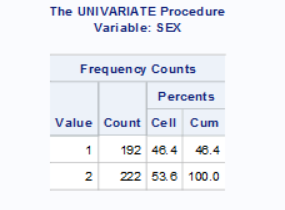
**Oral Lipid Tolerance Test**

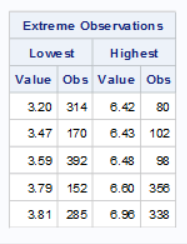
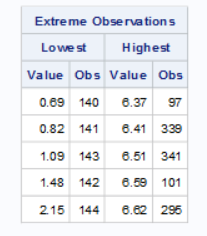
The assignment2.csv is an extract of data from an Oral Lipid Tolerance Test. It contains the observations based on the response of participants after consuming high fat drink. It has several observations based on which the participants were tested on, such as, ID, time, SEX AGE BMI, etc.

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.

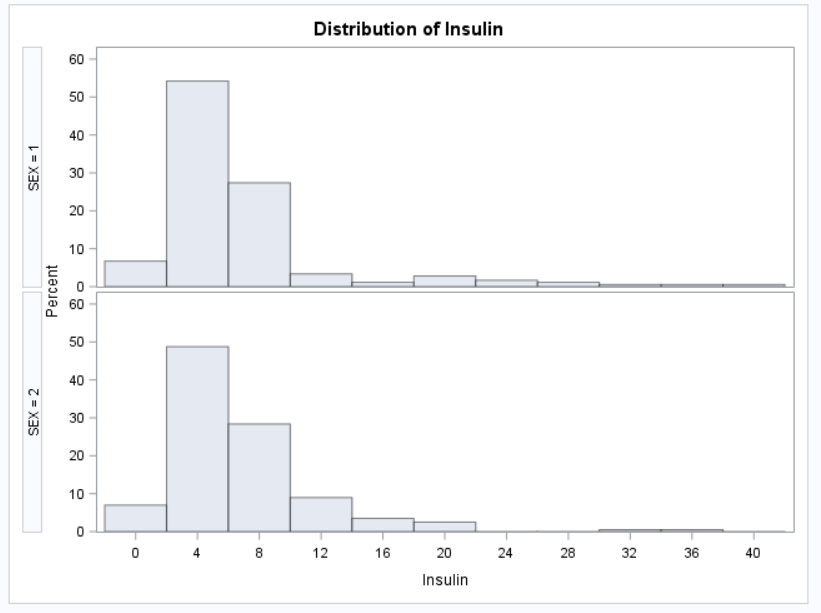
From the frequency command, we can see that among 414 participants, there were 192 males and 222 females. We can also cross verify this by looking at the quantiles table of sex variable which points out that more than 50% of the participants are female.



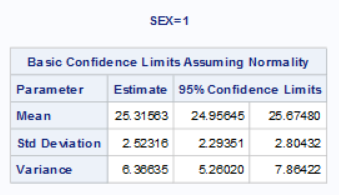
From the extreme observation tables, we see that the observations of ApoA1, ApoB and Insulin levels had huge variation among the lowest and highest levels. The values of other variables seem to increase gradually. There was not much difference in Non-essential fatty acid as the mean value was around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

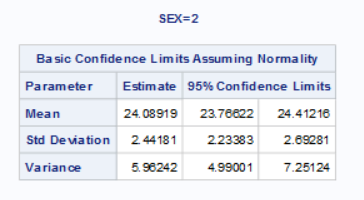
Observations of Glucose Observations of Cholesterol



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.

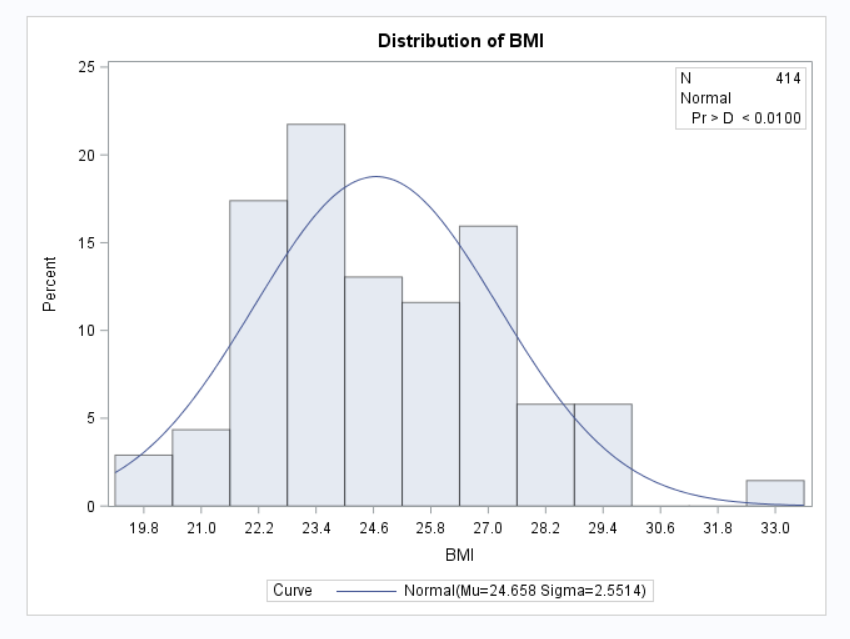


Confidence Interval of male participants

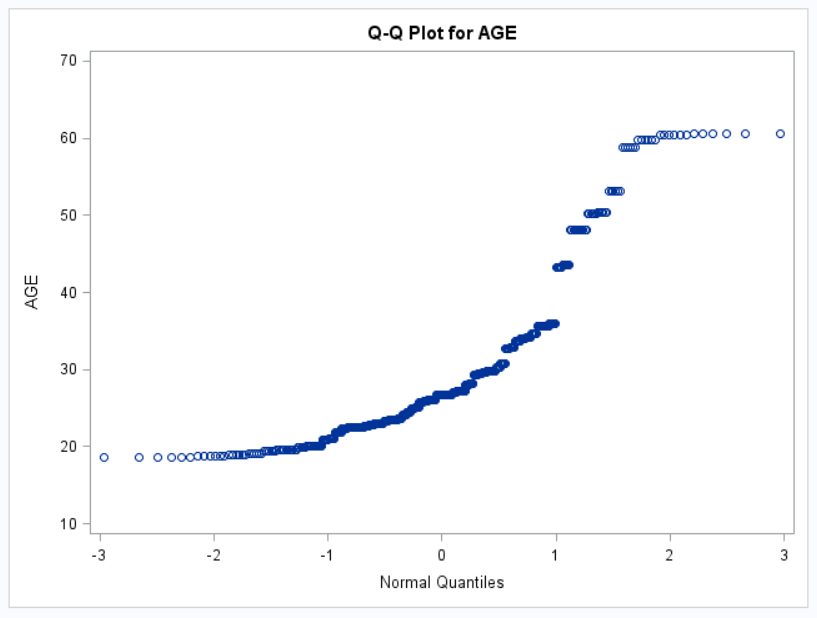


Confidence Interval of female participants

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.

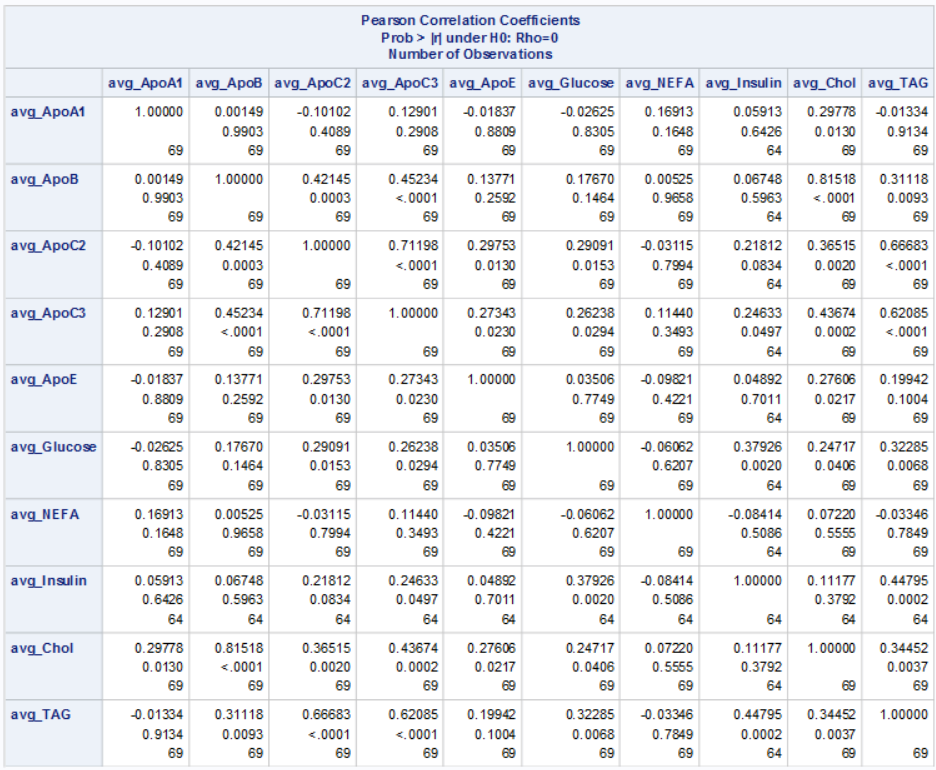


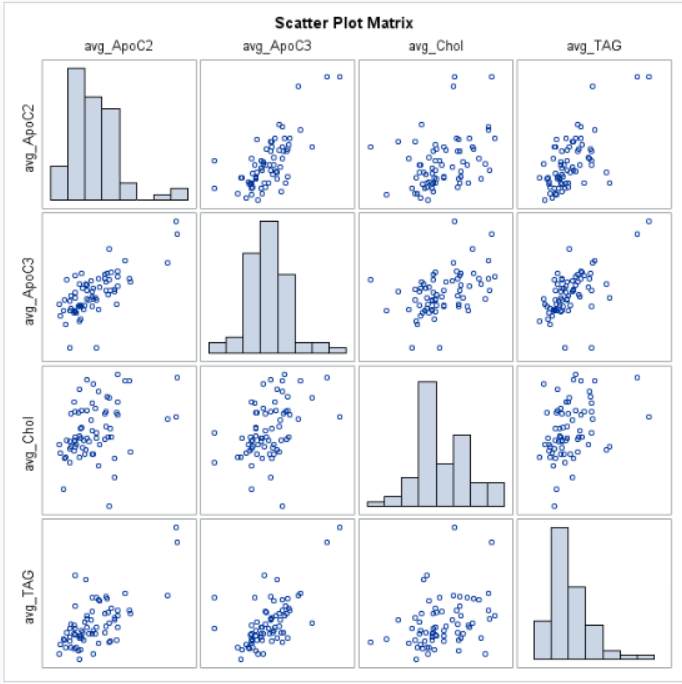
From the analysis on Q-Qplot and above descriptive statistics, we see that there are no outliers in the given dataset. The datapoints are stretched to a certain extent but it cannot be taken as outliers as the values play a significant role in the dataset. So, there are no outliers in the dataset.



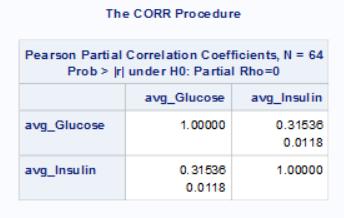
Using correlation function, we can correlate the blood chemistry of several variables. From the table below, we see that the variable ApoA1 is not much correlated with any of the other variables. Variable ApoB is well correlated with ApoC2, TAG & ApoC3 with highly positive correlated with Cholesterol. 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.

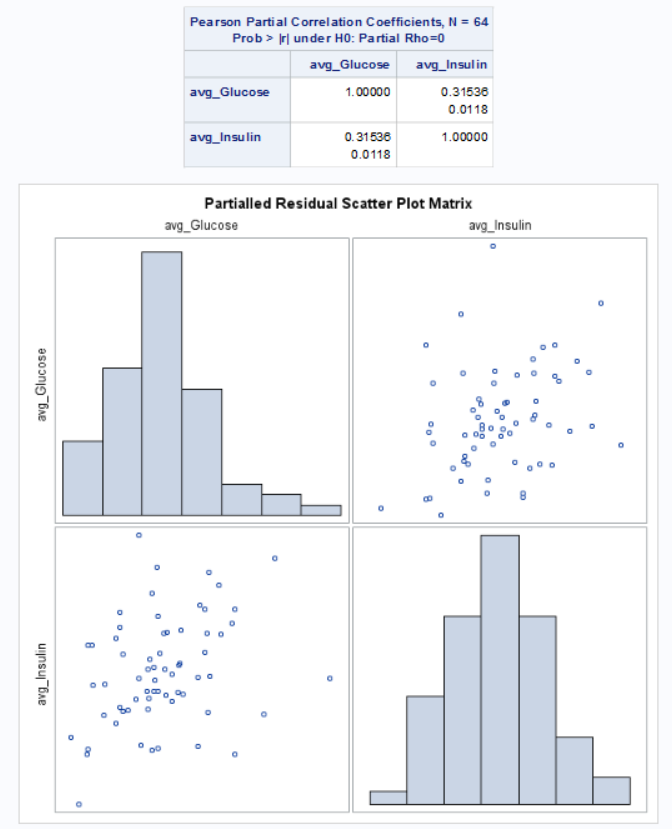




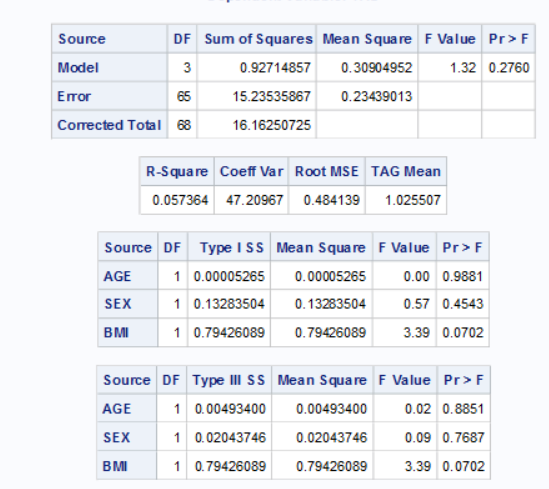
This plot shows that ApoC3 is positively correlated with ApoC2, Cholesterol and TAG as there is an upward pull of data points among the variables.



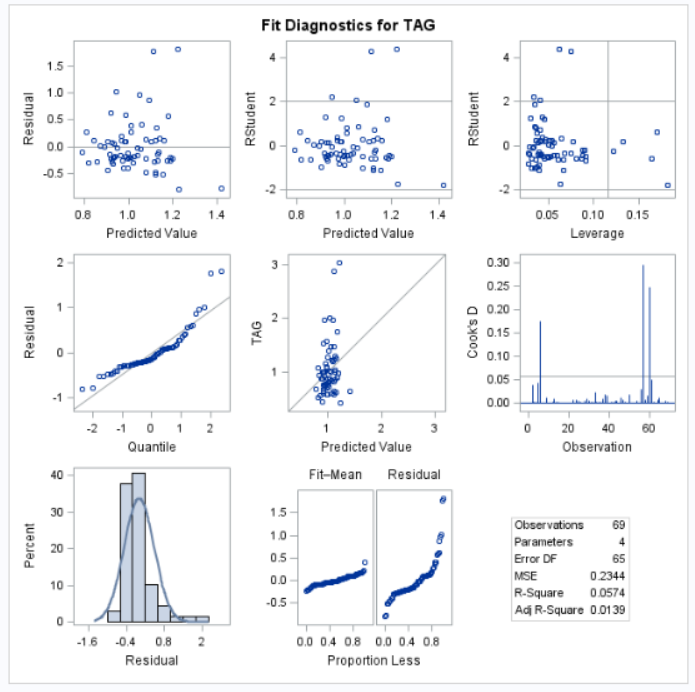
From the above table, we see that the variable Glucose and Insulin are positively correlated with a p value less than 0.05 which states that both the variables are weakly positively related with each other at 95% confidence level. We can see that there is an upward pull of the datapoints among glucose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



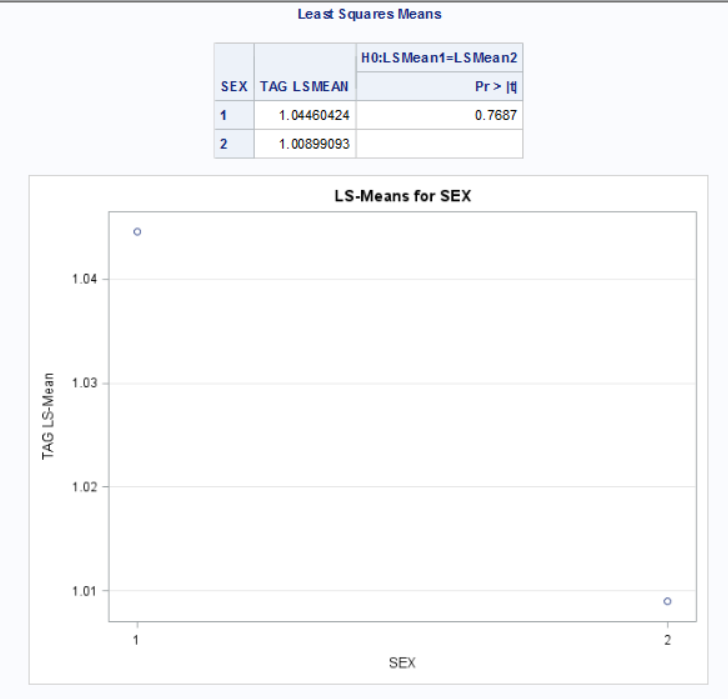
We can build a model on Triglyceride variable with predicted variables – AGE, SEX and BMI for time t=0 to check their relationship and the efficiency of the model. From the output values, our R - Square and adjusted R-square values are very less which states that this is not a good model. The p values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

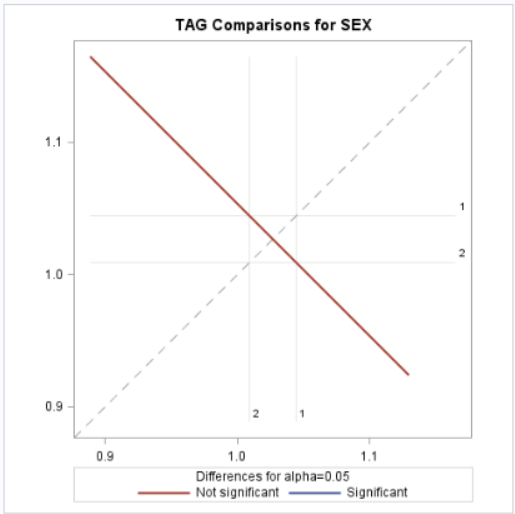


From the plots, we see that the values are spread randomly and few values are stretched away from the normal intervals of the data points. So, there is not much relationship with the variables among each other. The Q-Qplot also shows that the data are not linear and does not converge to a straight line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

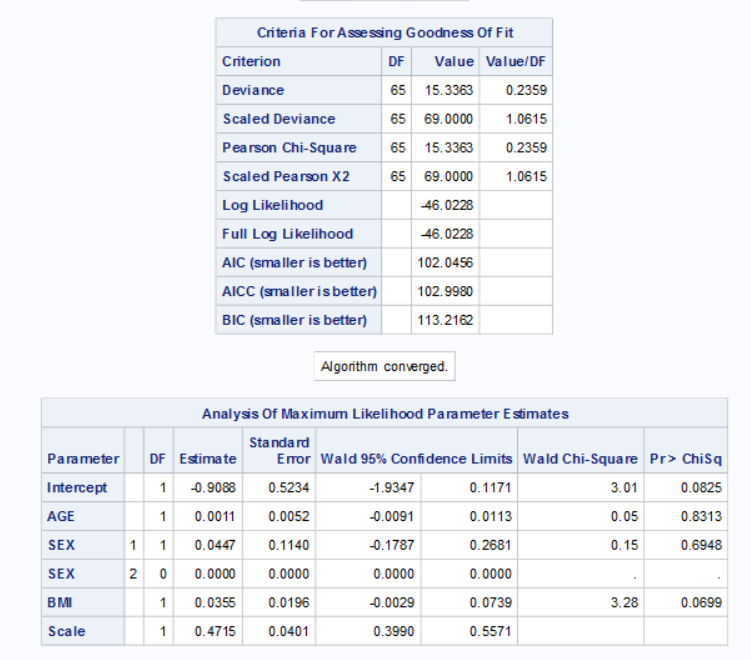


From the least square means values of male and female participants, we can see that the mean values differ for both the gender as the least squares mean of male participants is 1.0446 when compared with that of female participants which is 1.008. This difference of mean values can also be seen from the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



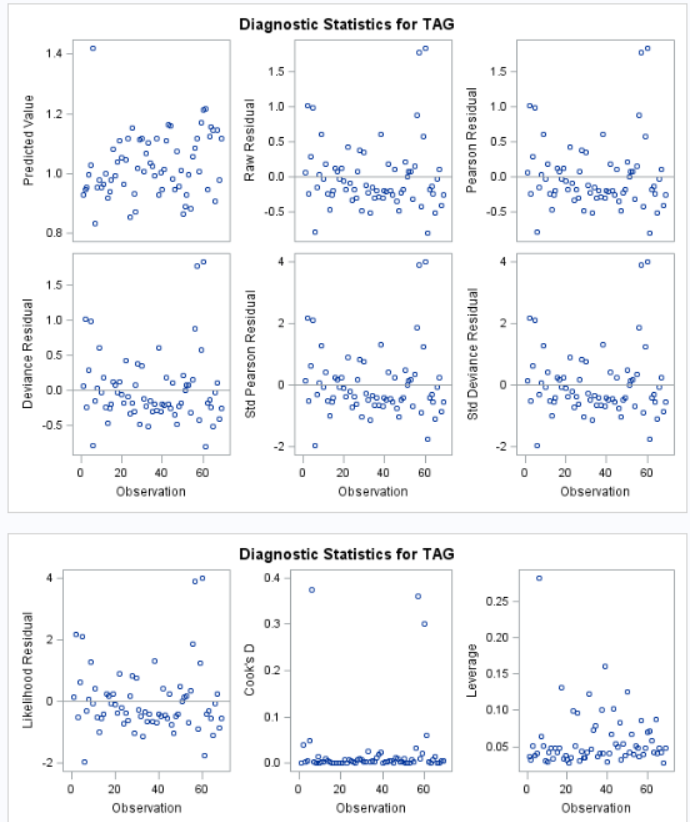


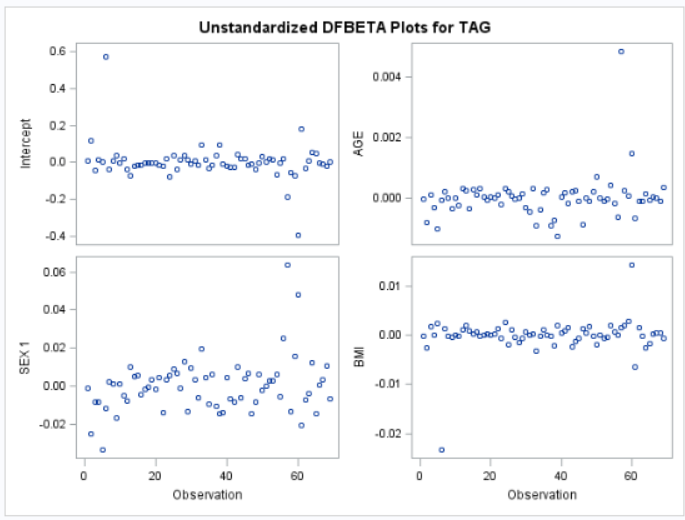
When creating a model using ‘genmod’ function on Triglyceride variable and predicted variables – AGE, SEX and BMI with log-normal distribution on the response variable for time t=0, we get several information about the model such as goodness of Fit, Maximum Likelihood analysis and diagnostic statistics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

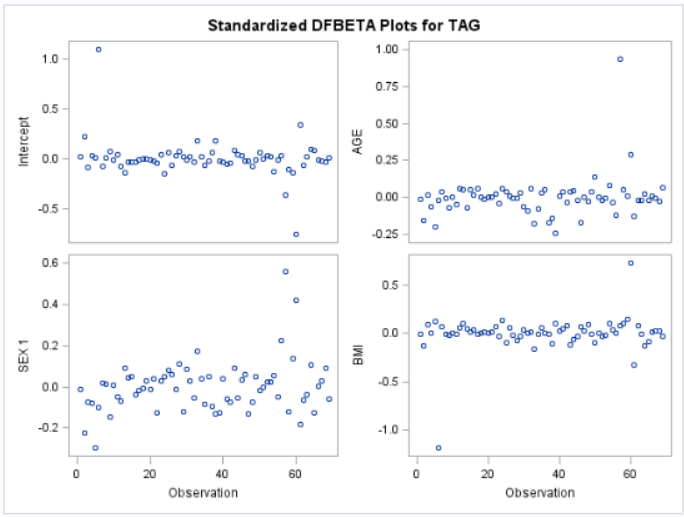


The model provides the AIC, BIC values that are helpful in comparing among different models and arriving at a better model. It provides the estimate of each predictor variables on the model with confidence limits of each variable. From this estimate, we can understand the affect of the predictor variables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

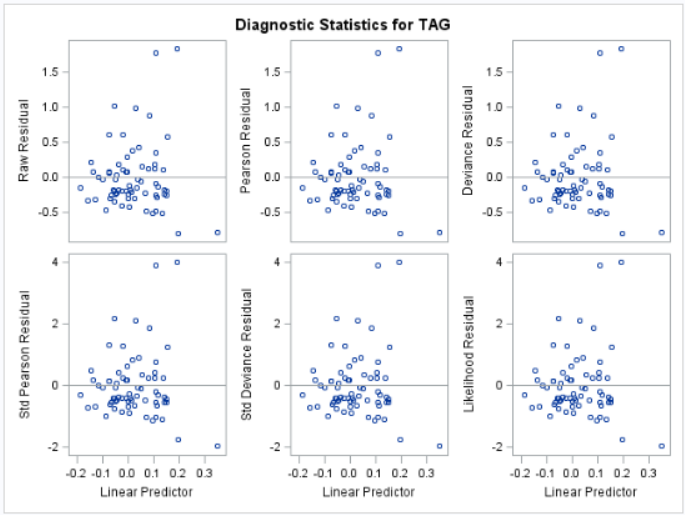
The model also standardizes the data points of the variables for producing an unbiased result. This can be seen from the plots below.





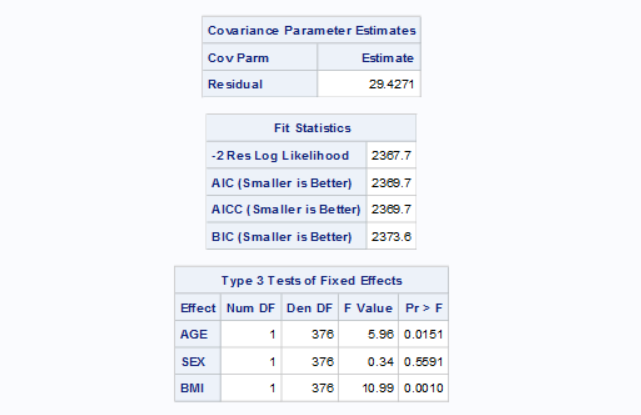


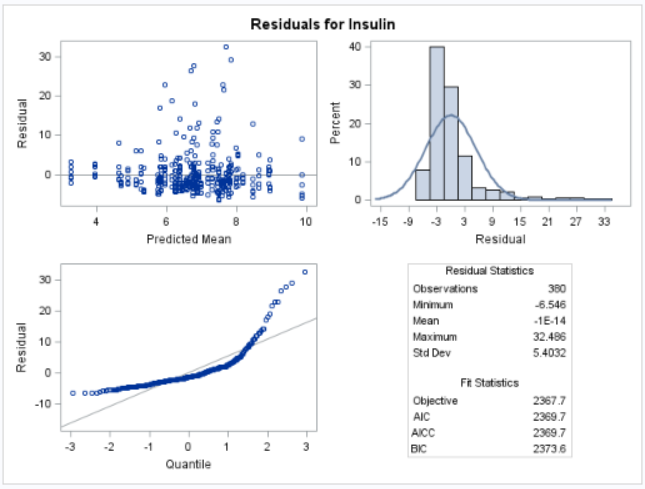
From the diagnostic statistics, we see that this model is also not a best model as the data points are weekly related to the response variable and there is no positive trend on the linear predictors.

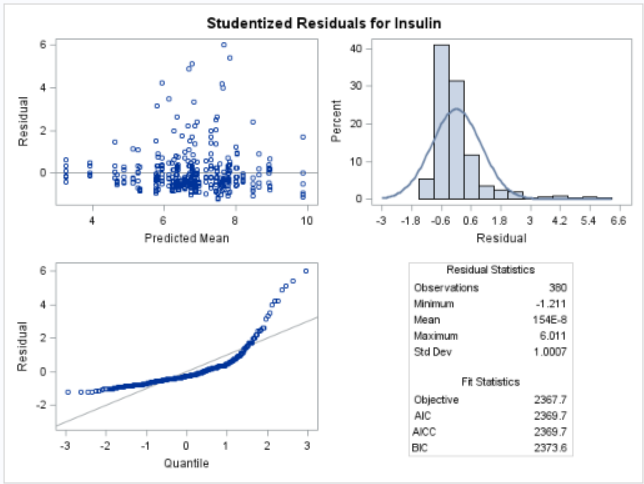


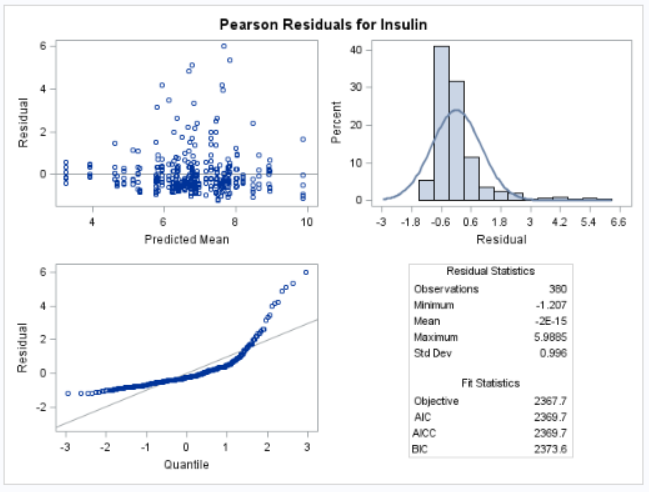
We can fit a general linear mixed model with all the timepoints with Insulin as response variable and age, sex and bmi as predictor variables and compare the change in model with varying covariance structures Autoregressive(1) and Heterogeneous AR(1). By creating a general linear model, we get the AIC 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.

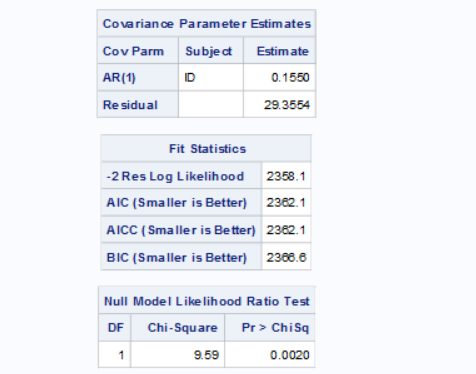




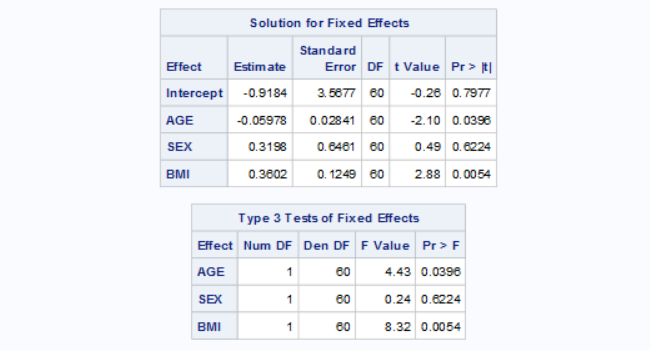


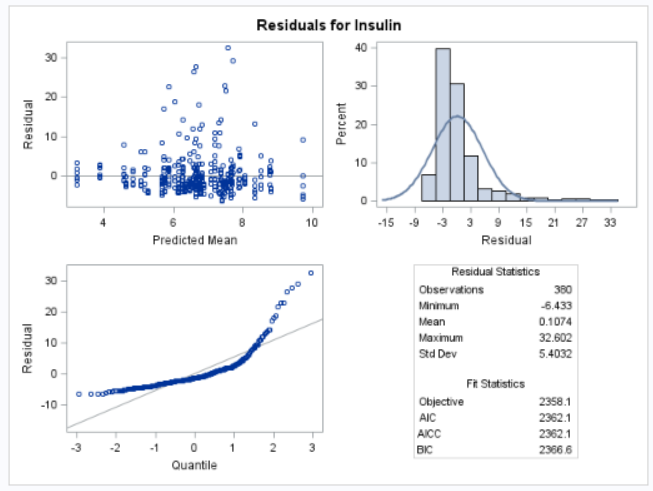


Fitting a general linear mixed model with all the timepoints with Insulin as response variable and age, sex and bmi as predictor variables and covariance structures Autoregressive(1), we see that the residual of this model is less than the previous model. Also, the AIC value of this model is 2362.1 which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

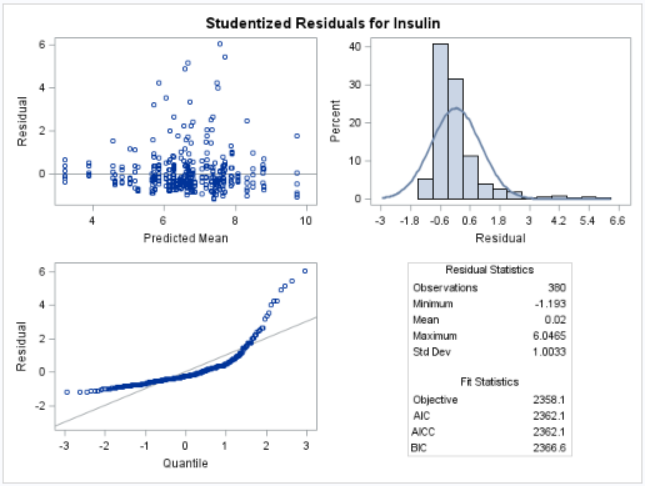


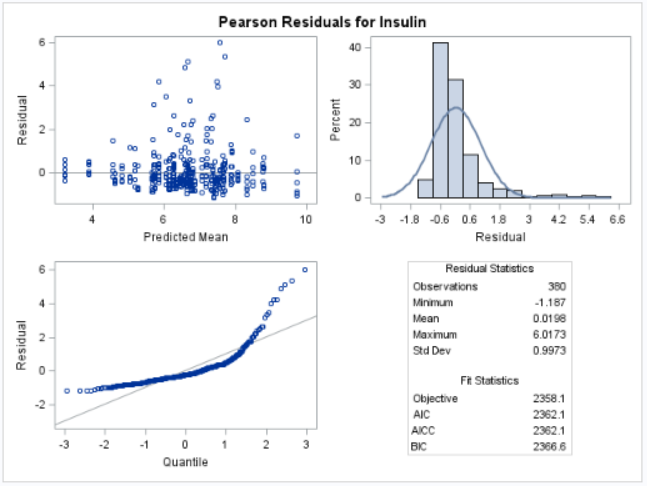
The estimates shows that the predictor variables SEX and BMI have a value of 0.3198 & 0.3602 respectively which is a positive influence on the response variable Insulin.

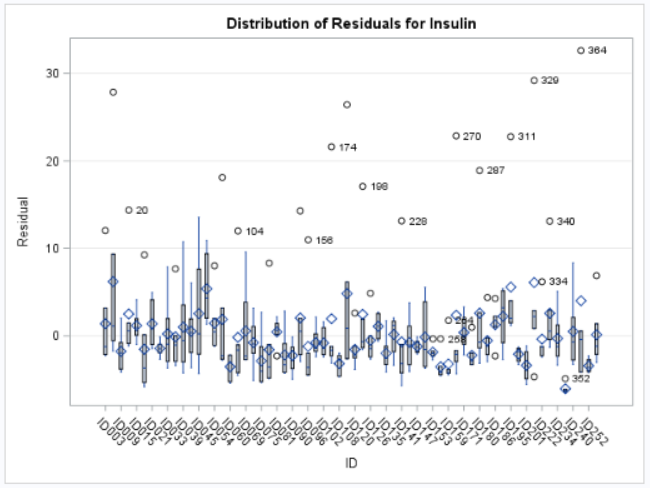




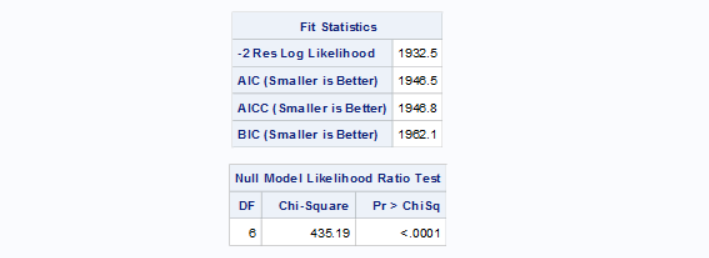
The residual plots also shows that this model is better than the previous model with a lower residual value.



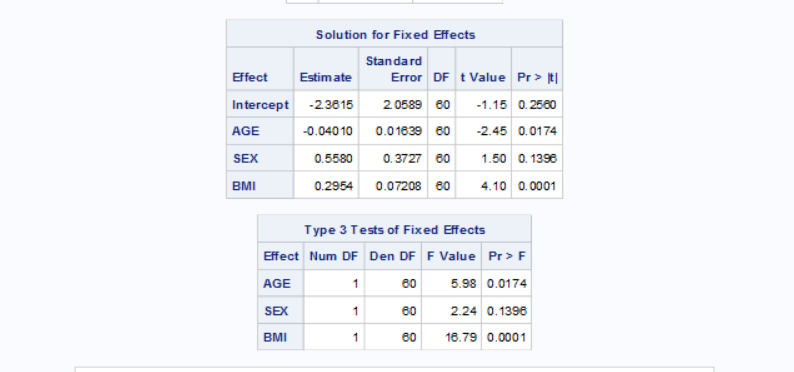




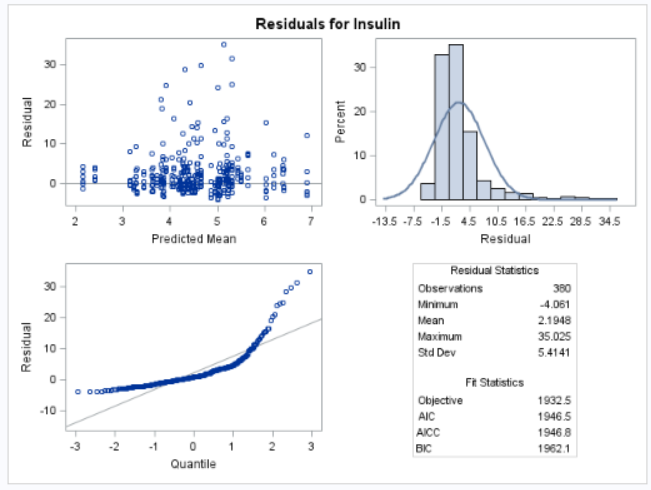
Fitting a general linear mixed model with all the timepoints with Insulin as response variable and age, sex and bmi as predictor variables and covariance structures heterogeneous AR(1), we get the AIC value as 1946.5 which is less than both the models defined above. So, this model can be considered a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

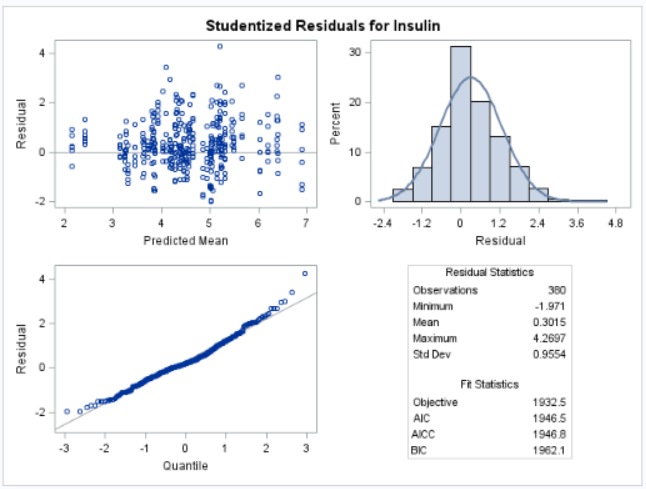


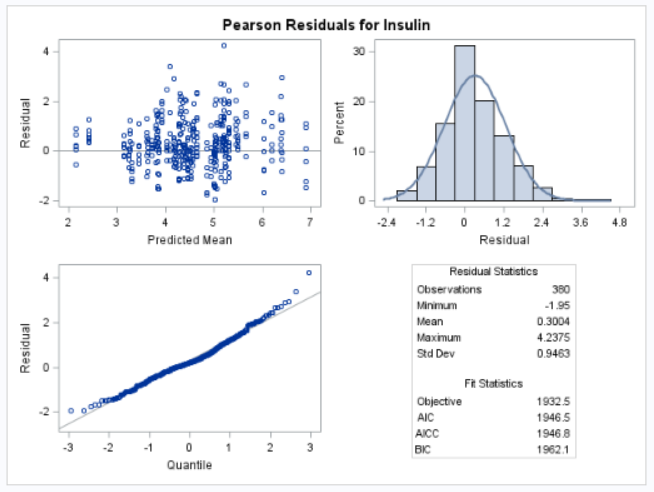
From the estimates of the predictor variables, the estimate of SEX and BMI has a significant role on the response variable Insulin with BMI being highly significant at 95% confidence level.

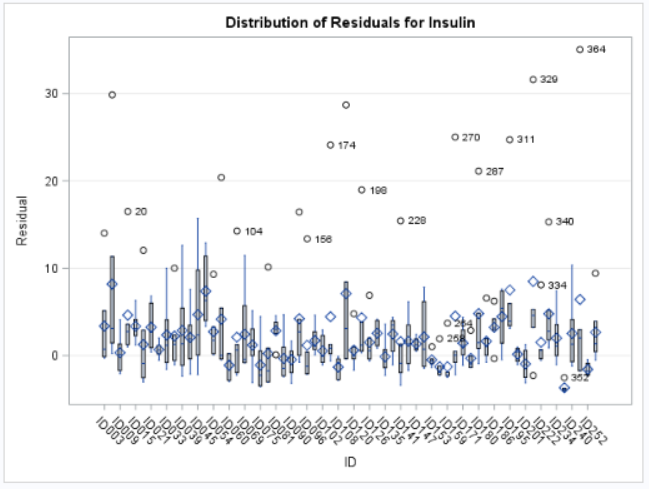


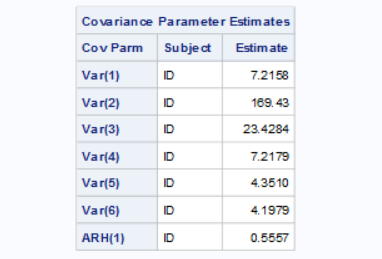
We can see that BMI is highly significant from the above table. The below plots of the residuals for Insulin, we can see that the residual converges to a straight line in the Q-Qplot of Pearson Residuals plots for Insulin.











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.