Analysis Report

This report is structured as follows.

Contents

Sample Characteristics

In the study of liver fibrosis among 283 participants, varying stages of the disease were analyzed in conjunction with several medical conditions and demographic factors. The analysis reveals that 50.9% of the participants are diabetic, 5.3% are pre-diabetic, and 43.8% do not have diabetes. Regarding hypertension, 56.5% of the participants have the condition, whereas 43.5% do not. Cholesterol levels are evenly split among the participants, with 50.5% reporting high cholesterol and 49.5% without. Notably, obesity is prevalent, with 89.7% of the sample being obese.

Category	Level	Count	Percentage
Type.II.Diabetic	No	124	43.8
	Pre-diabetic	15	5.3
	Yes	144	50.9
Hypertension	No	123	43.5
	Yes	160	56.5
High.Cholesterol	No	140	49.5
	Yes	143	50.5
Obesity	No	29	10.3
	Yes	253	89.7
Liver.Biopsy.Fibrosis.Stage	Stage 0	43	15.1
	Stage 1	66	23.2
	Stage 2	74	26.1
	Stage 3	76	26.8
	Stage 4	25	8.8
Gender	Female	208	73.2
	Male	76	26.8

The distribution of liver fibrosis stages indicates a higher concentration in the middle stages: Stage 2 and Stage 3 contain 26.1% and 26.8% of the participants, respectively, compared to 15.1% in Stage 0 and 8.8% in Stage 4. The prevalence of Type II Diabetes and Hypertension increases in more advanced stages of liver fibrosis, suggesting a potential link between these conditions and the severity of fibrosis. Specifically, the proportion of diabetic participants increases from 9.9% in Stage 0 to 18.7% in Stage 3, and similarly, hypertension prevalence rises from 9.5% in Stage 0 to 18.7% in Stage 3. High cholesterol also shows a comparable pattern, with 14.8% in Stage 2 increasing slightly to 18.4% in Stage 3.

Variable	Level	Stage 0 N	Stage 0 %	Stage 1 N	Stage 1 %	Stage 2 N	Stage 2 %	Stage 3 N	Stage 3 %	Stage 4 N	Stage 4 %
Type.II.Diabetic	No	28	9.9	32	11.3	36	12.7	22	7.8	6	2.1
	Pre-diabetic	1	0.4	3	1.1	8	2.8	0	0.0	3	1.1
	Yes	14	4.9	31	11.0	30	10.6	53	18.7	16	5.7
Hypertension	No	27	9.5	34	12.0	32	11.3	22	7.8	8	2.8
	Yes	16	5.7	32	11.3	42	14.8	53	18.7	17	6.0
High.Cholesterol	No	26	9.2	37	13.1	42	14.8	23	8.1	12	4.2
	Yes	17	6.0	29	10.2	32	11.3	52	18.4	13	4.6
Obesity	No	0	0.0	8	2.8	6	2.1	12	4.3	3	1.1
	Yes	43	15.2	57	20.2	68	24.1	63	22.3	22	7.8

Descriptive Statistics

The following tables present the mean, standard errors, standard deviations, skewness and kurtosis of the variables under study.

Variable	N	Mean	SEM	SD	Skewness	Kurtosis
FIB_4	198	1.254	0.044	0.614	1.046	4.038
ELF	31	9.495	0.134	0.746	0.346	3.411
FAST	278	0.478	0.015	0.246	-0.004	1.983
Fibroscan.CAP	279	334.806	2.563	42.810	-1.849	15.191
Fibroscan.E._kPA_	279	15.676	1.189	19.856	11.511	163.656

Variable	Stage 0	Stage 1	Stage 2	Stage 3	Stage 4
FIB_4	0.899	1.024	1.134	1.318	1.711
ELF	9.469	9.424	9.354	9.958	
FAST	0.320	0.401	0.535	0.553	0.561
Fibroscan.CAP	332.395	336.569	340.194	338.253	307.417
Fibroscan.E._kPA_	12.744	17.400	13.254	17.105	19.054

The FIB-4 index, calculated from 198 observations, has a mean of 1.254 with a standard error of the mean (SEM) of 0.044 and a standard deviation (SD) of 0.614. It exhibits a skewness of 1.046 and a kurtosis of 4.038, indicating a distribution that is moderately skewed to the right with a leptokurtic peak. Across fibrosis stages, the FIB-4 index increases progressively from 0.899 in Stage 0 to 1.711 in Stage 4, reflecting worsening liver fibrosis.

The ELF score, derived from 31 observations, averages at 9.495 with a SEM of 0.134 and an SD of 0.746. It has a lower skewness of 0.346 and a kurtosis of 3.411, suggesting a fairly symmetric and moderately peaked distribution. Across stages, the ELF scores show minor fluctuations with the highest average score observed in Stage 3 (9.958).

The FAST score, calculated from 278 observations, has a mean of 0.478, an SEM of 0.015, and an SD of 0.246. Its skewness is near zero (-0.004), indicating a symmetric distribution, and a kurtosis of 1.983 suggests a platykurtic distribution. The scores increase with advancing fibrosis stages, starting from 0.320 in Stage 0 to 0.561 in Stage 4.

The Fibroscan CAP, from 279 observations, shows a mean of 334.806 with a SEM of 2.563 and an SD of 42.810. It is negatively skewed (-1.849) and has a high kurtosis of 15.191, indicating a distribution that is heavily tailed and skewed to the left. The CAP values generally increase through the fibrosis stages, except for a drop in Stage 4 (307.417).

Lastly, the Fibroscan Elasticity (kPA) also based on 279 observations, shows a mean of 15.676, an SEM of 1.189, and an extraordinarily high SD of 19.856. This measure exhibits extreme positive skewness (11.511) and very high kurtosis (163.656), suggesting a distribution with extreme values influencing the mean and variability. The elasticity measures show variations across stages but generally increase in later stages, reaching 19.054 in Stage 4.

Outliers were examined using Z-scores and a particular case showed a Z-score of 14.6 for kPA and -7.8 for CAP, which is distant from suggested thresholds of 4 or 3.5 to be assigned as an outlier. This case was removed from further analyses.

ANOVA

A one way Analysis of Variance was employed to test if the mean differences across stages are statistically significant. The results are shown below.
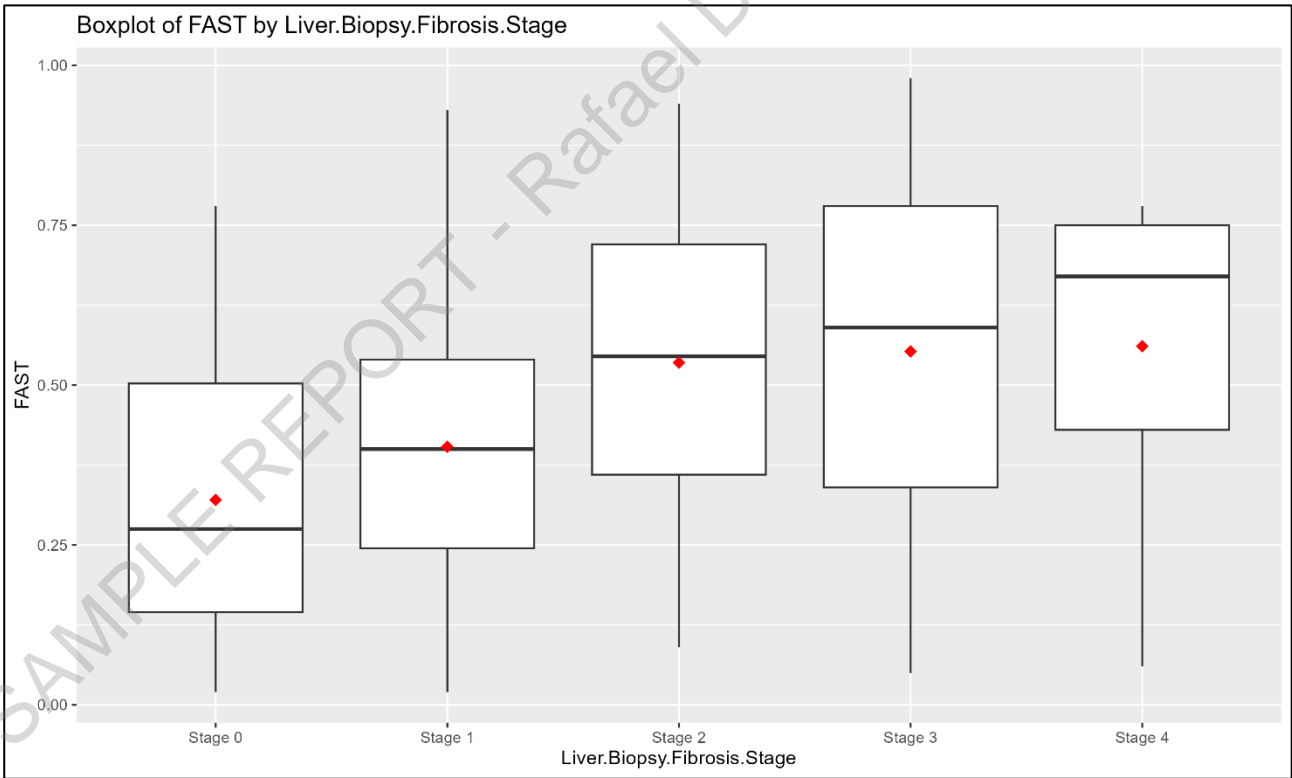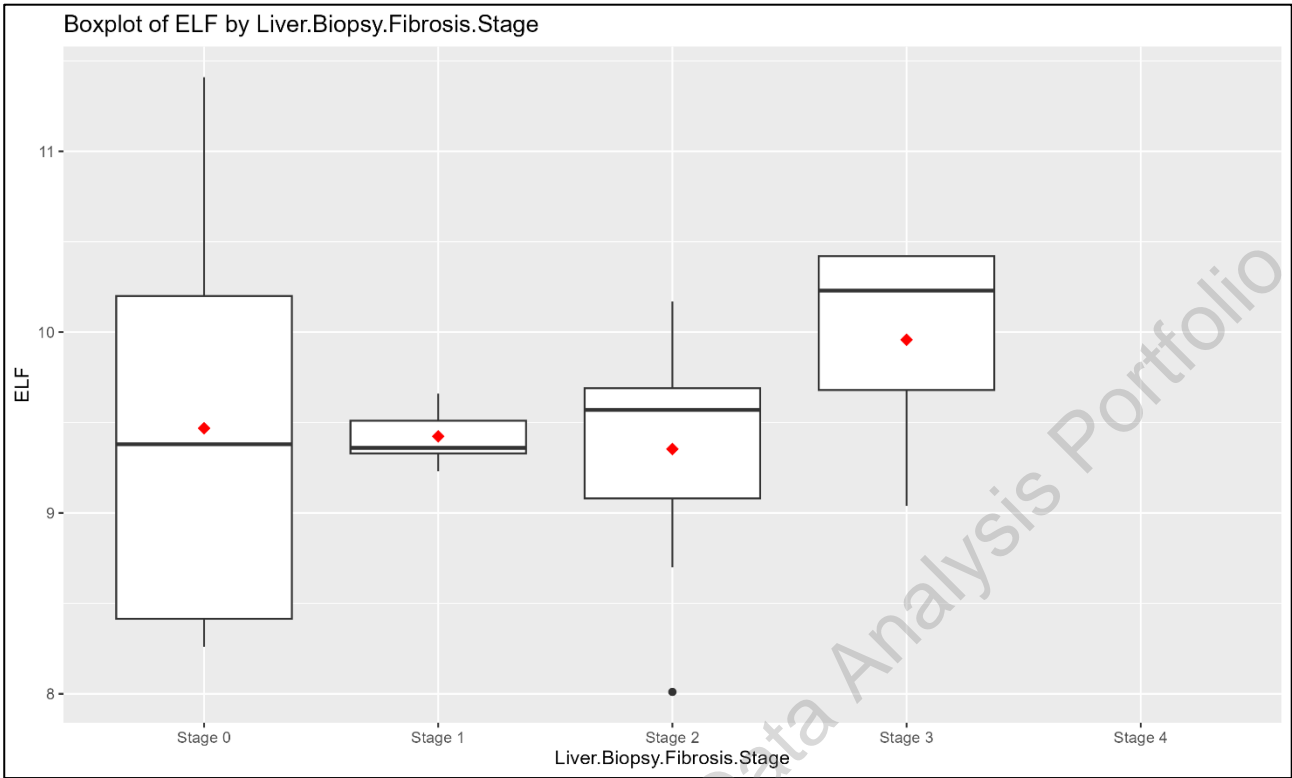
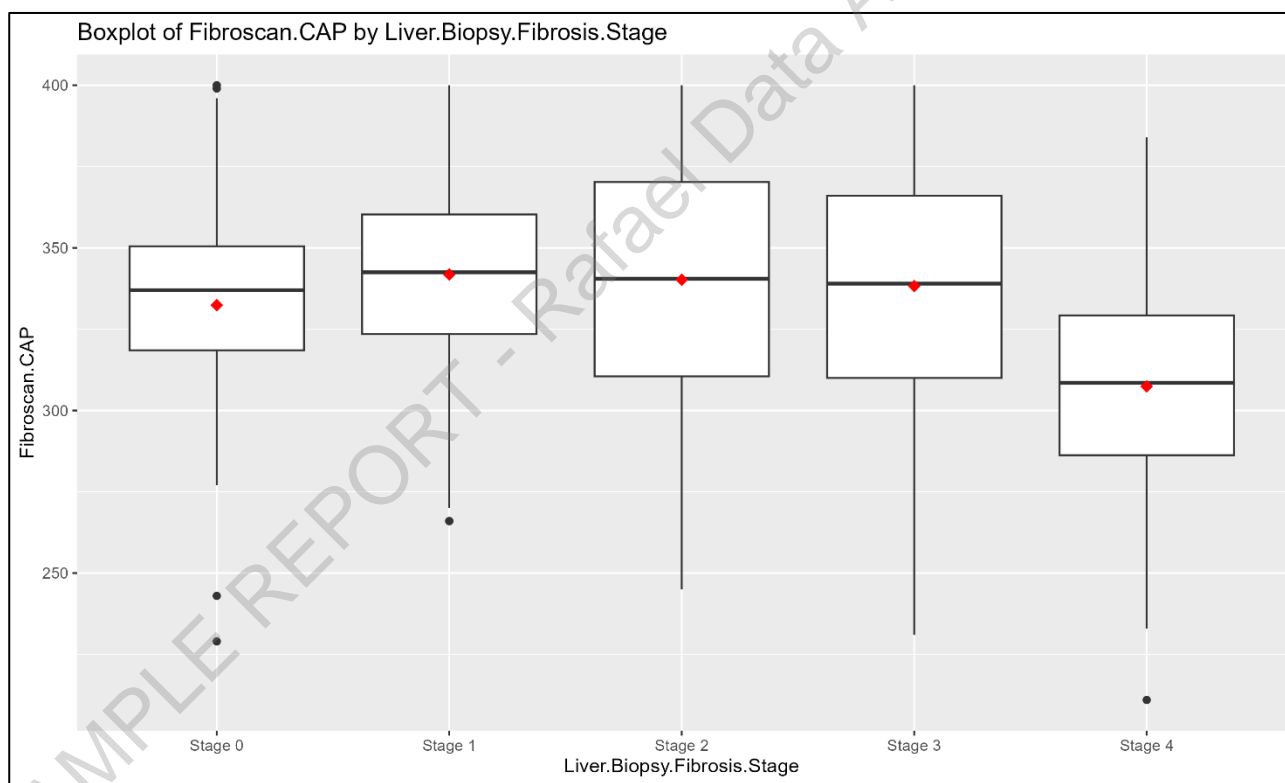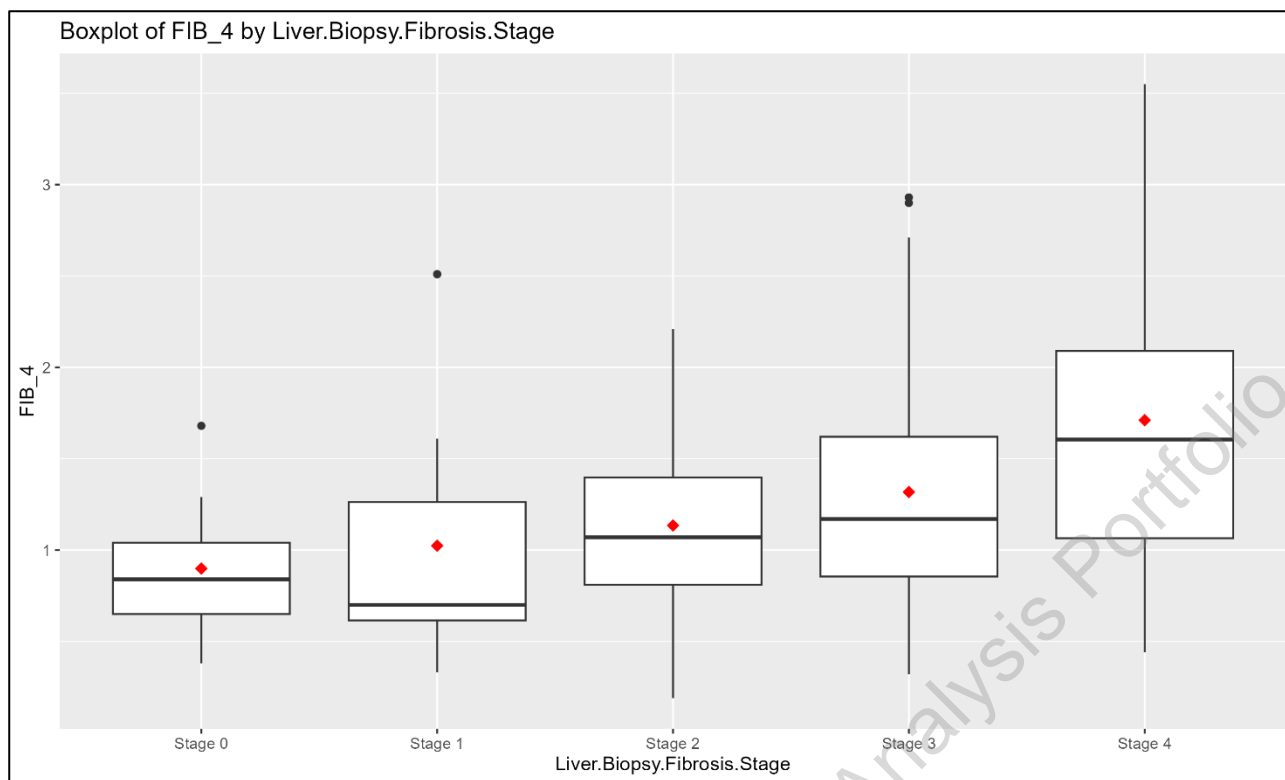| Variable	Sum_Sq	Mean_Sq	Df	F	p
FIB_4	8.353	2.088	4	6.109	0.000
ELF	1.337	0.446	3	0.783	0.514
FAST	2.248	0.562	4	10.543	0.000
Fibroscan.CAP	21438.033	5359.508	4	3.009	0.019
Fibroscan.E._kPA_	1412.212	353.053	4	0.894	0.468

The FIB-4 index exhibited notable statistical significance with a sum of squares (Sum_Sq) of 8.353, a mean square (Mean_Sq) of 2.088 across 4 degrees of freedom (Df), and an F-value of 6.109 ($p < .001$). This indicates significant variability in FIB-4 index values across liver fibrosis stages.

Similarly, the FAST score demonstrated significant mean differences across stages with a Sum_Sq of 2.248, a Mean_Sq of 0.562, also over 4 degrees of freedom, and an F-value of 10.543 ($p < .001$). This suggests that the FAST score varies significantly with the progression of liver fibrosis.

The Fibroscan CAP also showed statistical significance, with a substantial total Sum_Sq of 21438.033 and a Mean_Sq of 5359.508 across 4 degrees of freedom. The F-value for this measure was 3.009, with a p-value of .019.

To better visualize the differences across stages, the following box plots were created for each measure.

Boxplot of ELF by Liver.Biopsy.Fibrosis.Stage

Boxplot of FAST by Liver.Biopsy.Fibrosis.Stage

Boxplot of FIB_4 by Liver.Biopsy.Fibrosis.Stage

Boxplot of Fibroscan.CAP by Liver.Biopsy.Fibrosis.Stage

Boxplot of Fibroscan.E._kPA_ by Liver.Biopsy.Fibrosis.Stage
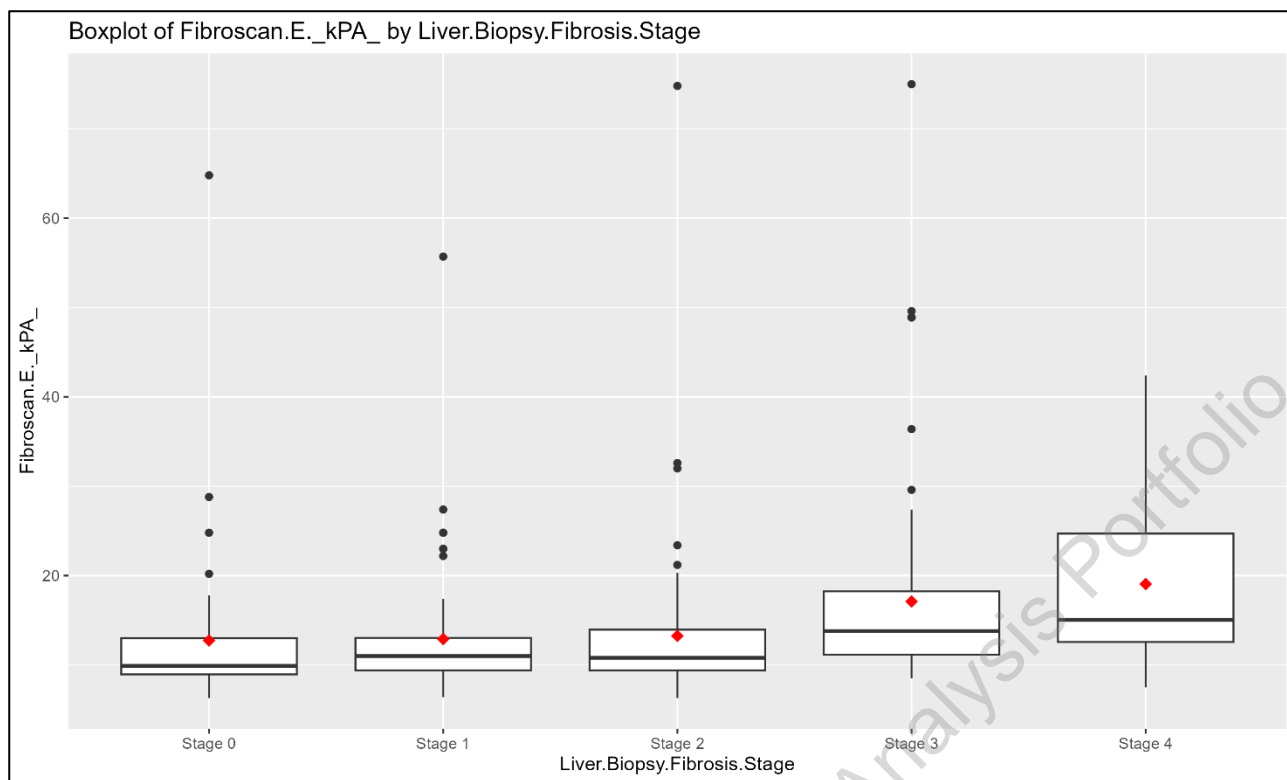
Models

In an assessment of liver fibrosis across various stages, multinomial logistic regression models were employed using the VGAM package in R. These models facilitated an exploration of factors influencing the progression of liver fibrosis from Stage 0, which served as the reference category. The analyses included city-specific controls to account for potential locational effects, with Brownsville designated as the reference category against Edinburg. This approach ensured that the potential influence of geographic location on liver fibrosis outcomes was addressed.

Fib

The model focusing on the Fibroscan index (FIB) demonstrated notable fit metrics, with a deviance of 497.526, a log likelihood of -248.763, and a residual deviance of 804.127. The model went through 6 iterations and achieved a rank of 12, indicating robust model performance.

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
(Intercept):1	0.071	1.127	0.063	0.949	1.074
(Intercept):2	0.808	0.981	0.823	0.410	2.243
(Intercept):3	-0.240	1.011	-0.237	0.813	0.787
(Intercept):4	-4.188	1.524	-2.747	0.006	0.015
FIB_4:1	0.692	1.000	0.692	0.489	1.997
FIB_4:2	1.099	0.880	1.248	0.212	3.000
FIB_4:3	1.661	0.881	1.884	0.060	5.262
FIB_4:4	2.549	0.926	2.753	0.006	12.797
Edinburg:1	-0.245	0.881	-0.278	0.781	0.783
Edinburg:2	0.274	0.764	0.359	0.720	1.315
Edinburg:3	0.749	0.786	0.953	0.340	2.116
Edinburg:4	2.375	1.279	1.857	0.063	10.751

Coefficient analysis for the FIB variable revealed the effects of this biomarker across different stages of fibrosis compared to Stage 0. The most significant findings were in the upper stages of fibrosis:

In Stage 4, the coefficient for FIB was 2.549, which was statistically significant (p = .006), suggesting a strong and significant increase in the odds of being in Stage 4 fibrosis compared to Stage 0, with an odds ratio of 12.797.

Similarly, Stage 3 showed a coefficient of 1.661 (p = .060), indicating a nearly significant result and an odds ratio of 5.262, highlighting a marked increase in the likelihood of being in Stage 3 relative to Stage 0.

ELF

The ELF model exhibited moderate fit statistics with a deviance of 80.069, a log likelihood of -40.035, and a residual deviance of 89.705. The model reached convergence after 4 iterations and had a rank of 9. No significant coefficients were observed for ELF across any stages, indicating that ELF values did not significantly vary by stage in this analysis.

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
(Intercept):1	1.243	7.240	0.172	0.864	3.466
(Intercept):2	2.792	6.454	0.433	0.665	16.312
(Intercept):3	-8.280	7.787	-1.063	0.288	0.000
ELF:1	-0.059	0.765	-0.076	0.939	0.943
ELF:2	-0.224	0.681	-0.328	0.743	0.800
ELF:3	0.805	0.815	0.987	0.323	2.236
Edinburg:1	-0.912	1.097	-0.831	0.406	0.402
Edinburg:2	-0.350	1.047	-0.334	0.738	0.705
Edinburg:3	0.179	1.449	0.124	0.902	1.196

FAST

The FAST model presented robust fit statistics with a deviance of 800.891, a log likelihood of -400.445, and a residual deviance of 1099.398. The model underwent 6 iterations and achieved a rank of 12. Notably, significant findings were observed in the coefficients:

Stage 4: The FAST:4 coefficient was 4.565, indicating a significant increase (p < .001) in the likelihood of being in Stage 4 fibrosis compared to Stage 0, with an odds ratio of 96.065.

Stage 2: The FAST:2 coefficient was 4.049, also significant (p < .001), with an odds ratio of 57.339, highlighting its predictive value across advancing stages.

Stage 3: Similarly, the FAST:3 coefficient was 4.388 (p < .001), showing a significant likelihood of being in Stage 3, with an odds ratio of 80.472.

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
(Intercept):1	0.436	0.507	0.860	0.390	1.547
(Intercept):2	-0.995	0.562	-1.770	0.077	0.370
(Intercept):3	-1.562	0.596	-2.620	0.009	0.210
(Intercept):4	-4.227	1.215	-3.480	0.001	0.015
FAST:1	1.604	0.921	1.741	0.082	4.972
FAST:2	4.049	0.927	4.370	0.000	57.339
FAST:3	4.388	0.930	4.719	0.000	80.472
FAST:4	4.565	1.194	3.822	0.000	96.065
Edinburg:1	-0.816	0.458	-1.784	0.074	0.442
Edinburg:2	-0.240	0.481	-0.499	0.618	0.786
Edinburg:3	0.299	0.506	0.591	0.555	1.349
Edinburg:4	1.874	1.099	1.705	0.088	6.515

Fibroscan

The Fibroscan model demonstrated substantial fit with a deviance of 807.557, a log likelihood of -403.779, and a residual deviance of 1254.592. After 6 iterations, the model achieved a rank of 16. Significant coefficients were:

Stage 4: For Fibroscan CAP, the coefficient for Stage 4 was -0.018, which was significant (p = .011) with an odds ratio of 0.982, suggesting a slight decrease in odds compared to Stage 0.

Stage 4: Fibroscan Elasticity (kPA) at Stage 4 also showed significance with a coefficient of 0.074 (p = .025), indicating an increase in odds with an odds ratio of 1.077.

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
(Intercept):1	-1.787	1.919	-0.931	0.352	0.168
(Intercept):2	-1.373	1.839	-0.746	0.455	0.253
(Intercept):3	-1.029	1.814	-0.567	0.571	0.358
(Intercept):4	2.430	2.447	0.993	0.321	11.359
Fibroscan.CAP:1	0.009	0.006	1.479	0.139	1.009
Fibroscan.CAP:2	0.006	0.006	1.103	0.270	1.006
Fibroscan.CAP:3	0.002	0.005	0.316	0.752	1.002
Fibroscan.CAP:4	-0.018	0.007	-2.550	0.011	0.982
Fibroscan.E._kPA_:1	-0.004	0.036	-0.115	0.909	0.996
Fibroscan.E._kPA_:2	0.007	0.033	0.217	0.829	1.007
Fibroscan.E._kPA_:3	0.057	0.030	1.895	0.058	1.059
Fibroscan.E._kPA_:4	0.074	0.033	2.248	0.025	1.077
Edinburg:1	-0.934	0.458	-2.041	0.041	0.393
Edinburg:2	-0.341	0.463	-0.738	0.461	0.711
Edinburg:3	0.233	0.492	0.473	0.636	1.262
Edinburg:4	1.776	1.102	1.612	0.107	5.905

Comorbidities

The model demonstrated a deviance of 748.252, a log likelihood of -374.126, and a residual deviance of 989.088.

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
(Intercept):1	17.529	1249.063	0.014	0.989	
(Intercept):2	16.864	1249.063	0.014	0.989	
(Intercept):3	16.354	1249.063	0.013	0.990	
(Intercept):4	-0.839	1568.179	-0.001	1.000	
Type.II.Diabetic:1	0.477	0.434	1.099	0.272	1.611
Type.II.Diabetic:2	0.364	0.425	0.858	0.391	1.439
Type.II.Diabetic:3	1.234	0.437	2.823	0.005	3.434
Type.II.Diabetic:4	1.450	0.604	2.400	0.016	4.264
Hypertension:1	0.563	0.442	1.275	0.202	1.756
Hypertension:2	0.741	0.428	1.729	0.084	2.097
Hypertension:3	1.063	0.445	2.390	0.017	2.896

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
Hypertension:4	1.401	0.638	2.195	0.028	4.060
High.Cholesterol:1	0.288	0.439	0.656	0.512	1.334
High.Cholesterol:2	0.120	0.428	0.280	0.779	1.127
High.Cholesterol:3	0.927	0.440	2.108	0.035	2.527
High.Cholesterol:4	0.081	0.576	0.141	0.888	1.085
Obesity:1	-17.050	1249.063	-0.014	0.989	0.000
Obesity:2	-16.553	1249.063	-0.013	0.989	0.000
Obesity:3	-17.578	1249.063	-0.014	0.989	0.000
Obesity:4	-17.586	1249.063	-0.014	0.989	0.000
Edinburg:1	-1.087	0.500	-2.174	0.030	0.337
Edinburg:2	-0.616	0.504	-1.221	0.222	0.540
Edinburg:3	-0.136	0.551	-0.247	0.805	0.873
Edinburg:4	16.153	948.170	0.017	0.986	1.04E+07

The analysis of comorbidities revealed that certain conditions significantly impacted the progression to higher stages of fibrosis:

Type II Diabetes:

For Stage 3, the coefficient of 1.234 was highly significant (p = .005), with an odds ratio of 3.434, suggesting a more than threefold increase in the likelihood of advancing to this stage compared to Stage 0 when Type II Diabetes is present.

In Stage 4, the presence of Type II Diabetes further increased the odds, with a coefficient of 1.450 (p = .016) and an odds ratio of 4.264, indicating a strong influence on progressing to the most severe stage of fibrosis.

Hypertension:

Hypertension also showed significant results, particularly in Stage 3, where the coefficient of 1.063 (p = .017) translated to an odds ratio of 2.896.

Stage 4 hypertension showed a coefficient of 1.401 (p = .028), suggesting a significant increase in the likelihood of being diagnosed with Stage 4 fibrosis, with an odds ratio of 4.060.

High Cholesterol:

Stage 3 was significantly affected by high cholesterol, where the coefficient was 0.927 (p = .035), yielding an odds ratio of 2.527, indicating a substantive increase in the risk of advancing to this fibrosis stage.

Control Variable (Edinburg):

The inclusion of Edinburg as a variable revealed a significant decrease in the odds for Stage 1 fibrosis with a coefficient of -1.087 (p = .030) and an odds ratio of 0.337, suggesting that residing in Edinburg compared to Brownsville significantly reduces the likelihood of being in Stage 1 fibrosis.

Obesity

The frequency table indicates that among 284 observed cases, the distribution of obesity levels is as follows: 41.9% in Class 3 Obesity, 29.6% in Class 2 Obesity, and 18.3% in Class 1 Obesity. Smaller proportions fall into other categories, with 9.2% classified as Overweight, 0.4% as Healthy, and 0.7% as Unknown (missing values). The healthy and unknown cases were filtered out of the dataset due to poor representativeness.

Level	Count	%
Class 1 Obesity	52	18.3
Class 2 Obesity	84	29.6
Class 3 Obesity	119	41.9
Healthy	1	0.4
Overweight	26	9.2
Unknown	2	0.7

The model results provided robust fit statistics with a deviance of 843.621, a log likelihood of -421.811, and a residual deviance of 1098. The model has 1108 residual degrees of freedom, 1124 total degrees of freedom, 15 iterations, and a rank of 16. These statistics suggest that the model has a moderate level of fit. The coefficients can be found below.

Index	Estimate	Std. Error	z value	Pr(>|z|)	OddsRatios
(Intercept):1	0.405	0.456	0.888	0.374	1.500
(Intercept):2	0.811	0.425	1.908	0.056	2.250
(Intercept):3	0.223	0.474	0.470	0.638	1.250
(Intercept):4	-0.693	0.612	-1.132	0.258	0.500
Class 2 Obesity:1	0.245	0.579	0.423	0.672	1.278
Class 2 Obesity:2	-0.405	0.565	-0.717	0.473	0.667
Class 2 Obesity:3	0.182	0.603	0.302	0.762	1.200
Class 2 Obesity:4	0.773	0.732	1.057	0.291	2.167
Class 3 Obesity:1	-0.405	0.543	-0.746	0.456	0.667
Class 3 Obesity:2	-0.481	0.505	-0.951	0.341	0.618
Class 3 Obesity:3	0.225	0.544	0.413	0.679	1.252
Class 3 Obesity:4	-0.833	0.786	-1.059	0.290	0.435
Overweight:1	15.137	543.718	0.028	0.978	3.75E+06
Overweight:2	14.578	543.718	0.027	0.979	2.14E+06
Overweight:3	15.676	543.718	0.029	0.977	6.43E+06
Overweight:4	15.389	543.718	0.028	0.977	4.82E+06

Examining the coefficients for the obesity-related coefficients, Class 2 Obesity and Class 3 Obesity produced statistically insignificant results, with p-values far exceeding the typical threshold of 0.05. This pattern suggests that the various obesity classes may not be significant predictors of liver fibrosis stage in this model, compared to lower classes. The overweight category yielded extremely high odds ratios and coefficients, which is an

indication of poor representativeness of this category across come fibrosis stages. Consequently, p-values are insignificant too.

Overall, while the frequency distribution provides a clear picture of the case breakdown, the model results show limited significance for predicting liver fibrosis stage from obesity levels.