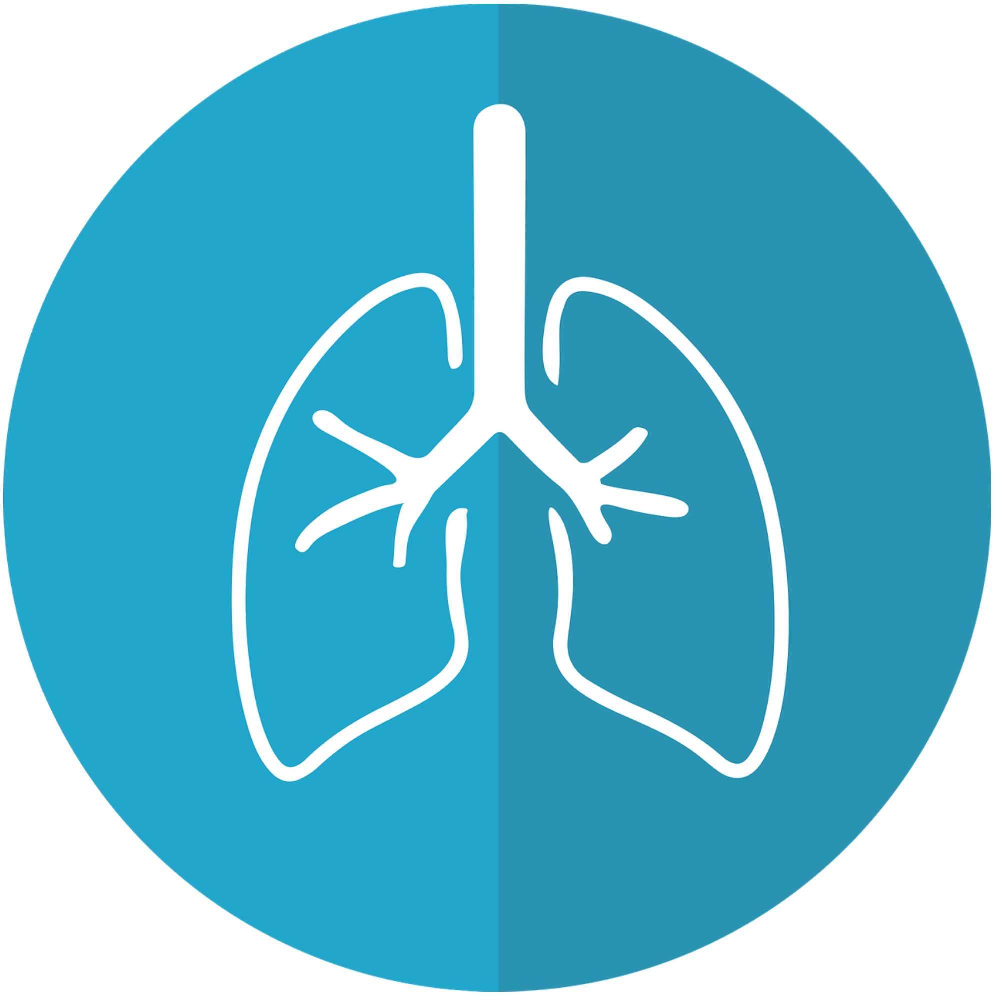


Survival Analysis Project

Lung Cancer Dataset

Dhruv Aggarwal (2033102)

Bsc (H) Statistics

Sem - VI

Biostatistics and Survival Analysis

Exploring Dataset

Our study consists of a dataset of 133 patients who have received a certain treatment for Lung Cancer. Each row represents a patient, and the columns provide various pieces of information about the patient, such as their cancer type, survival time in days, status (whether they survived or not), Karnofsky Score, months from diagnosis, age, and prior therapy.

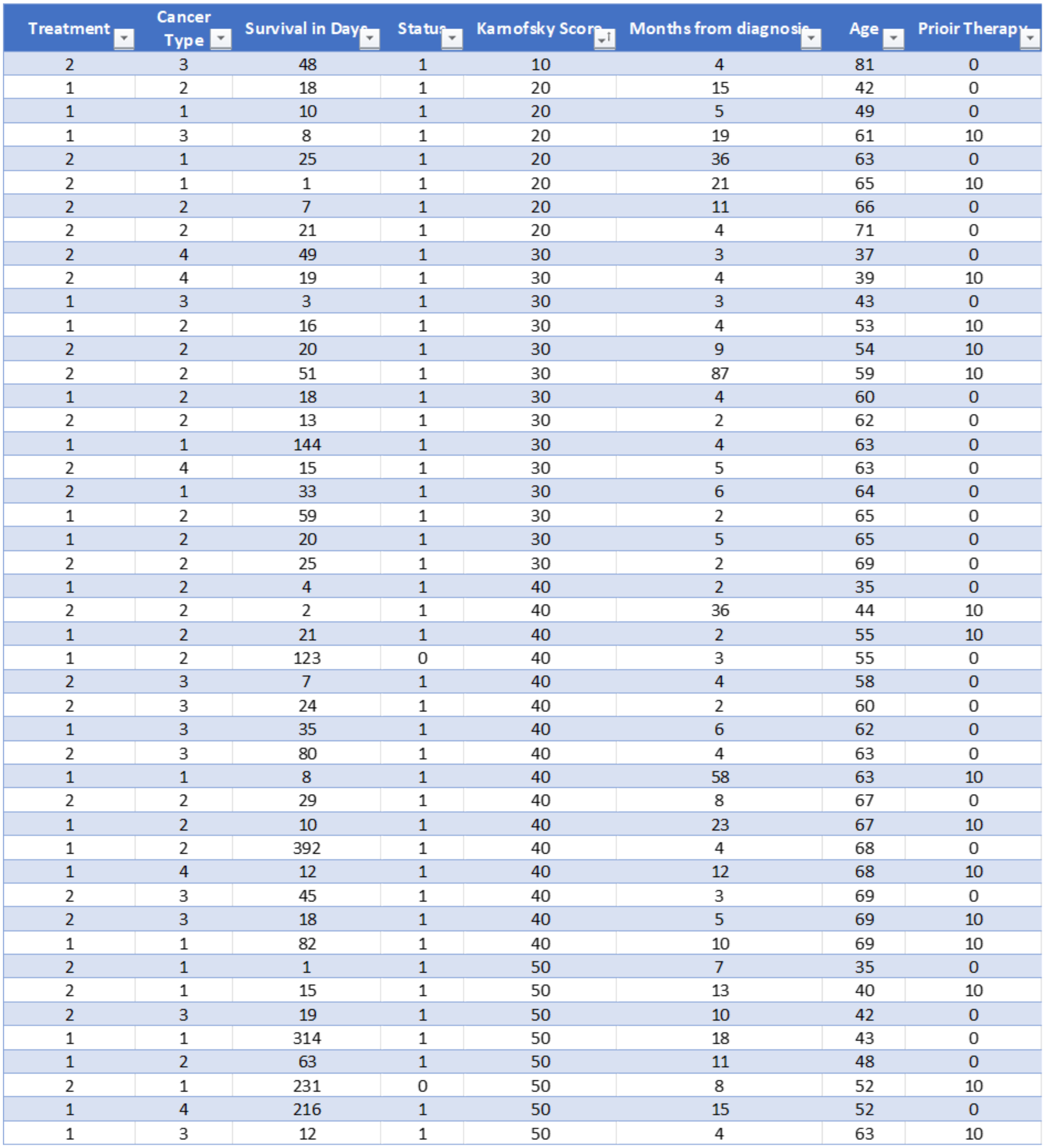
- Treatment: It indicates which treatment is received by the patient (1) or (2)
- Cancer Type: It indicates the type of lung cancer from which patient is suffering (1) = Squamous, (2) = Small Cells, (3) = Adeno Cells, (4) = Large Cells
- Survival in days: the number of days the patient survived after receiving the treatment.
- Status: whether the patient survived (1) or not (0)
- Karnofsky score: a measure of the patient's ability to perform normal activities of daily living.
- Months from diagnosis: the number of months that have passed since the patient was diagnosed with their condition.
- Age: the patient's age at the time of treatment
- Prior therapy: whether the patient received any prior therapy before receiving the current treatment (0 for no prior therapy, 10 for prior therapy)

Objectives

Major Objectives of our study is

- To Estimate the survival function
- Assess the presence of an underlying known probability distribution for the survival time data.
- Compare the severity of various types of lung cancer.
- Compare the efficacy of two treatments
- Investigating the impact of the therapy sessions on the survival time
- To investigate impact of time between treatment and diagnosis on the survival time
- Efficacy of the treatment against each cancer type
- To study impact of Age and Karnofsky Score on Survival Time

Dataset



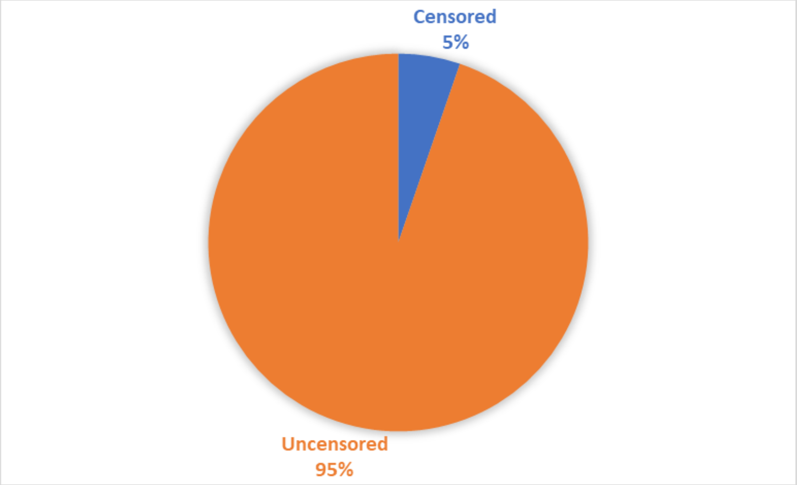
Treatment	Cancer Type	Survival in Days	Status	Kamofsky Score	Months from diagnosis	Age	Prioir Therapy
2	3	48	1	10	4	81	0
1	2	18	1	20	15	42	0
1	1	10	1	20	5	49	0
1	3	8	1	20	19	61	10
2	1	25	1	20	36	63	0
2	1	1	1	20	21	65	10
2	2	7	1	20	11	66	0
2	2	21	1	20	4	71	0
2	4	49	1	30	3	37	0
2	4	19	1	30	4	39	10
1	3	3	1	30	3	43	0
1	2	16	1	30	4	53	10
2	2	20	1	30	9	54	10
2	2	51	1	30	87	59	10
1	2	18	1	30	4	60	0
2	2	13	1	30	2	62	0
1	1	144	1	30	4	63	0
2	4	15	1	30	5	63	0
2	1	33	1	30	6	64	0
1	2	59	1	30	2	65	0
1	2	20	1	30	5	65	0
2	2	25	1	30	2	69	0
1	2	4	1	40	2	35	0
2	2	2	1	40	36	44	10
1	2	21	1	40	2	55	10
1	2	123	0	40	3	55	0
2	3	7	1	40	4	58	0
2	3	24	1	40	2	60	0
1	3	35	1	40	6	62	0
2	3	80	1	40	4	63	0
1	1	8	1	40	58	63	10
2	2	29	1	40	8	67	0
1	2	10	1	40	23	67	10
1	2	392	1	40	4	68	0
1	4	12	1	40	12	68	10
2	3	45	1	40	3	69	0
2	3	18	1	40	5	69	10
1	1	82	1	40	10	69	10
2	1	1	1	50	7	35	0
2	1	15	1	50	13	40	10
2	3	19	1	50	10	42	0
1	1	314	1	50	18	43	0
1	2	63	1	50	11	48	0
2	1	231	0	50	8	52	10
1	4	216	1	50	15	52	0
1	3	12	1	50	4	63	10

Source : http://lib.stat.cmu.edu/datasets/veteran

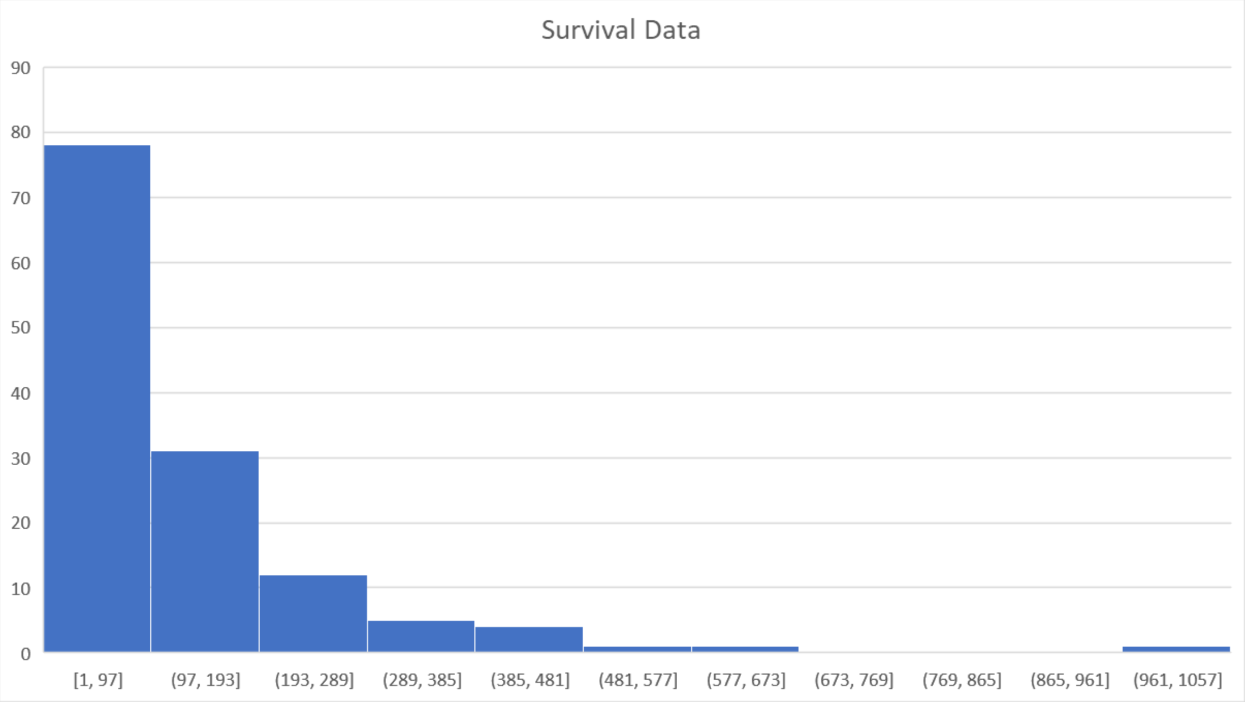
Analysis

- ## Censored Vs Uncensored Observation

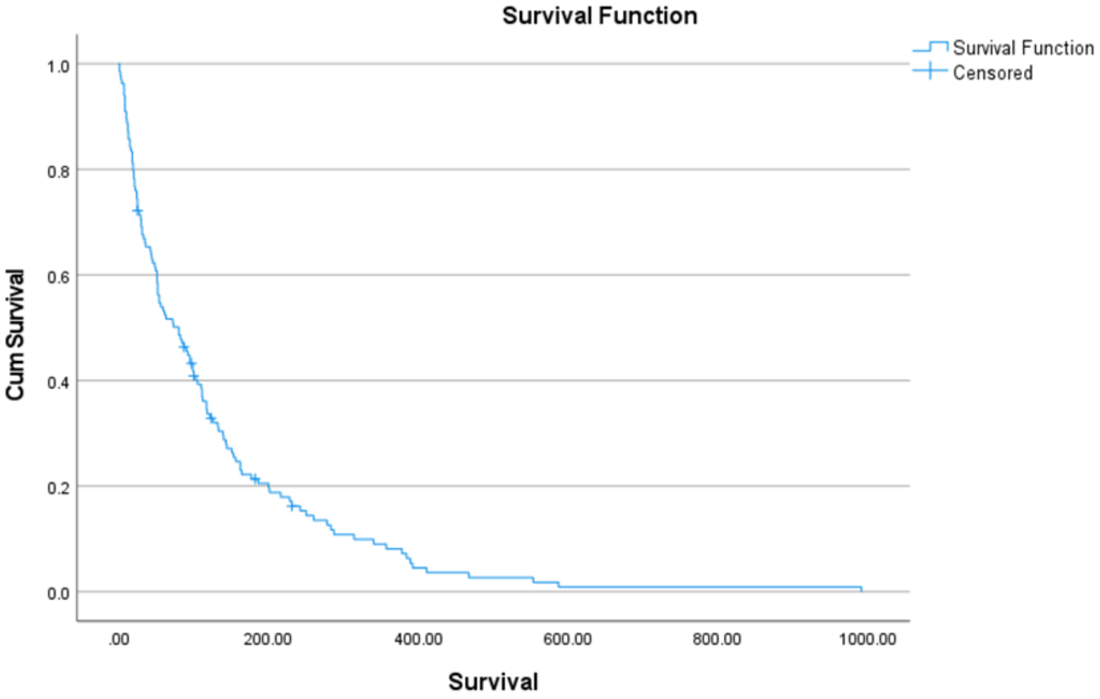
The given data is randomly progressive Type I Censored Data

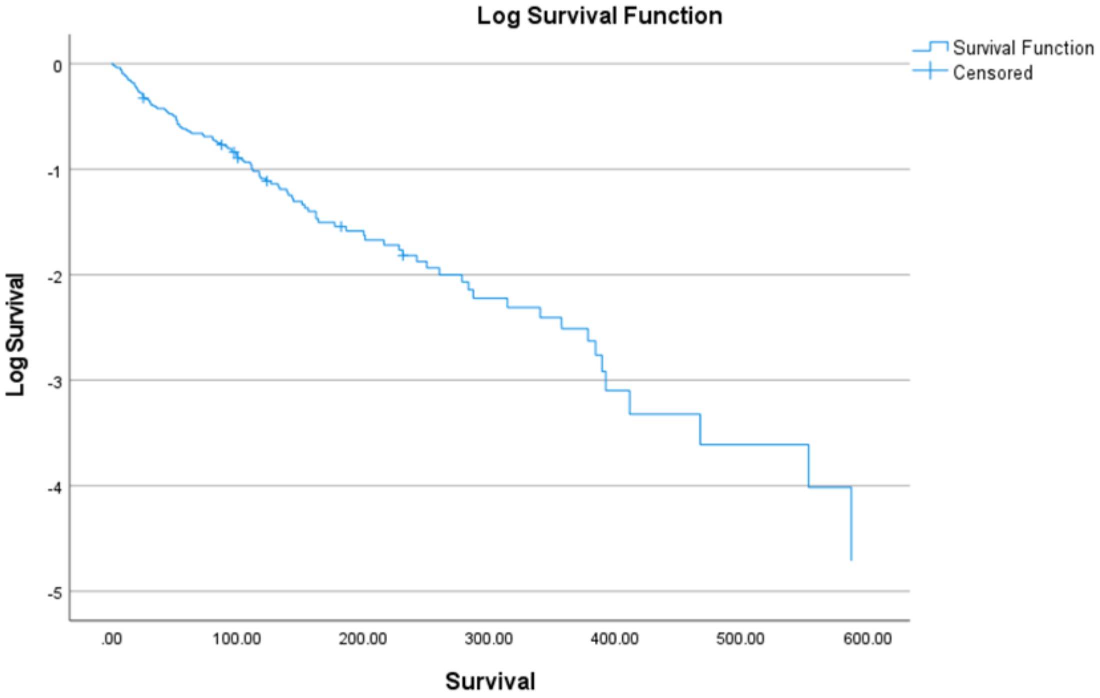


- ## Histogram of Survival Data



- **Estimated Survival Function using kaplan meier**





From above graphs we can indicate that most of the deaths occur in first 100 days of the treatment received

Underlying Probability Distribution

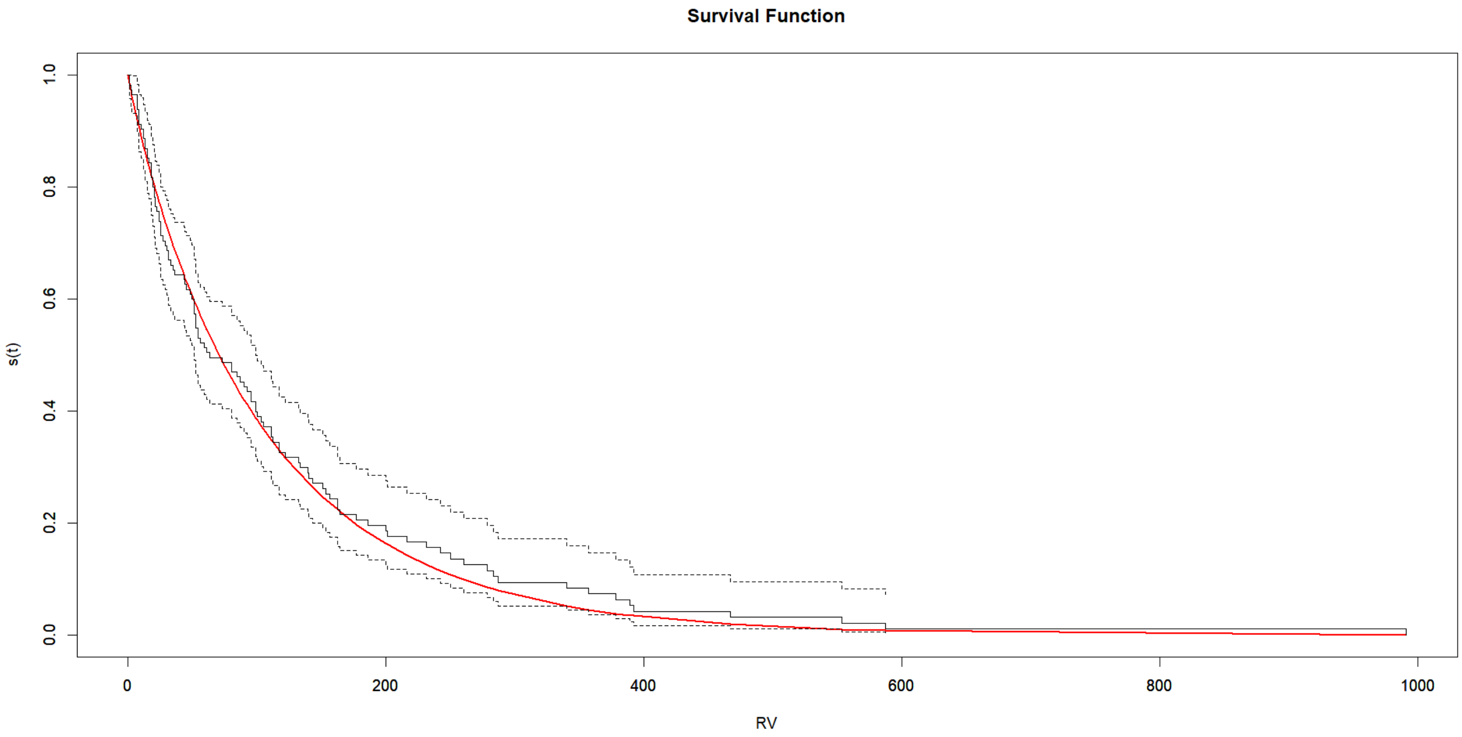
From the Above graphs we can infer that log of survival function is not completely a straight line so Weibull Distribution will be good to handle Non Linear Log of Survival Functions:

Estimated Model Parameters For Weibull Distribution

Shape : .923

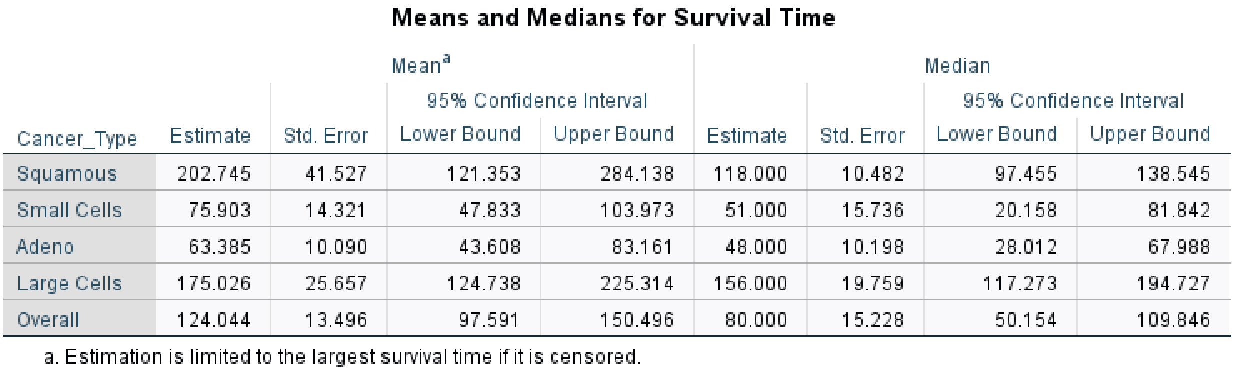
Scale : 104.871

Comparing the Theoretical and Empirical Survival Function



So from the above graph we can infer that since our Estimated Survival Function using Kaplan Meier Method and Theoretical Survival function are overlapping over each other we can say that our Survival Time follows Weibull (Shape=.923 , Scale=104.847).

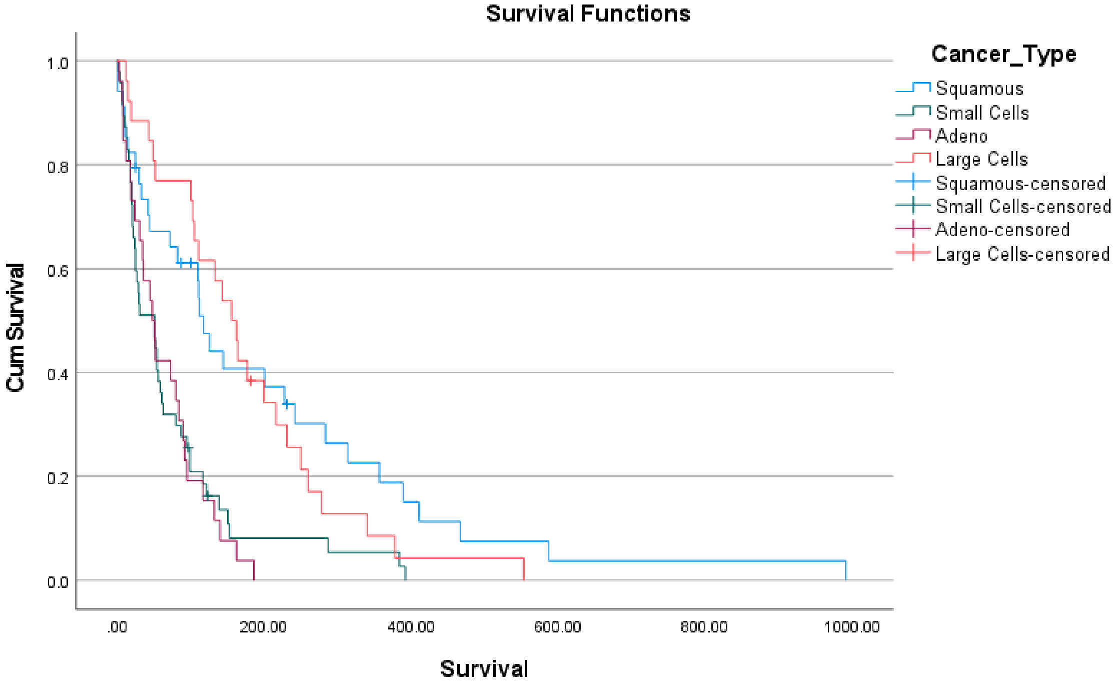
Comparing Severity of Different Type of Lung Cancers



Means and Medians for Survival Time

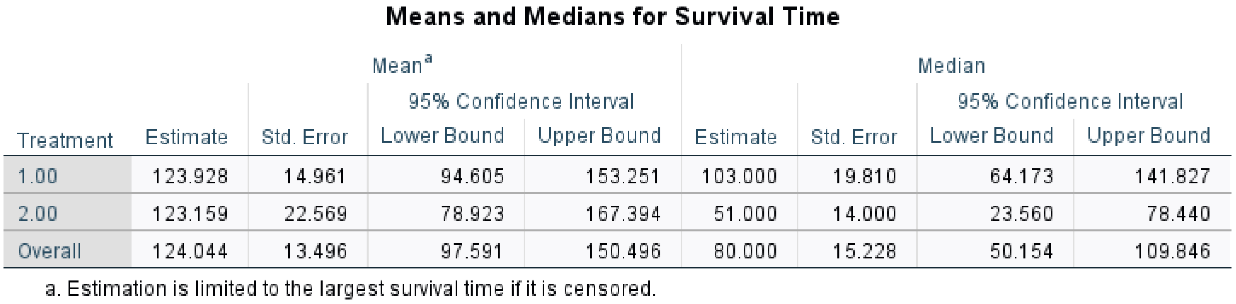
	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
Cancer_Type	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
Squamous	202.745	41.527	121.353	284.138	118.000	10.482	97.455	138.545
Small Cells	75.903	14.321	47.833	103.973	51.000	15.736	20.158	81.842
Adeno	63.385	10.090	43.608	83.161	48.000	10.198	28.012	67.988
Large Cells	175.026	25.657	124.738	225.314	156.000	19.759	117.273	194.727
Overall	124.044	13.496	97.591	150.496	80.000	15.228	50.154	109.846

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



From Survival Functions and Estimated Mean and Median for survival time for different Lung Cancer Types we conclude that **Small Cell and Adeno type** of lung cancer are more severe than the Squamous and Large Cells Cancer

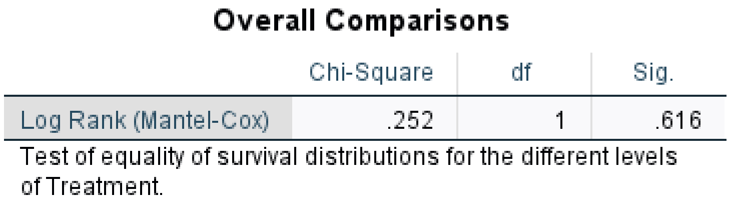
• Comparing efficacy of treatments for treating Lung Cancer



Means and Medians for Survival Time

	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
Treatment	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
1.00	123.928	14.961	94.605	153.251	103.000	19.810	64.173	141.827
2.00	123.159	22.569	78.923	167.394	51.000	14.000	23.560	78.440
Overall	124.044	13.496	97.591	150.496	80.000	15.228	50.154	109.846

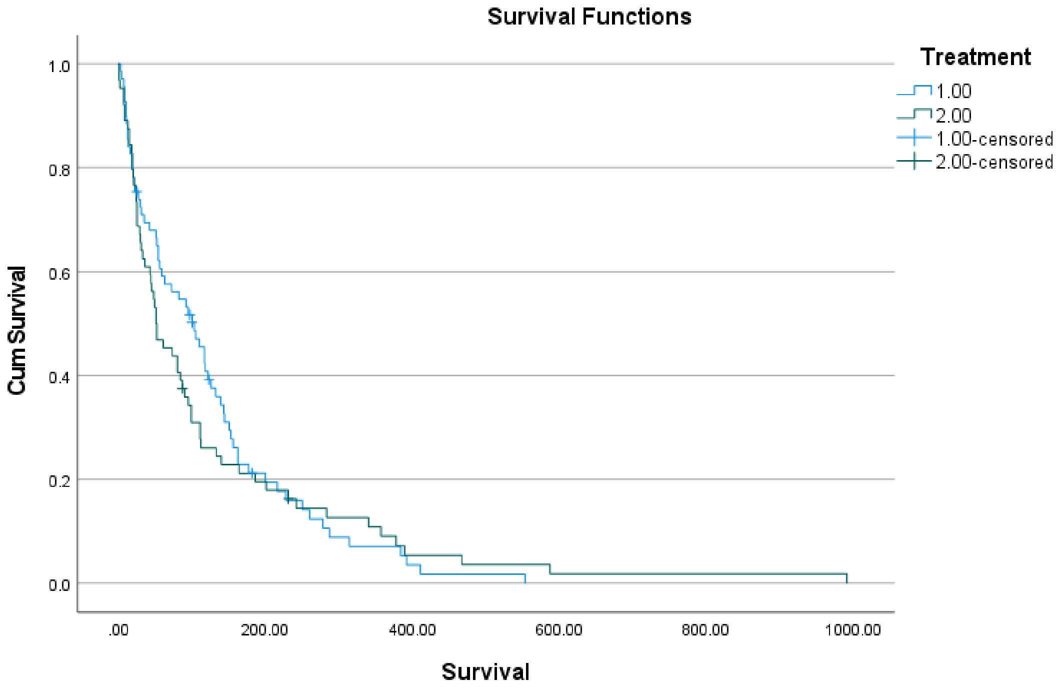
a. Estimation is limited to the largest survival time if it is censored.



Overall Comparisons

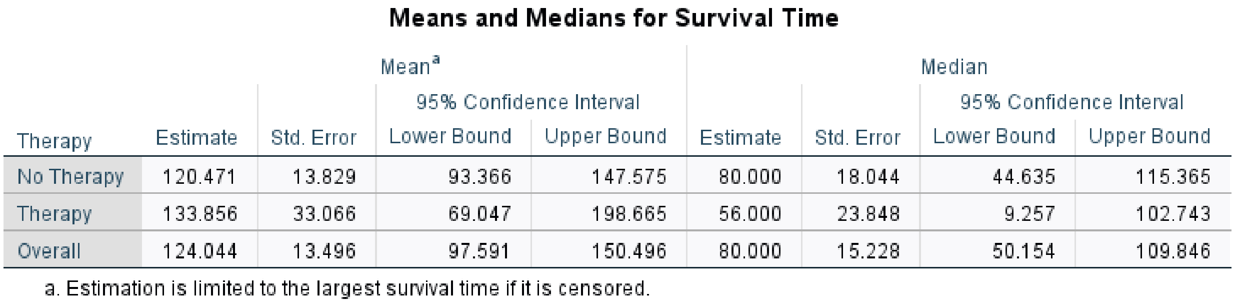
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.252	1	.616

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



From the Survival Functions and Estimated Mean and Median survival time of lung patient under different treatments we conclude that there is no such significant difference among the treatments

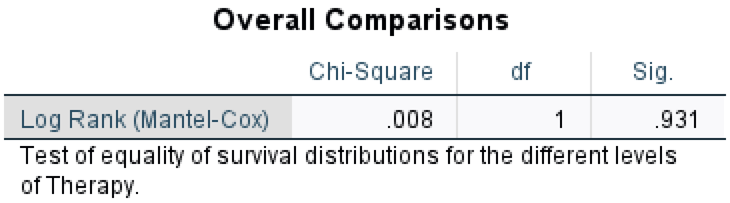
- ## Investigating impact of therapy sessions on survival time



Means and Medians for Survival Time

	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
Therapy	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
No Therapy	120.471	13.829	93.366	147.575	80.000	18.044	44.635	115.365
Therapy	133.856	33.066	69.047	198.665	56.000	23.848	9.257	102.743
Overall	124.044	13.496	97.591	150.496	80.000	15.228	50.154	109.846

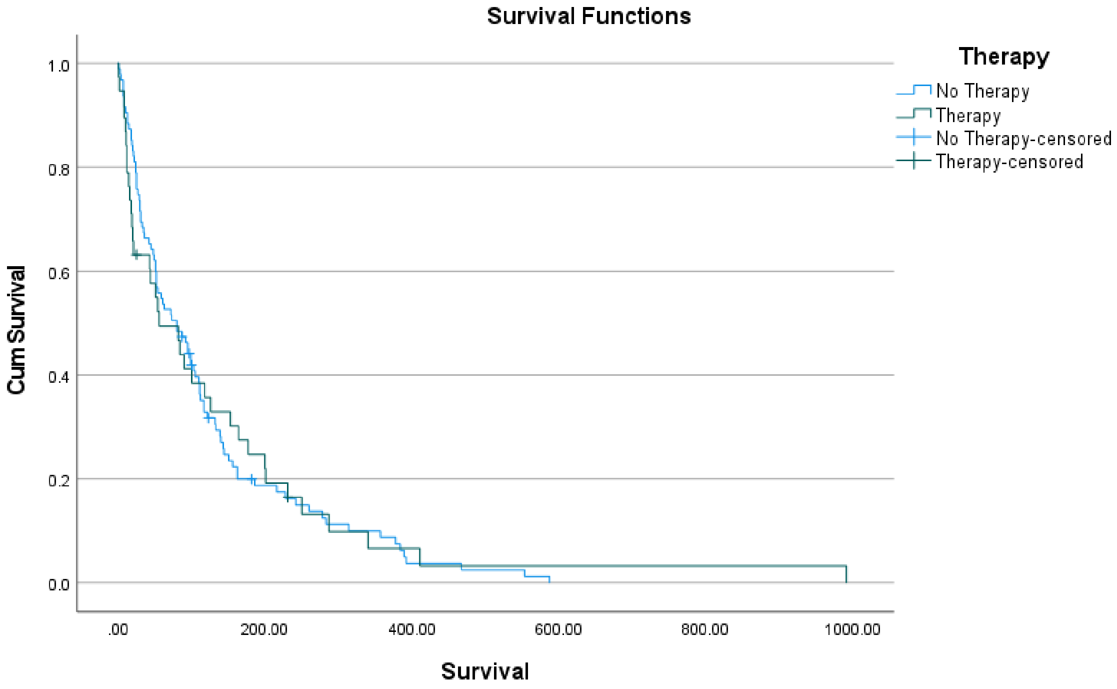
a. Estimation is limited to the largest survival time if it is censored.



Overall Comparisons

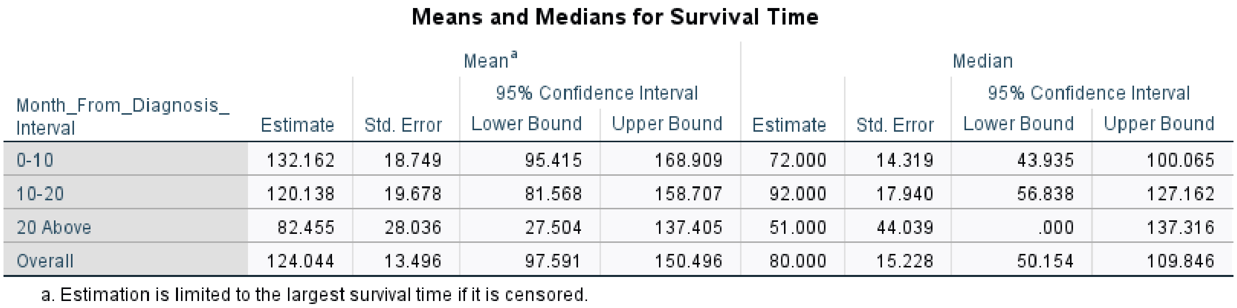
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.008	1	.931

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



From the Survival Functions and Estimated Mean and Median survival time of lung patients who are taking therapy sessions and the one with no therapy sessions we conclude that Therapy Sessions has no significant impact on the survival time

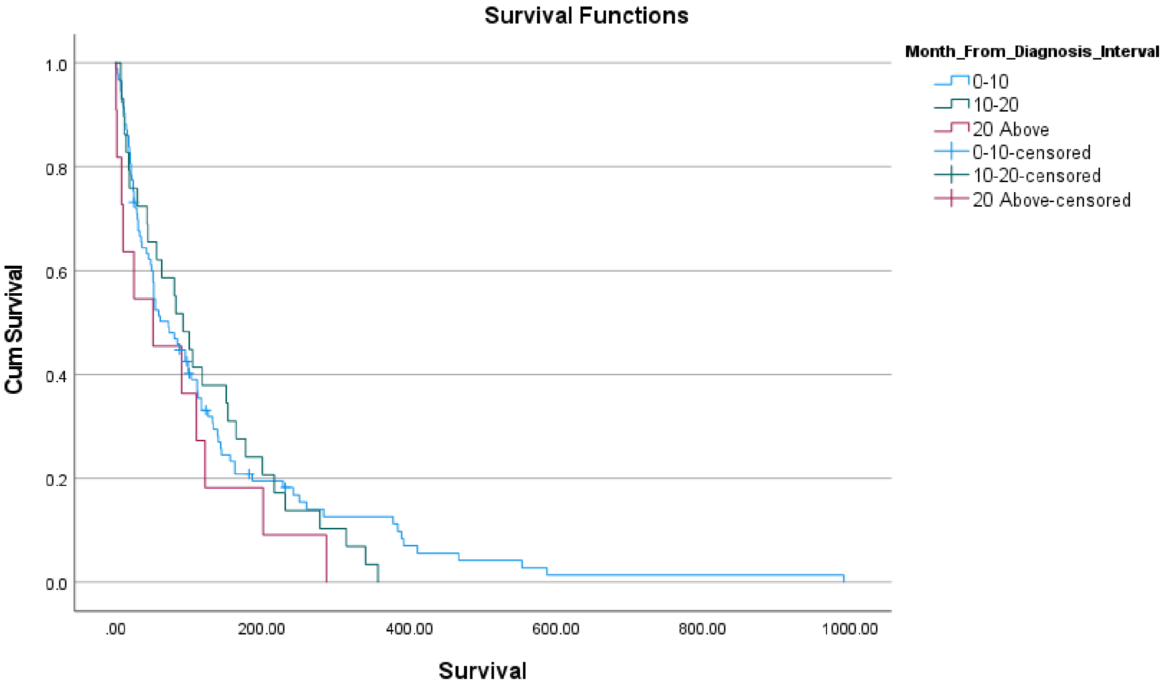
- ## Investigating impact of time elapsed between treatment and diagnosis on survival time



Means and Medians for Survival Time

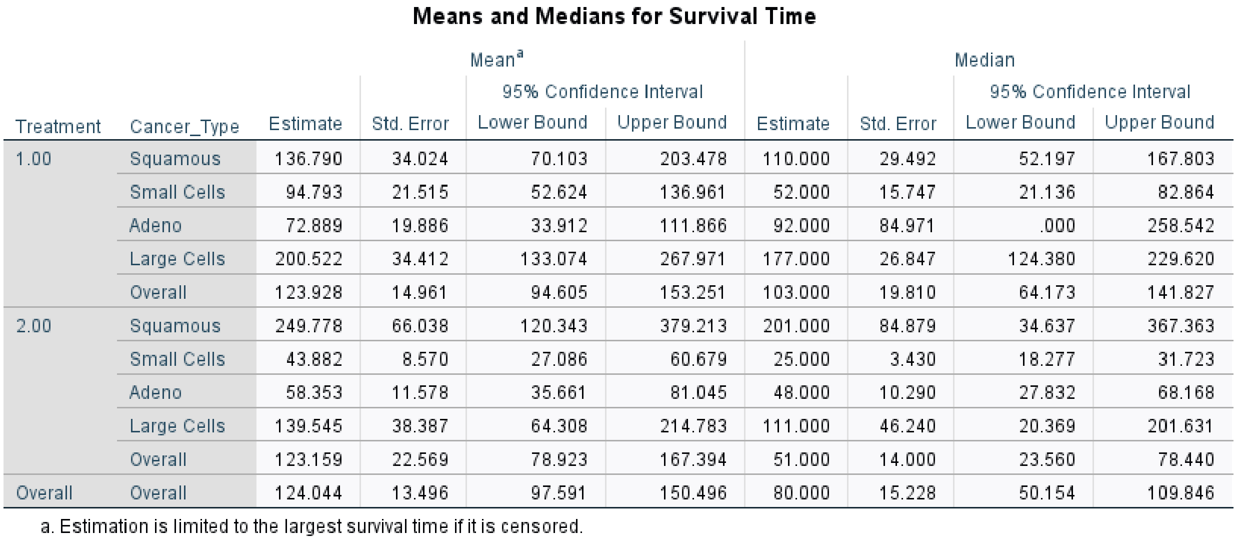
Month_From_Diagnosis_Interval	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
0-10	132.162	18.749	95.415	168.909	72.000	14.319	43.935	100.065
10-20	120.138	19.678	81.568	158.707	92.000	17.940	56.838	127.162
20 Above	82.455	28.036	27.504	137.405	51.000	44.039	.000	137.316
Overall	124.044	13.496	97.591	150.496	80.000	15.228	50.154	109.846

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



From the Estimated Survival Function and the Mean and Median of Survival time we conclude that the patients who have waited more after the diagnoses of the lung cancer have comparatively less survival time than those who immediately begin with the treatment after diagnosing.

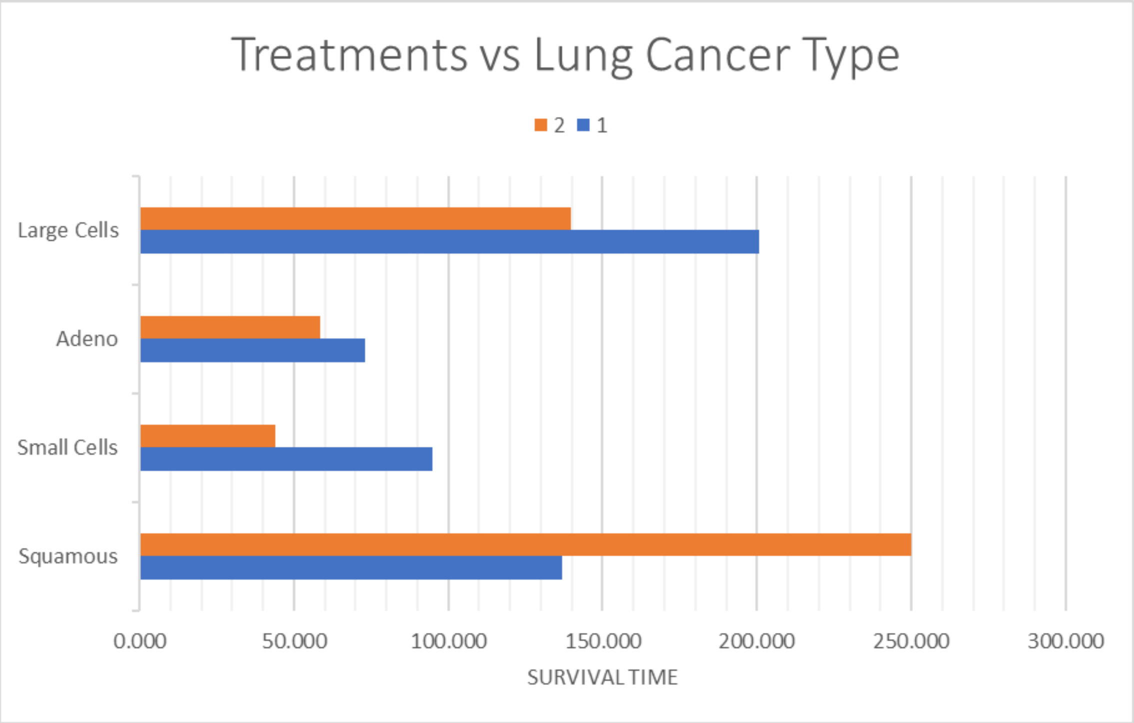
- **Investigating effect of treatment on survival time due to different Lung Cancer Type**

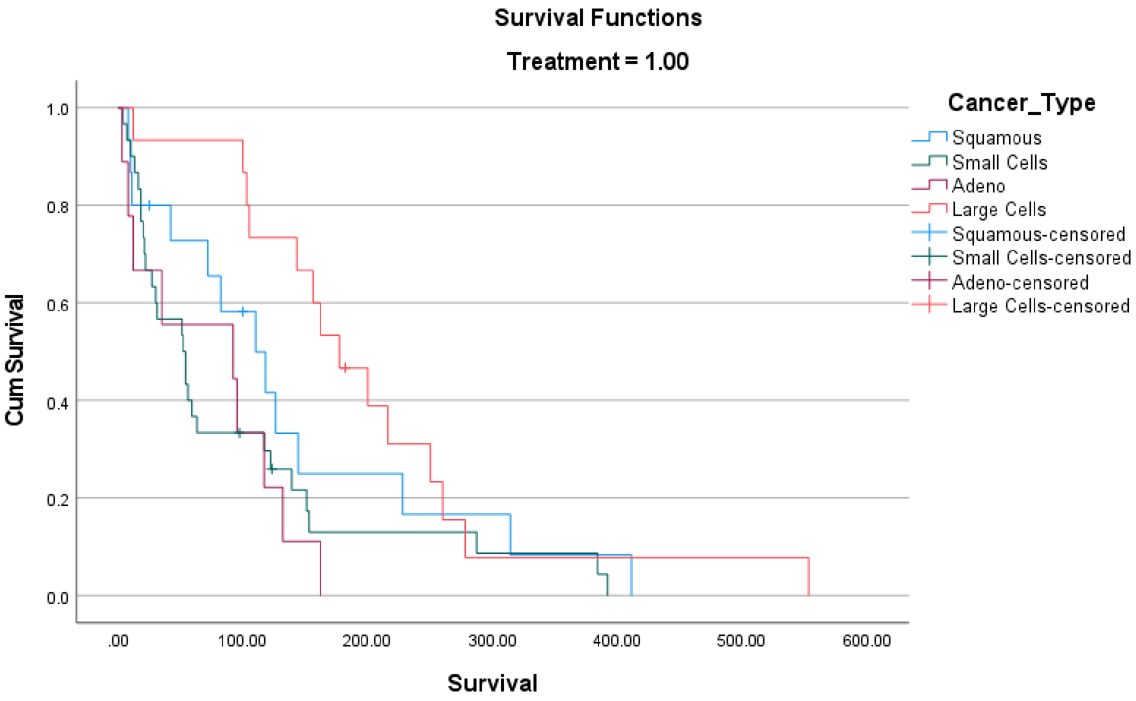


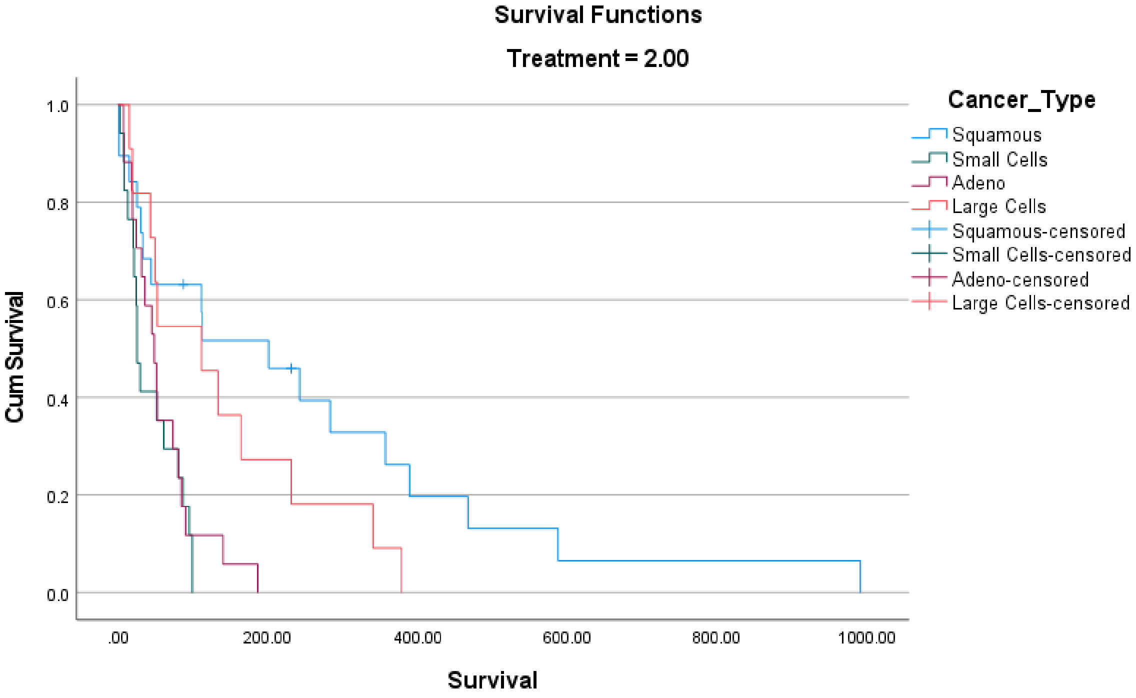
Means and Medians for Survival Time

		Mean[a]				Median			
				95% Confidence Interval				95% Confidence Interval	
Treatment	Cancer_Type	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
1.00	Squamous	136.790	34.024	70.103	203.478	110.000	29.492	52.197	167.803
	Small Cells	94.793	21.515	52.624	136.961	52.000	15.747	21.136	82.864
	Adeno	72.889	19.886	33.912	111.866	92.000	84.971	.000	258.542
	Large Cells	200.522	34.412	133.074	267.971	177.000	26.847	124.380	229.620
	Overall	123.928	14.961	94.605	153.251	103.000	19.810	64.173	141.827
2.00	Squamous	249.778	66.038	120.343	379.213	201.000	84.879	34.637	367.363
	Small Cells	43.882	8.570	27.086	60.679	25.000	3.430	18.277	31.723
	Adeno	58.353	11.578	35.661	81.045	48.000	10.290	27.832	68.168
	Large Cells	139.545	38.387	64.308	214.783	111.000	46.240	20.369	201.631
	Overall	123.159	22.569	78.923	167.394	51.000	14.000	23.560	78.440
Overall	Overall	124.044	13.496	97.591	150.496	80.000	15.228	50.154	109.846

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

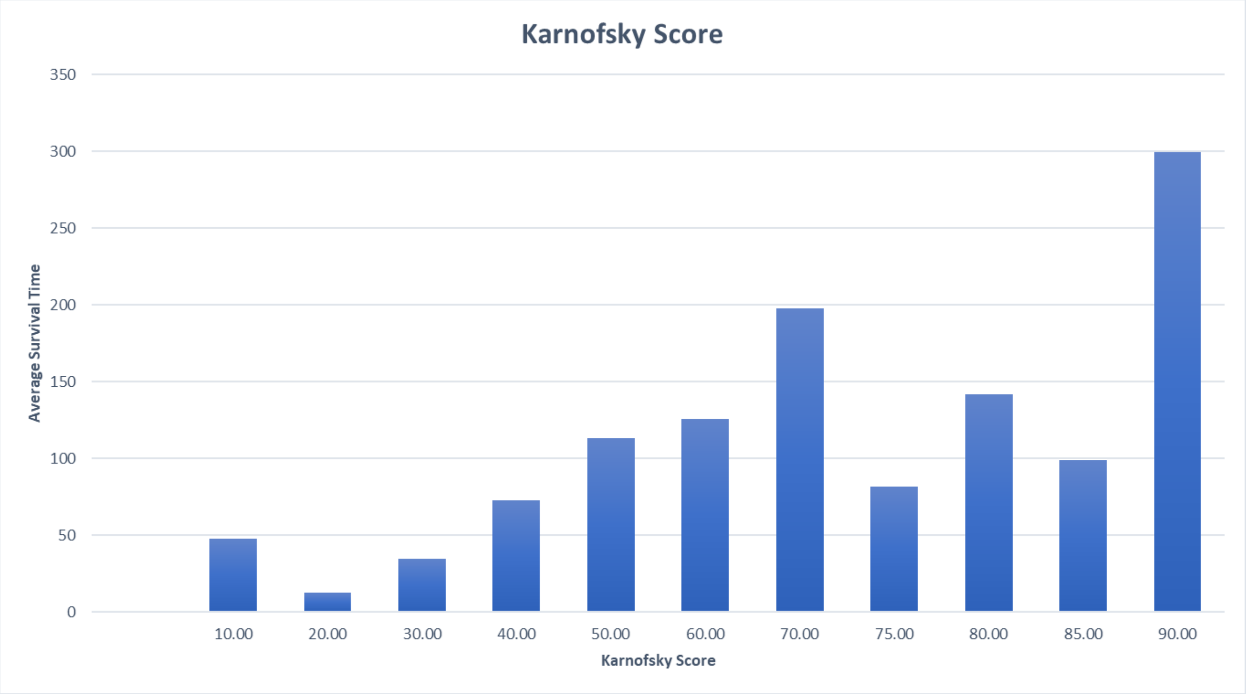




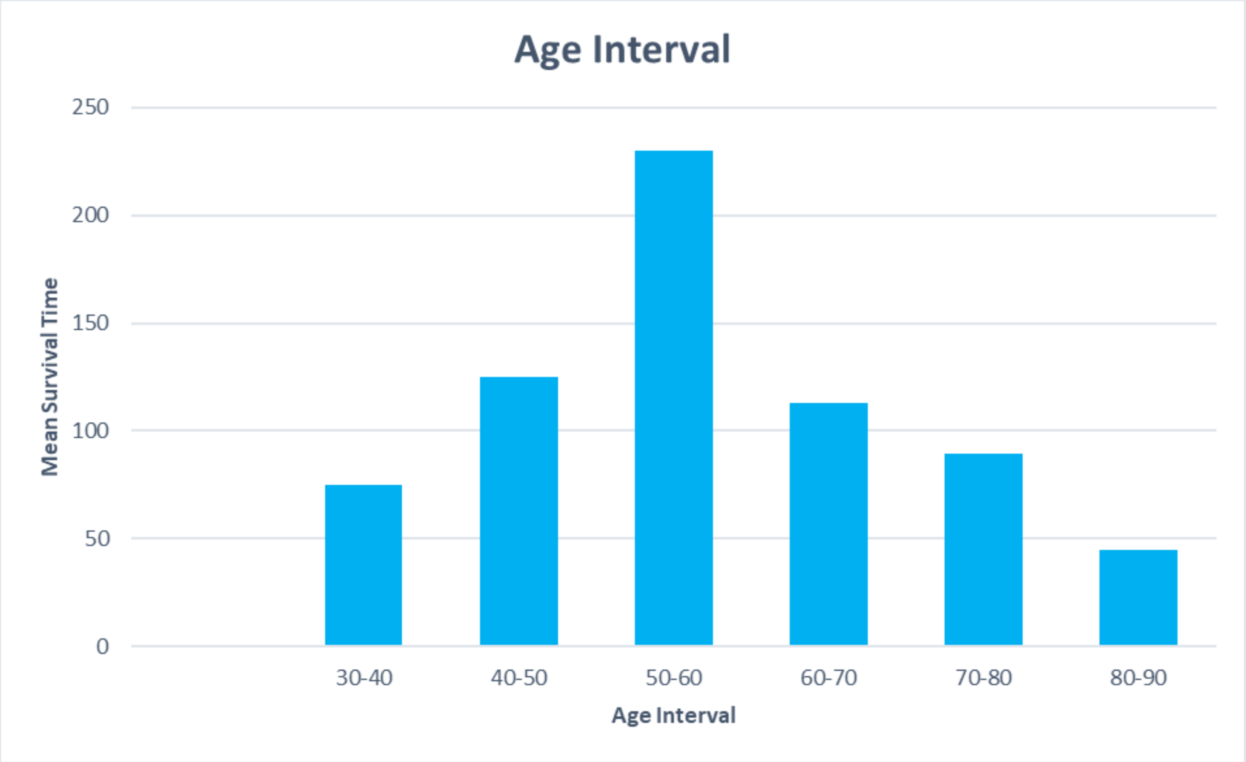


So From Above Estimates we conclude that Treatment 1 is more effective against Large Cells, Adeno and Small Cells. However treatment 2 is more effective against Squamous.

• Impact of Age and Karnofsky Score on survival time



So as the Karnofsky Score decreases survival time of the patient also decreases



So as the age increases we are able to see that as age increases survival time decreases

Concluding Remarks

- From the histogram and estimated survival function we can infer that most of the deaths occur in first 200 days of the treatment received
- We infer that our Survival Time follows Weibull (Shape=.923 , Scale=104.847).
- We also infer that the Small Cells and Adeno Cells Lung Cancer are more deadly than the Sqamous and Large Cells cancer
- We also infer that the there is no significant difference in the survival time due to both the treatments
- We also infer that giving therapy is not significantly effective against Lung Cancer
- We also infer that paitents concerning about their health and took immediate action regarding the
- At first glance the treatment seems to be no significantly different but when we deep dive into the comparison of treatment against each cancer type we found that treatment 2 is effective against the Squamous Cell and treatment 1 is more effective against Large Cells , Small Cells and Adeno Cancer type.

THANK
You!