

## Oral Lipid Tolerance Test

The assignment2.csv is an extract of data from an Oral Lipid Tolerance Test. It contains the observations based on the response of participants after consuming high fat drink. It has several observations based on which the participants were tested on, such as, ID, time, SEX AGE BMI, etc.

Using Univariate command on all variables, we observe statistical information of each variable with information on mean, standard deviation, quantiles, lowest and highest observations of the variable and count of empty values of the variable. From the statistical report on age, we see that 75% of the participants are at the age of 33.75 or below with half of them under the age of 27. From the BMI Quantiles table, we can see that 75% of the participants had the BMI of about 26.5 or below and 25% had above that and up till 33.3. So, three-fourth of the participants with an age of 33.75 or below, have the BMI of 26.5 or below. Assuming an average height of 165 cms of all the participants, this BMI would be slightly more than normal limit, indicating obese. But, since we are not provided with the information on participants much, we cannot justify this statement.

From the frequency command, we can see that among 414 participants, there were 192 males and 222 females. We can also cross verify this by looking at the quantiles table of sex variable which points out that more than 50% of the participants are female.

The UNIVARIATE Procedure			
Variable: SEX			
Frequency Counts			
Value	Count	Percents	
		Cell	Cum
1	192	46.4	46.4
2	222	53.6	100.0

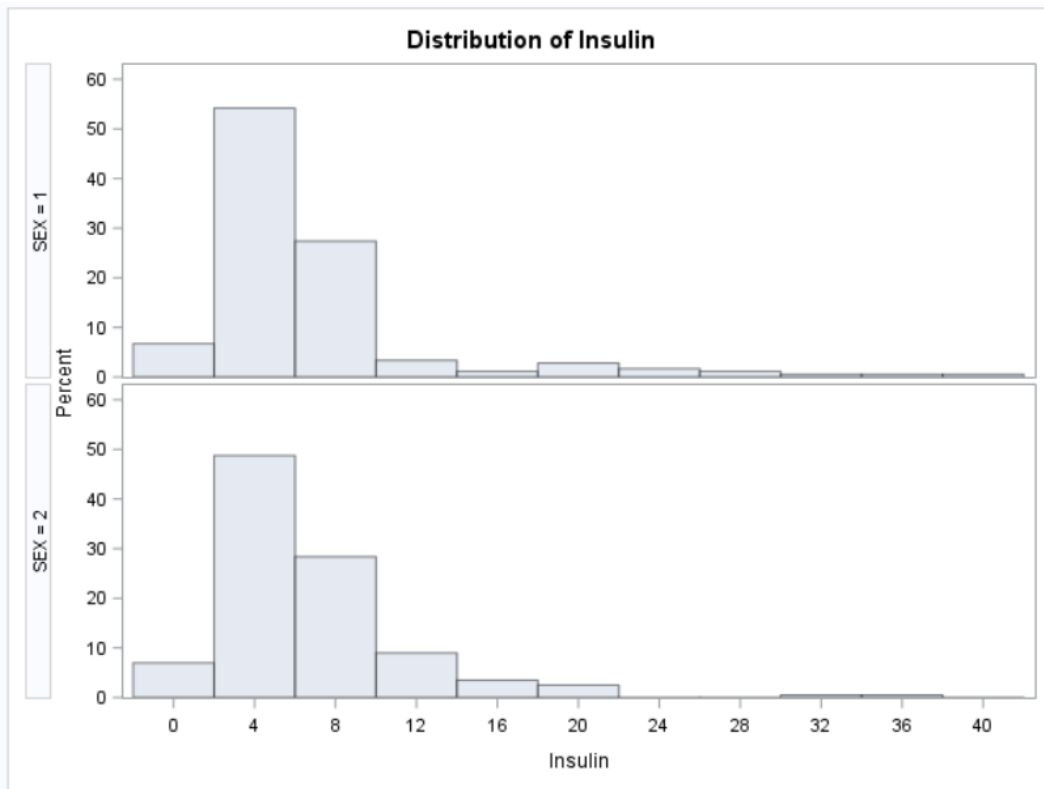
From the extreme observation tables, we see that the observations of ApoA1, ApoB and Insulin levels had huge variation among the lowest and highest levels. The values of other variables seem to increase gradually. There was not much difference in Non-essential fatty acid as the mean value was around 0.53 with a sd of 0.28. The glucose and cholesterol level also increased gradually.

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
3.20	314	6.42	80
3.47	170	6.43	102
3.59	392	6.48	98
3.79	152	6.60	356
3.81	285	6.96	338

Observations of Glucose

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.69	140	6.37	97
0.82	141	6.41	339
1.09	143	6.51	341
1.48	142	6.59	101
2.15	144	6.62	295

Observations of Cholesterol



From the group by on gender with Insulin variable, we see that the Insulin level of male was higher when compared with that of female participants and this is verified by the above plot. Using cibasic on BMI, we see that the average BMI of male participants was between 24.95 and 25.67 and for female participants, it was between 23.76 to 24.41.

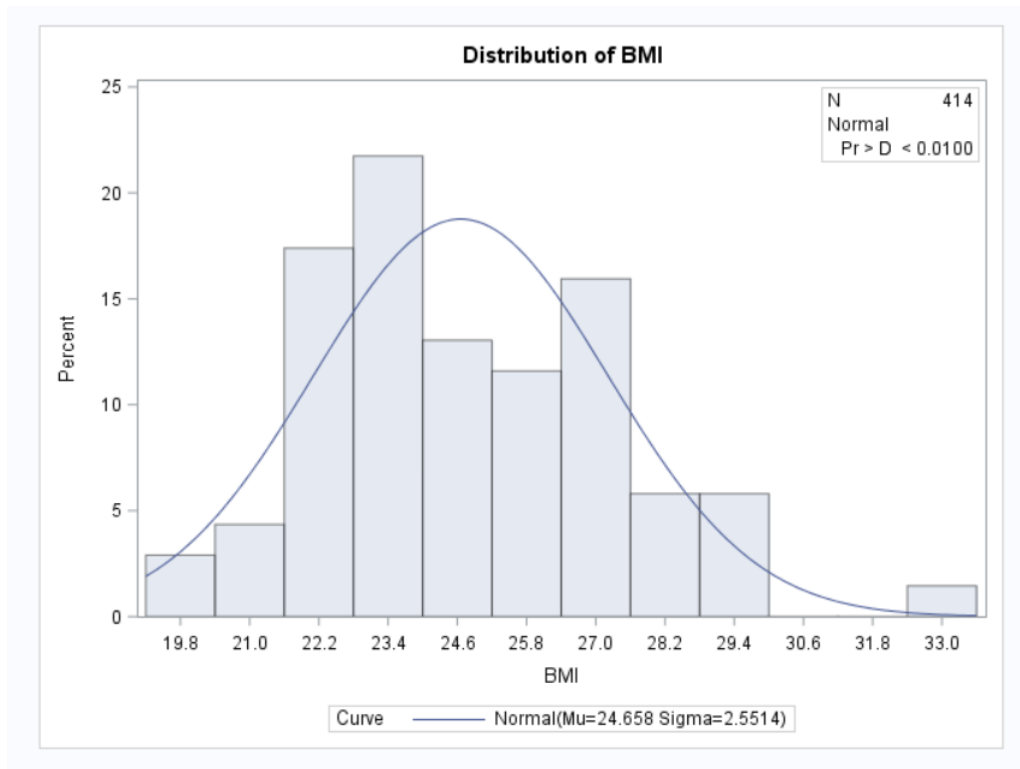
SEX=1			
Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	25.31563	24.95645	25.67480
Std Deviation	2.52316	2.29351	2.80432
Variance	6.36635	5.26020	7.86422

Confidence Interval of male participants

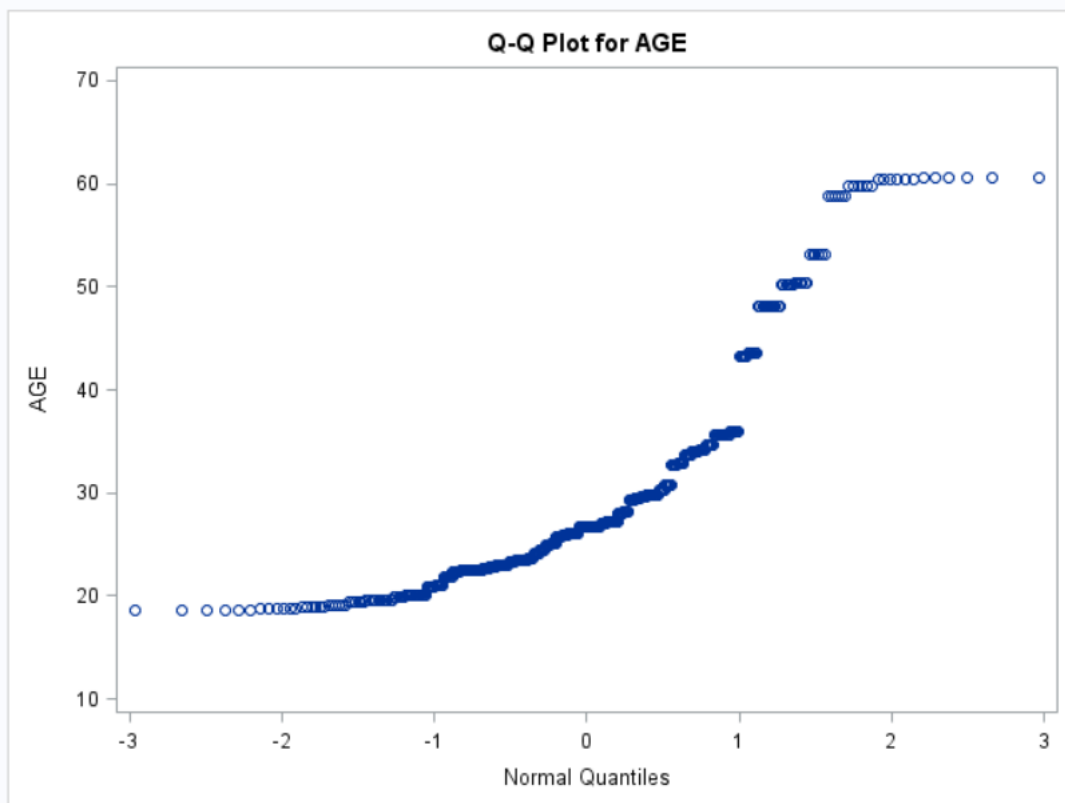
SEX=2			
Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	24.08919	23.76622	24.41216
Std Deviation	2.44181	2.23383	2.69281
Variance	5.96242	4.99001	7.25124

Confidence Interval of female participants

So, the BMI of male participants was slightly more than that of female participants and the BMI values of whole dataset are evenly spread in normal distribution.



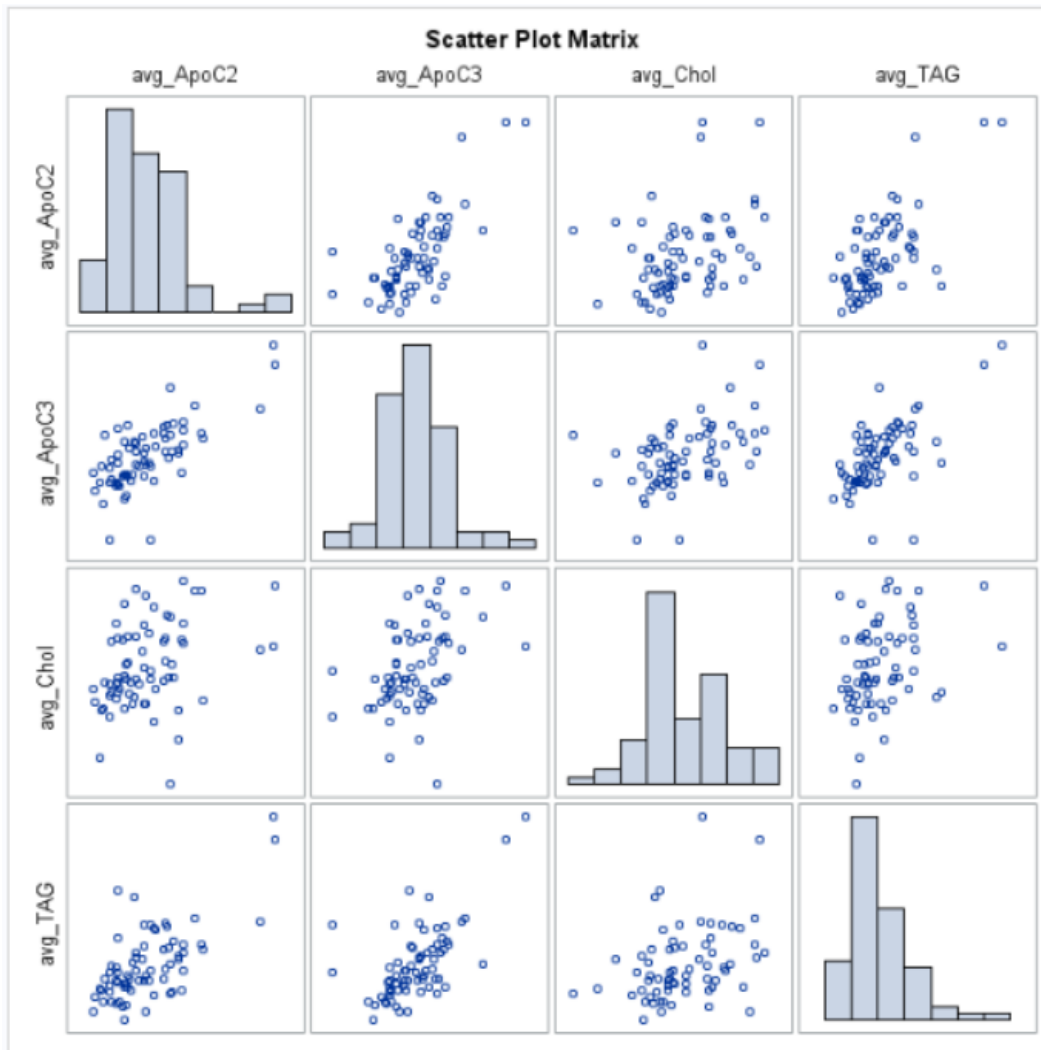
From the analysis on Q-Qplot and above descriptive statistics, we see that there are no outliers in the given dataset. The datapoints are stretched to a certain extent but it cannot be taken as outliers as the values play a significant role in the dataset. So, there are no outliers in the dataset.



Using correlation function, we can correlate the blood chemistry of several variables. From the table below, we see that the variable ApoA1 is not much correlated with any of the other variables. Variable ApoB is well correlated with ApoC2, TAG & ApoC3 with highly positive correlated with Cholesterol. ApoC2 and ApoC3 is well correlated with TAG and ApoC3 & ApoC2 are also highly correlated among each other. The relationships between ApoC2 and ApoC3 on Triglyceride are highly significant to each other as its p value is less than 0.0001. The variable ApoE is slightly correlated with ApoC2, ApoC3 and Cholesterol. Glucose levels have a relationship with ApoC2, ApoC3, Insulin, Cholesterol and Triglyceride. NEFA doesn't seem to have any relation with other variables as the correlation values are very less. Insulin and Triglyceride are well related to each other. As mentioned above, cholesterol and Apolipoprotein B are highly related to each other. So, any change in one variable will affect another. Also, the variables can be derived from one another and both provide similar data.

Since ApoC2, ApoC3, Cholesterol and Tag have higher correlation with each other, we can create a scatter plot of these variables for showing their relation. Also, we can create a scatter plot between Glucose and Insulin for verifying their relation among each other.

Pearson Correlation Coefficients Prob >  r  under H0: Rho=0 Number of Observations										
	avg_ApoA1	avg_ApoB	avg_ApoC2	avg_ApoC3	avg_ApoE	avg_Glucose	avg_NEFA	avg_Insulin	avg_Cholesterol	avg_TAG
avg_ApoA1	1.00000 69	0.00149 0.9903 69	-0.10102 0.4089 69	0.12901 0.2908 69	-0.01837 0.8809 69	-0.02625 0.8305 69	0.16913 0.1648 69	0.05913 0.6426 64	0.29778 0.0130 69	-0.01334 0.9134 69
avg_ApoB	0.00149 0.9903 69	1.00000 69	0.42145 0.0003 69	0.45234 <.0001 69	0.13771 0.2592 69	0.17670 0.1464 69	0.00525 0.9658 69	0.06748 0.5963 64	0.81518 <.0001 69	0.31118 0.0093 69
avg_ApoC2	-0.10102 0.4089 69	0.42145 0.0003 69	1.00000 69	0.71198 <.0001 69	0.29753 0.0130 69	0.29091 0.0153 69	-0.03115 0.7994 69	0.21812 0.0834 64	0.36515 0.0020 69	0.66683 <.0001 69
avg_ApoC3	0.12901 0.2908 69	0.45234 <.0001 69	0.71198 <.0001 69	1.00000 69	0.27343 0.0230 69	0.26238 0.0294 69	0.11440 0.3493 69	0.24633 0.0497 64	0.43674 0.0002 69	0.62085 <.0001 69
avg_ApoE	-0.01837 0.8809 69	0.13771 0.2592 69	0.29753 0.0130 69	0.27343 0.0230 69	1.00000 69	0.03506 0.7749 69	-0.09821 0.4221 69	0.04892 0.7011 64	0.27606 0.0217 69	0.19942 0.1004 69
avg_Glucose	-0.02625 0.8305 69	0.17670 0.1464 69	0.29091 0.0153 69	0.26238 0.0294 69	0.03506 0.7749 69	1.00000 69	-0.06062 0.6207 69	0.37926 0.0020 64	0.24717 0.0406 69	0.32285 0.0068 69
avg_NEFA	0.16913 0.1648 69	0.00525 0.9658 69	-0.03115 0.7994 69	0.11440 0.3493 69	-0.09821 0.4221 69	-0.06062 0.6207 69	1.00000 69	-0.08414 0.5086 64	0.07220 0.5555 69	-0.03346 0.7849 69
avg_Insulin	0.05913 0.6426 64	0.06748 0.5963 64	0.21812 0.0834 64	0.24633 0.0497 64	0.04892 0.7011 64	0.37926 0.0020 64	-0.08414 0.5086 64	1.00000 64	0.11177 0.3792 64	0.44795 0.0002 64
avg_Cholesterol	0.29778 0.0130 69	0.81518 <.0001 69	0.36515 0.0020 69	0.43674 0.0002 69	0.27606 0.0217 69	0.24717 0.0406 69	0.07220 0.5555 69	0.11177 0.3792 64	1.00000 69	0.34452 0.0037 69
avg_TAG	-0.01334 0.9134 69	0.31118 0.0093 69	0.66683 <.0001 69	0.62085 <.0001 69	0.19942 0.1004 69	0.32285 0.0068 69	-0.03346 0.7849 69	0.44795 0.0002 64	0.34452 0.0037 69	1.00000 69

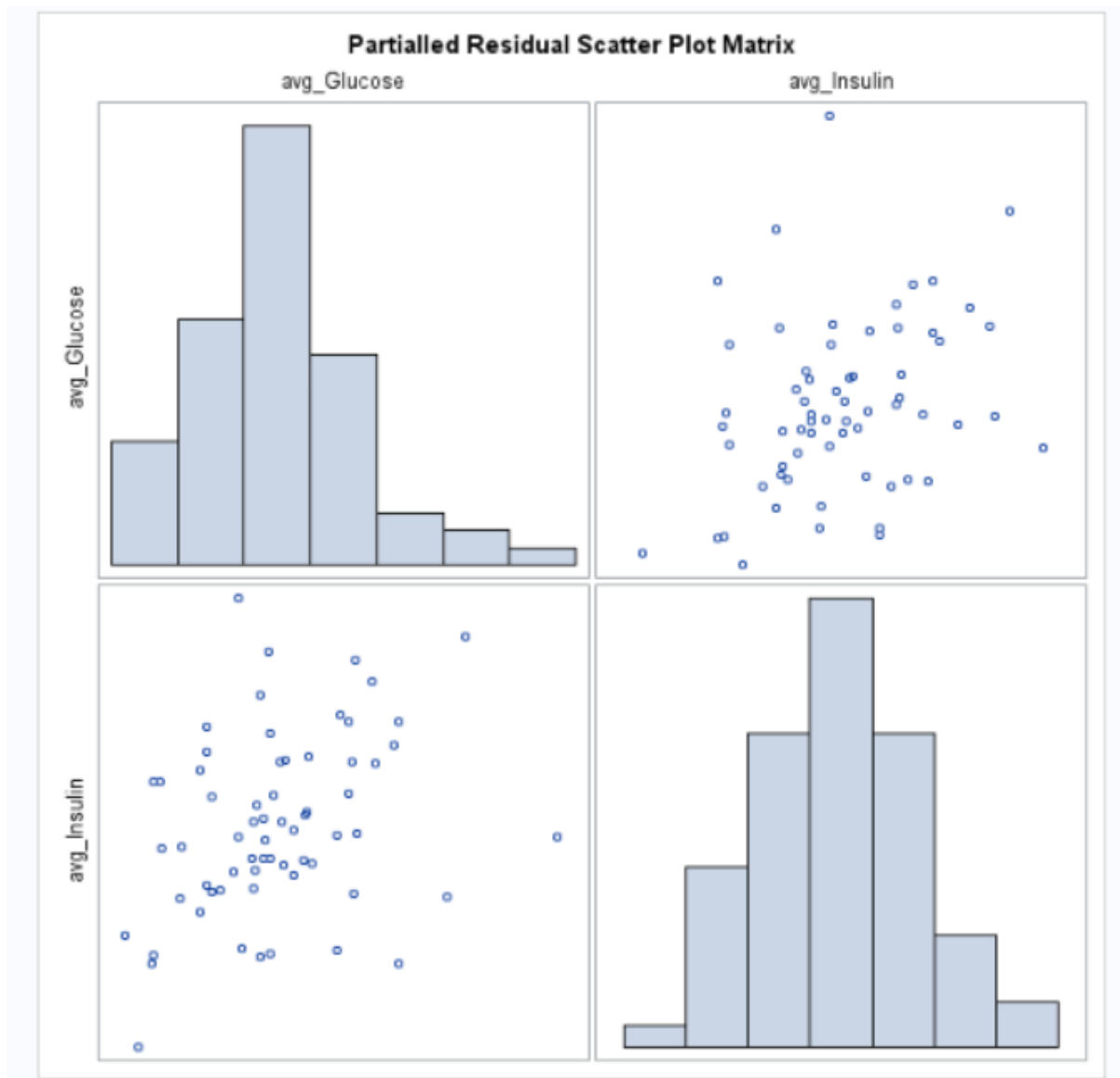


This plot shows that ApoC3 is positively correlated with ApoC2, Cholesterol and TAG as there is an upward pull of data points among the variables.

The CORR Procedure		
Pearson Partial Correlation Coefficients, N = 64 Prob >  r  under H0: Partial Rho=0		
	avg_Glucose	avg_Insulin
avg_Glucose	1.00000	0.31536 0.0118
avg_Insulin	0.31536 0.0118	1.00000

From the above table, we see that the variable Glucose and Insulin are positively correlated with a p value less than 0.05 which states that both the variables are weakly positively related with each other at 95% confidence level. We can see that there is an upward pull of the datapoints among glucose and

Insulin from the below scatter plot but the values are shattered which shows that both Insulin and Glucose have a weakly relationship.



We can build a model on Triglyceride variable with predicted variables – AGE, SEX and BMI for time  $t=0$  to check their relationship and the efficiency of the model. From the output values, our R - Square and adjusted R-square values are very less which states that this is not a good model. The p values of the estimate are also higher than 0.05 which shows that variables are not highly significant to the response variable Triglyceride.

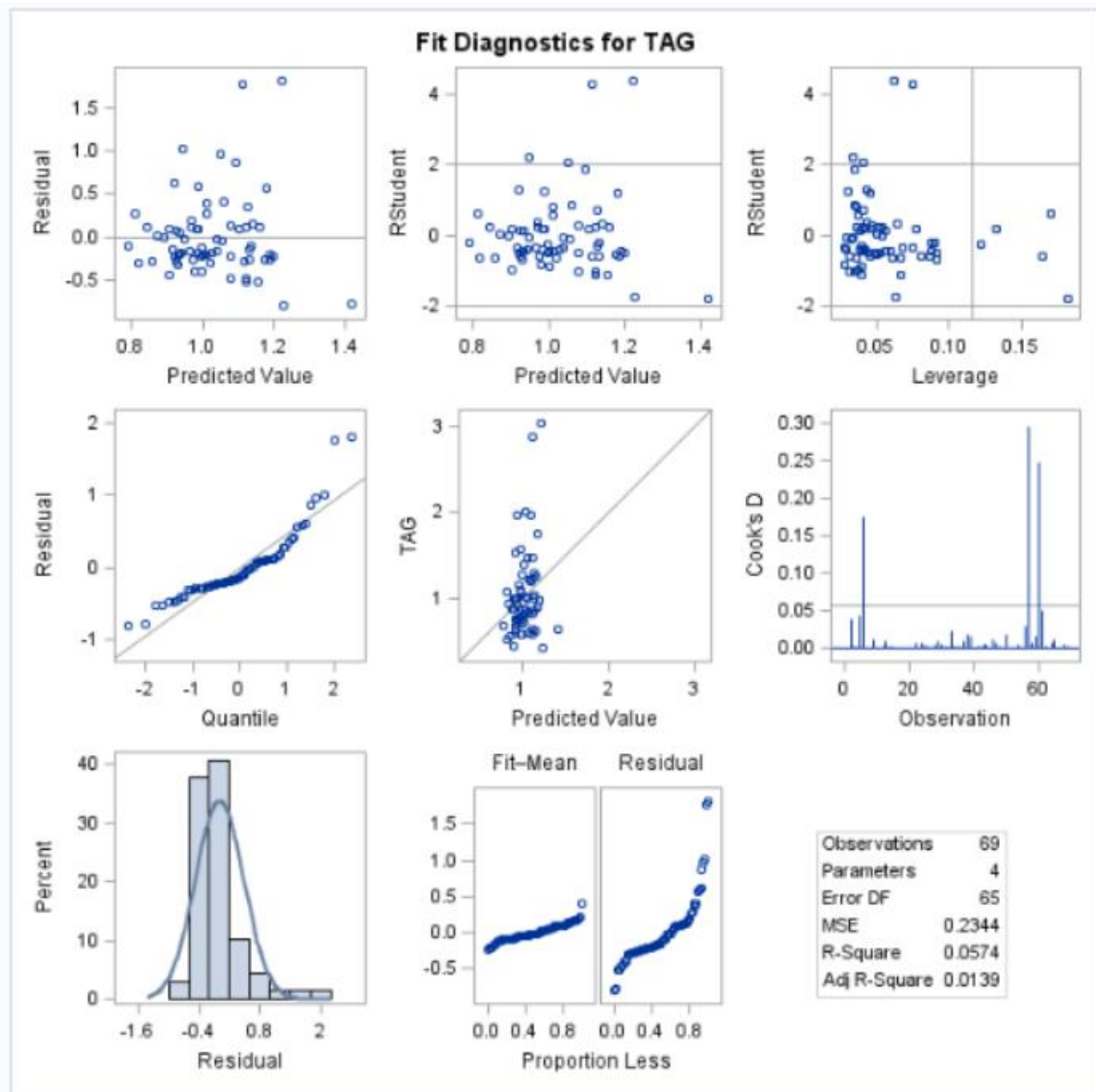
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.92714857	0.30904952	1.32	0.2760
Error	65	15.23535867	0.23439013		
Corrected Total	68	16.16250725			

R-Square	Coeff Var	Root MSE	TAG Mean
0.057364	47.20967	0.484139	1.025507

Source	DF	Type I SS	Mean Square	F Value	Pr > F
AGE	1	0.00005265	0.00005265	0.00	0.9881
SEX	1	0.13283504	0.13283504	0.57	0.4543
BMI	1	0.79426089	0.79426089	3.39	0.0702

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	1	0.00493400	0.00493400	0.02	0.8851
SEX	1	0.02043746	0.02043746	0.09	0.7687
BMI	1	0.79426089	0.79426089	3.39	0.0702

From the plots, we see that the values are spread randomly and few values are stretched away from the normal intervals of the data points. So, there is not much relationship with the variables among each other. The Q-Qplot also shows that the data are not linear and does not converge to a straight line exactly.

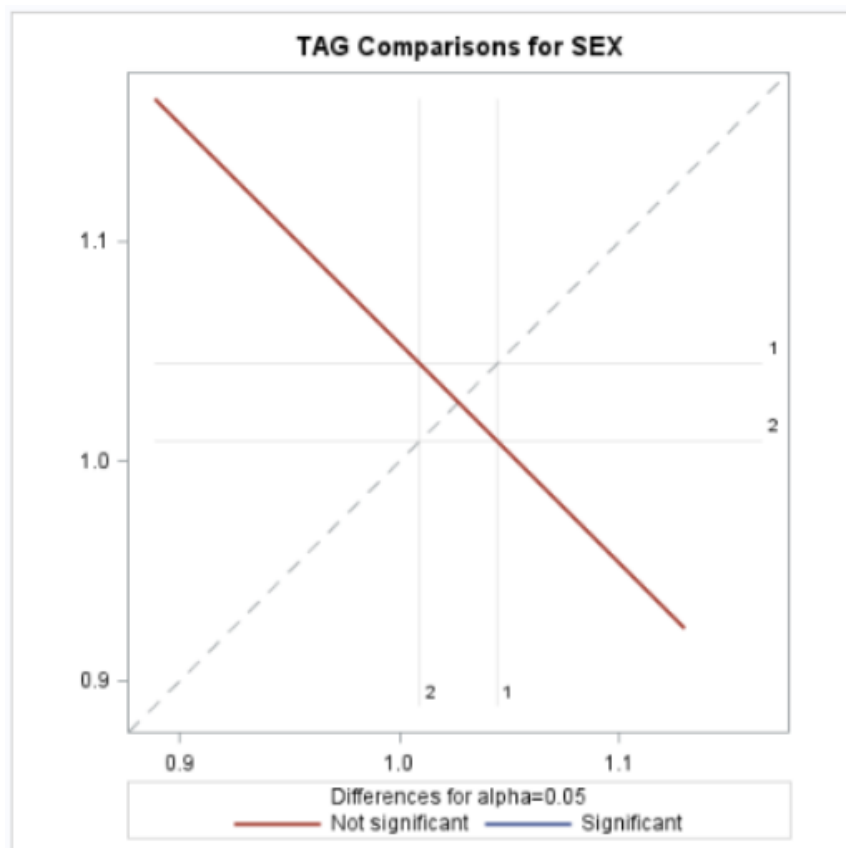
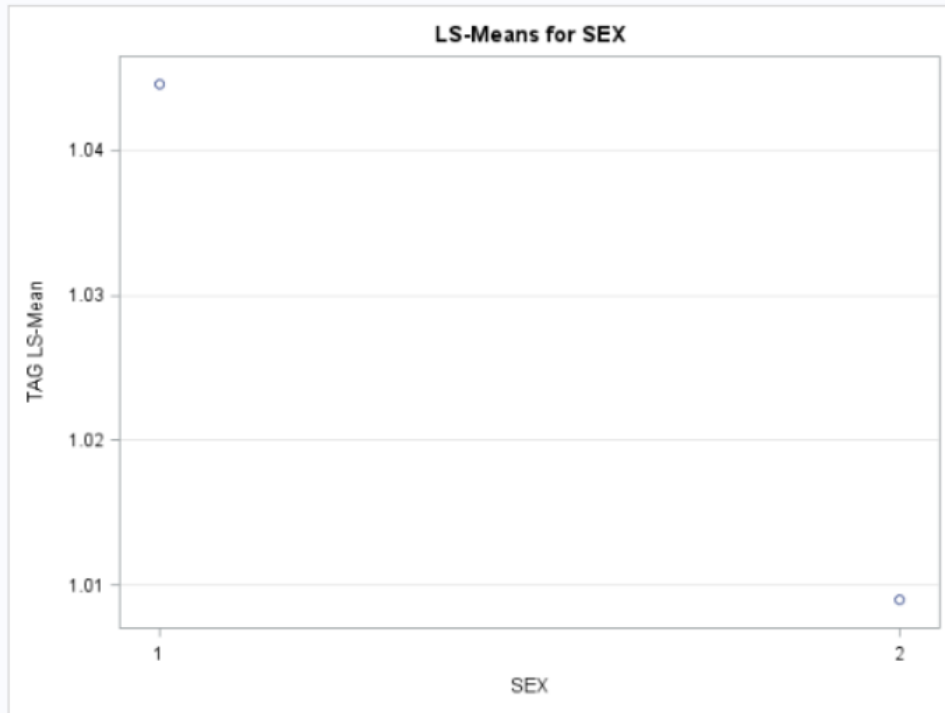


From the least square means values of male and female participants, we can see that the mean values differ for both the gender as the least squares mean of male participants is 1.0446 when compared with that of female participants which is 1.008. This difference of mean values can also be seen from the plot provided below. The corresponding p value is 0.7 which shows that the Triglyceride comparison is not significant at 95% confidence level.



# Least Squares Means

SEX	TAG LSMEAN	H0:LSMean1=LSMean2
		Pr >  t
1	1.04460424	0.7687
2	1.00899093	



When creating a model using 'genmod' function on Triglyceride variable and predicted variables – AGE, SEX and BMI with log-normal distribution on the response variable for time t=0, we get several information about the model such as goodness of Fit, Maximum Likelihood analysis and diagnostic statistics of the response variable.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	65	15.3363	0.2359
Scaled Deviance	65	69.0000	1.0615
Pearson Chi-Square	65	15.3363	0.2359
Scaled Pearson X2	65	69.0000	1.0615
Log Likelihood		-46.0228	
Full Log Likelihood		-46.0228	
AIC (smaller is better)		102.0456	
AICC (smaller is better)		102.9980	
BIC (smaller is better)		113.2162	

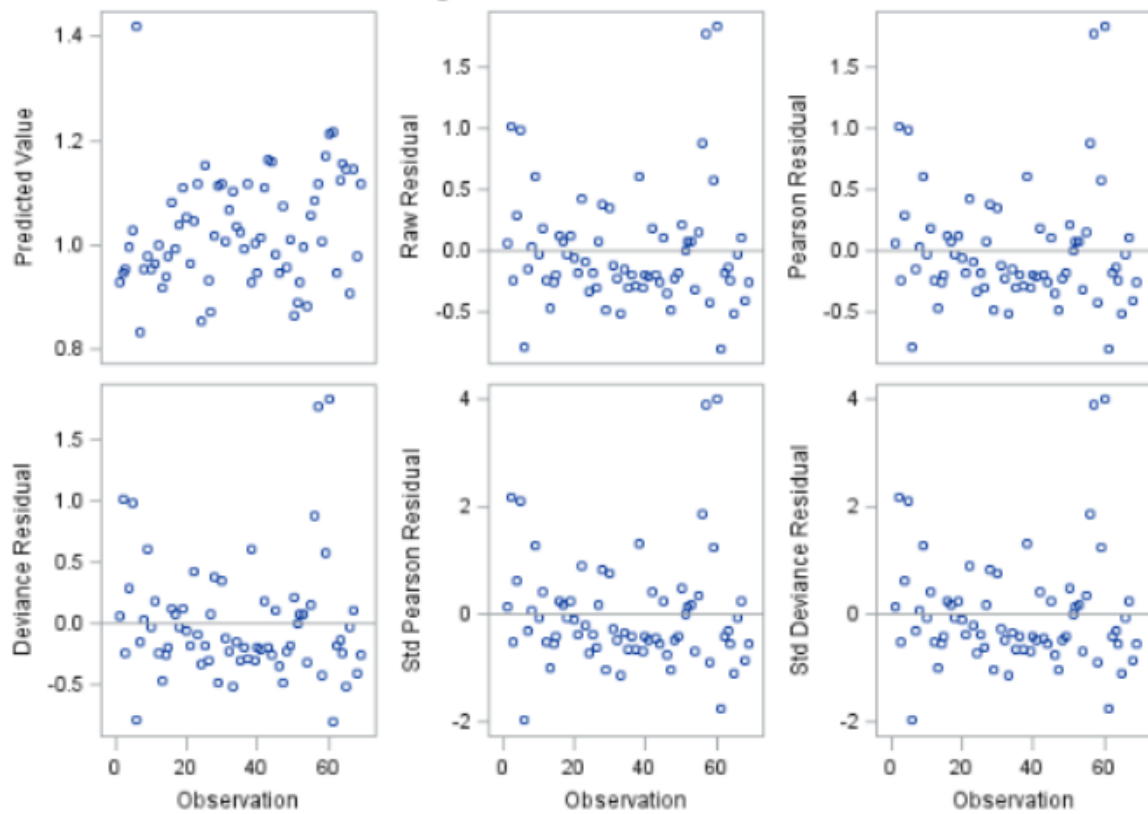
Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr> ChiSq
Intercept	1	-0.9088	0.5234	-1.9347	0.1171	3.01	0.0825
AGE	1	0.0011	0.0052	-0.0091	0.0113	0.05	0.8313
SEX	1 1	0.0447	0.1140	-0.1787	0.2681	0.15	0.6948
SEX	2 0	0.0000	0.0000	0.0000	0.0000	.	.
BMI	1	0.0355	0.0196	-0.0029	0.0739	3.28	0.0699
Scale	1	0.4715	0.0401	0.3990	0.5571		

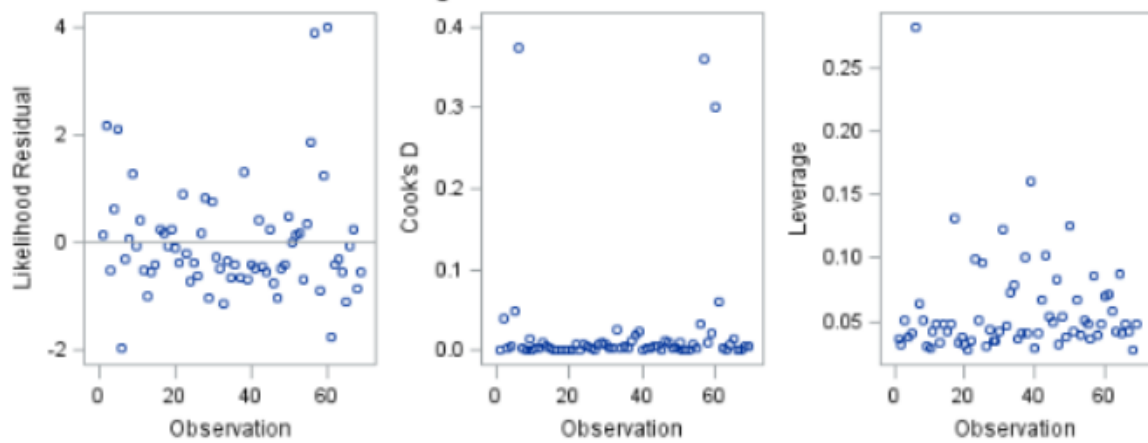
The model provides the AIC, BIC values that are helpful in comparing among different models and arriving at a better model. It provides the estimate of each predictor variables on the model with confidence limits of each variable. From this estimate, we can understand the affect of the predictor variables on the response variable.

The model also standardizes the data points of the variables for producing an unbiased result. This can be seen from the plots below.

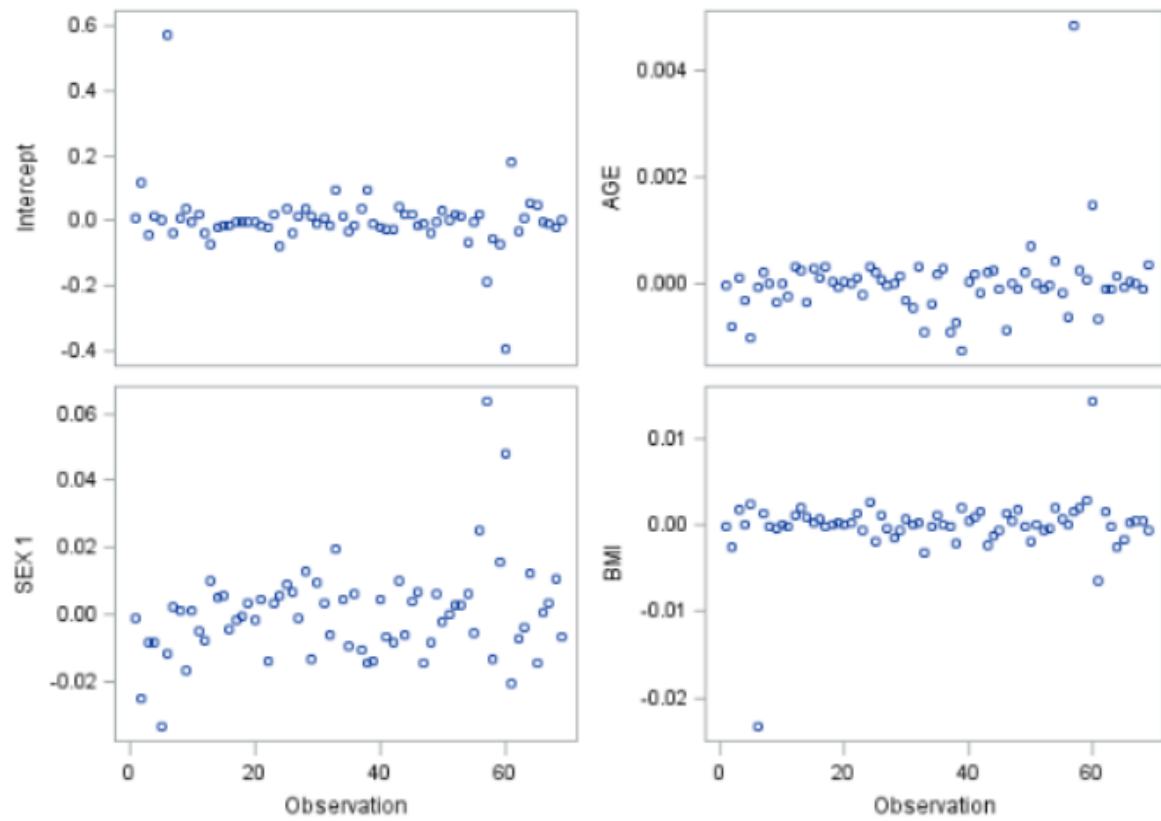
**Diagnostic Statistics for TAG**



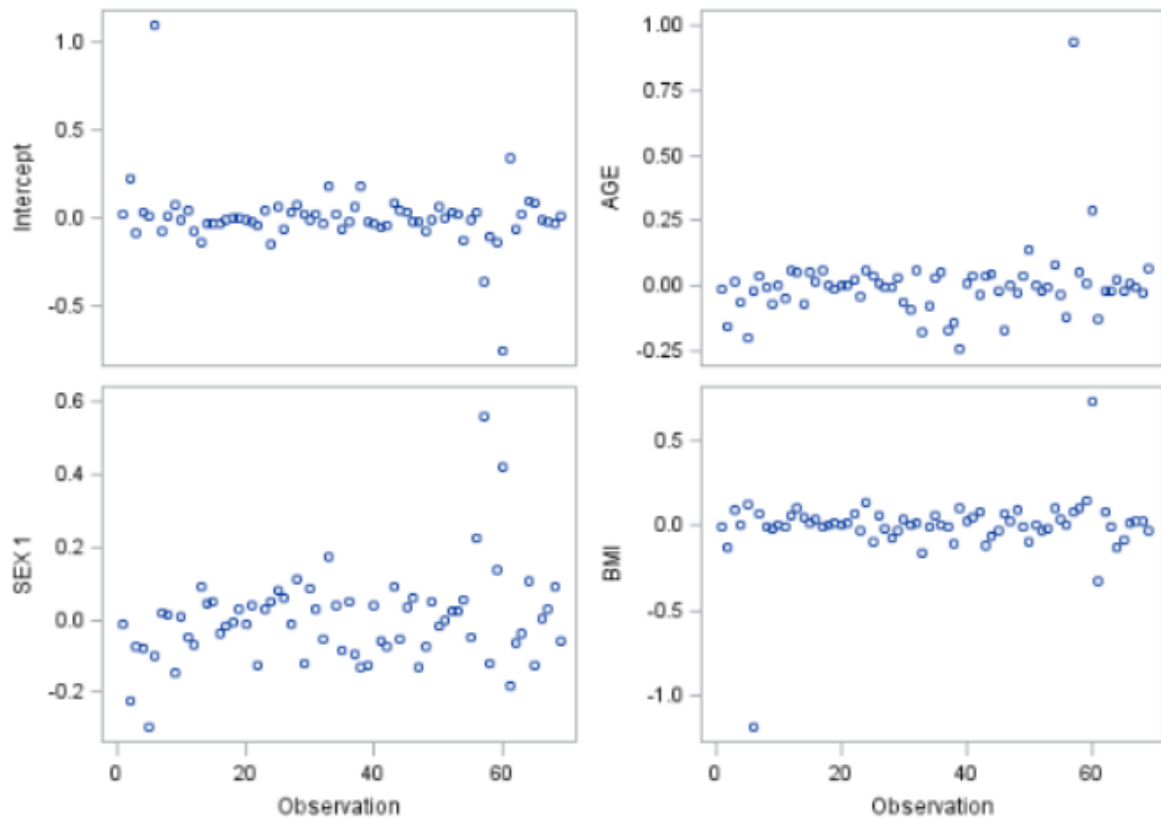
**Diagnostic Statistics for TAG**



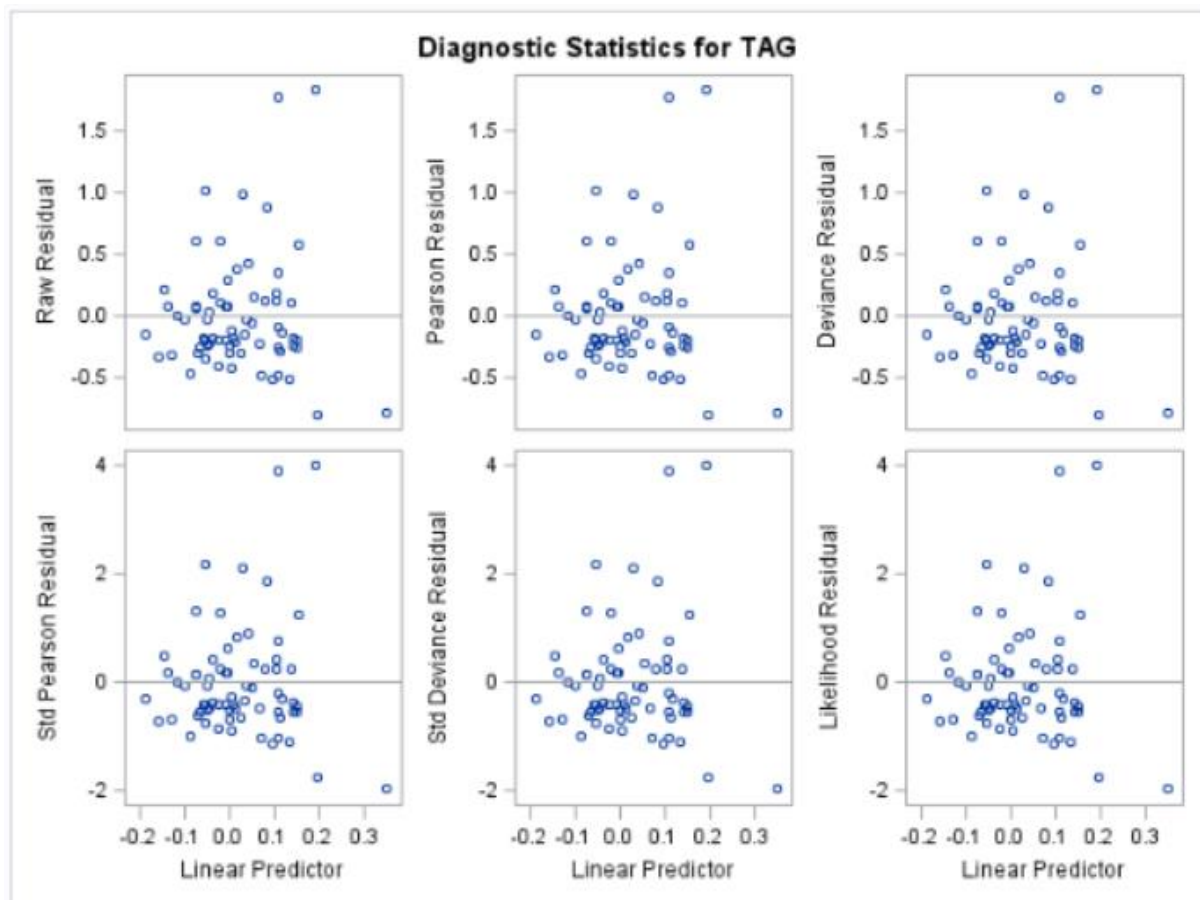
**Unstandardized DFBETA Plots for TAG**



**Standardized DFBETA Plots for TAG**



From the diagnostic statistics, we see that this model is also not a best model as the data points are weakly related to the response variable and there is no positive trend on the linear predictors.



We can fit a general linear mixed model with all the timepoints with Insulin as response variable and age, sex and bmi as predictor variables and compare the change in model with varying covariance structures Autoregressive(1) and Heterogeneous AR(1). By creating a general linear model, we get the AIC value of 2369.7. This model provides three types of residual results which shows the convergence level of predictor variables and these residual helps in checking the efficiency of the model.

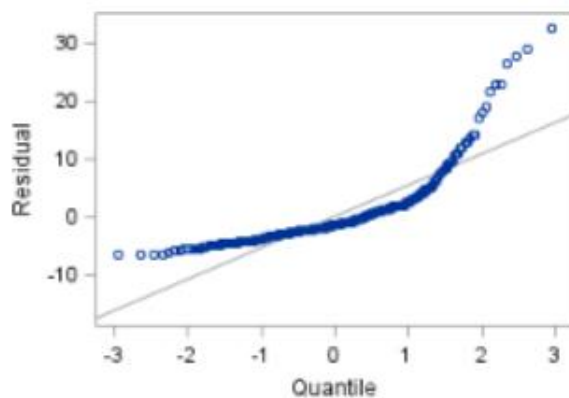
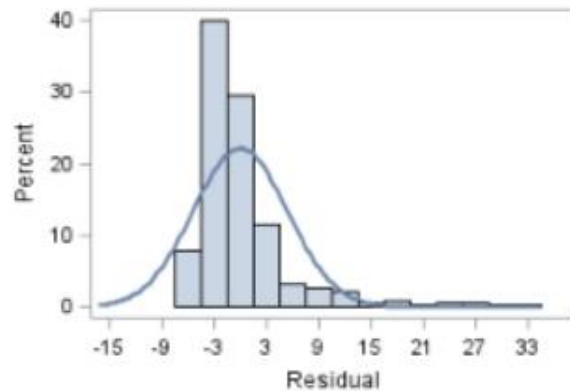
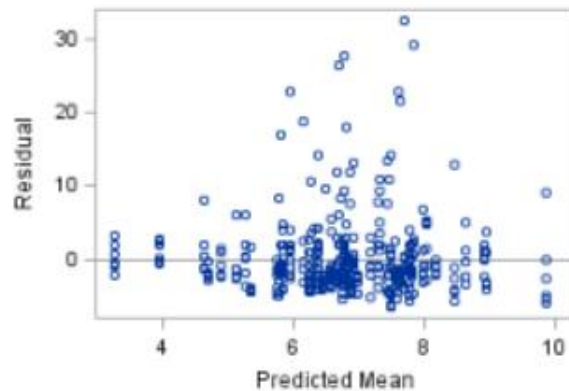
We can compare the AIC and BIC values of the three models and conclude the better model among the three.

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	29.4271

Fit Statistics	
-2 Res Log Likelihood	2367.7
AIC (Smaller is Better)	2369.7
AICC (Smaller is Better)	2369.7
BIC (Smaller is Better)	2373.6

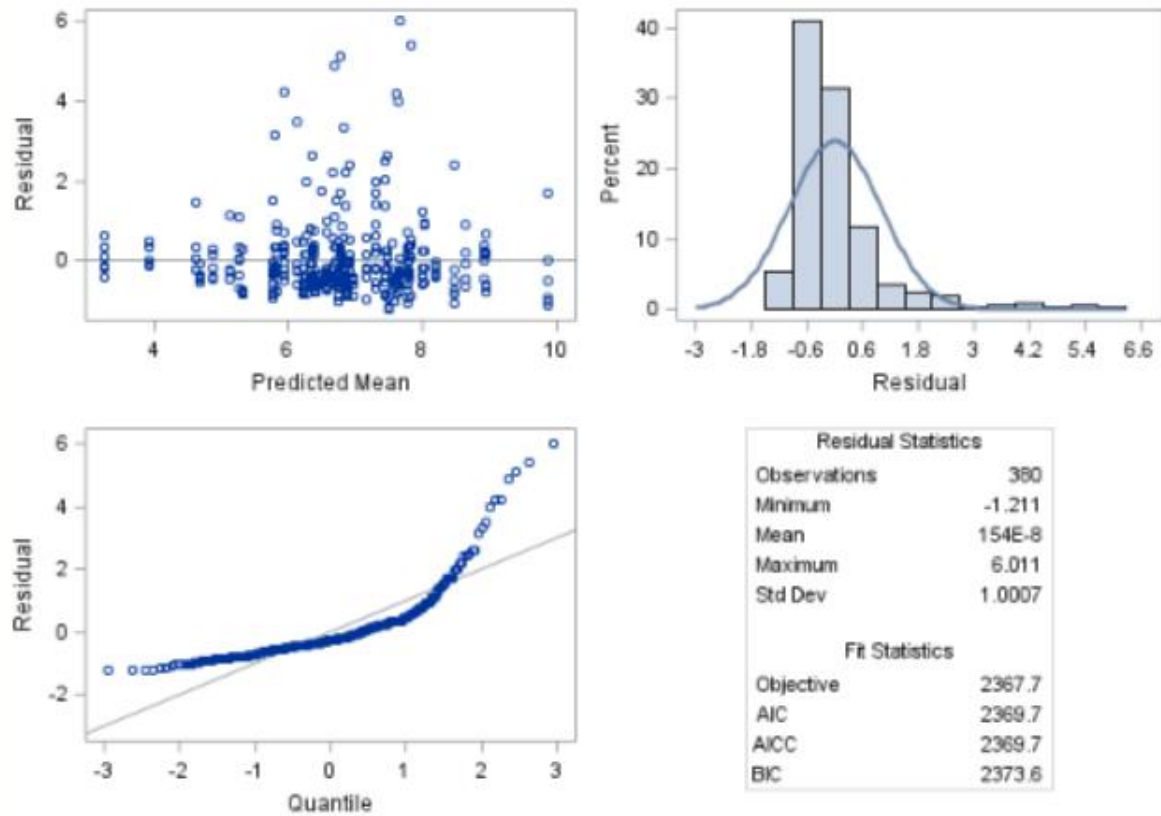
Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	376	5.96	0.0151
SEX	1	376	0.34	0.5591
BMI	1	376	10.99	0.0010

### Residuals for Insulin

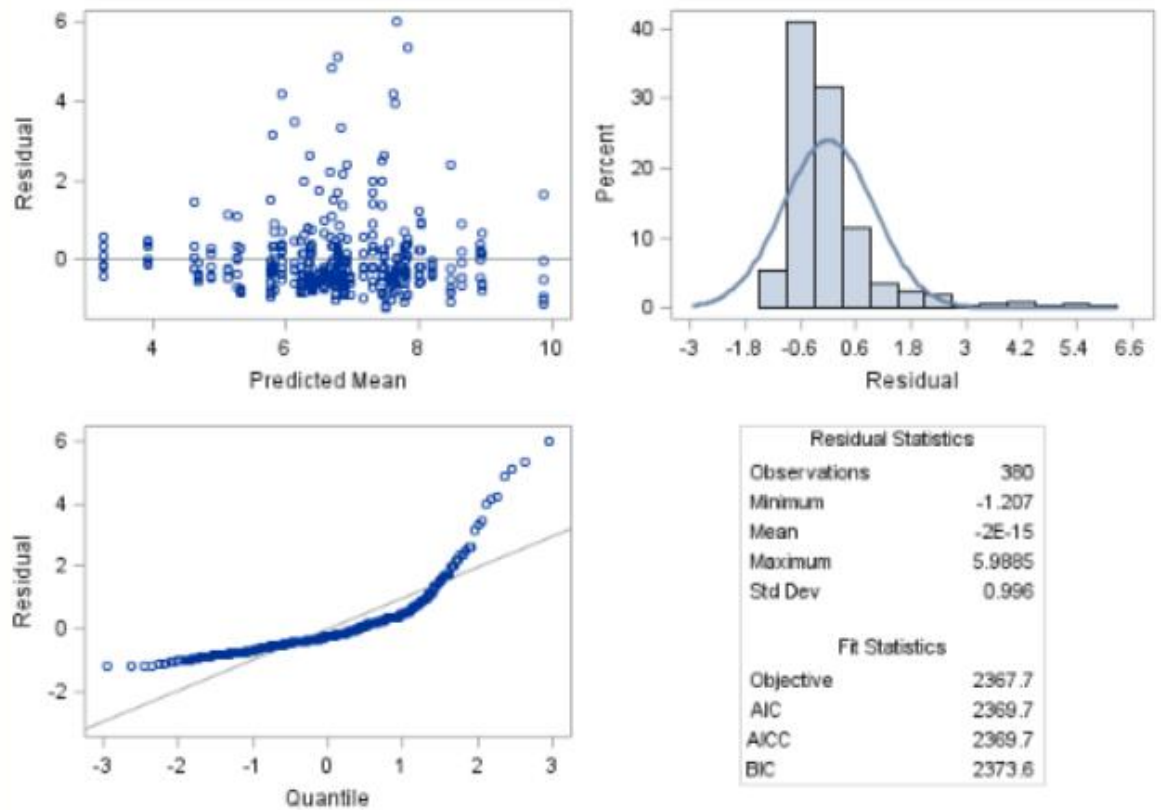


Residual Statistics	
Observations	380
Minimum	-6.546
Mean	-1E-14
Maximum	32.486
Std Dev	5.4032
Fit Statistics	
Objective	2367.7
AIC	2369.7
AICC	2369.7
BIC	2373.6

### Studentized Residuals for Insulin



### Pearson Residuals for Insulin





Fitting a general linear mixed model with all the timepoints with Insulin as response variable and age, sex and bmi as predictor variables and covariance structures Autoregressive(1), we see that the residual of this model is less than the previous model. Also, the AIC value of this model is 2362.1 which is lesser than the previous model. So, this model can be considered as a better model than the previous one.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
AR(1)	ID	0.1550
Residual		29.3554

Fit Statistics	
-2 Res Log Likelihood	2358.1
AIC (Smaller is Better)	2362.1
AICC (Smaller is Better)	2362.1
BIC (Smaller is Better)	2366.6

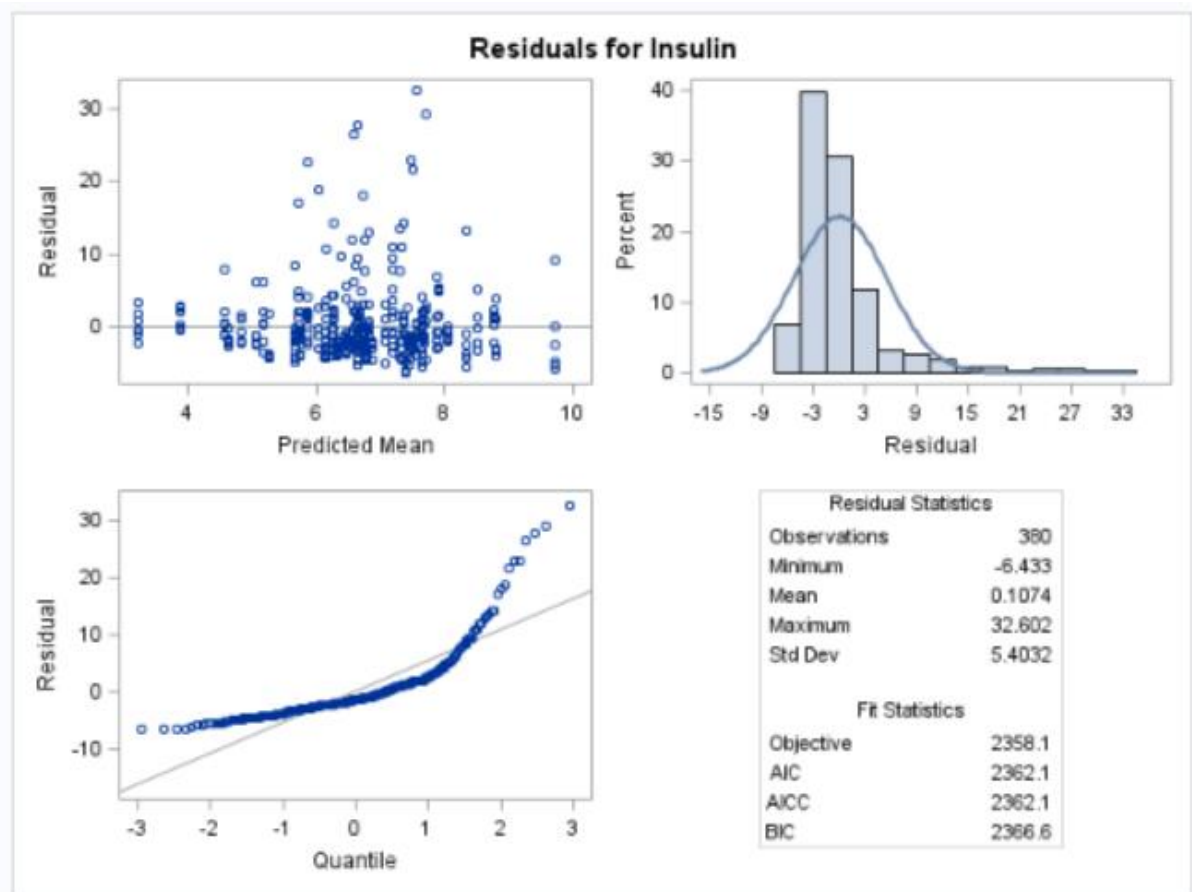
Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
1	9.59	0.0020

The estimates shows that the predictor variables SEX and BMI have a value of 0.3198 & 0.3602 respectively which is a positive influence on the response variable Insulin.

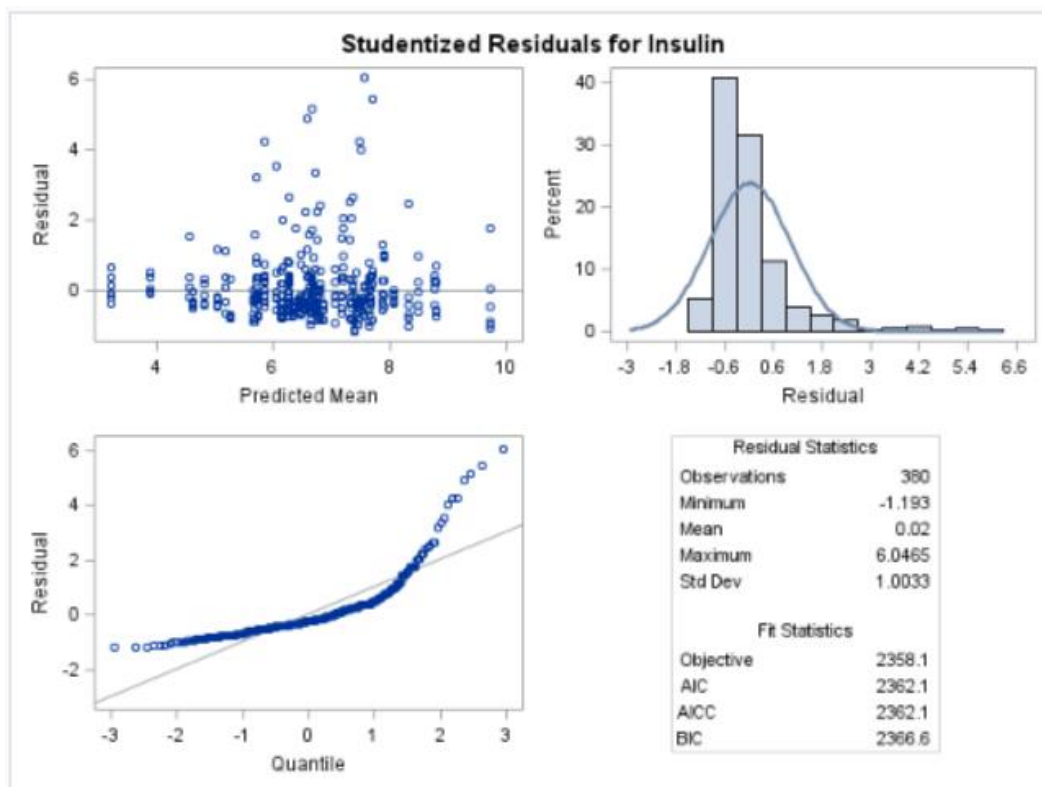
Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-0.9184	3.5677	60	-0.26	0.7977
AGE	-0.05978	0.02841	60	-2.10	0.0396
SEX	0.3198	0.6461	60	0.49	0.6224
BMI	0.3602	0.1249	60	2.88	0.0054

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	60	4.43	0.0396
SEX	1	60	0.24	0.6224
BMI	1	60	8.32	0.0054

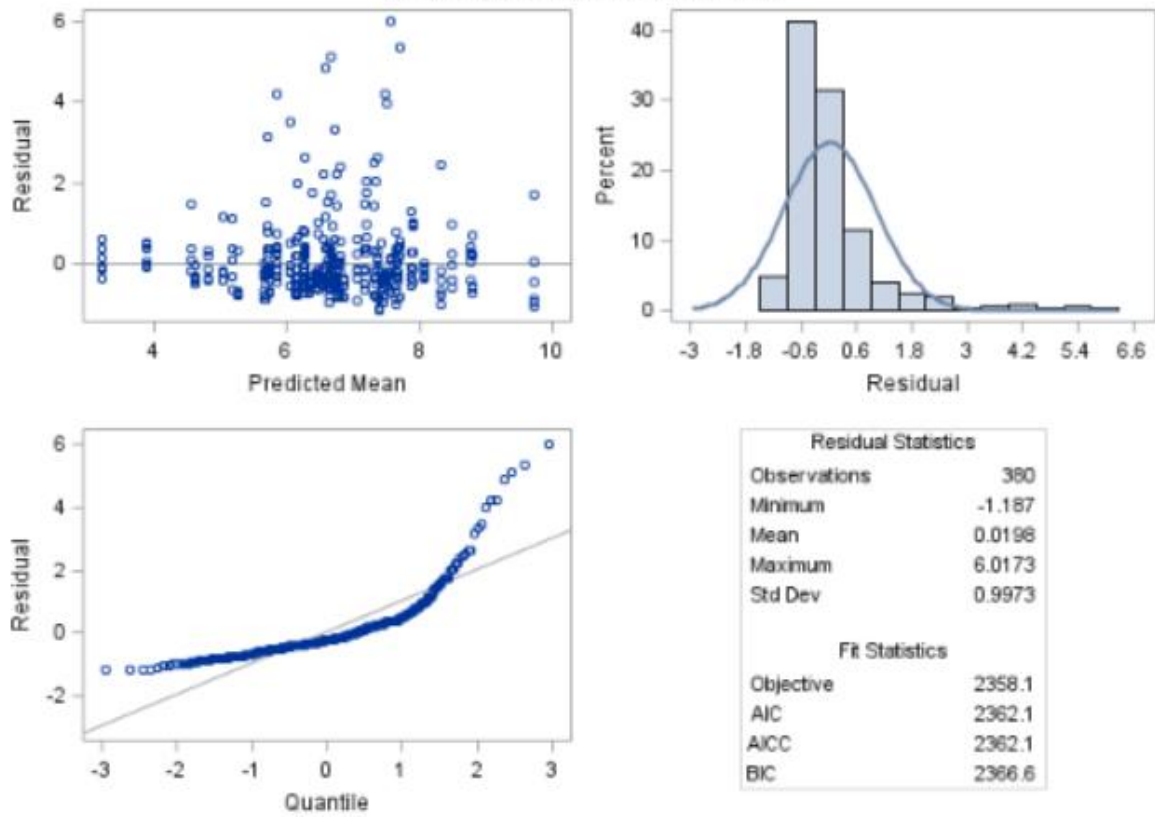




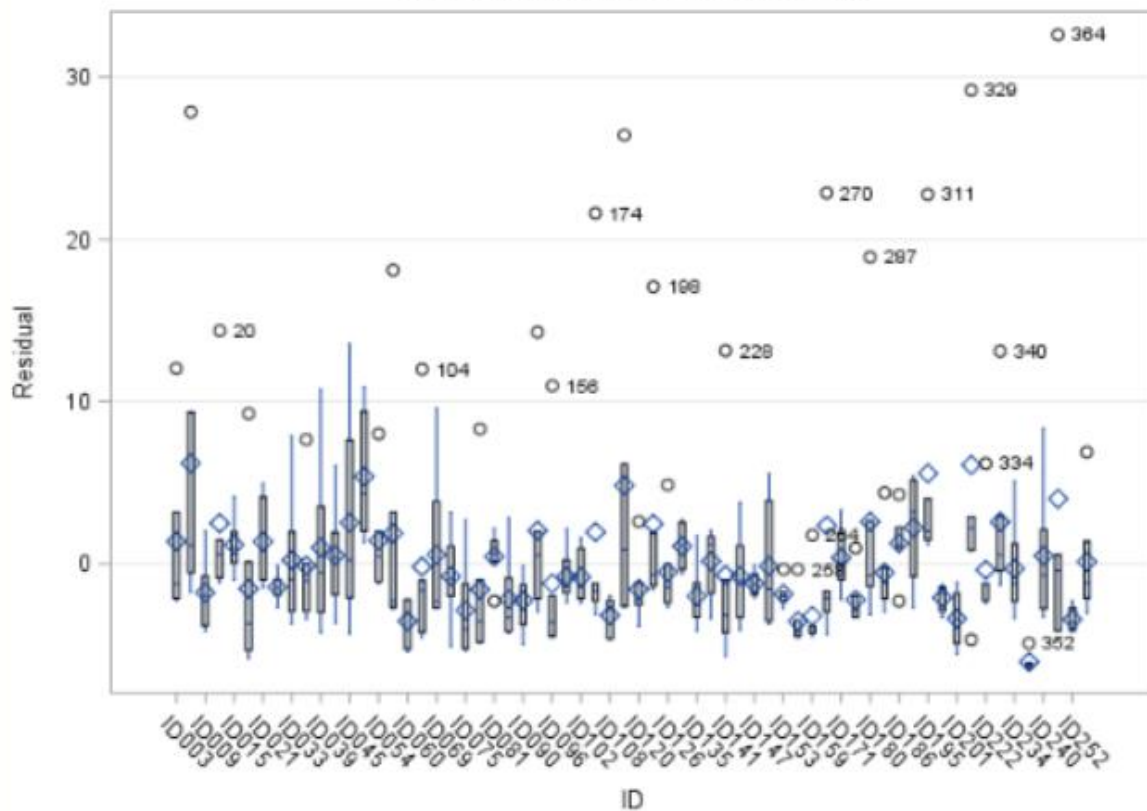
The residual plots also shows that this model is better than the previous model with a lower residual value.



**Pearson Residuals for Insulin**



**Distribution of Residuals for Insulin**



Fitting a general linear mixed model with all the timepoints with Insulin as response variable and age, sex and bmi as predictor variables and covariance structures heterogeneous AR(1), we get the AIC value as 1946.5 which is less than both the models defined above. So, this model can be considered a better model among three models.

Fit Statistics	
-2 Res Log Likelihood	1932.5
AIC (Smaller is Better)	1946.5
AICC (Smaller is Better)	1946.8
BIC (Smaller is Better)	1962.1

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
6	435.19	<.0001

From the estimates of the predictor variables, the estimate of SEX and BMI has a significant role on the response variable Insulin with BMI being highly significant at 95% confidence level.

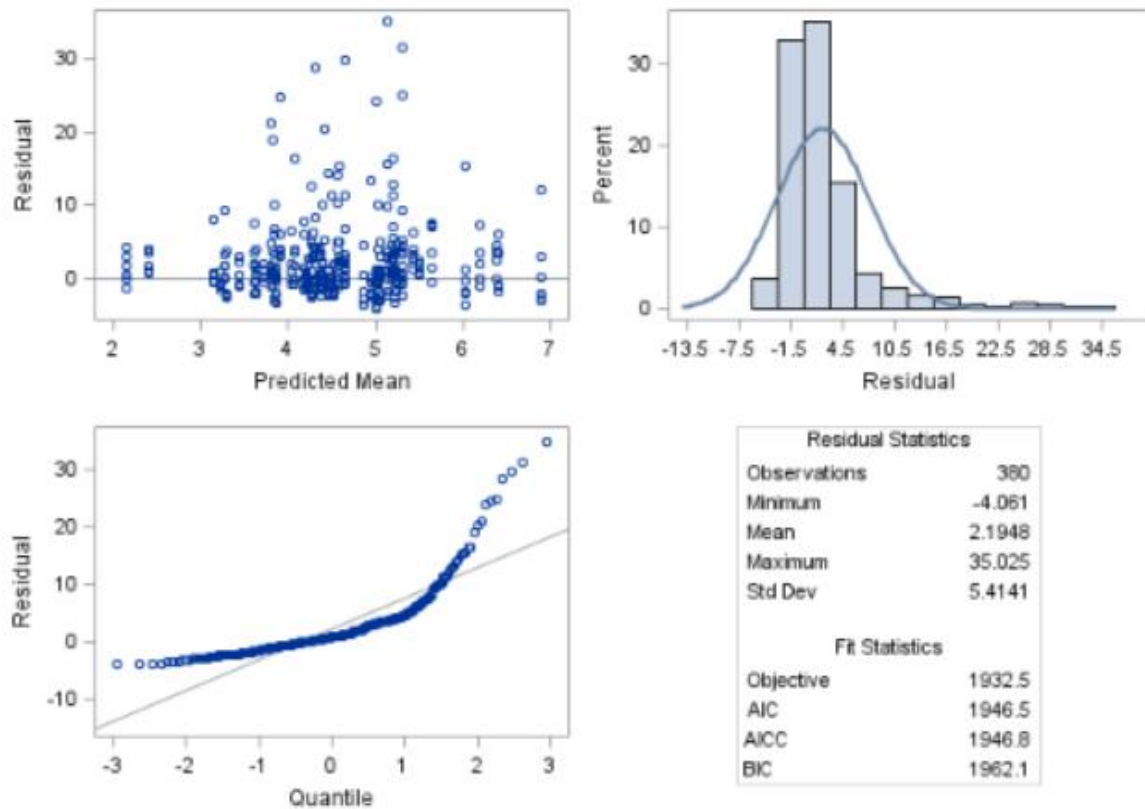
Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-2.3615	2.0589	60	-1.15	0.2560
AGE	-0.04010	0.01639	60	-2.45	0.0174
SEX	0.5580	0.3727	60	1.50	0.1396
BMI	0.2954	0.07208	60	4.10	0.0001

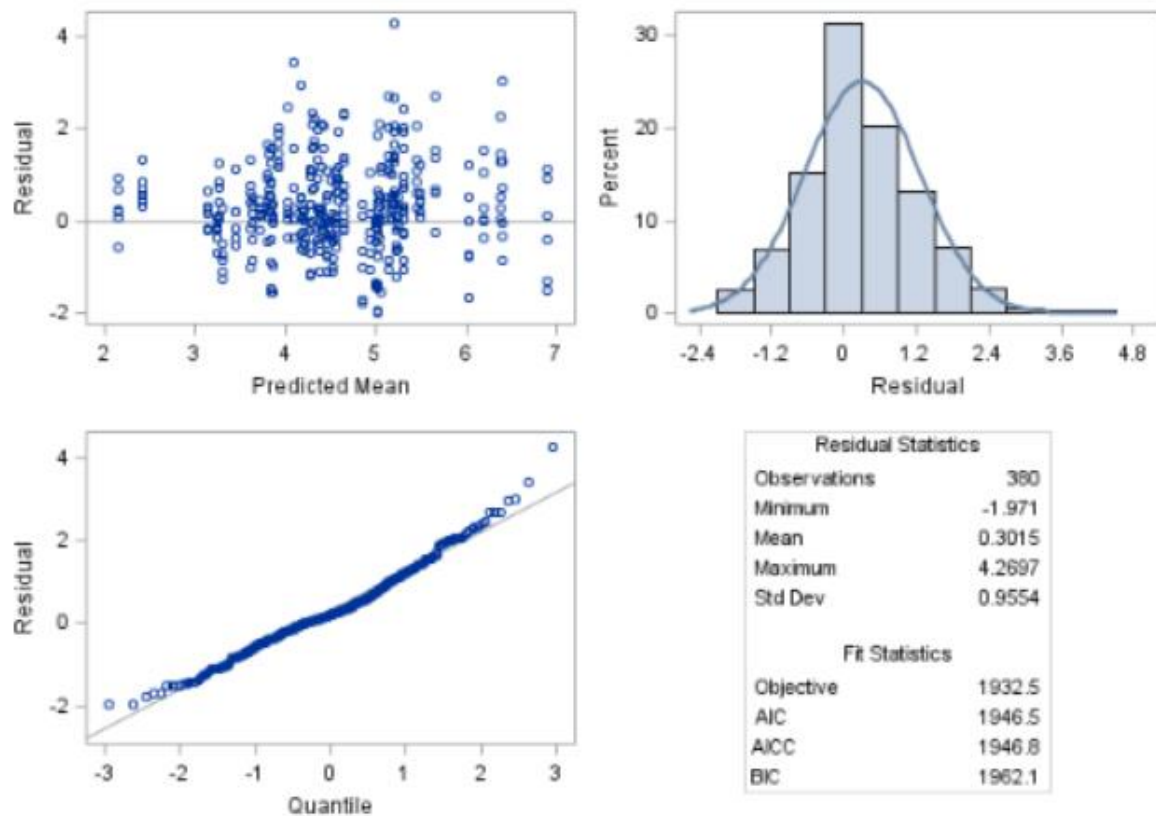
Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	60	5.98	0.0174
SEX	1	60	2.24	0.1396
BMI	1	60	16.79	0.0001

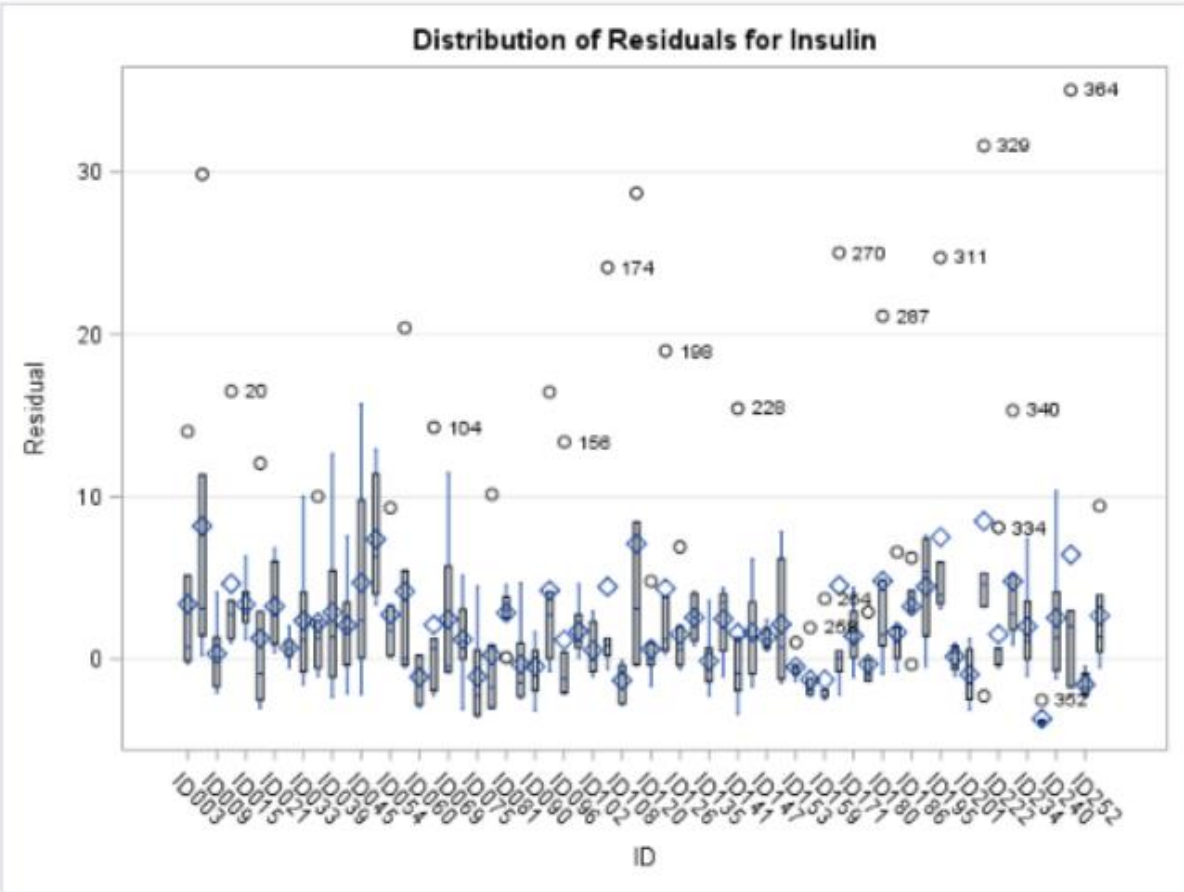
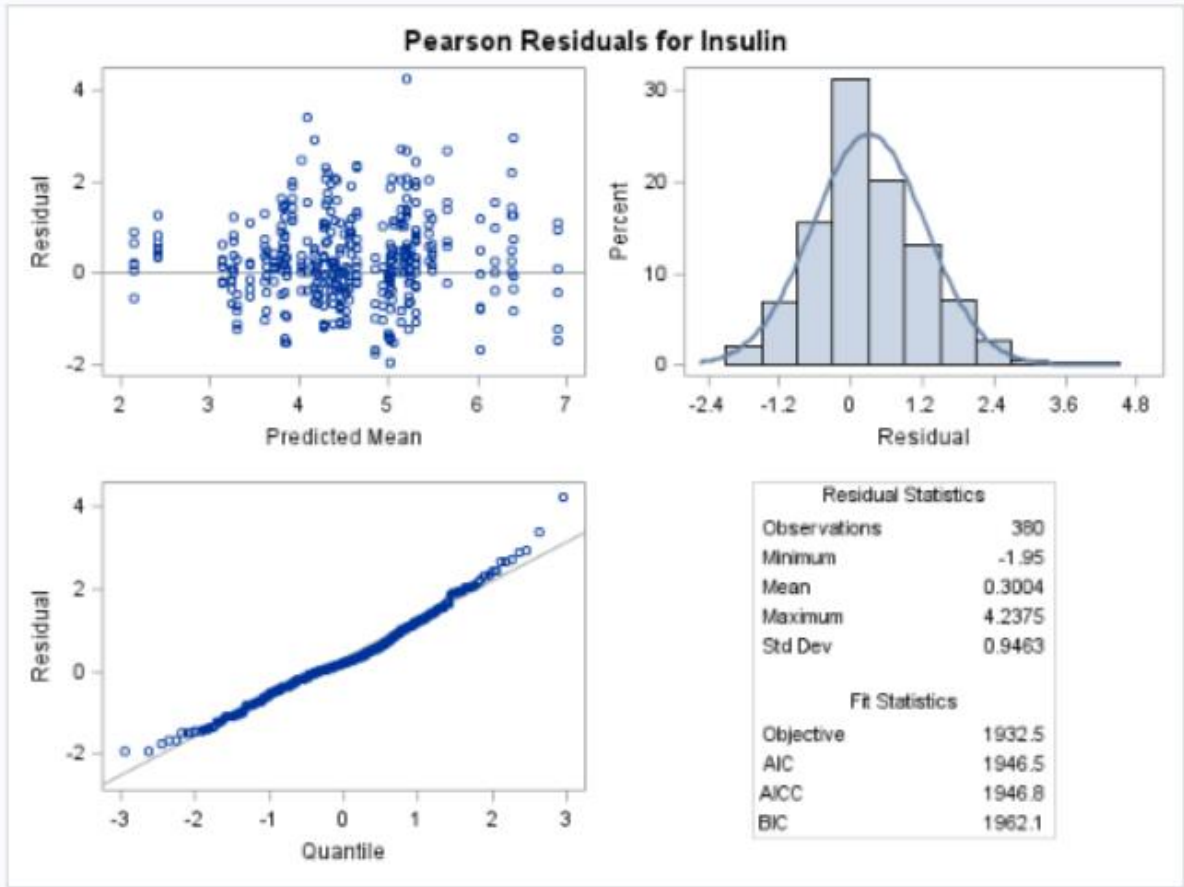
We can see that BMI is highly significant from the above table. The below plots of the residuals for Insulin, we can see that the residual converges to a straight line in the Q-Qplot of Pearson Residuals plots for Insulin.

### Residuals for Insulin



### Studentized Residuals for Insulin





Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Var(1)	ID	7.2158
Var(2)	ID	169.43
Var(3)	ID	23.4284
Var(4)	ID	7.2179
Var(5)	ID	4.3510
Var(6)	ID	4.1979
ARH(1)	ID	0.5557

So, we conclude that the better model is the one with covariance structure heterogeneous AR(1) as it has the lowest AIC values and less residual values compared to the other two models and from the above table, we can see that the variance are heterogeneous across time.