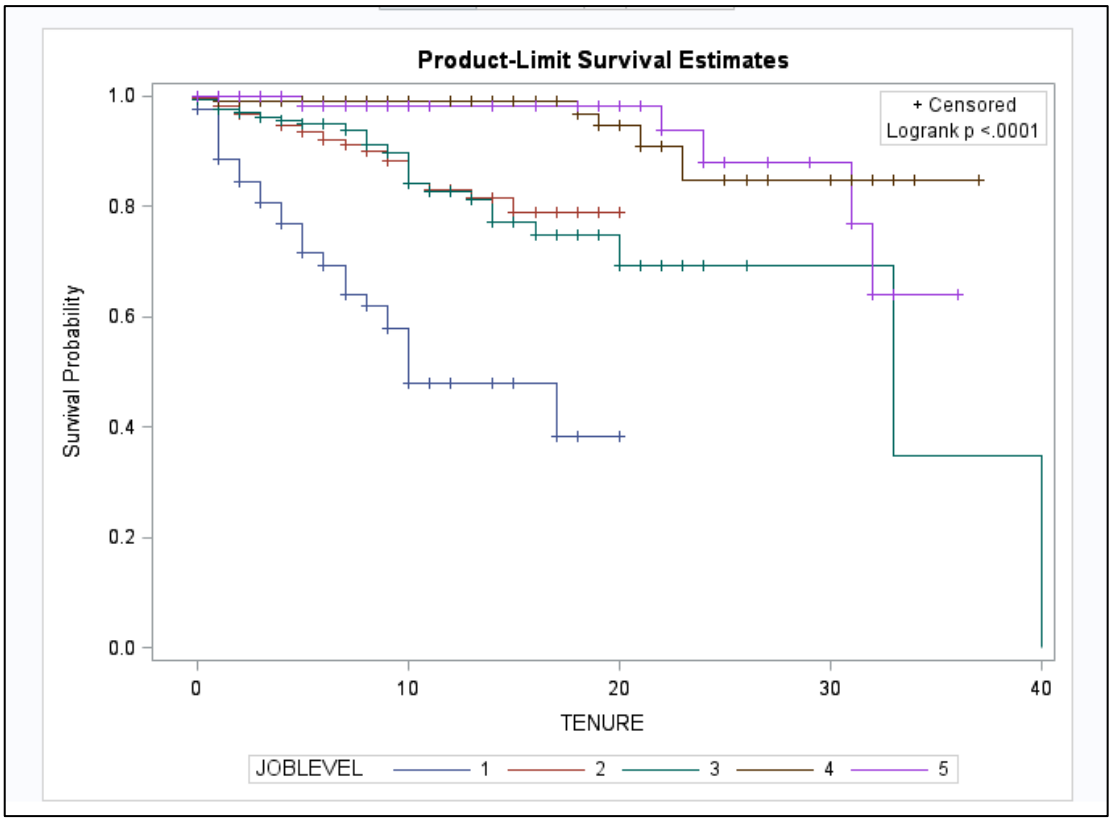
The Log rank Adjustment table on the LHS shows Strata comparison. Here the significance of Tukey-Kramer p-value suggests that the survival curve between the two strata levels is different! The survival curves of between the following Job Roles is different:

- Job Level 1 &2
- Job Level 1 &3
- Job Level 1 &4
- Job Level 1 &5

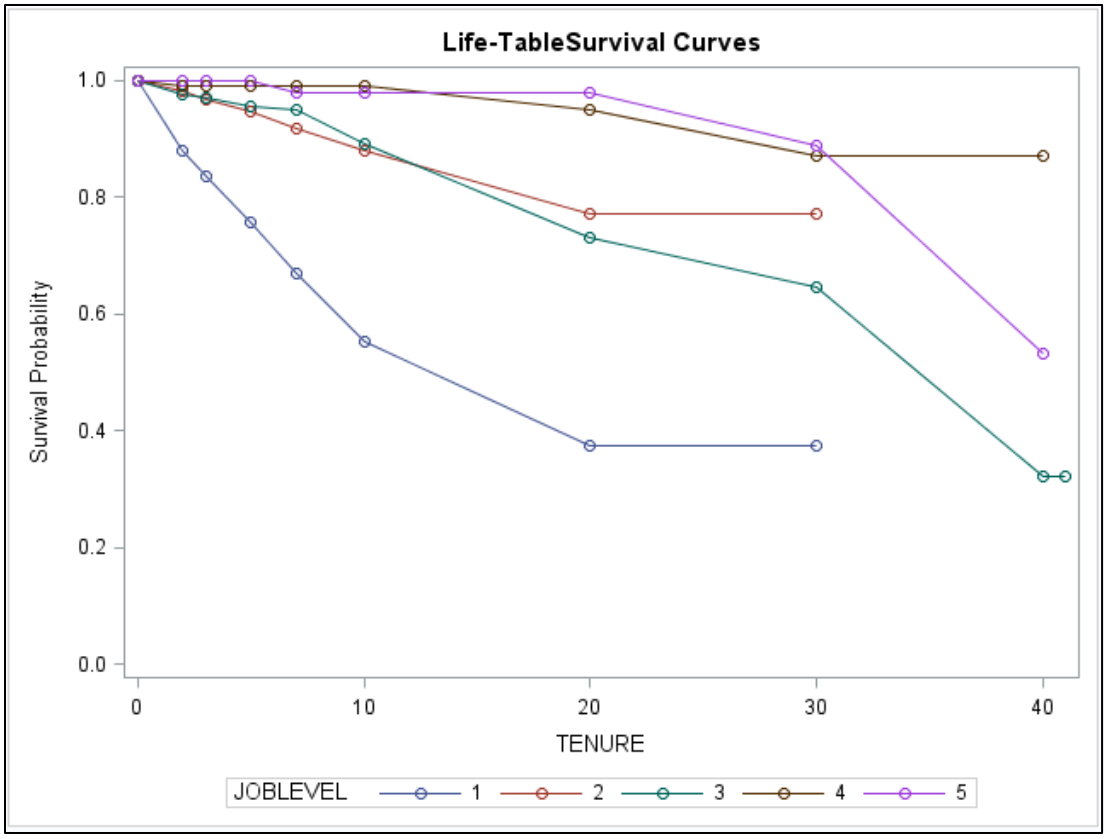
This clearly shows that the survival curve of employees having junior positions is very different from the other management level groups.

The survival plot clearly shows that as the management level of the employee increases, the survival plot becomes less steep. The steeper the survival curve is, the faster is the attrition over tenure.

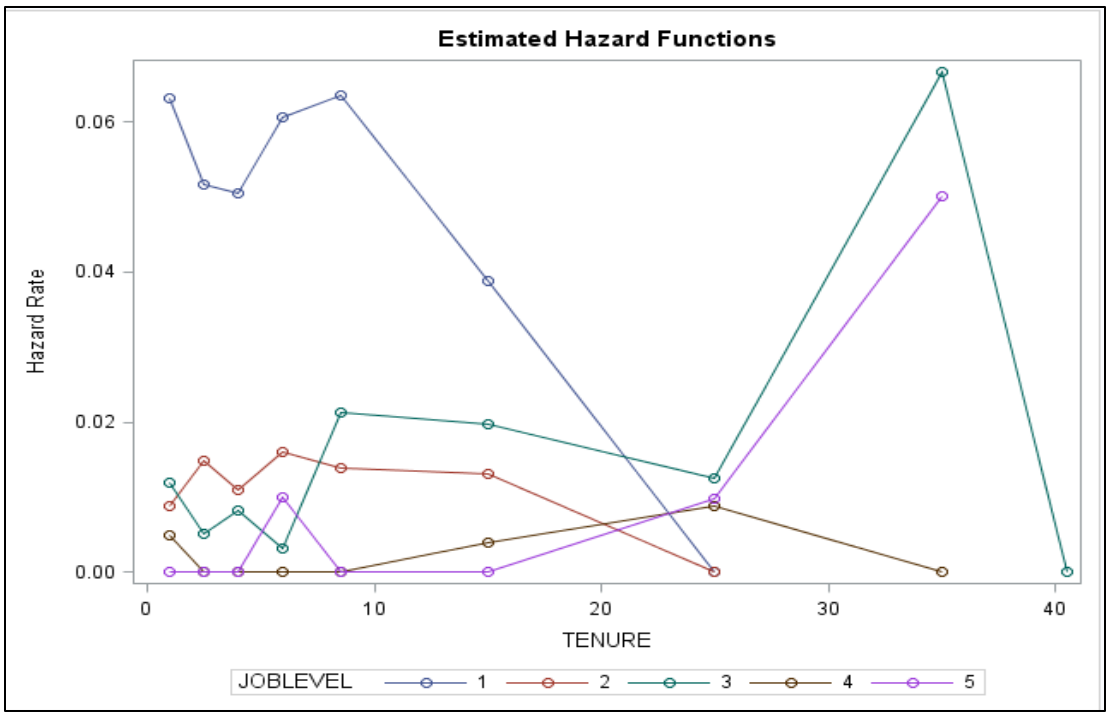
We can conclude that employees having junior positions having job level=1 attrite the fastest over time and as the seniority of the employee increases, attrition lessens. In 1 year, 15% of the junior employees leave the company.



Survival Curve: Life Table Method



Estimated Hazard Plot



The hazard for employees with JobLevel=1 is the highest from the very start till year 20. **Senior most management level employees have 0 risk of attrition for the initial 6 years !**

JOB ROLE- "LABORATORY TECHNICIAN" & INCOME GROUP

Since Laboratory technician is one of the largest job roles (17.62% of employees) in the company & their attrition is pretty high, we decided to look at the impact of income on their survival!

Summary of the Number of Censored and Uncensored Values

Stratum	income	Total	Failed	Censored	Percent Censored
1	high	25	3	22	88.00
2	low	130	38	92	70.77
3	medium	104	21	83	79.81
Total		259	62	197	76.06

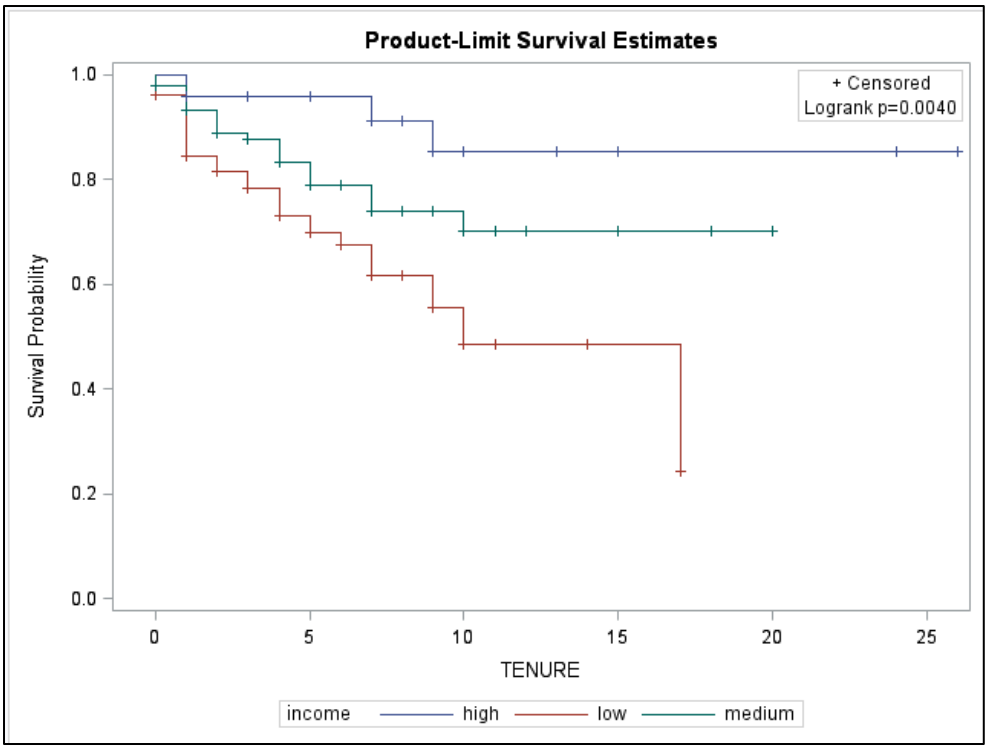
The summary table on the LHS shows that low income group among the Laboratory Technicians have the highest attrition as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.** Also around 50% of them belong to the low income group.

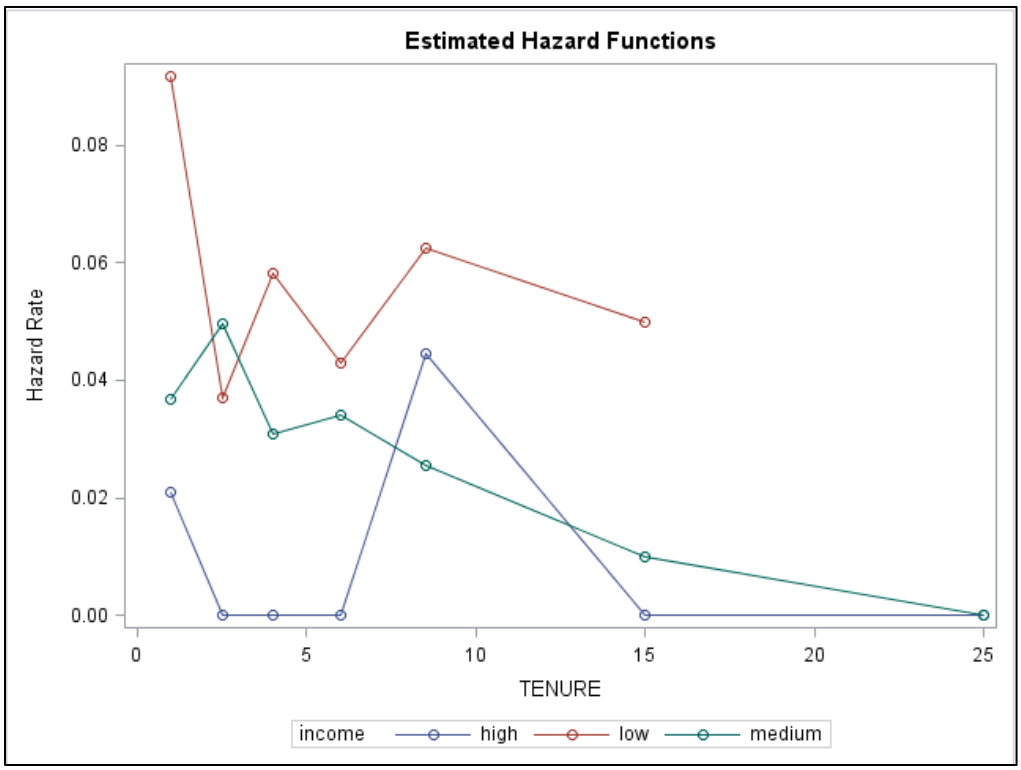
The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant <=0.05 at 5% level of significance. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the income groups!**

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	11.0283	2	0.0040
Wilcoxon	9.0354	2	0.0109
-2Log(LR)	17.3662	2	0.0002

The survival plot on the RHS shows that Low Income Laboratory Technicians attrite at a faster rate over time than the medium and high income ones. **In 1 year, 10% of the low income Laboratoty Technicians leave the company.**





The hazard for Low income LTs decreases sharply after year 1. For medium income LTs, the hazard increases after year 1 and after year 2 it decreases and thereafter decreases gradually. For high income LTs the hazard decreases after year 1 and thereafter remains constant till year 5 and then increases sharply till year 10 and then decreases sharply till year 15.

JOB ROLE- "RESEARCH SCIENTIST" & INCOME GROUP

Since Research Scientist is one of the largest job roles (19.86% of employees) in the company, we decided to look at the impact of income on their survival!

Summary of the Number of Censored and Uncensored Values

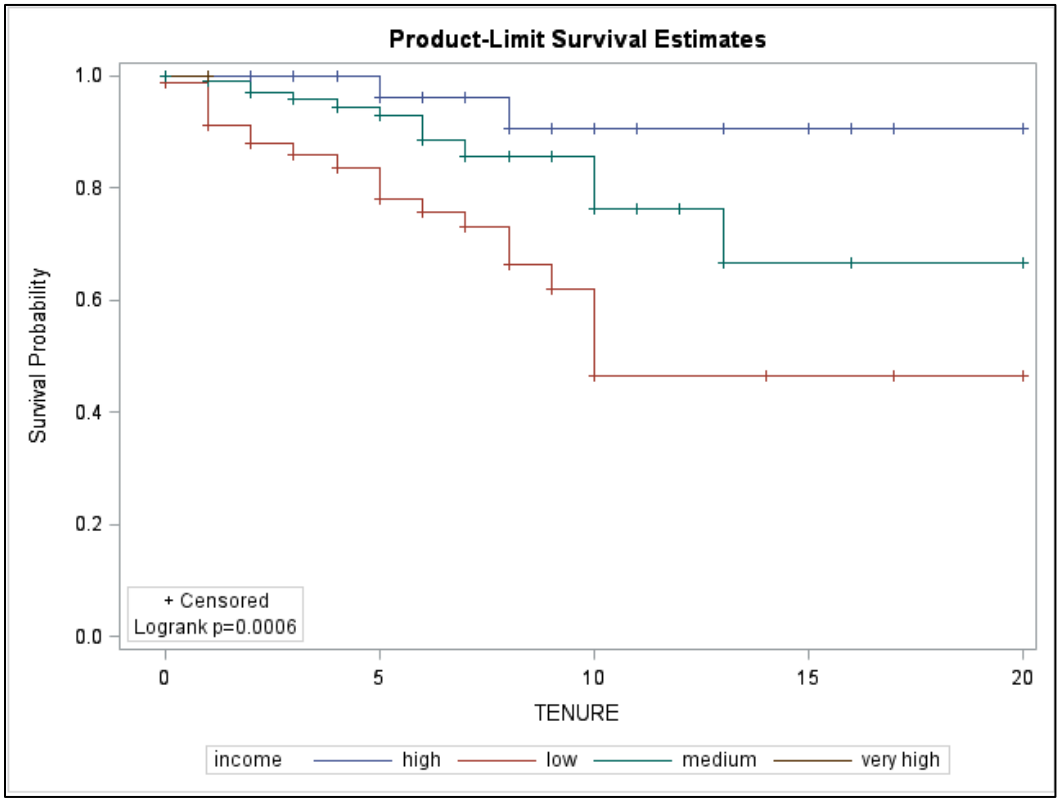
Stratum	income	Total	Failed	Censored	Percent Censored
1	high	30	2	28	93.33
2	low	152	33	119	78.29
3	medium	109	12	97	88.99
4	very high	1	0	1	100.00
Total		292	47	245	83.90

The summary table on the LHS shows that low income group among the Research Scientists have the highest attrition as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.** Also **around 86% of them belong to the low& medium income groups**.

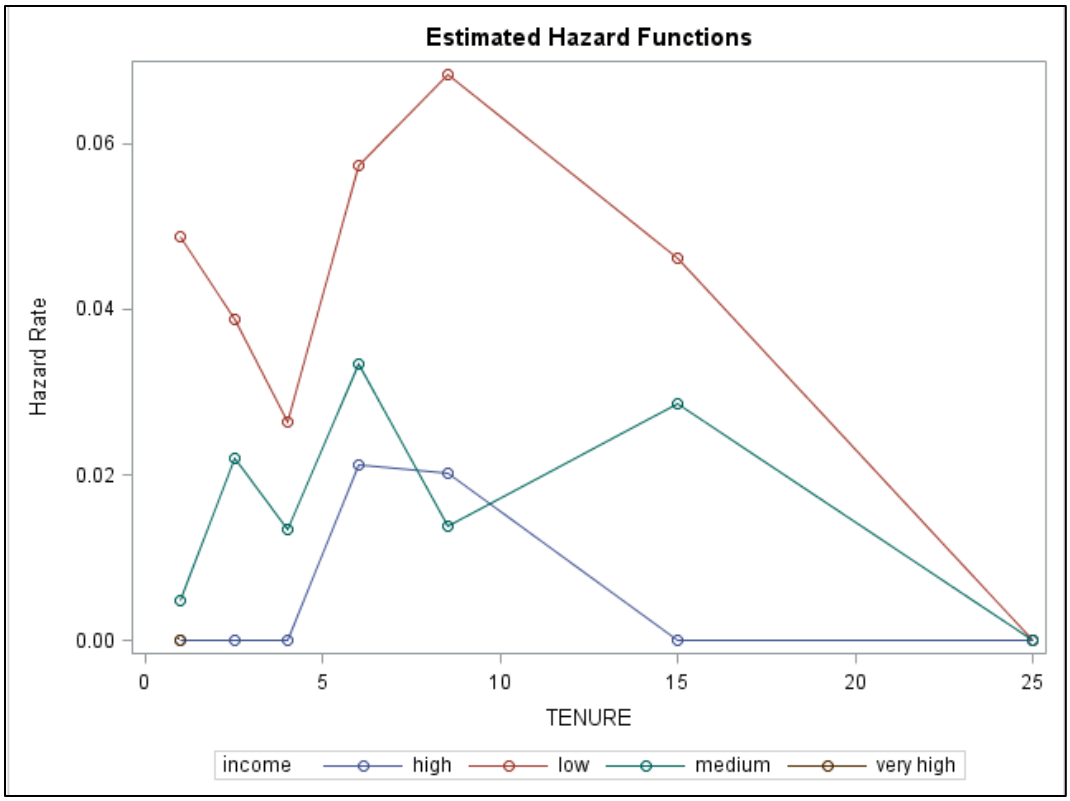
The probability values of the Chi square statistics for all the three tests as shown on the RHS is significant <=0.05 at 5% level of significance. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is different across the income groups!**

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	17.4400	3	0.0006
Wilcoxon	15.6159	3	0.0014
-2Log(LR)*	17.9105	3	0.0005



The survival plot on the LHS shows that Low Income Research Scientists attrite at a faster rate over time than the medium and high income ones**. <u>In 1 year, 10% of the low income Research Scientists leave the company.</u>**



The hazard for low income Research Scientists decrease till year 5 and then picks up till year 10 after which it starts to decline again. The attrition risk for high income RSs is 0 constant till year 5 after which it increases and thereafter declines. **Low income RS's have the highest hazard and high income RS's have 0 hazard for the initial 5 years!**

JOB ROLE- "SALES EXECUTIVE" & INCOME GROUP

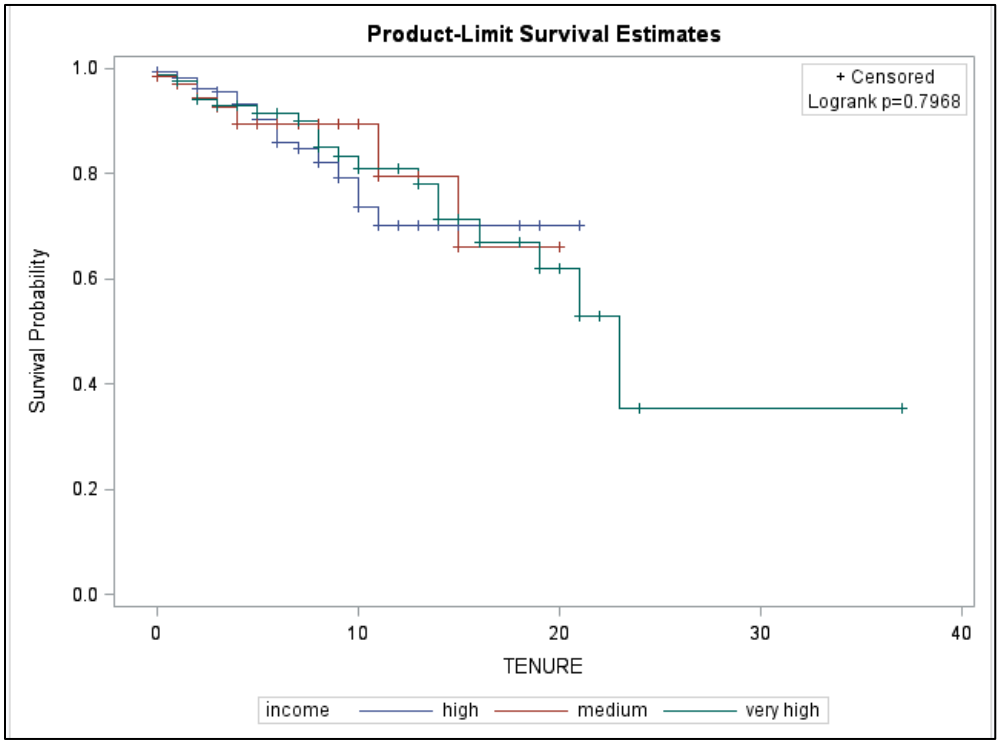
Since Sales Executive is the largest job role (22.18% of employees) in the company, we decided to look at the impact of income on their survival!

Summary of the Number of Censored and Uncensored Values					
Stratum	income	Total	Failed	Censored	Percent Censored
1	high	165	28	137	83.03
2	medium	73	9	64	87.67
3	very high	88	20	68	77.27
Total		326	57	269	82.52

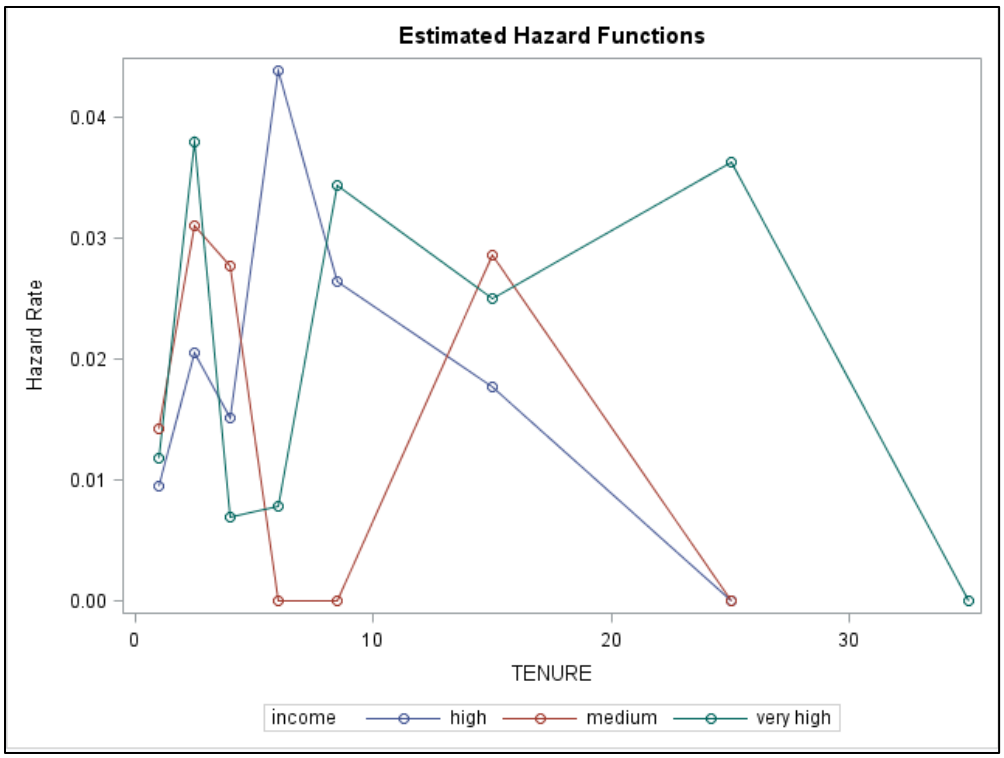
The summary table on the LHS shows that very high income group among the Sales executives have the highest attrition as inferred from the "*Percent censored*" column. **Lower value of percent censored implies higher attrition.**

The probability values of the Chi square statistics for all the three tests as shown on the RHS is insignificant >=0.05 at 5% level of significance. Here the Null hypothesis is that survival curve is similar over strata and the Alternate Hypothesis is that it is NOT! Since the p-value is significant, we can reject the Null hypothesis and **safely conclude that the survival curve is NOT different across the income groups!**

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.4544	2	0.7968
Wilcoxon	0.2375	2	0.8880
-2Log(LR)	0.2815	2	0.8687



The survival plot also shows that attrition pattern is not significantly different among the income groups!



Hazard rate is highest for very high income group at year 4 and decreases sharply thereafter. Again increases sharply after year 8. **High income SEs & Very high income SEs have the lowest & highest hazard rate respectively for the initial 4 years.**