ACM 661 Harsha Kankanamge PROJECT 1

The data set WHAS500.SAV represents a sample of 500 subjects admitted to hospitals with acute myocardial infarction (MI) in Worcester, Massachusetts. Using the variables AGE (age at hospital admission), LENFOL (total length of follow-up - days between date of last follow-up and hospital admission date) and FSTAT (vital status at last follow-up),

1. Conduct a descriptive analysis of the selected variables in the data set and summarize the results.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
age at hospital admission	500	30	104	69.85	14.491
initial heart rate in beats per minute	500	35	186	87.02	23.586
initial systolic blood pressure in mm/Hg	500	57	244	144.70	32.295
body mass index	500	13.05	44.84	26.6138	5.40566
total length of follow-up	500	1	2358	882.44	705.665
Valid N (listwise)	500				

Mean and Std. deviation of variable “age at hospital admission” are respectively 69.85 and 14.491. As well as it has maximum 104 and minimum 30.

Mean and Std. deviation of variable “initial heart rate in beats per minute” are respectively 87.02 per minutes and 23.586. As well as it has maximum 186 per minutes and minimum 35 per minutes.

Mean and Std. deviation of variable “initial systolic blood pressure in mm/Hg” are respectively 144.70 and 32.295. As well as it has maximum 244 and minimum 57.

Mean and Std. deviation of variable “body mass index” are respectively 26.6138 and 5.40566. As well as it has maximum 44.84 and minimum 13.05.

Mean and Std. deviation of variable “total length of follow-up” are respectively 882.44 and 705.665. As well as it has maximum 2358 and minimum 1.

vital status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Alive	285	57.0	57.0	57.0
	Dead	215	43.0	43.0	100.0
	Total	500	100.0	100.0	

In the sample 285 people are alive out of 500 people when fallow up time end. In general, 57% people are alive when fallow up time end. 43% people are dead when fallow up time end. Number of survival people are higher than dead people when fallow up time end. There are not missing values in this column.

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	300	60.0	60.0	60.0
	female	200	40.0	40.0	100.0
	Total	500	100.0	100.0	

In the sample 300 people are males out of 500 people. In general, 60% of the entire dataset represent males. 40% represent the females in the dataset. There are not missing values in this column.

2. Recode AGE into a new variable with four groups including the following age intervals: 30-59, 60-69, 70-79, 80 and above. Use total length of follow-up in months and do a descriptive analysis of the new variables. Display time intervals 0 through 80 by 5.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
follow-up time in months	500	.03	78.60	29.4145	23.52217
Valid N (listwise)	500				

Mean and Std. deviation of variable “follow-up time in months” are respectively 29.4145 and 23.52217. As well as it has maximum 78.60 and minimum 0.03.

Age_group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	30-59	138	27.6	27.6	27.6
	60-69	86	17.2	17.2	44.8
	70-79	114	22.8	22.8	67.6
	80+	162	32.4	32.4	100.0
	Total	500	100.0	100.0	

In the dataset 27.6% represent 30-59 age group, 17.2% represent 60-69 age group, 22.8% represent 70-79 age group and 32.4% represent 80+ age group. There is no missing value. In the data set majority of people are in the 80+ age group. Lowest proportion in 60-69 age group.

3. Conduct a life table analysis with the whole group of 500 subjects. Estimate and interpret the median survival time. Graph the estimated survival function. Describe the survival distribution.

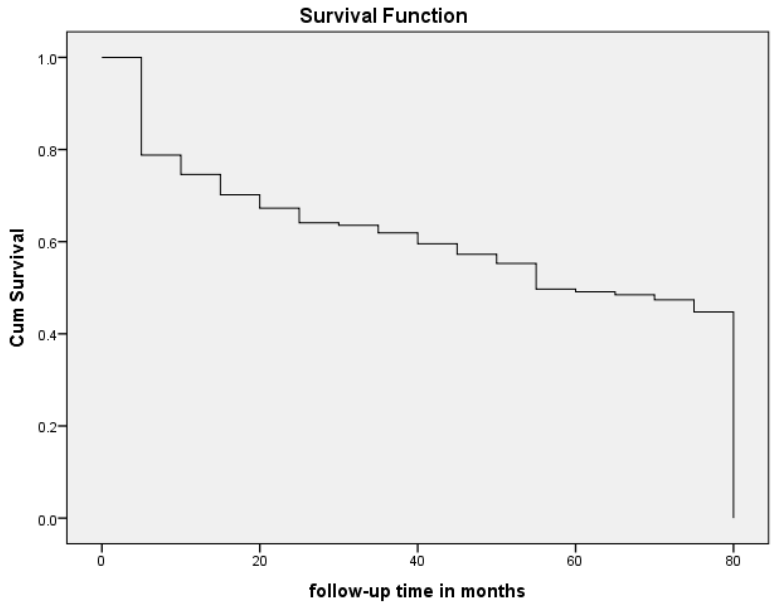
Life Table[a]

Interval Start Time	Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	Cumulative Proportion Surviving at End of Interval	Std. Error of Cumulative Proportion Surviving at End of Interval	Probability Density	Std. Error of Probability Density	Hazard Rate	Std. Error of Hazard Rate
0	500	0	500.000	106	.21	.79	.79	.02	.042	.004	.05	.00
5	394	0	394.000	21	.05	.95	.75	.02	.008	.002	.01	.00
10	373	38	354.000	21	.06	.94	.70	.02	.009	.002	.01	.00
15	314	46	291.000	12	.04	.96	.67	.02	.006	.002	.01	.00
20	256	8	252.000	12	.05	.95	.64	.02	.006	.002	.01	.00
25	236	0	236.000	2	.01	.99	.64	.02	.001	.001	.00	.00
30	234	0	234.000	6	.03	.97	.62	.02	.003	.001	.01	.00
35	228	39	208.500	8	.04	.96	.60	.02	.005	.002	.01	.00
40	181	47	157.500	6	.04	.96	.57	.02	.005	.002	.01	.00
45	128	25	115.500	4	.03	.97	.55	.03	.004	.002	.01	.00
50	99	0	99.000	10	.10	.90	.50	.03	.011	.003	.02	.01
55	89	0	89.000	1	.01	.99	.49	.03	.001	.001	.00	.00
60	88	30	73.000	1	.01	.99	.48	.03	.001	.001	.00	.00
65	57	24	45.000	1	.02	.98	.47	.03	.002	.002	.00	.00
70	32	28	18.000	1	.06	.94	.45	.04	.005	.005	.01	.01
75	3	0	3.000	3	1.00	.00	.00	.00	.090	.008	.40	.00

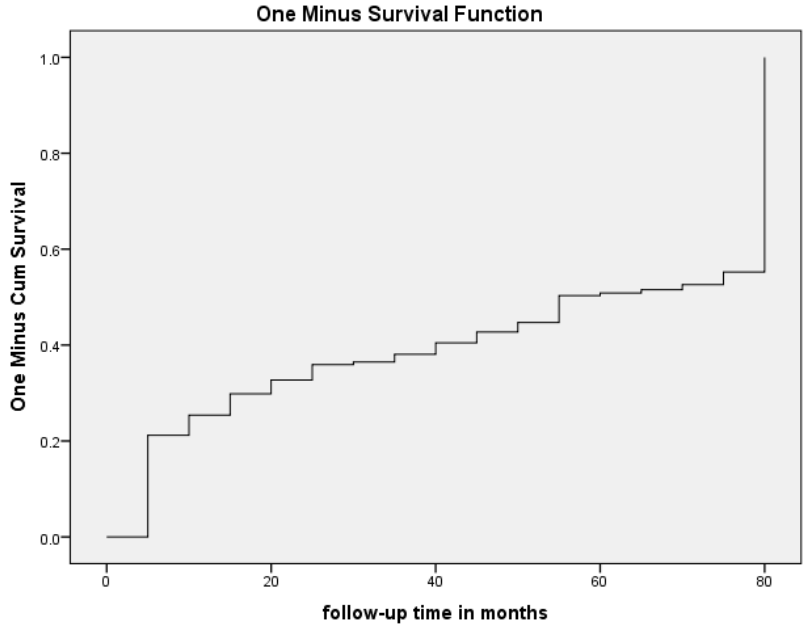
a. The median survival time is 54.7269

Estimated median survival time: 54.7269

Median survival is a statistic that refers to how long patients survive with a acute myocardial infarction. When 54.7269 month pass half of the patients (cumulative) are expected to be alive (50% of people estimated to be dead by 54.7269 month). It means that the chance of surviving beyond 54.7269-month is 50 percent. It gives an approximate indication of survival as well as the prognosis of a group of patients with myocardial infarction.



Within first five month of observation period about 20% of patients were dead. Which is the highest proportion of termination within the 5 month periods of time for entire observation period. As well as 45 to 50-month period about 10% of patients were dead. Cumulative survival function is keep decreasing with observation time.



4. Repeat the life table analysis with the four age intervals.

a)Test the overall survival and pairwise survivals.

Overall Comparisons[a]

Wilcoxon (Gehan) Statistic	df	Sig.
102.741	3	.000

a. Comparisons are exact.

H0: $S_1(t)=S_2(t)=S_3(t)=S_4(t)$

H1: Two or more survival functions are different.

Willcoxon (Gehan) Statistic = 102.741 and p-value = 0.000

p-value (0.000)<0.05 Reject H0

At least one $S_i(t)$ (probability that an individual survives beyond time 't') is significantly different in four age groups. Which means two or more survival functions are significantly different.

pairwise survivals

Pairwise Comparisons[a]

(I) Age_group	(J) Age_group	Wilcoxon (Gehan) Statistic	df	Sig.
1	2	2.825	1	.093
	3	40.620	1	.000
	4	81.124	1	.000
2	1	2.825	1	.093
	3	15.942	1	.000
	4	40.997	1	.000
3	1	40.620	1	.000
	2	15.942	1	.000
	4	8.940	1	.003
4	1	81.124	1	.000
	2	40.997	1	.000
	3	8.940	1	.003

a. Comparisons are exact.

H0: Si(t)=Sj(t) I,j = 1,2,3,4

H1: Si(t)≠Sj(t) i≠j

Age Group 30-59 and 60-69

Willcoxon (Gehan) Statistic = 2.825 and p-value = 0.093

p-value (0.093)>0.05 Do not reject H0 (borderline significant)

Survival functions are not significantly (borderline significant) different from each other in Age Group 30-59 and 60-69.

In between all other age groups

Survival functions are significantly different from each other in between all other age groups.

p-values< 0.05 Reject H0

b)Test whether the survival experience for the middle two age groups is the same or different from the youngest and oldest age groups.

Pairwise Comparisons[a]

(I) Age_group	(J) Age_group	Wilcoxon (Gehan) Statistic	df	Sig.
1	2	2.825	1	.093
	3	40.620	1	.000
	4	81.124	1	.000
2	1	2.825	1	.093
	3	15.942	1	.000
	4	40.997	1	.000
3	1	40.620	1	.000
	2	15.942	1	.000
	4	8.940	1	.003
4	1	81.124	1	.000
	2	40.997	1	.000
	3	8.940	1	.003

a. Comparisons are exact.

H0: S2(t) = S1(t)

H1: S2(t) ≠ S1 (t)

Willcoxon (Gehan) Statistic = 2.825 and p-value = 0.093

p-value (0.093)>0.05 Do not reject H0 (borderline significant)

Survival functions are not significantly different from each other in Age Group 30-59 and 60-69.

H0: S2(t) = S4(t)

H1: S2(t) ≠ S4 (t)

Willcoxon (Gehan) Statistic = 40.997 and p-value = 0.000

p-value (0.000) < 0.05 Reject H0

Survival functions are significantly different from each other in Age Group 60-69 and 80+.

H0: S3(t) = S1(t)

H1: S3(t) ≠ S1(t)

Willcoxon (Gehan) Statistic = 40.620 and p-value = 0.000

p-value (0.000) < 0.05 Reject H0

Survival functions are significantly different from each other in Age Group 70-79 and 30-59.

H0: S3(t) = S4(t)

H1: S3(t) ≠ S4(t)

Willcoxon (Gehan) Statistic = 8.94 and p-value = 0.003

p-value (0.003) < 0.05 Reject H0

$S_i(t)$ (probability that an individual survives beyond time 't') is significantly different from each other in Age Group 70-79 and 80+.

c) Report the number of subjects who had the event in the four age groups.

Age_group * vital status Crosstabulation

Count

		vital status		
		Alive	Dead	Total
Age_group	30-59	118	20	138
	60-69	67	19	86
	70-79	50	64	114
	80+	50	112	162
Total		285	215	500

In age group 30-59, 20 people had the event out off 138 at end of the observation period. Which is comparably less proportion of event occurred.

In age group 60-69, 19 people had the event out of 86 at end of the observation period.

In age group 70-79, 64 people had the event out of 114 at end of the observation period.

In age group 80+, 112 people had the event out of 162 at end of the observation period. These people are at high risk.

People who are in 80+ age group have a highest risk of been bead. As well as people who are in age group 70-79 also have high risk of been dead. People who are in 30-59 and 60-69 age group have a lower risk of been bead.

Age_group * vital status Crosstabulation

			vital status		
			Alive	Dead	Total
Age_group	30-59	Count	118	20	138
		% within Age_group	85.5%	14.5%	100.0%
	60-69	Count	67	19	86
		% within Age_group	77.9%	22.1%	100.0%
	70-79	Count	50	64	114
		% within Age_group	43.9%	56.1%	100.0%
	80+	Count	50	112	162
		% within Age_group	30.9%	69.1%	100.0%
Total		Count	285	215	500
		% within Age_group	57.0%	43.0%	100.0%

D)Estimate and interpret the median survival time of the four groups.

Median Survival Time

First-order Controls		Med Time
Age_group	30-59	76.9204
	60-69	76.5307
	70-79	41.3347
	80+	12.9732

Estimated median survival time for age group 30-59 is: 76.92

At 76.92 month which the value of the cumulative survival function is expected to be 0.5 in the 30-59 age group. It means that the chance of surviving beyond 76.92-month is 50 percent in the 30-59 age group.

Estimated median survival time for age group 60-69 is: 76.53

At 76.53 month which the value of the cumulative survival function is expected to be 0.5 in the 60-69 age group. It means that the estimated chance of surviving beyond 76.53-month is 50 percent in the 60-69 age group.

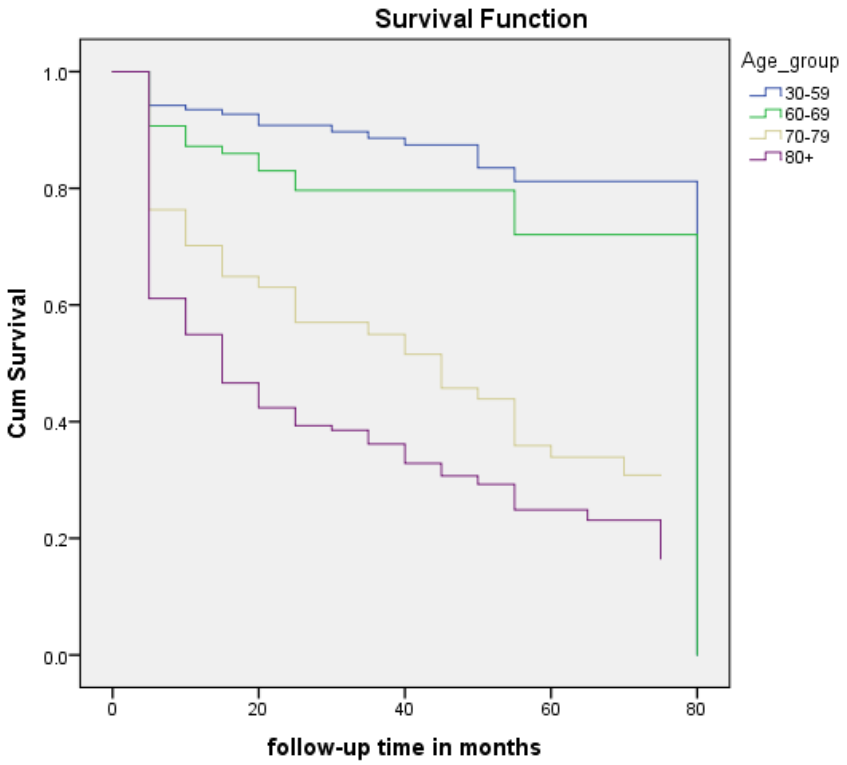
Estimated median survival time for age group 70-79 is: 41.34

At 41.34 month which the value of the cumulative survival function is expected to be 0.5 in the 70-79 age group. It means that the estimated chance of surviving beyond 41.34-month is 50 percent in the 70-79 age group.

Estimated median survival time for age group 80+ is: 12.9732

At 12.9732 month which the value of the cumulative survival function is expected to be 0.5 in the 80+ age group. It means that the estimated chance of surviving beyond 12.9732-month is 50 percent in the 80+ age group.

e) Graph and interpret the estimated survival function of the four groups.



80+ age group

Within first five month of observation period about 40% of patients were dead in 80+ age group. Which is the highest proportion of termination within the 5 month periods of time for entire observation period in the 80+ age group. As well as 80+ group has lowest cumulative proportion of survival time for entire observation period compare with other age groups. Cumulative survival proportion is keep decreasing with observation time.

70-79 age group

Within first five month of observation period about 25% of patients were dead in 70-79 age group. Which is the highest proportion of termination within the 5 month periods of time for entire observation period in the 70-79 age group. As well as 70-79 age group has second lowest cumulative proportion of survival time for entire observation period compare with other age groups. Cumulative survival proportion is keep decreasing with observation time.

60-69 age group

Within first five month of observation period and 50-55 period about 10% of patients were dead in 60-69 age group. Which are the highest proportion of termination within the 5 month periods of time for entire observation period in the 60-69 age group. As well as 60-69 age group has second highest cumulative proportion of survival time for entire observation period compare with other age groups. In observation time period months 25 to 50 and 55 to 80 no deaths are reported in this age group.

30-59 age group

Within first five month of observation period about 6% of patients were dead in 30-59 age group. Which are the highest proportion of termination within the 5 month periods of time for entire observation period in the 30-59 age group. As well as 30-59 age group has highest cumulative proportion of survival time for entire observation period compare with other age groups. In observation time period months 55 to 80 no deaths are reported in this age group.

5. Conduct a Kaplan-Meier survival analysis with the whole group of 500 subjects. Estimate and interpret the quartile survival times. Graph the estimated survival function.

Case Processing Summary

Total N	N of Events	Censored	
		N	Percent
500	215	285	57.0%

Means and Medians for Survival Time

Mean[a]				Median			
		95% Confidence Interval				95% Confidence Interval	
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
47.240	1.605	44.096	50.385	54.233	5.319	43.809	64.658

a. Estimation is limited to the largest survival time if it is censored.

Percentiles

25.0%		50.0%		75.0%	
Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
78.433	2.649	54.233	5.319	9.833	2.107

Q1=25% = 78.433

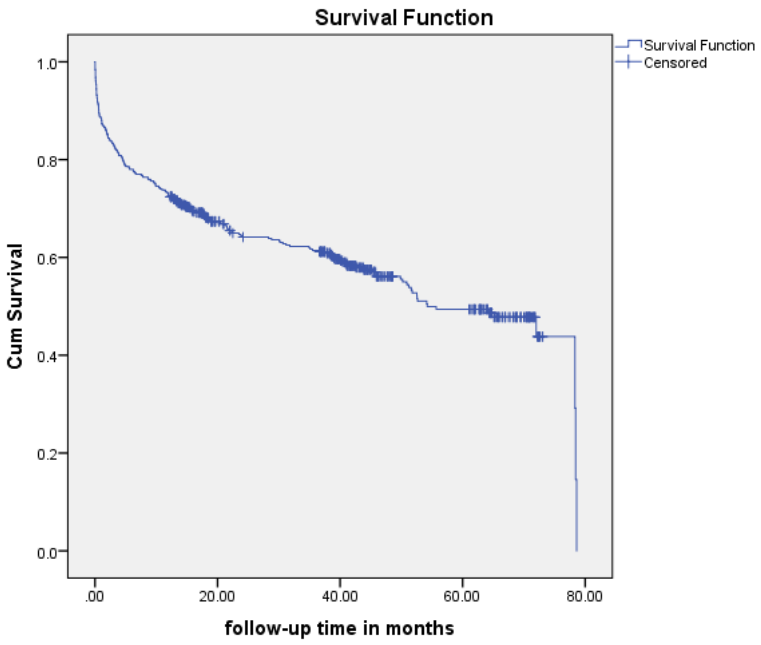
In 78.433 months' cumulative survivals is 25% or less.

Q2=50% = 54.233

In 54.233 months' cumulative survivals is 50% or less.

Q3=75% = 9.833

In 9.833 months' cumulative survivals is 75% or less.



Within first 5 months' cumulative survival proportion decrease by 20%. As well as cumulative survival keep decreasing with time. A lot of patients were withdrawn (censored) from the observation in between 10-25, 35-50 and 60-75 these months.

6)Repeat the Kaplan-Meier survival analysis with the four age intervals.

a)Mention the # and percent of events in each group.

Case Processing Summary

Age_group	Total N	N of Events	Censored	
			N	Percent
30-59	138	20	118	85.5%
60-69	86	19	67	77.9%
70-79	114	64	50	43.9%
80+	162	112	50	30.9%
Overall	500	215	285	57.0%

In age group 30-59, 20 people had the event out off 138 at end of the observation period. Which means 14.5% had the event (dead) at the end of the observation period in this age group. It is comparably lower than other groups.

In age group 60-69, 19 people had the event out of 86 at end of the observation period. Which means 22.1% had the event (dead) at the end of the observation period in this age group

In age group 70-79, 64 people had the event out of 114 at end of the observation period. Which means 56.1% had the event at the end of the observation period in this age group. They at high risk of event.

In age group 80+, 112 people had the event out of 162 at end of the observation period. Which means 69.1% had the event at the end of the observation period in this age group. They have lowest survival percentage.

People who are in 80+ age group have a highest risk of been bead. As well as people who are in age group 70-79 also have high risk of been dead. People who are in 30-59 and 60-69 age group have a lower risk of been bead.

b)Test the overall survival and pairwise survivals using the Log rank Test.

Overall survival using the Log rank Test

H0: $S_1(t)=S_2(t)=S_3(t)=S_4(t)$

H1: Two or more survival functions are different

Percentiles

	25.0%		50.0%		75.0%	
Age_group	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
30-59	78.600	.	78.433	18.350	78.433	18.350
60-69	78.333	.	78.333	.	51.600	4.695
70-79			40.567	6.808	5.533	3.028
80+	64.200	6.161	12.833	2.977	1.633	.662
Overall	78.433	2.649	54.233	5.319	9.833	2.107

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	120.128	3	.000

Test of equality of survival distributions for the different levels of Age_group.

Log Rank (Mantel-Cox) Statistic = 120.128 and p-value = 0.000

p-value (0.000)<0.05 Reject H0

At least one $S_i(t)$ (probability that an individual survives beyond time 't') is significantly different in four age groups. Which means two or more survival functions are significantly different.

pairwise survivals using the Log rank Test

Percentiles

	25.0%		50.0%		75.0%	
Age_group	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
30-59	78.600	.	78.433	18.350	78.433	18.350
60-69	78.333	.	78.333	.	51.600	4.695
70-79			40.567	6.808	5.533	3.028
80+	64.200	6.161	12.833	2.977	1.633	.662
Overall	78.433	2.649	54.233	5.319	9.833	2.107

Pairwise Comparisons

		30-59		60-69		70-79		80+	
	Age_group	Chi-Square	Sig.	Chi-Square	Sig.	Chi-Square	Sig.	Chi-Square	Sig.
Log Rank (Mantel-Cox)	30-59			3.187	.074	47.969	.000	93.885	.000
	60-69	3.187	.074			19.639	.000	45.327	.000
	70-79	47.969	.000	19.639	.000			7.955	.005
	80+	93.885	.000	45.327	.000	7.955	.005		

H0: Si(t)=Sj(t) I,j = 1,2,3,4

H1: Si(t)≠Sj(t) i≠j

Age Group 30-59 and 60-69

Chi-Square Statistic = 3.187 and p-value = 0.074

p-value (0.074)>0.05 Do not reject H0 (borderline significant)

$S_i(t)$ (probability that an individual survives beyond time 't') is not significantly different from each other in Age Group 30-59 and 60-69. In general survival experience are borderline significant in between these groups.

In between all other age groups

Survival functions are significantly different from each other in between all other age groups. In general survival experience are significantly different from each other in between all other groups.

p-values< 0.05 Reject H0

c)Test whether the survival experience for the middle two age groups is the same or different from the youngest and oldest age groups.

Pairwise Comparisons

		30-59		60-69		70-79		80+	
	Age_group	Chi-Square	Sig.	Chi-Square	Sig.	Chi-Square	Sig.	Chi-Square	Sig.
Log Rank (Mantel-Cox)	30-59			3.187	.074	47.969	.000	93.885	.000
	60-69	3.187	.074			19.639	.000	45.327	.000
	70-79	47.969	.000	19.639	.000			7.955	.005
	80+	93.885	.000	45.327	.000	7.955	.005		

H0: S2(t) = S1(t)

H1: S2(t) ≠ S1 (t)

Chi-Square test Statistic = 3.187 and p-value = 0.074

p-value (0.074)>0.05 Do not reject H0 (borderline significant)

Survival functions are borderline significantly different from each other in Age Group 30-59 and 60-69. In general survival experience are borderline significant in between these groups.

H0: S2(t) = S4(t)

H1: S2(t) ≠ S4 (t)

Chi-Square test Statistic = 45.327 and p-value = 0.000

p-value (0.000) < 0.05 Reject H0

Survival functions are significantly different from each other in Age Group 60-69 and 80+. Survival experience are significantly different in between these groups.

H0: S3(t) = S1(t)

H1: S3(t) ≠ S1(t)

Chi-Square test Statistic = 47.969 and p-value = 0.000

p-value (0.000) < 0.05 Reject H0

Survival functions are significantly different from each other in Age Group 70-79 and 30-59. Survival experience are significantly different in between these groups.

H0: S3(t) = S4(t)

H1: S3(t) ≠ S4(t)

Chi-Square test Statistic = 7.955 and p-value = 0.005

p-value (0.005) < 0.05 Reject H0

Survival functions are significantly different from each other in Age Group 70-79 and 80+. Survival experience are significantly different in between these groups.

d)Estimate and interpret the median survival times of the four groups.

Means and Medians for Survival Time

	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
Age_group	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
30-59	68.509	2.241	64.117	72.901	78.433	18.350	42.468	114.399
60-69	62.171	3.452	55.405	68.937	78.333	.000	.	.
70-79	38.298	2.901	32.613	43.984	40.567	6.808	27.222	53.911
80+	27.431	2.388	22.752	32.111	12.833	2.977	6.999	18.668
Overall	47.240	1.605	44.096	50.385	54.233	5.319	43.809	64.658

a. Estimation is limited to the largest survival time if it is censored.

Estimated median survival time for age group 30-59 is: 78.433

At 78.433 month which the value of the cumulative survival function is expected to be 0.5 in the 30-59 age group. It means that the estimated chance of surviving beyond 78.433-month is 50 percent or less in the 30-59 age group.

Estimated median survival time for age group 60-69 is: 78.333

When 78.333 month pass half the patients are expected to be alive in the 60-69 age group. It means that the estimated chance of surviving beyond 78.333-month is 50 percent or less in the 60-69 age group.

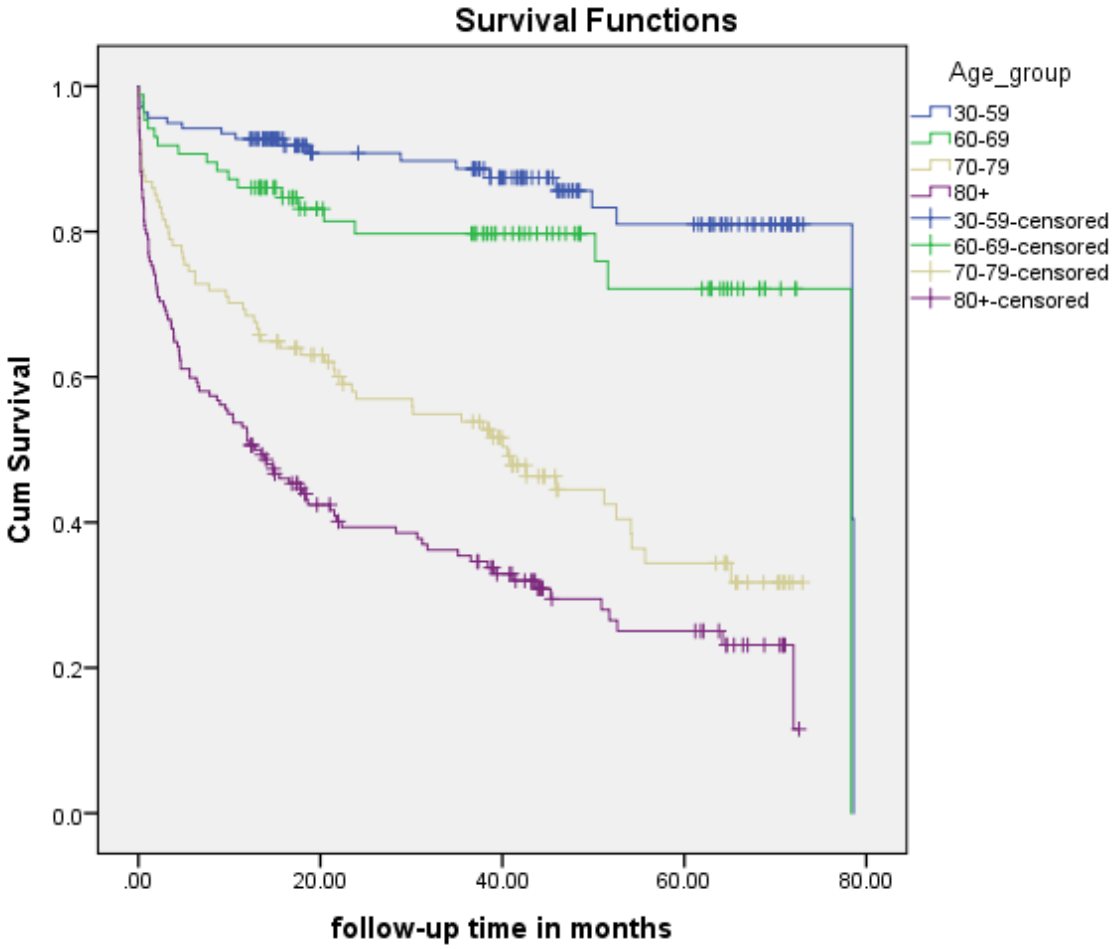
Estimated median survival time for age group 70-79 is: 40.567

When 40.567month pass half the patients are expected to be alive in the 70-79 age group. It means that the estimated chance of surviving beyond 40.567-month is 50 percent or less in the 70-79 age group.

Estimated median survival time for age group 80+ is: 12.833

When 12.833 month pass half the patients are expected to be alive in the 80+ age group. It means that the estimated chance of surviving beyond 12.833-month is 50 percent or less in the 80+ age group.

e) Graph and interpret the estimated survival function of the four groups.



80+ age group

Within first five month of observation period about 40% of patients were experienced the event (dead) in 80+ age group. Which is the highest proportion of termination within the 5 month periods of time for entire observation period in the 80+ age group. As well as 80+ group has lowest cumulative proportion of survival time for entire observation period compare with other age groups. Cumulative survival proportion is keep decreasing with observation time. A lot of patients were withdrawn (Censored) from the observation in around 15, 40, 65 months. Highest Censored amount recoded 15-25 and 35-45 months in this age group.

70-79 age group

Within first five month of observation period about 25% of patients were dead in 70-79 age group. Which is the highest proportion of termination within the 5 month periods of time for entire observation period in the 70-79 age group. As well as 70-79 age group has second lowest

cumulative proportion of survival time for entire observation period compare with other age groups. Cumulative survival proportion is keep decreasing with observation time. A lot of patients were withdrawn (Censored) from the observation in around 10-25, 35-45, 65-75 months.

60-69 age group

Within first five month of observation period and 50-55 period about 10% of patients were dead in 60-69 age group. Which are the highest proportion of termination within the 5 month periods of time for entire observation period in the 60-69 age group. As well as 60-69 age group has second highest cumulative proportion of survival time for entire observation period compare with other age groups. In observation time period months 25 to 50 and 55 to 80 no deaths are reported in this age group. Patients were withdrawn (Censored) from the observation in around 10-25, 35-45, 60-75 period of months

30-59 age group

Within first five month of observation period about 6% of patients were dead in 30-59 age group. Which are the highest proportion of termination within the 5 month periods of time for entire observation period in the 30-59 age group. As well as 30-59 age group has highest cumulative proportion of survival time for entire observation period compare with other age groups. In observation time period months 55 to 80 no deaths are reported in this age group. Patients were withdrawn (Censored) from the observation in around 10-20, 35-45, 60-75 period of months.