Research Topic: Exploring Predictors of Cirrhosis Patient Survival: An In-Depth Analysis Using Statistical Tests

Data Description

1. ID: unique identifier

2. N_Days: number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986

3. Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)

4. Drug: type of drug D-penicillamine or placebo

5. Age: age in [days]

6. Sex: M (male) or F (female)

7. Ascites: presence of ascites N (No) or Y (Yes)

8. Hepatomegaly: presence of hepatomegaly N (No) or Y (Yes)

9. Spiders: presence of spiders N (No) or Y (Yes)

10. Edema: presence of edema N (no edema and no diuretic therapy

Introduction:

Cirrhosis, a chronic liver condition that scars liver tissue, is a global health issue. Optimizing patient care and therapy requires understanding cirrhosis patient survival, variables. This study conducts a comprehensive statistical analysis with three goals to add to existing knowledge. First, we want to examine gender-based differences in cirrhosis patients' survival time to determine prognostic factors.

Second, we will examine average survival time across illness phases to understand cirrhosis progression. Finally, we will explore the associations between major clinical variables—Albumin, Bilirubin, and Prothrombin—in the context of survival time. This multimodal research addresses significant gaps in current knowledge and may inform targeted therapies and improve cirrhosis patient management.

Li and Liu (2004) developed data-depth-based nonparametric tests for multivariate locations and scales. Advanced statistical methods for multivariate data analysis examined in this Statistical Science article. The research adds data depth metrics to the statistical toolbox, revealing multivariate dataset distributional features.

Etikan (2018) discusses critical survival analysis statistical test selection in "Choosing statistical tests for survival analysis. They provides useful recommendations on selecting statistical approaches for survival data analysis, improving biostatistics procedures.

Research Objective

- To conduct a statistical analysis to examine the disparities in age among cirrhosis patients based on gender.
- To assess whether there exist statistically significant variations in the average survival time across various phases of the disease.
- To conduct a statistical analysis in order to ascertain the presence of significant correlations between the variables of Albumin, Bilirubin, and prothrombin, within the specific context of survival time.

Hypothesis

- There is no significant difference in the age of male and female of cirrhosis patients.
- There is no significant correlation between Albumin, Bilirubin and prothrombin

- There is no significant difference in the mean survival time among difference disease stages.

Normality Analysis

Normality test for age

Null Hypothesis

H_0: Data about age is fellow the normal distribution

Alternative Hypothesis

H_1: Data about age is not fellow the normal distribution

Fitted Normal Distribution for Age

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.03305165	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.06870264	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.58593726	Pr > A-Sq	0.131

In our case Since P value is not less than 5%. So we will not reject Ho and concluded that data is normal about age.

Normality test for Bilirubin

Null Hypothesis

H_0: Data about Bilirubin is fellow the normal distribution

Alternative Hypothesis

H_1: Data about Bilirubin is not fellow the normal distribution

Fitted Normal Distribution for Bilirubin

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.2569359	Pr > D	<0.010
Cramer-von Mises	W-Sq	7.3535051	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	39.0901771	Pr > A-Sq	<0.005

In our case Since P value is less than 5%. So we will reject Ho and concluded that data is not normal about Bilirubin

Normality test for Albumin

Null Hypothesis

H_0: Data about Albumin is fellow the normal distribution

Alternative Hypothesis

H_1: Data about Albumin is not fellow the normal distribution

Fitted Normal Distribution for Albumin

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.07124768	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.23188297	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	1.51762395	Pr > A-Sq	<0.005

In our case Since P value is less than 5%. So we will reject Ho and concluded that data is not normal about Albumin

Normality test for Prothrombin

Null Hypothesis

H_0: Data about prothrombin is fellow the normal distribution

Alternative Hypothesis

H_1: Data about prothrombin is not fellow the normal distribution

Fitted Normal Distribution for Prothrombin

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.12142217	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.01661650	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	6.48376825	Pr > A-Sq	<0.005

In our case Since P value is less than 5%. So we will reject Ho and concluded that data is not normal about Prothrombin.

Two Sample T test

Null Hypothesis (H0):

There is no significant difference in the age of male and female of cirrhosis patients.

Alternative Hypothesis (H1):

There is a significant difference in the age of male and female of cirrhosis patients.

Variable: Age
Sex = F

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.989813	Pr < W	0.0106
Kolmogorov-Smirnov	D	0.039077	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.106644	Pr > W-Sq	0.0937
Anderson-Darling	A-Sq	0.81326	Pr > A-Sq	0.0370

Variable: Age
Sex = M

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.979036	Pr < W	0.5955
Kolmogorov-Smirnov	D	0.107324	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.05618	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.332912	Pr > A-Sq	>0.2500

Variable: Age

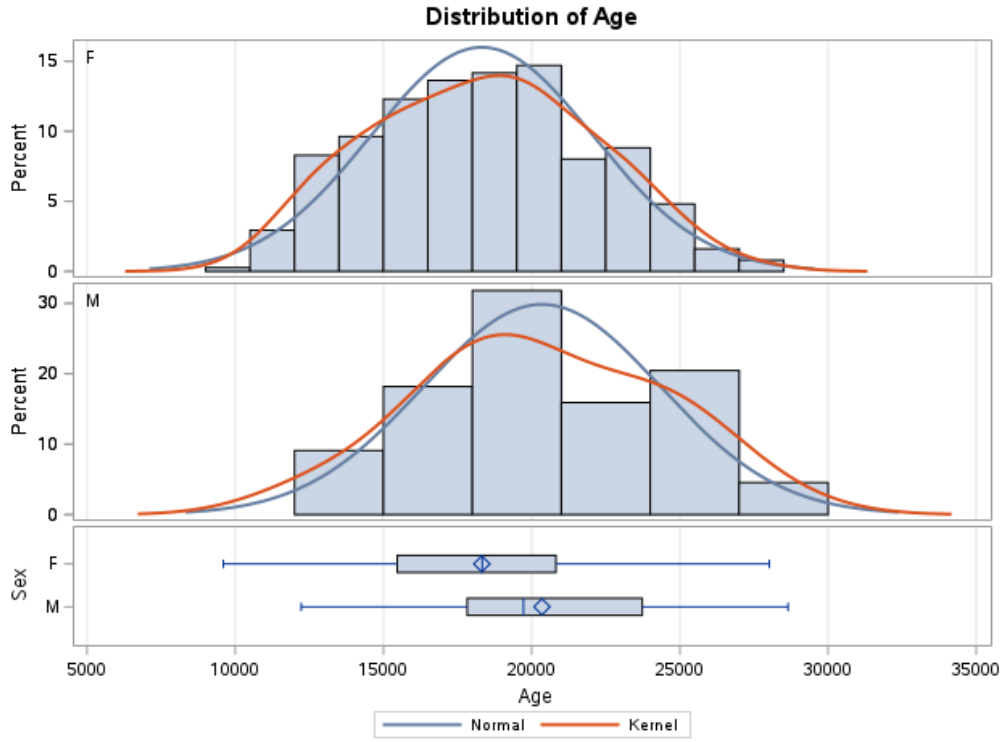
Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
F		374	18319.8	3740.4	193.4	9598.0	28018.0
M		44	20348.3	4009.6	604.5	12227.0	28650.0
Diff (1-2)	Pooled		-2028.5	3769.1	600.7		
Diff (1-2)	Satterthwaite		-2028.5		634.7		

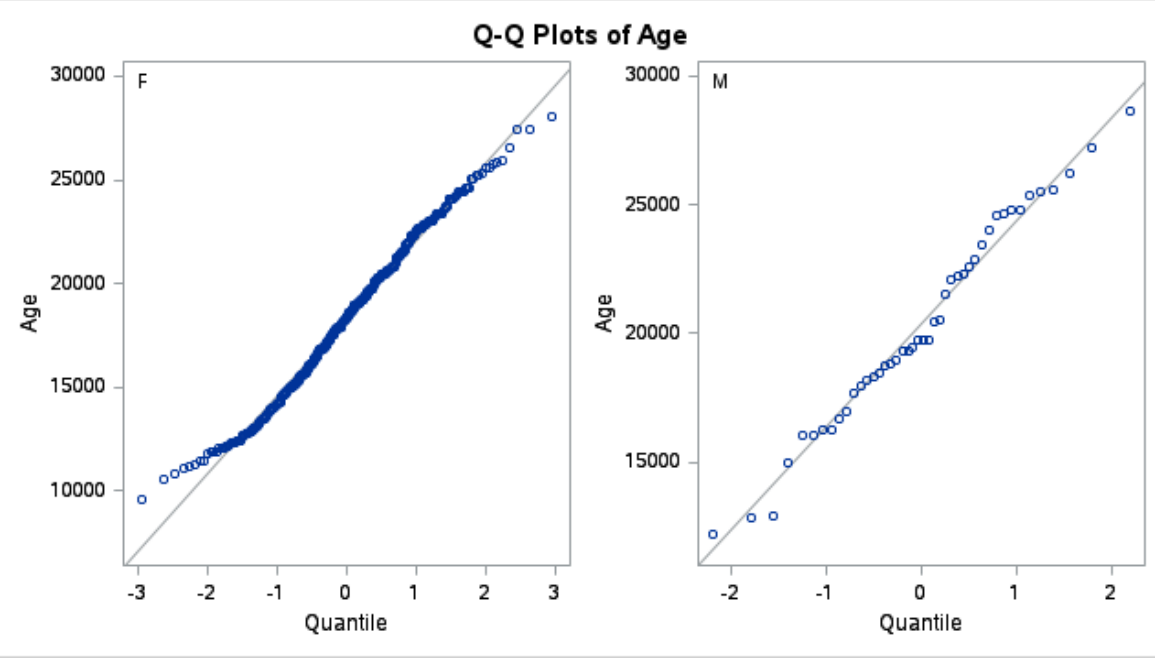
Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
F		18319.8	17939.5	18700.1	3740.4	3490.2	4029.6
M		20348.3	19129.3	21567.4	4009.6	3312.9	5080.3
Diff (1-2)	Pooled	-2028.5	-3209.3	-847.7	3769.1	3529.5	4043.9
Diff (1-2)	Satterthwaite	-2028.5	-3302.0	-755.1			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	416	-3.38	0.0008
Satterthwaite	Unequal	52.192	-3.20	0.0024

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	43	373	1.15	0.4961

According to the findings of the T-test, there is a statistically significant gap between the ages of patients suffering from cirrhosis who are males and those who are females ($p<0.001$). Based on these findings, it appears that the ages of patients suffering from cirrhosis greatly differ depending on which gender they are.





Correlation Analysis

Null Hypothesis (H0):

There is no significant correlation between Albumin, Bilirubin and prothrombin

Alternative Hypothesis (H1):

There is significant correlation between Albumin, Bilirubin and prothrombin

3 Variables:	Bilirubin Albumin Prothrombin

Pearson Correlation Coefficients, N = 309 Prob > |r| under H0: Rho=0			
	Bilirubin	Albumin	Prothrombin
Bilirubin	1.00000	-0.33165 <.0001	0.36107 <.0001
Albumin	-0.33165 <.0001	1.00000	-0.23021 <.0001
Prothrombin	0.36107 <.0001	-0.23021 <.0001	1.00000

Spearman Correlation Coefficients, N = 309 Prob > |r| under H0: Rho=0			
	Bilirubin	Albumin	Prothrombin
Bilirubin	1.00000	-0.35708 <.0001	0.29721 <.0001
Albumin	-0.35708 <.0001	1.00000	-0.20531 0.0003
Prothrombin	0.29721 <.0001	-0.20531 0.0003	1.00000

According to the findings of the correlation analysis, there is a statistically significant negative association between bilirubin and albumin. This suggests that as bilirubin levels rise, albumin levels often fall as result of the trend. Additionally, there is a statistically significant positive connection between bilirubin and prothrombin, which indicates that greater bilirubin levels related with elevated prothrombin levels. This association suggests that higher bilirubin levels connected with higher prothrombin levels. Additionally, a statistically significant inverse correlation can shown between albumin and prothrombin, which suggests that when albumin levels fall, prothrombin levels have a tendency to rise. This finding supported by the data presented in the previous section.

Kruskal Wallis Test

Null Hypothesis

H0: There is no significant difference in the median survival time among difference disease stages.

Alternative Hypothesis (H1):

H0: There is significant difference in the median survival time among difference disease stages.

Wilcoxon Scores (Rank Sums) for Variable Prothrombin Classified by Variable Stage					
Stage	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
3	119	15951.00	18445.0	763.291389	134.042017
2	66	8196.50	10230.0	642.858198	124.189394
4	108	21520.50	16740.0	747.911116	199.263889
1	16	2227.00	2480.0	347.563277	139.187500
Average scores were used for ties.					

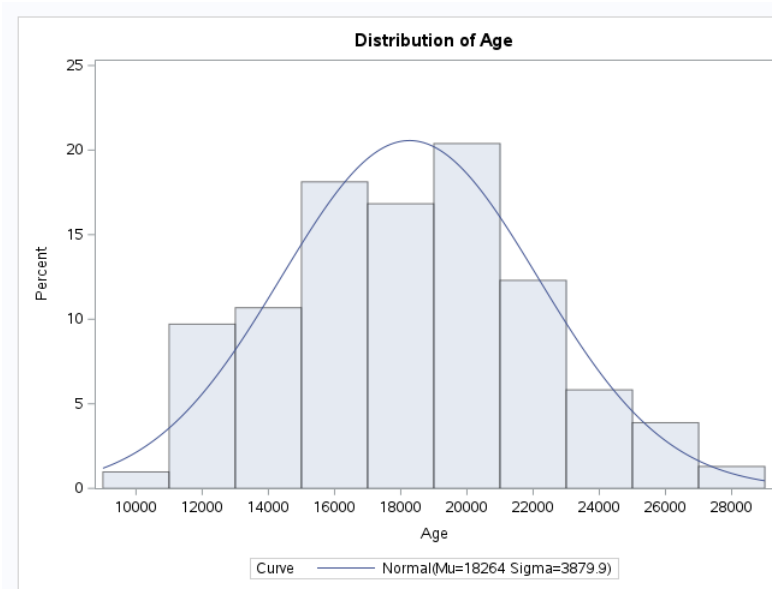
Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
41.5114	3	<.0001

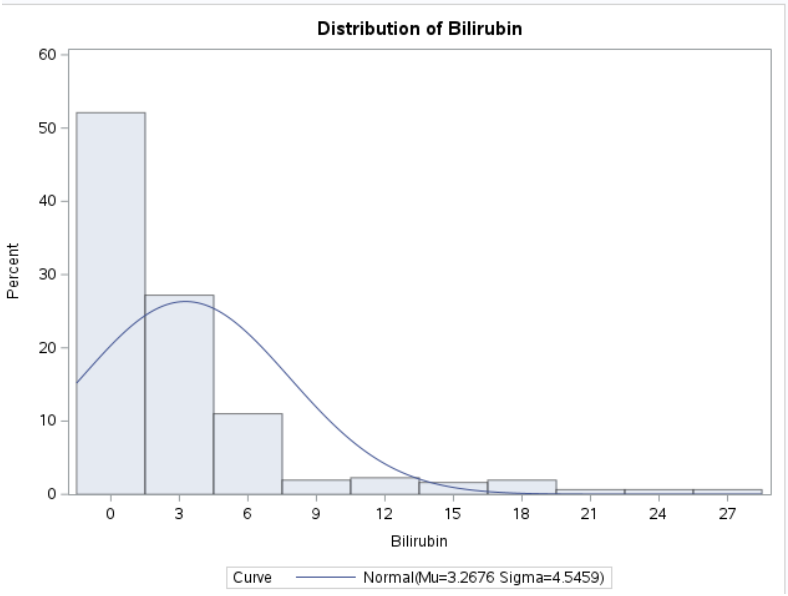
Pairwise Two-Sided Multiple Comparison Analysis			
Dwass, Steel, Critchlow-Fligner Method			
Variable: Prothrombin			
Stage	Wilcoxon Z	DSCF Value	Pr > DSCF
3 vs. 2	0.5698	0.8059	0.9411
3 vs. 4	-5.3971	7.6326	<.0001
3 vs. 1	-0.1943	0.2748	0.9974
2 vs. 4	-5.5996	7.9190	<.0001
2 vs. 1	-0.3755	0.5310	0.9820
4 vs. 1	2.3406	3.3102	0.0890

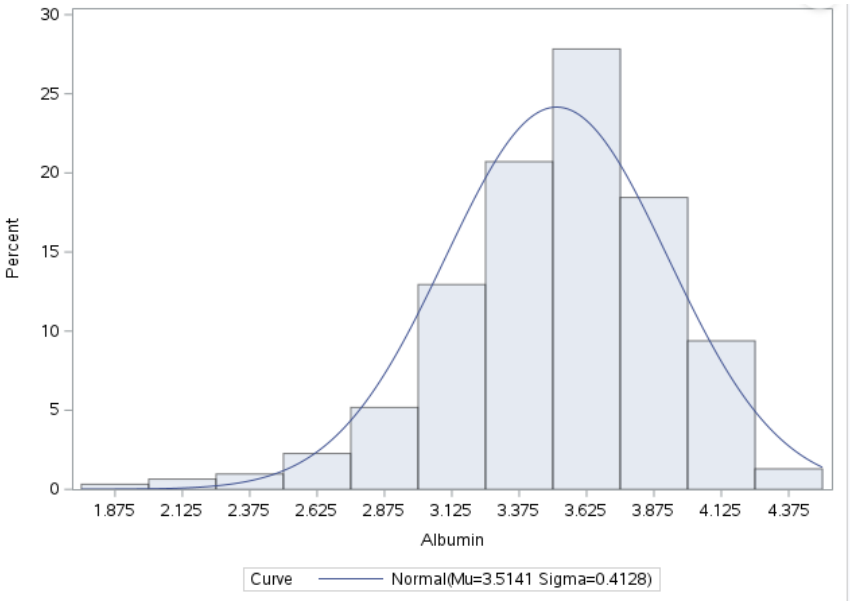
The Kruskal-Wallis test for several stages of the disease demonstrates that there is a statistically significant variation in the mean amount of time patients survive between these stages ($p < 0.01$). According to the results of the pairwise comparisons, there is also no statistically significant difference in the amount of time spent alive between stages 3 and 2, stages 3 and 1, stages 2 and 1 ($p > 0.05$). Nevertheless, there is a statistically significant gap in the amount of time spent alive between every other possible pairing of illness phases ($p <0.001$).

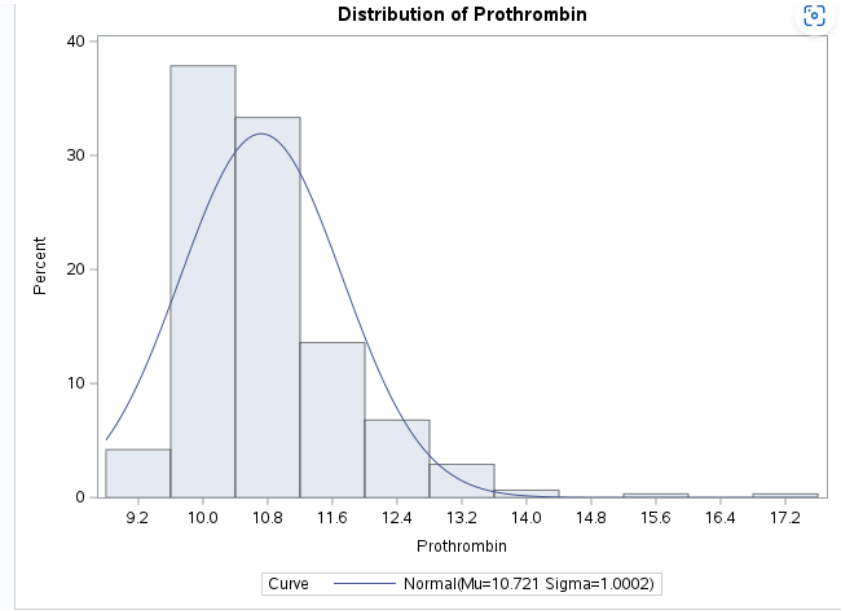
TASK C

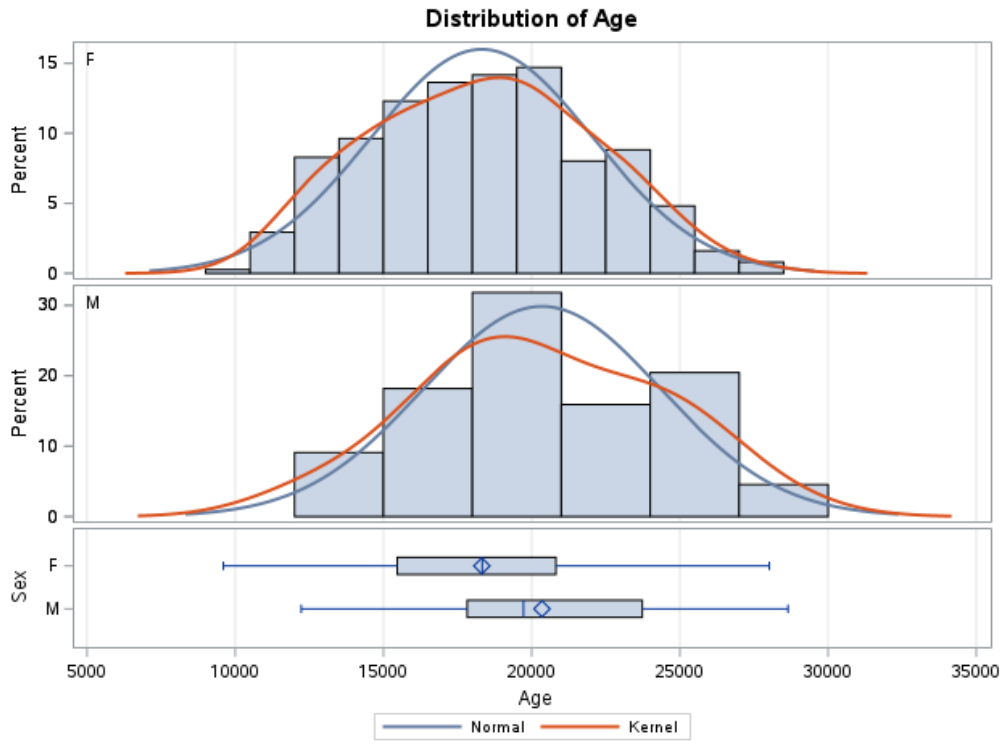
Data visualization

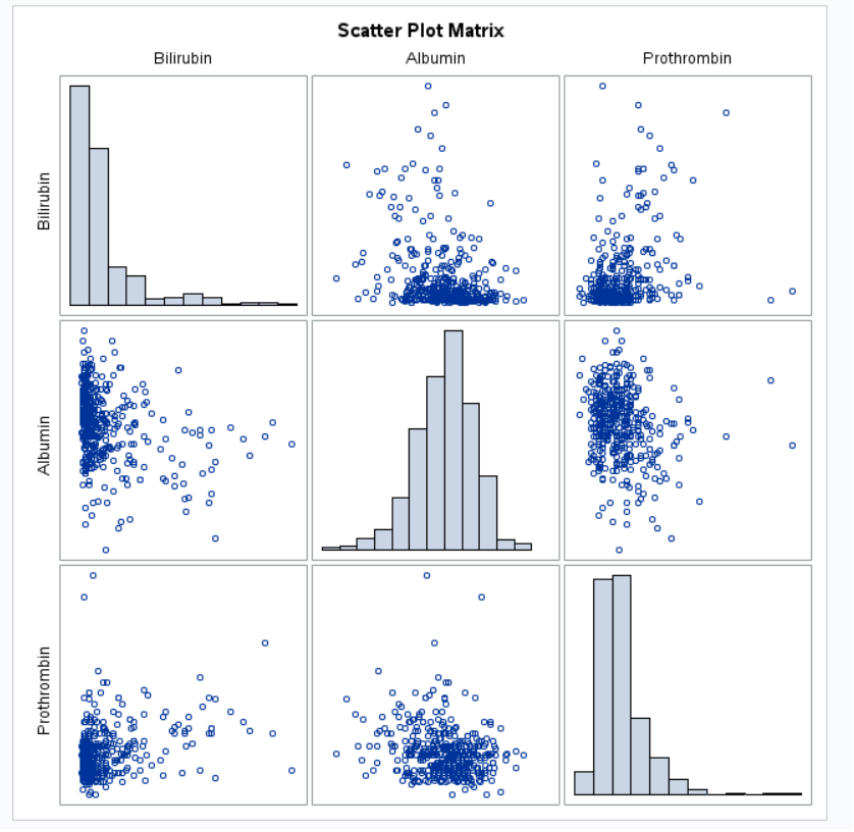

Distribution of Age
Percent
25
20
15
10
5
0
10000
12000
14000
16000
18000
20000
22000
24000
26000
28000
Age
Curve
Normal(Mu=18264 Sigma=3879.9)

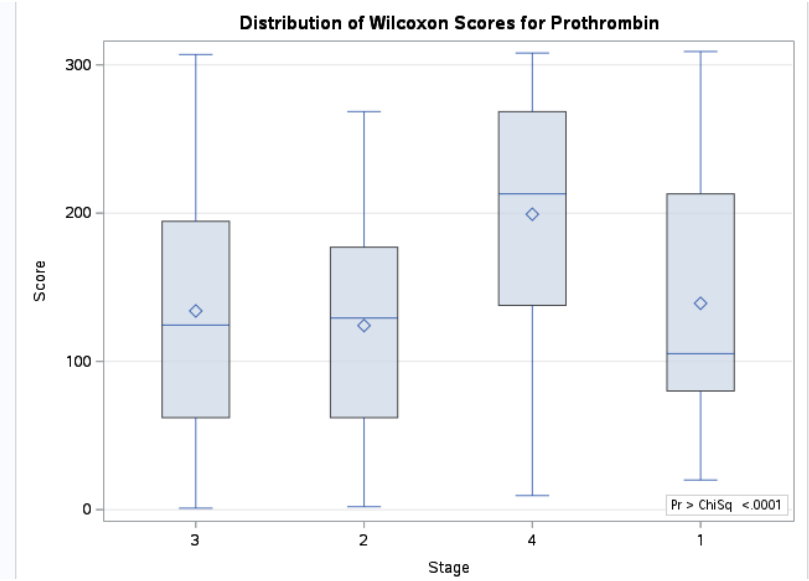

Distribution of Bilirubin
Percent
60
50
40
30
20
10
0
0
3
6
9
12
15
18
21
24
27
Bilirubin
Curve
Normal(Mu=3.2676 Sigma=4.5459)


Percent
30
25
20
15
10
5
0
1.875
2.125
2.375
2.625
2.875
3.125
3.375
3.625
3.875
4.125
4.375
Albumin
Curve
Normal(Mu=3.5141 Sigma=0.4128)


Distribution of Prothrombin
Percent
40
30
20
10
0
9.2
10.0
10.8
11.6
12.4
13.2
14.0
14.8
15.6
16.4
17.2
Prothrombin
Curve
Normal(Mu=10.721 Sigma=1.0002)


Distribution of Age
F
M
Percent
Sex
F
M
5000
10000
15000
20000
25000
30000
35000
Age
Normal
Kernel


Scatter Plot Matrix
Bilirubin
Albumin
Prothrombin



Conclusion:

In Conclusion the T-test, which show that there are significant gender-based disparities in age, this study emphasizes how important it is to take into account patients' ages as a vital aspect when dealing with cirrhosis patients. The findings of the association between Bilirubin, Albumin, and Prothrombin have shed light on the complicated linkages that exist between these clinical variables, hence providing useful insights into the convoluted course of cirrhosis. In addition, the stage of the disease has a major impact on the amount of time a patient expected to live; nevertheless, there was no significant difference in survival time between stages 3 and 2, stages 3 and 1, and stages 2 and 1. Nevertheless, the statistically significant changes in survival time that were found in other pairwise comparisons shed light on the necessity of individualized therapies and highlight the significance of taking into account the stage of the disease in order to maximize the likelihood of a positive outcome for patients being treated for cirrhosis. These findings make a sizeable contribution to a more nuanced knowledge of the factors influencing the outcomes of cirrhosis patients, so paving the way for treatment strategies that are more specifically focused and efficient.

Reference:

Li, J., & Liu, R. Y. (2004, November 1). New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth. Statistical Science, 19(4). https://doi.org/10.1214/088342304000000594

Etikan, L. (2018, October 17). Choosing statistical tests for survival analysis. Biometrics & Biostatistics International Journal, 7(5). https://doi.org/10.15406/bbij.2018.07.00249