The provided hospital dataset consist of contains 4 variables and 7288 observations. Those four variables include ID, AGE, BMI and FEV.

From the summary statistics, we can see that there are outliers in BMI and FEV variable in the dataset.

BMI extreme observations table

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
12.20	6472	45.50	1257
12.70	4379	63.20	696
12.70	4378	99.90	1616
12.80	4382	1214.34	1042
13.06	3430	1654.38	3937

FEV extreme observations table

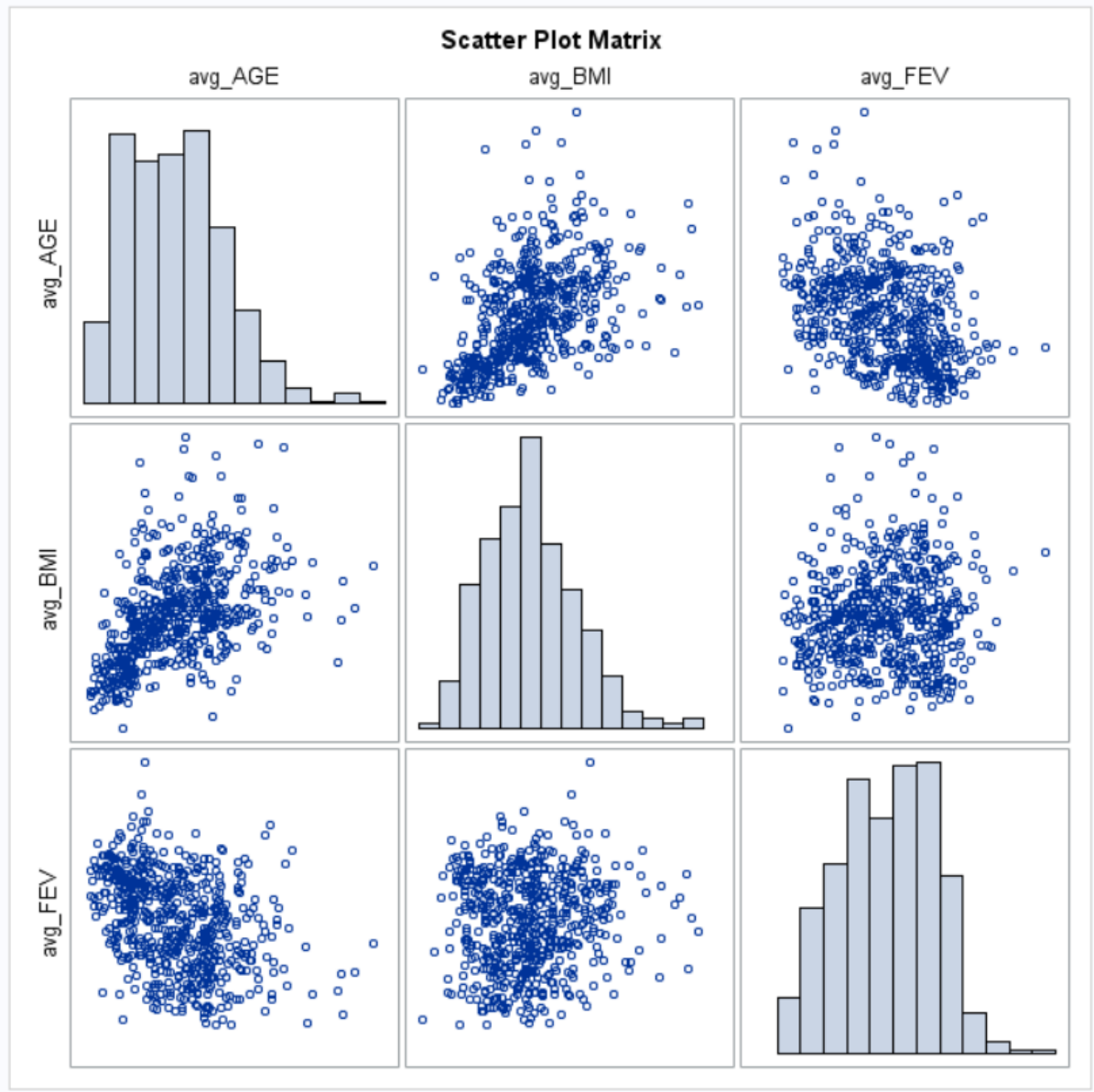
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
4.43	7099	157	1777
10.00	7124	158	1775
11.00	5901	165	1773
11.78	3806	165	1776
13.00	5254	581	6191

From these summary statistics tables of BMI and FEV, we see that there are 2 outliers in BMI(obs 1042 & 3937) and one outlier in FEV(obs 6191). So we will have to remove these three observations before moving forward.

After removing these and creating a correlation matrix between AGE, BMI and FEV, we see that age and BMI are positively correlated with each other and has a p value less than 0.001 which shows

that these variables are highly significant at 95% confidence level. The correlation between AGE and FEV is negatively correlated at 95% confidence level. BMI and FEV are not related as the correlation value is very small and almost equal to zero.

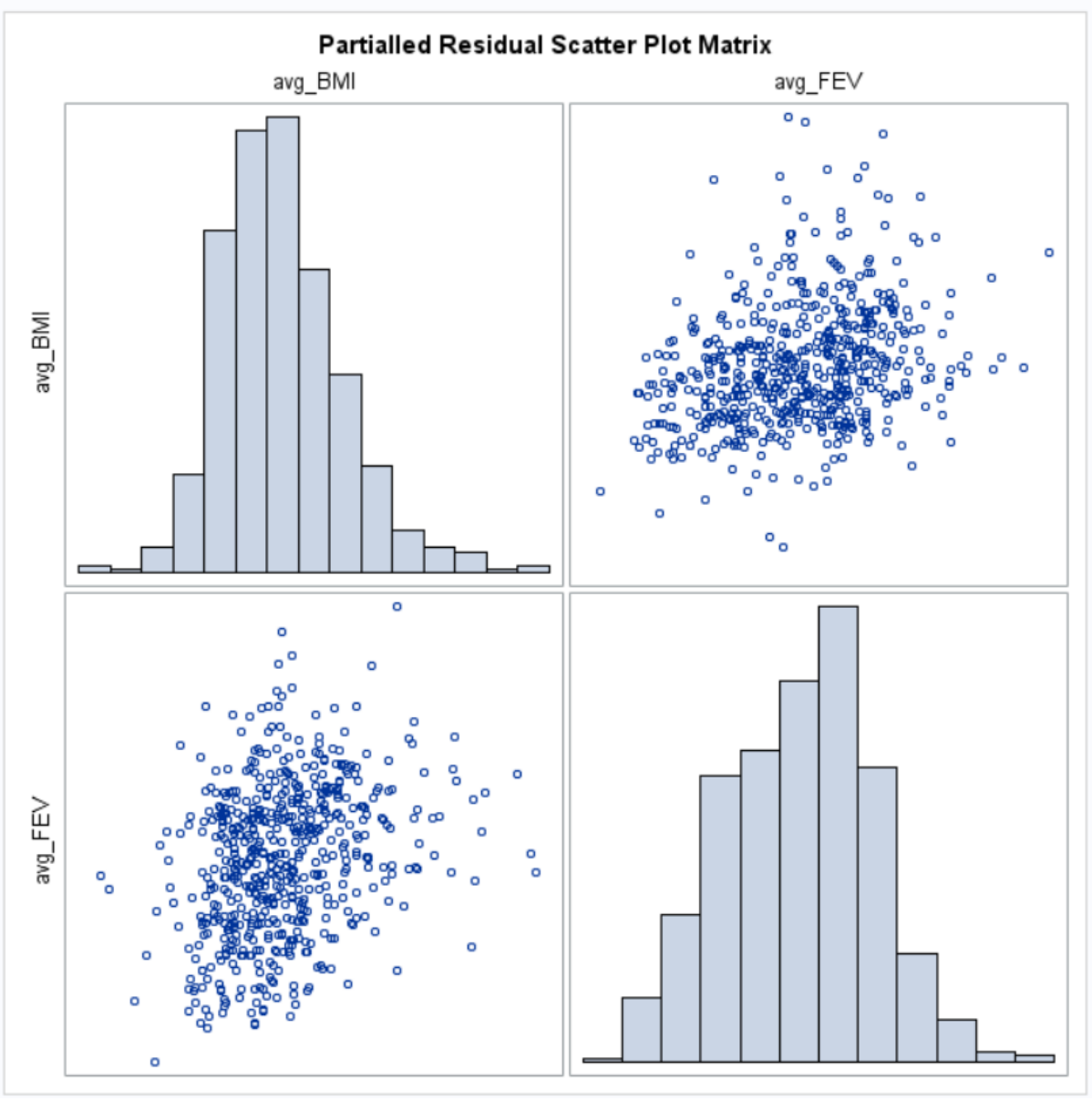
Pearson Correlation Coefficients, N = 589 Prob > |r| under H0: Rho=0			
	avg_AGE	avg_BMI	avg_FEV
avg_AGE	1.00000	0.48185 <.0001	-0.38027 <.0001
avg_BMI	0.48185 <.0001	1.00000	0.05972 0.1477
avg_FEV	-0.38027 <.0001	0.05972 0.1477	1.00000



The plot also shows the same relation as mentioned before.

From partial correlation on BMI and FEV with AGE, we can see that both the variables are slightly related to each other with AGE as the medium. So, any changes in BMI will affect the FEV variable. These two variables are related with each other at 95% confidence interval with their age.

Pearson Partial Correlation Coefficients, N = 589 Prob > |r| under H0: Partial Rho=0		
	avg_BMI	avg_FEV
avg_BMI	1.00000	0.29979 <.0001
avg_FEV	0.29979 <.0001	1.00000

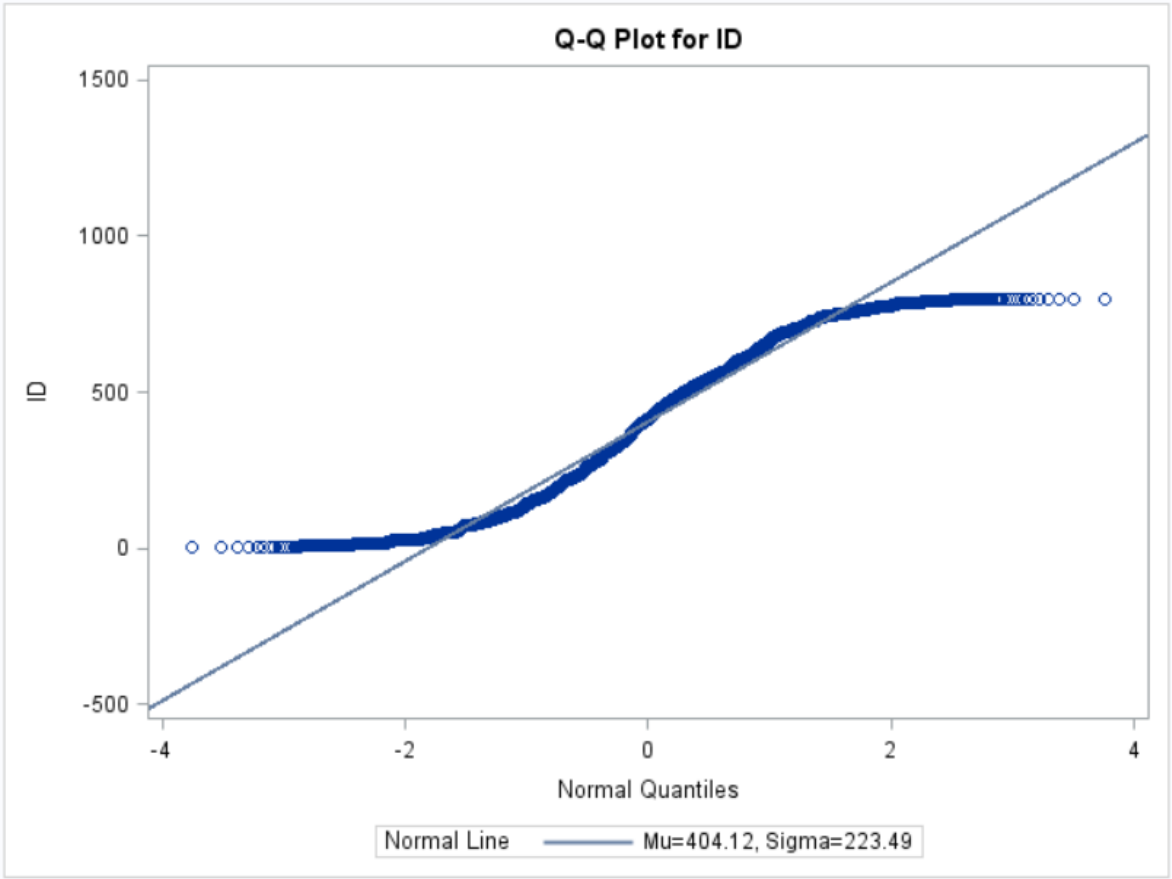


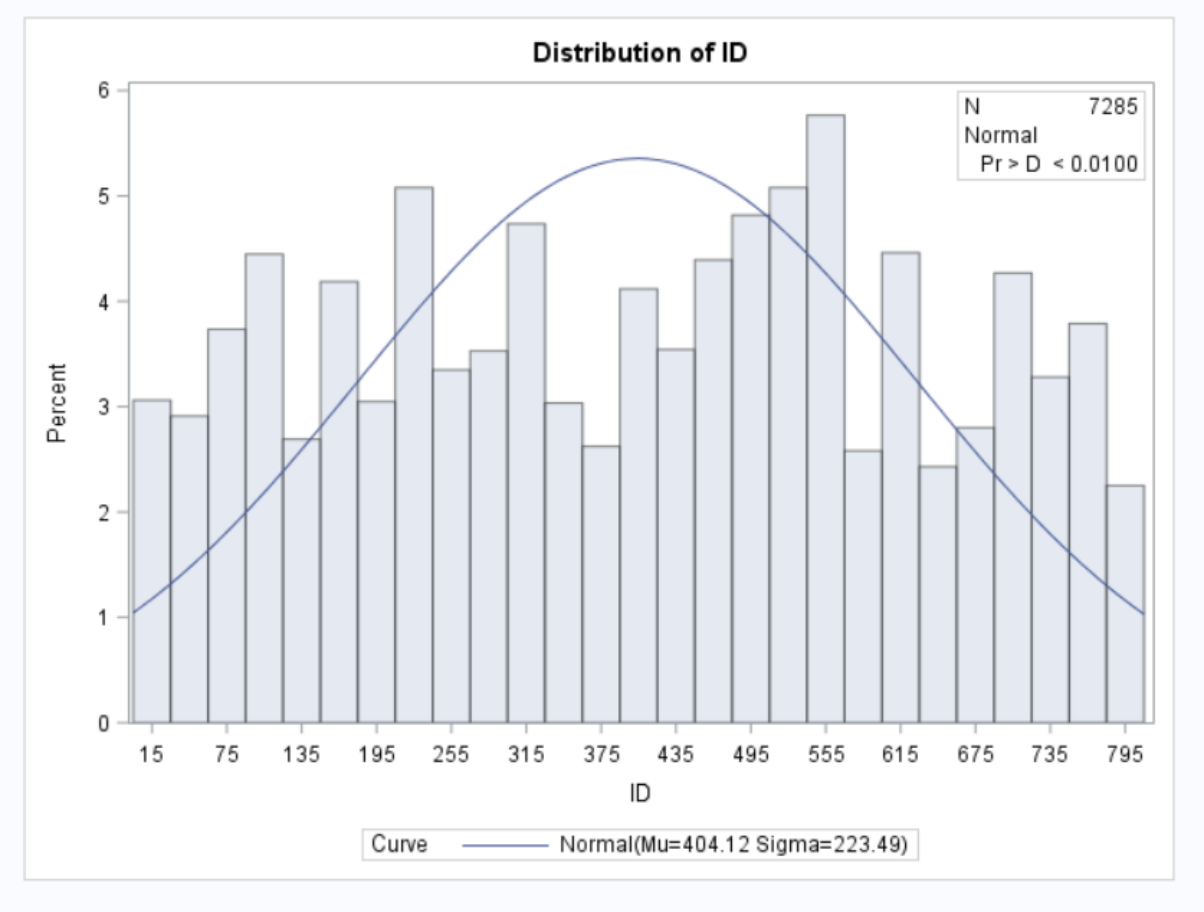
The plot also proves that the varibales BMI and FEV are slightly related to each other.

From the GLM model on AGE and BMI with FEV, we see that AGE and BMI have significant influence on the lung function of the patients and any change in these variables will impact the lungs functionality.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	59.94776481	1.57939307	37.96	<.0001
AGE	-1.20789665	0.03236153	-37.33	<.0001
BMI	1.71839059	0.08311414	20.68	<.0001

The residual versus fitted values show that the variables are linearly related to each other as the qq plot converges with the line of the model.





The above plot shows that the model values are normally distributed as there is a bell shaped curve in the plot above.