Problem 1:

Full model :
Uric Acid level $= 92.046 + 1.422 \cdot$ Diastolic Blood pressure $+ 4.593 \cdot$ High-density lipoprotein cholesterol $- 6.459 \cdot$ Total cholesterol $+ 99.701 \cdot$ Triglycerides level in body fat $+ 0.424 \cdot$ Alcohol intake

Best model using R^2_{adj} criterion is the model with predictors Dia, choles, trig and alco:
Uric Acid level $= 98.230 + 1.432 \cdot$Diastolic Blood pressure$- 6.195 \cdot$Total cholesterol$+ 98.582 \cdot$Triglycerides level in body fat$+$ $0.431 \cdot$ Alcohol intake

Problem 2:

1. Test for linearity (Lack Of Fit Test): H_0 : Linear model holds vs. H_A : Linear model does not hold.
   $\overline{\text{SSLF}} = 9,214,581$; $\overline{\text{SSPE}} = 25,088$; $F_{obs} = 0.37$; p-value $=0.8994 > 0.05 \implies$ fail to rejectH_0. Therefore, we can conclude that linear model holds.

2. Residual Plots for Homogeneity assumptions: All four predictors' residual plot show no pattern but taking a closer look can help us infer that there may be presence of outliers that may be influential on further analysis. The residual plot for predictor alcohol intake shows a slightly funnel shaped pattern on a closer look implying that this variable may have non-constant variance.
   Finally, the absolute residual plot shows no obvious pattern and we can conclude that there is no heteroscedasticity. But, we perform the following tests to check our assumption numerically:

   - Brown-Forsythe test (Does not depend on normality of errors):
     H_0 : Constant Variance vs. H_A : Non - constant variance

Results table			
Predictor	Test statistic, F_{BF}	p-value	Conclusion
Diastolic BP	15.36	$< 0.0001 < \alpha = 0.05$	Reject H_0
Total Cholestrol	6.63	$0.0102 > \alpha$	Fail to Reject H_0
Triglyceride levels	58.80	$< 0.0001 < \alpha$	Reject H_0
Alcohol Intake	20.87	$< 0.00001 < \alpha$	Reject H_0

     Therefore, since we fail to reject H_0 for only total cholestrol, we can conclude that there is noon-constant variance in the error terms and hence a transformation may be required for the other three predictor variables.

   - Breush-Pagan test (Depends on normality of errors):
     H_0 : Constant Variance vs. H_A : Non - constant variance

Results table			
Predictor	Test statistic, χ^2_{BP}	p-value	Conclusion
Full model	129.7	$< 0.0001 < \alpha = 0.05$	Reject H_0
Diastolic BP	1.63	$0.2022 > \alpha$	Fail to Reject H_0
Total Cholestrol	3.77	$0.0523 > \alpha$	Fail to Reject H_0
Triglyceride level	85.35	$< 0.0001 < \alpha$	Reject H_0
Alcohol Intake	52.83	$< 0.0001 < \alpha$	Reject H_0

     Notice that in the Brown-forsythe test, even predictor "chol" was concluded to have non-constant variance; This is a contardiction. But, when we take a closer look at the residual plot for this variable, there is no pattern and we can strongly conclude that total cholestrol has constant variance and these contradictory results may be due to the presence of influrntial points or outliers.
     Therefore, since we reject H_0 for the test with the full model and with individual predictors triglyceride levels and alcohol intake, we can conclude that there is non-constant variance in the model caused by these variables. And our assumption from the residual plot for was mostly correct.

3. Test for normality:

   - QQ plot: The QQ plot of residuals looks like it is very slightly skewed to the right but most data points lie on the 45^o line. There are a few points at the ends which depicts presence of outliers. So, a transformation

may be required for the response variable. (This was checked by performing BoxCox transformation and resulted in $\lambda = 0 \implies$ log transformation.)

- Shapiro-Wilks test:
  H_0 : Response Uric acid levels is normally distributed. vs H_A : Uric acid levels is not normally distributed.
  $W_{obs} = 0.8996$ and $p =< 0.0001 < \alpha = 0.05$. Hence, we reject H_0 and conclude that response variable Uric acid levels is not normally distributed and requires the aforementioned transformation.

4. Test for presence of outliers:
   We use the Bonferroni method to calculate $t_i = t_{(1-\frac{0.05}{2\cdot 998}, 998-5-1)=992}$. Using this method, observations $267, 477, 483$ are outliers.

5. Test to determine influential points:
   Influential points were determined based on $h_{ii} = 1$, $|DFFITS_i| > 1$ and $F_{(p,n-p)} \approx 20$ percentile. This resulted in 122 observations being influential. All of them had $h_{ii} = 1$

6. Test for presence of multicollinearity:

Collinearity Diagnostics			
Predictor	VIF	Tolerance	Condition Index
Diastolic BP	$1.095 < 10$	$0.913 > 0.1$	$2.177 < 30$
Total Cholestrol	$1.127 < 10$	$0.886 > 0.1$	$4.104 < 30$
Triglyceride level	$1.136 < 10$	$0.880 > 0.1$	$11.136 < 30$
Alcohol Intake	$1.049 < 10$	$0.953 > 0.1$	$18.343 < 30$

   From the table we can clearly see that since none of the assumptions are violated, there is no presence of multicollinearity (also see from correlation coefficients table)

Problem 3:

WLS model:
Uric Acid level = 77.791+1.680·Diastolic Blood pressure−3.858·Total cholesterol+86.598·Triglycerides level in body fat+ $0.435 \cdot$ Alcohol intake.
After iterating 3 times, we achieved convergence.

Results					
Iteration	b_0	b_1 (DBP)	b_2 (Chol)	b_3 (Trig)	b_4 (Alc)
0 (OLS)	98.230	1.432	-6.195	98.582	0.431
1 (WLS)	77.791	1.680	-3.858	86.598	0.435
2	75.263	1.693	-3.483	85.840	0.444
3	75.071	1.695	-3.467	85.760	0.447
4	75.044	1.695	-3.467	85.760	0.447

We can see from the above table that after the third iteration, the parameter estimates are the same (except for the intercept). From the plots of absolute residuals, variables like triglyceride level and alcohol intake seemed to have non-constant variance and their respective parameter estimates have improved because of WLS method and repeated iteration until convergence. Therefore, iterating the process of estimating weights improved the estimates.

Problem 4:

Results					
Iteration	b_0	b_1 (DBP)	b_2 (Chol)	b_3 (Trig)	b_4 (Alc)
0 (OLS)	98.230	1.432	-6.195	98.582	0.431
1 (WLS)	86.581	1.647	-7.061	94.752	0.373
2	85.756	1.689	-7.162	92.557	0.351
3	85.856	1.699	-7.118	91.406	0.343

From the above table, we can notice that the values for the iterations vary from that of problem 3 above. (For residuals and weights see output tables).
For the bisquare weighted function, as weight goes down, residuals go up and this is reflected in the output attached. Also notice that parameter estimates don't change drastically because of the presence of very few (3) outliers. Since

our values of residuals are quite high, we can confirm that there are many influential data points.
Therefore, overall since the parameter coefficients are not changing substantially after multiple iterations, we can conclude that the model has adjusted to outliers.

Problem 5:

1. Resampling of residuals (fixed X sampling)

   - Histogram and Q-Q plot of the bootstrap distribution of $\hat{\beta}_1$ [*SEE Relevant SAS outputs*]:

     The histogram resembles mostly a normal distribution; it is bell-shaped but not as symmetric. Maybe more bootstrap samples will achieve more symmetry.
     The Q-Q plot indicates normality because all the data points lie on the 45^0 line with very few outliers at the ends, which is not unusual.
   - Bias and standard error of $\hat{\beta}_1$: 0.002921 and 0.0686 respectively.
   - 2.5^{th} and 97.5^{th} percentiles of the sampling distribution of $\hat{\beta}_1$: $[104.037, 104.304]$
   - 2.5^{th} and 97.5^{th} percentiles of the sampling distribution of $\hat{\beta}_1 - \beta_1 = \text{Bias}$: $[0.01913, 0.01913]$
   - 95% CI for β_1 using:
     normal approximation: $[104.038, 104.308]$
     basic bootstrap: $[104.042, 104.309]$
     percentile bootstrap: $[104.037, 104.304]$

   Notice that the 2.5^{th} and 97.5^{th} percentiles of the sampling distribution of $\hat{\beta}_1 - \beta_1 = \text{Bias}$ are identical. This indicates that there is no variability in bias across the bootstrap samples. Also, notice that bias is very small $\implies$ almost unbiasedness.

2. Resampling of (X, Y)

   - Histogram and Q-Q plot of the bootstrap distribution of $\hat{\beta}_1$ [*SEE Relevant SAS outputs*]:

     The histogram resembles a normal distribution; it is bell-shaped and very close to symmetry.
     The Q-Q plot indicates normality because all the data points lie on the 45^0 line with extremely few outliers at the ends, which is not unusual.
   - Bias and standard error of $\hat{\beta}_1$: -0.4298 and 8.7411 respectively.
   - 2.5^{th} and 97.5^{th} percentiles of the sampling distribution of $\hat{\beta}_1$: $[87.3494, 104.304]$
   - 2.5^{th} and 97.5^{th} percentiles of the sampling distribution of $\hat{\beta}_1 - \beta_1 = \text{Bias}$: $[-16.8234, 17.7369]$
   - 95% CI for β_1 using:
     normal approximation: $[86.6103, 120.876]$
     basic bootstrap: $[86.4359, 120.996]$
     percentile bootstrap: $[87.3494, 121.910]$

3. Comparison:
   Looking at the results of the above methods, we see that the second method has more wider confidence intervals and percentiles compared to the first method of resampling. A negative bias indicates underestimation in method 2. The first method has a smaller SE whereas, the latter has a much larger SE.
   Therefore, for this dataset, the fixed X resampling method is more reliable for providing CI's and parameter estimates of $\hat{\beta}_1$ because it provides more precise and stable estimates compared to resampling of (X, Y).

   In the simple linear regression model, $\hat{\beta}_1 = 104.172$, $\text{SE}(\hat{\beta}_1) = 3.803$ and $95\%CI = [96.709, 111.636]$. These results match very closely with the first method of resampling.

Problem 6:

Univariate Logistic regression models for all predictors:

Results					
Predictors	Model Fit, $logit$(**Class**) =	H_0	Wald test Statistic, Z^*_{obs}	p-value	Conclusion
Clump Thickness	$5.0637 - 0.9191 \cdot$ Cl Thic	$\beta_1 = 0$	158.2321	< 0.0001	Reject H_0, significant
Size Uniformity	$5.1743 - 1.5980 \cdot$ Size Unif	$\beta_2 = 0$	143.3401	< 0.0001	Reject H_0, significant
Shape uniformity	$5.1642 - 1.4726 \cdot$ Shape Unif	$\beta_3 = 0$	158.2321	< 0.0001	Reject H_0, significant
Marginal Adhesion	$3.2732 - 1.0439 \cdot$ Marg Adh	$\beta_4 = 0$	137.8512	< 0.0001	Reject H_0, significant
Epithelial Size	$5.0321 - 1.4604 \cdot$ Ep Size	$\beta_5 = 0$	147.0733	< 0.0001	Reject H_0, significant
Bare Nucleoli	$3.5188 - 0.8554 \cdot$ B Nuc	$\beta_6 = 0$	146.6061	< 0.0001	Reject H_0, significant
Bland Chromatin	$5.2797 - 1.3652 \cdot$ B Chr	$\beta_7 = 0$	135.4301	< 0.0001	Reject H_0, significant
Normal Nucleoli	$2.9770 - 0.9267 \cdot$ N Nuc	$\beta_8 = 0$	125.0152	< 0.0001	Reject H_0, significant
Mitoses	$2.4479 - 1.3425 \cdot$ Mito	$\beta_9 = 0$	56.1508	< 0.0001	Reject H_0, significant

All predictors are significant individually by logistic regression at a significance level of $\alpha < 0.1$.

Multiple regression logistic model with all significant predictors:
$logit$(**Class**) $= 10.067 - 0.52 \cdot$ Cl Thic $- 0.00015 \cdot$ Size Unif $- 0.33 \cdot$ Shape Unif $- 0.32 \cdot$ Marg Adh $- 0.092 \cdot$ Ep Size $- 0.38 \cdot$ B Nuc $- 0.44 \cdot$ B Chr $- 0.21 \cdot$ N Nuc $- 0.53 \cdot$ Mito

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$
$H_A :$ not all $\beta_k = 0$ in H_0
$-2\text{Log}L = 103.062$,
LRT, $G^2 = 779.1853$
p $=< 0.0001$. Therefore, we reject H_0 and state that all predictors are jointly significant.

Let us look at Wald test statistic values for which p > 0.1 for the full model above and drop each variable in order of decreasing p-value. The predictor variables with high p-value (in order) are as follows: Size Uniformity ($p = 0.9994$), Epithelial Size ($p = 0.5527$), Shape uniformity ($p = 0.1448$) and Mitoses ($p = 0.1028$).

1. When we first dropped size uniformity, we obtained a p value of < 0.0001 for the LRT and $-2\text{Log}L = 103.062$ $\implies$ difference between $-2\text{Log}L = 0$. So, the model did not change at all. Hence, we can drop this variable.

2. Next, we drop epithelial size and obtained the same result as above for the LRT. $-2\text{Log}L = 103.415$, therefore the difference between the above model and this model is $-2\text{Log}L = 0.353$, again is not substantially changing the model. So this variable can also be removed.

3. Next we drop shape uniformity and again receive same p-value as above model for the LRT, $-2\text{Log}L = 108.866$, taking the difference, $108.866 - 103.415 = 5.451$ $\implies$ that removing shape uniformity substantially changed the model so we retain it in the model and don't drop it.

4. Lastly, we drop mitoses (after retaining shape uniformity) and receive a p-value of < 0.0001 for the LRT. $-2\text{Log}L = 107.310$, taking the difference, $107.310 - 103.415 = 3.895$ $\implies$ the model changes substantially. Therefore, we must keep this variable because it helps the model perform better.

Therefore, the final model with significant predictors is:
$logit$(**Class**) $= 9.96 - 0.52 \cdot$ Cl Thic $- 0.35 \cdot$ Shape Unif $- 0.34 \cdot$ Marg Adh $- 0.38 \cdot$ B Nuc $- 0.45 \cdot$ B Chr $- 0.22 \cdot$ N Nuc $- 0.53 \cdot$ Mito. For this model, $G^2 = 778.832$, p $= < 0.0001$ and, $-2\text{Log}L = 103.415$.

Comparing this with the original model, the original LRT test statistic is very close to the best model, same for the likelihood ratios.

Coefficient interpretation of the above model:

1. For each unit increase in clump thickness, the odds of a tumor being benign decreases by $1 - e^{-0.527} = 1 - 0.5903 = 0.4096 \approx 41\%$. Therefore, larger clump thickness indicates stronger association with the tumor being malignant.
2. For each unit increase in shape uniformity, the odds of a tumor being benign decreases by $1 - e^{-0.3584} = 1 - 0.6987 = 0.3012 = 30.12\%$. Less uniform cell shape indicates higher tumor malignancy.
3. For each unit increase in marginal adhesion, the odds of a tumor being benign decreases by $1 - e^{-0.3416} = 1 - 0.7106 = 0.2893 = 28.93\%$. Reduced marginal adhesion indicates higher tumor malignancy.
4. For each unit increase in bare nucleoli, the odds of a tumor being benign decreases by $1 - e^{-0.3870} = 1 - 0.6790 = 0.3209 \approx 32\%$. More abnormal nucleoli are strongly linked to tumor malignancy.
5. For each unit increase in bland chromatin, the odds of a tumor being benign decreases by $1 - e^{-0.4581} = 1 - 0.6324 = 0.3675 = 36.75\%$. More bland chromatin indicates higher tumor malignancy.
6. For each unit increase in normal nucleoli, the odds of a tumor being benign decreases by $1 - e^{-0.2245} = 1 - 0.7989 = 0.2010 \approx 20\%$. More normal nucleoli higher tumor benignancy.
7. For each unit increase in mitoses, the odds of a tumor being benign decreases by $1 - e^{-0.5315} = 1 - 0.5877 = 0.4122 \approx 41\%$. Increased mitoses higher tumor malignancy.

Stepwise selection:
$\overline{logit(\textbf{Class}) = -9.74 + 0.61 \cdot \text{Cl Thic} + 0.36 \cdot \text{Shape Unif} + 0.33 \cdot \text{Marg Adh} + 0.37 \cdot \text{B Nuc} + 0.46 \cdot \text{B Chr} + 0.24 \cdot \text{N Nuc}}$.

In this model, predictor mitoses has been dropped indicating that a difference in the log likelihood ratios of 3.895 does not change the model substantially and hence can be dropped. This model is very similar to our final model but with mitoses dropped. Therefore, we can use this model suggested by the stepwise selection method and conclude that the predictors in this model have a statistically significant impact on predicting Class = 0.

Relevant SAS outputs:

1:

Full model

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	10118200	2023640	217.34	<.0001
Error	992	9236375	9310.86164		
Corrected Total	997	19354575			

Root MSE	96.49281	R-Square	0.5228
Dependent Mean	330.10120	Adj R-Sq	0.5204
Coeff Var	29.23128		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	92.04641	24.36508	3.78	0.0002
DBP	1	1.42445	0.22632	6.29	<.0001
HDL	1	4.59383	7.72338	0.59	0.5521
Chol	1	-6.45949	2.62444	-2.46	0.0140
Trig	1	99.70139	4.16454	23.94	<.0001
Alc	1	0.42497	0.04156	10.23	<.0001

Model selection with Adjusted R^2

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Adjusted R-Square Selection Method

Number of Observations Read	998
Number of Observations Used	998

Number in Model	Adjusted R-Square	R-Square	C(p)	BIC	Variables in Model
4	0.5207	0.5226	4.3538	9127.0536	DBP Chol Trig Alc
5	0.5204	0.5228	6.0000	9128.7134	DBP HDL Chol Trig Alc
3	0.5184	0.5199	8.0909	9130.7492	DBP Trig Alc
4	0.5179	0.5199	10.0579	9132.7164	DBP HDL Trig Alc
3	0.5018	0.5033	42.5347	9164.3399	Chol Trig Alc
4	0.5017	0.5037	43.6134	9165.3908	HDL Chol Trig Alc
2	0.5012	0.5022	42.8678	9164.6459	Trig Alc
3	0.5009	0.5024	44.4155	9166.1418	HDL Trig Alc
4	0.4704	0.4725	108.5703	9225.7450	DBP HDL Chol Trig
3	0.4651	0.4667	118.4880	9234.6307	DBP Chol Trig
3	0.4635	0.4651	121.8495	9237.6297	DBP HDL Trig
2	0.4599	0.4610	128.4891	9243.5150	DBP Trig
3	0.4398	0.4415	171.0373	9280.5095	HDL Chol Trig
2	0.4358	0.4370	178.4158	9286.7741	HDL Trig
2	0.4313	0.4325	187.7316	9294.6417	Chol Trig
1	0.4290	0.4296	191.6626	9297.9350	Trig
4	0.2440	0.2471	577.1514	9577.8574	DBP HDL Chol Alc
3	0.2297	0.2320	606.4329	9596.2133	DBP HDL Alc
3	0.1964	0.1988	675.5180	9638.2454	HDL Chol Alc
3	0.1815	0.1839	706.3700	9656.4583	DBP Chol Alc

2:

Lack of fit test

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Analysis of Variance

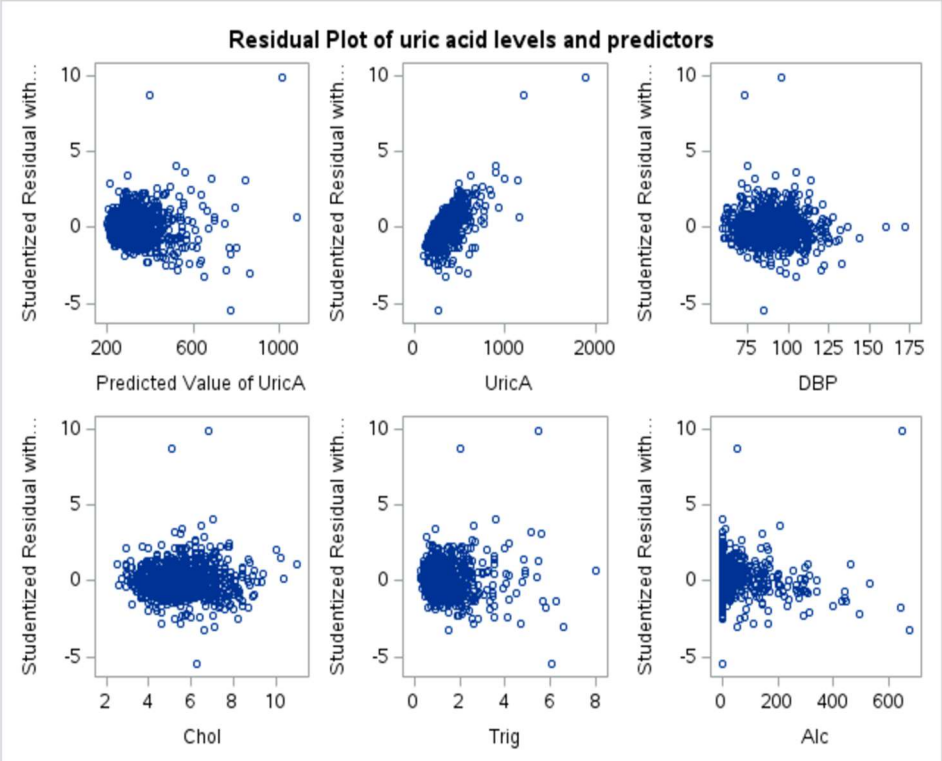
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10114906	2528727	271.77	<.0001
Error	993	9239669	9304.80238		
Lack of Fit	992	9214581	9288.89190	0.37	0.8994
Pure Error	1	25088	25088		
Corrected Total	997	19354575			

Root MSE	96.46140	R-Square	0.5226
Dependent Mean	330.10120	Adj R-Sq	0.5207
Coeff Var	29.22177		

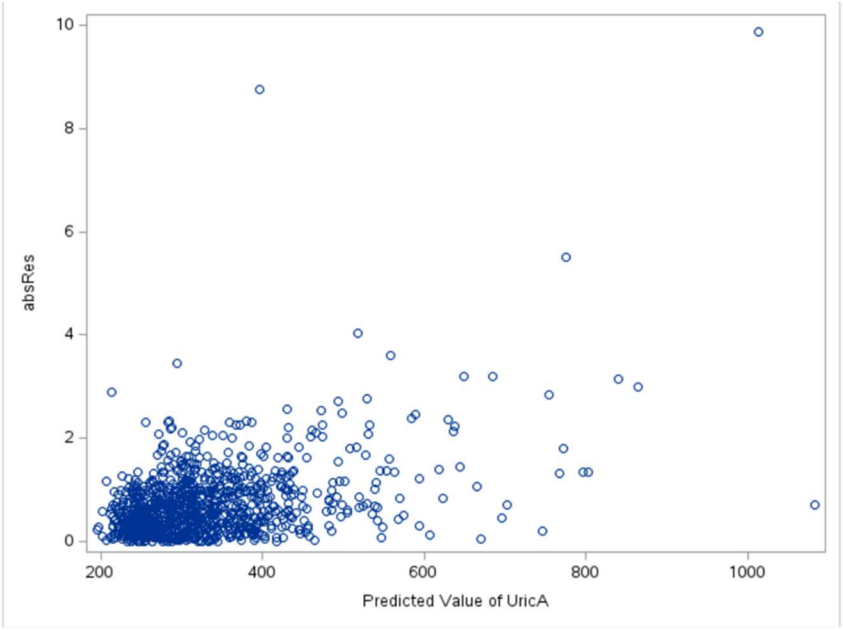
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	98.23003	22.02861	4.46	<.0001
DBP	1	1.43222	0.22587	6.34	<.0001
Chol	1	-6.19569	2.58585	-2.40	0.0168
Trig	1	98.58245	3.71421	26.54	<.0001
Alc	1	0.43157	0.04003	10.78	<.0001

Residual plot of Response vs. best model predictors



Absolute Residual Plot



Results of the Breush-Pagan Test (full model followed by the best model's predictors):

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
UricA	5	993	9239669	9304.8	96.4614	0.5226	0.5207

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
b0	98.23003	22.0286	4.46	<.0001
b1	1.432223	0.2259	6.34	<.0001
b2	-6.19569	2.5858	-2.40	0.0168
b3	98.58245	3.7142	26.54	<.0001
b4	0.431572	0.0400	10.78	<.0001

Number of Observations		Statistics for System	
Used	998	Objective	9258
Missing	0	Objective*N	9239669

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
UricA	White's Test	411.5	14	<.0001	Cross of all vars
	Breusch-Pagan	129.7	4	<.0001	DBP, Chol, Trig, Alc, 1

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
UricA	5	993	9239669	9304.8	96.4614	0.5226	0.5207

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
b0	98.23003	22.0286	4.46	<.0001
b1	1.432223	0.2259	6.34	<.0001
b2	-6.19569	2.5858	-2.40	0.0168
b3	98.58245	3.7142	26.54	<.0001
b4	0.431572	0.0400	10.78	<.0001

Number of Observations		Statistics for System	
Used	998	Objective	9258
Missing	0	Objective*N	9239669

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
UricA	Breusch-Pagan	1.63	1	0.2022	DBP, 1

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
UricA	5	993	9239669	9304.8	96.4614	0.5226	0.5207

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
b0	98.23003	22.0286	4.46	<.0001
b1	1.432223	0.2259	6.34	<.0001
b2	-6.19569	2.5858	-2.40	0.0168
b3	98.58245	3.7142	26.54	<.0001
b4	0.431572	0.0400	10.78	<.0001

Number of Observations		Statistics for System	
Used	998	Objective	9258
Missing	0	Objective*N	9239669

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
UricA	Breusch-Pagan	3.77	1	0.0523	Chol, 1

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
UricA	5	993	9239669	9304.8	96.4614	0.5226	0.5207

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
b0	98.23003	22.0286	4.46	<.0001
b1	1.432223	0.2259	6.34	<.0001
b2	-6.19569	2.5858	-2.40	0.0168
b3	98.58245	3.7142	26.54	<.0001
b4	0.431572	0.0400	10.78	<.0001

Number of Observations		Statistics for System	
Used	998	Objective	9258
Missing	0	Objective*N	9239669

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
UricA	Breusch-Pagan	85.35	1	<.0001	Trig, 1

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
UricA	5	993	9239669	9304.8	96.4614	0.5226	0.5207

Nonlinear OLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
b0	98.23003	22.0286	4.46	<.0001
b1	1.432223	0.2259	6.34	<.0001
b2	-6.19569	2.5858	-2.40	0.0168
b3	98.58245	3.7142	26.54	<.0001
b4	0.431572	0.0400	10.78	<.0001

Number of Observations		Statistics for System	
Used	998	Objective	9258
Missing	0	Objective*N	9239669

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
UricA	Breusch-Pagan	52.83	1	<.0001	Alc, 1

Results of the Brown - Forsythe Test:

Diastolic BP

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of R Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Group	1	7.9053	7.9053	15.36	<.0001
Error	996	512.7	0.5147		

Total Cholesterol

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of R Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Group	1	3.4435	3.4435	6.63	0.0102
Error	996	517.0	0.5191		

Triglyceride levels

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of R Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Group	1	29.1857	29.1857	58.80	<.0001
Error	996	494.4	0.4964		

Alcohol Intake

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of R Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Group	1	10.5983	10.5983	20.87	<.0001
Error	996	505.8	0.5078		

Test for Normality:

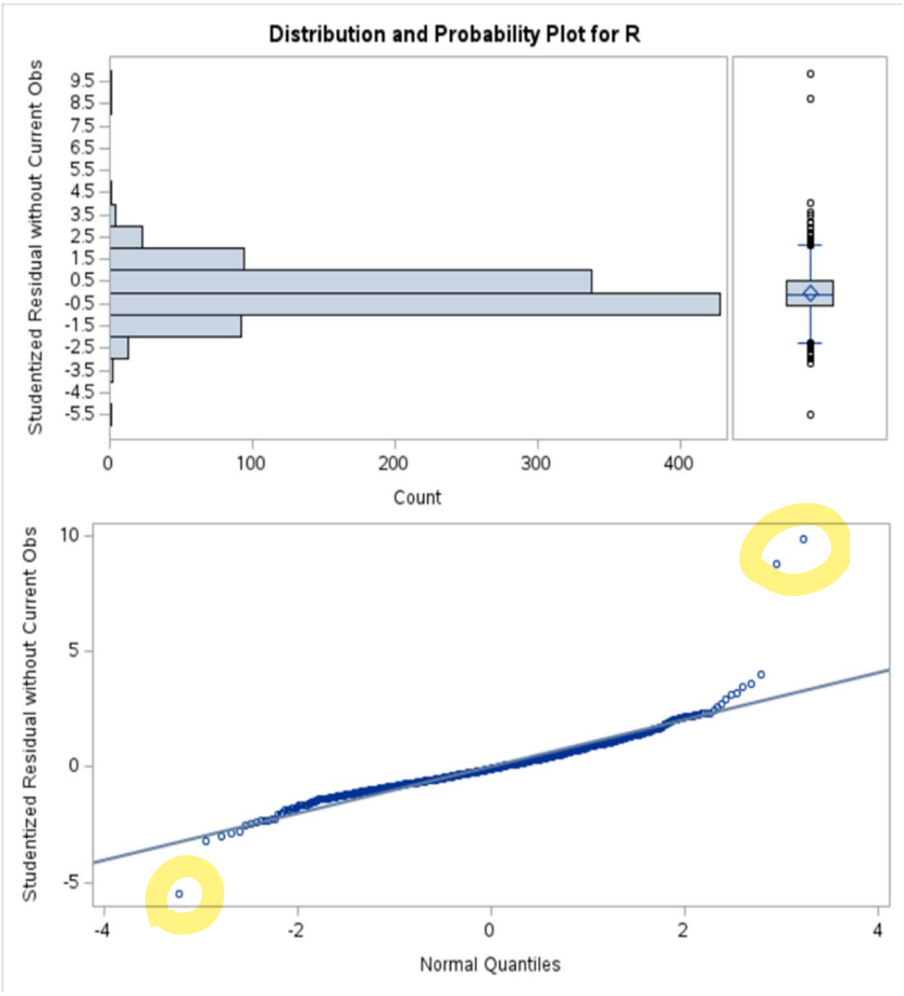
The UNIVARIATE Procedure
Variable: R (Studentized Residual without Current Obs)

Moments			
N	998	Sum Weights	998
Mean	0.00081713	Sum Observations	0.81550062
Std Deviation	1.01400792	Variance	1.02821207
Skewness	1.67983467	Kurtosis	15.3140046
Uncorrected SS	1025.1281	Corrected SS	1025.12743
Coeff Variation	124093.089	Std Error Mean	0.03209786

Basic Statistical Measures			
Location		Variability	
Mean	0.00082	Std Deviation	1.01401
Median	-0.07964	Variance	1.02821
Mode	.	Range	15.38549
		Interquartile Range	1.09519

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0.025458	Pr > |t|	0.9797
Sign	M	-37	Pr >= |M|	0.0208
Signed Rank	S	-14458.5	Pr >= |S|	0.1125

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.899635	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.065026	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.489844	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	9.503758	Pr > A-Sq	<0.0050



Outliers:

Obs	RStudent
267	8.7634
477	-5.5039
483	9.8816

Influential Observations:

Obs	HatDiagonal	hilev	DFFITS	dfflag	CooksD	Fpercent	DFB_DBP	b1flag	DFB_Chol	b2flag	DFB_Trig	b3flag	DFB_Alc	b4flag
1	0.0081	1	-0.0771	0	0.001	0.0003	0.0118	0	-0.0205	0	-0.0597	0	0.0130	0
7	0.0084	1	0.3325	0	0.022	0.0929	0.0568	0	0.0423	0	0.1206	0	0.2345	0
11	0.0087	1	-0.1577	0	0.005	0.0049	0.0281	0	-0.0391	0	-0.1248	0	0.0254	0
14	0.0084	1	-0.1683	0	0.006	0.0064	-0.0062	0	-0.0710	0	-0.1074	0	0.0306	0
22	0.0122	1	-0.0493	0	0.000	0.0000	0.0055	0	-0.0174	0	-0.0366	0	0.0071	0
28	0.0174	1	-0.0874	0	0.002	0.0005	-0.0279	0	-0.0284	0	0.0060	0	-0.0676	0
31	0.0645	1	0.1859	0	0.007	0.0095	0.0343	0	-0.0417	0	0.1714	0	0.0031	0
44	0.0124	1	-0.2673	0	0.014	0.0398	0.0550	0	0.0208	0	-0.2491	0	0.0436	0
46	0.0136	1	-0.1443	0	0.004	0.0035	0.0489	0	-0.0115	0	-0.1278	0	0.0035	0
47	0.0085	1	-0.0609	0	0.001	0.0001	-0.0318	0	-0.0419	0	0.0169	0	0.0105	0
49	0.0094	1	0.0503	0	0.001	0.0001	0.0361	0	0.0245	0	-0.0088	0	-0.0111	0
52	0.0085	1	-0.0866	0	0.001	0.0004	-0.0656	0	-0.0370	0	0.0214	0	0.0200	0
68	0.0080	1	0.0432	0	0.000	0.0000	0.0358	0	0.0131	0	-0.0070	0	-0.0035	0
74	0.0237	1	0.1648	0	0.005	0.0059	0.0137	0	0.1557	0	-0.0671	0	-0.0021	0
79	0.0085	1	-0.0053	0	0.000	0.0000	-0.0048	0	-0.0004	0	0.0014	0	0.0014	0
85	0.0127	1	-0.2574	0	0.013	0.0343	-0.0535	0	-0.0008	0	-0.0178	0	-0.2218	0
95	0.0105	1	-0.0066	0	0.000	0.0000	-0.0027	0	-0.0051	0	0.0028	0	0.0008	0
103	0.0121	1	-0.0964	0	0.002	0.0007	-0.0878	0	0.0362	0	-0.0044	0	0.0292	0
105	0.0125	1	0.0425	0	0.000	0.0000	0.0336	0	-0.0247	0	-0.0023	0	0.0027	0
106	0.0083	1	0.0731	0	0.001	0.0002	-0.0177	0	0.0659	0	-0.0296	0	-0.0005	0
110	0.0084	1	0.0368	0	0.000	0.0000	0.0305	0	0.0095	0	-0.0148	0	-0.0087	0
124	0.0710	1	-0.8859	0	0.156	3.9468	0.0457	0	-0.1503	0	0.0872	0	-0.8680	0
142	0.0147	1	0.0183	0	0.000	0.0000	-0.0020	0	-0.0027	0	-0.0040	0	0.0167	0
149	0.0456	1	-0.0467	0	0.000	0.0000	0.0053	0	-0.0087	0	0.0105	0	-0.0456	0
152	0.0258	1	0.0059	0	0.000	0.0000	0.0038	0	-0.0017	0	-0.0013	0	0.0032	0
160	0.0253	1	0.0314	0	0.000	0.0000	0.0044	0	-0.0070	0	0.0293	0	-0.0064	0
184	0.0087	1	0.1468	0	0.004	0.0037	0.1263	0	0.0221	0	-0.0616	0	-0.0359	0
193	0.0082	1	-0.0914	0	0.002	0.0006	-0.0084	0	-0.0802	0	0.0082	0	0.0094	0
194	0.0083	1	-0.0657	0	0.001	0.0001	-0.0526	0	0.0318	0	0.0171	0	0.0183	0
230	0.0087	1	-0.1276	0	0.003	0.0021	-0.0600	0	0.0051	0	-0.0872	0	0.0365	0
231	0.0200	1	0.0641	0	0.001	0.0001	0.0126	0	-0.0148	0	0.0583	0	-0.0141	0
233	0.0742	1	-0.5073	0	0.051	0.4936	0.1218	0	-0.1498	0	0.1074	0	-0.4908	0
244	0.0089	1	-0.0928	0	0.002	0.0006	0.0091	0	-0.0348	0	0.0277	0	-0.0817	0
245	0.0142	1	-0.0257	0	0.000	0.0000	0.0033	0	0.0053	0	0.0043	0	-0.0231	0
246	0.0160	1	-0.0950	0	0.002	0.0007	-0.0582	0	0.0497	0	-0.0134	0	-0.0408	0
258	0.0302	1	-0.1172	0	0.003	0.0015	0.0105	0	0.0179	0	0.0035	0	-0.1114	0
269	0.0088	1	0.0068	0	0.000	0.0000	-0.0005	0	-0.0012	0	-0.0004	0	0.0061	0
303	0.0177	1	-0.1130	0	0.003	0.0013	0.0332	0	0.0290	0	-0.1081	0	0.0110	0
311	0.0216	1	0.1598	0	0.005	0.0052	-0.0497	0	-0.0233	0	0.0891	0	0.1192	0
383	0.0165	1	0.2764	0	0.015	0.0455	0.0538	0	-0.1630	0	0.2378	0	-0.0245	0
390	0.0292	1	-0.2323	0	0.011	0.0230	0.0106	0	-0.0603	0	0.0119	0	-0.2207	0

390	0.0292	1	-0.2323	0	0.011	0.0230	0.0106	0	-0.0603	0	0.0119	0	-0.2207	0
397	0.0215	1	-0.0082	0	0.000	0.0000	0.0019	0	-0.0035	0	-0.0058	0	-0.0008	0
399	0.0083	1	0.0419	0	0.000	0.0000	-0.0019	0	-0.0143	0	-0.0115	0	0.0321	0
402	0.0162	1	0.1963	0	0.008	0.0118	0.0099	0	0.1866	0	-0.0613	0	0.0123	0
406	0.0106	1	0.0071	0	0.000	0.0000	0.0006	0	0.0021	0	-0.0030	0	0.0060	0
407	0.0081	1	-0.0591	0	0.001	0.0001	0.0342	0	-0.0470	0	0.0055	0	-0.0012	0
411	0.0206	1	0.0593	0	0.001	0.0001	-0.0276	0	0.0317	0	0.0349	0	-0.0017	0
421	0.0106	1	0.4178	0	0.034	0.2262	-0.1862	0	0.0645	0	0.3399	0	-0.0436	0
432	0.0127	1	0.0568	0	0.001	0.0001	-0.0248	0	-0.0183	0	0.0463	0	0.0190	0
440	0.0377	1	-0.5942	0	0.070	0.8961	0.0386	0	0.0565	0	-0.5716	0	0.0185	0
441	0.0113	1	-0.0899	0	0.002	0.0005	-0.0014	0	0.0091	0	0.0181	0	-0.0816	0
442	0.0298	1	0.0142	0	0.000	0.0000	0.0139	0	-0.0012	0	-0.0019	0	-0.0035	0
449	0.0293	1	0.5464	0	0.059	0.6489	0.0945	0	-0.1682	0	0.4904	0	0.0676	0
456	0.0367	1	-0.4369	0	0.038	0.2754	-0.0261	0	-0.0207	0	-0.0116	0	-0.4134	0
459	0.0094	1	-0.0140	0	0.000	0.0000	0.0042	0	0.0070	0	-0.0042	0	-0.0098	0
477	0.0336	1	-1.0261	1	0.205	6.4063	0.1845	0	0.1811	0	-1.0025	1	0.1376	0
483	0.0829	1	2.9702	1	1.608	82.9945	-0.5668	0	0.0981	0	1.3270	1	2.4954	1
490	0.0253	1	0.5172	0	0.053	0.5246	-0.1234	0	-0.1625	0	0.5013	0	-0.0715	0
495	0.0112	1	-0.0632	0	0.001	0.0001	0.0057	0	-0.0059	0	0.0169	0	-0.0590	0
499	0.0382	1	-0.2764	0	0.015	0.0458	0.1024	0	-0.0875	0	-0.0391	0	-0.2503	0
500	0.0154	1	-0.2944	0	0.017	0.0583	-0.2093	0	0.0600	0	-0.1542	0	0.0119	0
506	0.0159	1	0.0127	0	0.000	0.0000	-0.0001	0	0.0118	0	-0.0010	0	-0.0006	0
507	0.0109	1	-0.1107	0	0.002	0.0012	-0.0966	0	-0.0129	0	0.0503	0	0.0256	0
508	0.0198	1	0.2061	0	0.008	0.0143	-0.0653	0	-0.0087	0	0.1921	0	-0.0216	0
513	0.0095	1	0.1337	0	0.004	0.0025	0.0175	0	-0.0530	0	0.1196	0	-0.0337	0
523	0.0269	1	-0.2688	0	0.014	0.0409	-0.0335	0	-0.0724	0	0.0819	0	-0.2422	0
525	0.0116	1	-0.0401	0	0.000	0.0000	0.0187	0	0.0055	0	0.0040	0	-0.0340	0
528	0.0086	1	-0.0043	0	0.000	0.0000	-0.0033	0	0.0021	0	0.0011	0	-0.0005	0
533	0.0093	1	0.0603	0	0.001	0.0001	0.0145	0	0.0452	0	0.0084	0	-0.0091	0
534	0.0089	1	0.1517	0	0.005	0.0042	-0.0474	0	-0.0280	0	0.1376	0	0.0193	0
535	0.0293	1	-0.3126	0	0.020	0.0742	-0.0327	0	-0.0483	0	-0.2619	0	0.0514	0
544	0.0104	1	0.1096	0	0.002	0.0012	0.0800	0	-0.0720	0	0.0285	0	-0.0340	0
582	0.0299	1	-0.1250	0	0.003	0.0019	-0.0787	0	-0.0055	0	0.0350	0	-0.0773	0
583	0.0147	1	-0.0694	0	0.001	0.0002	0.0107	0	0.0123	0	0.0159	0	-0.0620	0
588	0.0142	1	-0.0960	0	0.002	0.0007	-0.0160	0	0.0041	0	0.0081	0	-0.0863	0
592	0.0137	1	-0.0304	0	0.000	0.0000	0.0013	0	0.0111	0	0.0057	0	-0.0247	0
605	0.0143	1	-0.0892	0	0.002	0.0005	0.0062	0	0.0286	0	-0.0099	0	-0.0775	0
625	0.0144	1	-0.0255	0	0.000	0.0000	-0.0237	0	0.0085	0	0.0046	0	0.0070	0
633	0.0138	1	-0.0868	0	0.002	0.0005	-0.0288	0	0.0155	0	0.0188	0	-0.0691	0
634	0.0115	1	-0.0921	0	0.002	0.0006	-0.0816	0	0.0214	0	0.0067	0	-0.0148	0
638	0.0085	1	-0.0254	0	0.000	0.0000	-0.0210	0	-0.0073	0	0.0086	0	0.0061	0
643	0.0121	1	-0.0260	0	0.000	0.0000	-0.0215	0	0.0151	0	-0.0013	0	0.0077	0

652	0.0091	1	-0.0512	0	0.001	0.0001	-0.0451	0	-0.0089	0	0.0143	0	0.0132	0
655	0.0086	1	0.0361	0	0.000	0.0000	-0.0086	0	0.0325	0	-0.0015	0	-0.0015	0
657	0.0084	1	-0.0259	0	0.000	0.0000	-0.0088	0	0.0002	0	-0.0196	0	0.0070	0
661	0.0152	1	0.0136	0	0.000	0.0000	-0.0054	0	0.0020	0	-0.0035	0	0.0124	0
662	0.0110	1	-0.2622	0	0.014	0.0368	-0.1602	0	-0.1040	0	-0.0706	0	0.0627	0
664	0.0089	1	0.0279	0	0.000	0.0000	-0.0133	0	0.0240	0	-0.0058	0	0.0006	0
680	0.0086	1	0.0118	0	0.000	0.0000	-0.0008	0	0.0108	0	-0.0051	0	-0.0004	0
720	0.0089	1	-0.1559	0	0.005	0.0047	0.0447	0	-0.0196	0	0.0303	0	-0.1440	0
724	0.0483	1	0.0177	0	0.000	0.0000	0.0137	0	0.0021	0	-0.0048	0	0.0078	0
727	0.0158	1	0.0489	0	0.000	0.0000	0.0017	0	0.0125	0	-0.0142	0	0.0448	0
736	0.0329	1	-0.2119	0	0.009	0.0160	0.0567	0	-0.0543	0	0.0255	0	-0.2052	0
738	0.0130	1	0.0345	0	0.000	0.0000	-0.0065	0	-0.0022	0	0.0322	0	-0.0056	0
771	0.0094	1	-0.1239	0	0.003	0.0019	0.0127	0	-0.0273	0	0.0355	0	-0.1136	0
795	0.0097	1	-0.0876	0	0.002	0.0005	0.0508	0	-0.0469	0	-0.0426	0	0.0031	0
800	0.0100	1	0.0701	0	0.001	0.0002	0.0036	0	0.0003	0	0.0171	0	0.0598	0
803	0.0082	1	-0.1501	0	0.004	0.0040	-0.1202	0	0.0733	0	-0.0390	0	0.0496	0
807	0.0084	1	-0.0227	0	0.000	0.0000	0.0065	0	-0.0203	0	0.0009	0	0.0008	0
818	0.0119	1	-0.1204	0	0.003	0.0017	0.0021	0	0.0193	0	0.0226	0	-0.1084	0
844	0.0114	1	-0.0120	0	0.000	0.0000	-0.0015	0	-0.0108	0	0.0055	0	0.0006	0
851	0.0216	1	0.1388	0	0.004	0.0030	-0.0594	0	0.0704	0	-0.0331	0	0.1169	0
880	0.0091	1	-0.0078	0	0.000	0.0000	-0.0064	0	0.0040	0	0.0006	0	-0.0004	0
886	0.0094	1	0.0158	0	0.000	0.0000	-0.0023	0	-0.0135	0	0.0086	0	-0.0030	0
894	0.0126	1	-0.0327	0	0.000	0.0000	-0.0221	0	0.0100	0	0.0066	0	-0.0158	0
895	0.0086	1	0.0979	0	0.002	0.0007	0.0050	0	-0.0699	0	0.0124	0	0.0523	0
900	0.0341	1	-0.2546	0	0.013	0.0331	0.0187	0	0.0224	0	-0.2443	0	0.0377	0
907	0.0091	1	-0.0849	0	0.001	0.0004	0.0076	0	-0.0792	0	0.0193	0	0.0039	0
910	0.0089	1	-0.0090	0	0.000	0.0000	0.0051	0	-0.0055	0	-0.0035	0	0.0003	0
919	0.0102	1	0.0324	0	0.000	0.0000	0.0270	0	-0.0137	0	0.0094	0	-0.0058	0
922	0.0088	1	-0.0794	0	0.001	0.0003	0.0084	0	0.0097	0	-0.0456	0	-0.0545	0
924	0.0320	1	-0.2424	0	0.012	0.0272	0.0278	0	0.0930	0	-0.2370	0	0.0232	0
928	0.0114	1	-0.0152	0	0.000	0.0000	0.0018	0	-0.0018	0	-0.0132	0	-0.0000	0
944	0.0160	1	-0.2595	0	0.013	0.0355	0.0730	0	0.0312	0	-0.0194	0	-0.2417	0
953	0.0318	1	0.2076	0	0.009	0.0147	-0.0145	0	0.0189	0	-0.0160	0	0.2029	0
969	0.0142	1	0.2948	0	0.017	0.0586	-0.0991	0	0.0755	0	0.2390	0	-0.0037	0
971	0.0212	1	0.1053	0	0.002	0.0010	0.0211	0	-0.0317	0	0.0962	0	-0.0234	0
976	0.0090	1	-0.1122	0	0.003	0.0013	-0.0486	0	0.0803	0	-0.0700	0	0.0175	0
979	0.0091	1	-0.0063	0	0.000	0.0000	0.0030	0	-0.0052	0	0.0024	0	-0.0002	0
981	0.0238	1	-0.4427	0	0.039	0.2882	-0.1553	0	0.2029	0	-0.3623	0	-0.0172	0
985	0.0285	1	0.2301	0	0.011	0.0221	-0.0035	0	-0.0897	0	0.2179	0	0.0218	0
988	0.0154	1	0.2605	0	0.014	0.0360	-0.0124	0	0.2491	0	-0.0885	0	-0.0033	0

Multicollinearity:

The CORR Procedure

5 Variables:	UricA DBP Chol Trig Alc

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
UricA	998	330.10120	139.32987	329441	99.00000	1885
DBP	998	87.99549	14.15340	87820	59.00000	172.00000
Chol	998	5.58136	1.25468	5570	2.50000	11.00000
Trig	998	1.29167	0.87666	1289	0.28000	7.98000
Alc	998	30.32200	78.16114	30261	0	672.11432

Pearson Correlation Coefficients, N = 998
Prob > |r| under H0: Rho=0

	UricA	DBP	Chol	Trig	Alc
UricA	1.00000	0.30571 <.0001	0.14660 <.0001	0.65545 <.0001	0.32941 <.0001
DBP	0.30571 <.0001	1.00000	0.17675 <.0001	0.20183 <.0001	0.18544 <.0001
Chol	0.14660 <.0001	0.17675 <.0001	1.00000	0.30137 <.0001	-0.04235 0.1813
Trig	0.65545 <.0001	0.20183 <.0001	0.30137 <.0001	1.00000	0.09345 0.0031
Alc	0.32941 <.0001	0.18544 <.0001	-0.04235 0.1813	0.09345 0.0031	1.00000

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10114906	2528727	271.77	<.0001
Error	993	9239669	9304.80238		
Corrected Total	997	19354575			

Root MSE	96.46140	R-Square	0.5226
Dependent Mean	330.10120	Adj R-Sq	0.5207
Coeff Var	29.22177		

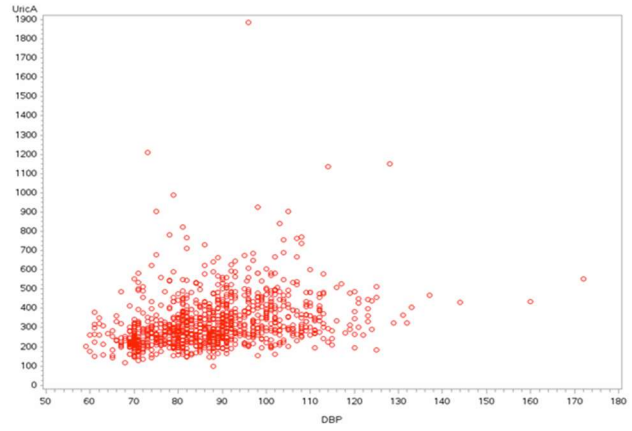
Parameter Estimates

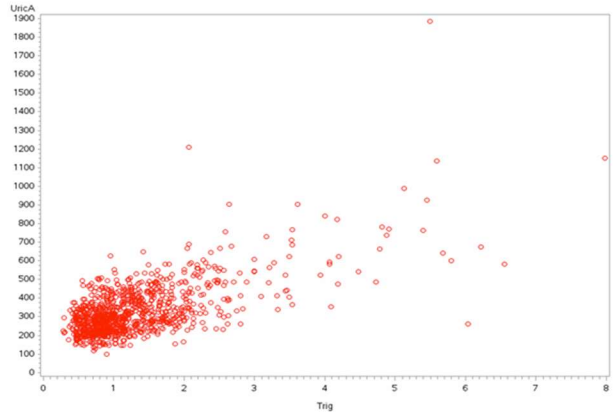
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	98.23003	22.02861	4.46	<.0001	.	0
DBP	1	1.43222	0.22587	6.34	<.0001	0.91321	1.09503
Chol	1	-6.19569	2.58585	-2.40	0.0168	0.88663	1.12787
Trig	1	98.58245	3.71421	26.54	<.0001	0.88027	1.13601
Alc	1	0.43157	0.04003	10.78	<.0001	0.95316	1.04914

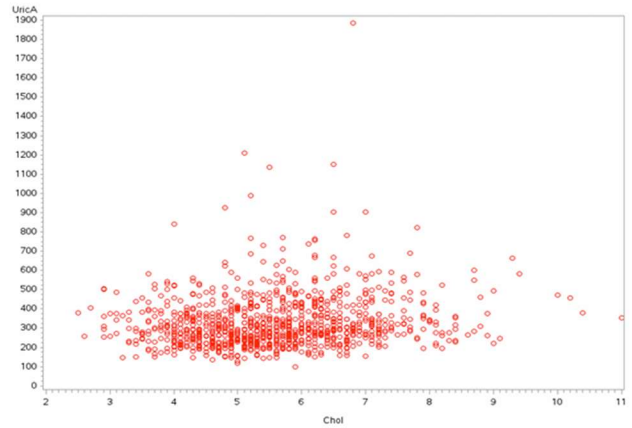
Collinearity Diagnostics

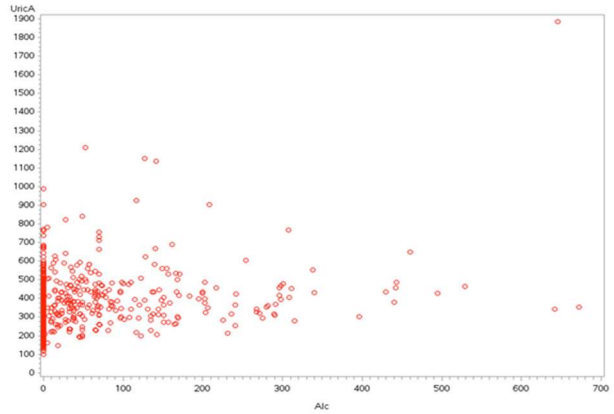
Number	Eigenvalue	Condition Index	Proportion of Variation				
			Intercept	DBP	Chol	Trig	Alc
1	3.90256	1.00000	0.00120	0.00145	0.00266	0.01479	0.01229
2	0.82275	2.17792	0.00039544	0.00029376	0.00118	0.00309	0.94699
3	0.23163	4.10464	0.00809	0.00767	0.00848	0.93269	0.00111
4	0.03147	11.13666	0.04167	0.19314	0.89848	0.02712	0.02703
5	0.01160	18.34382	0.94864	0.79744	0.08920	0.02231	0.01258

3:
Scatter Plot of response and predictors of best model:

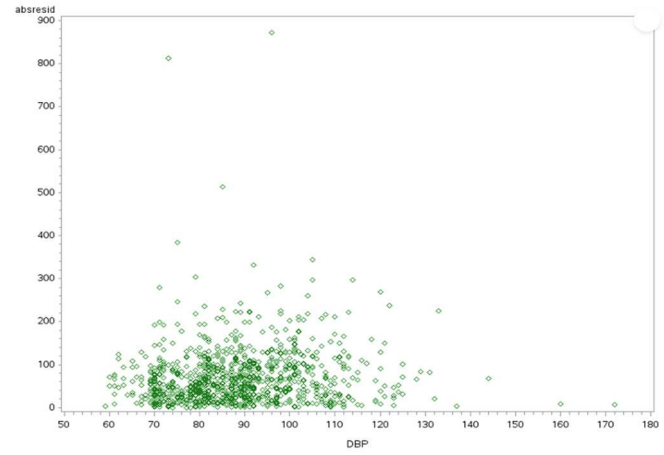


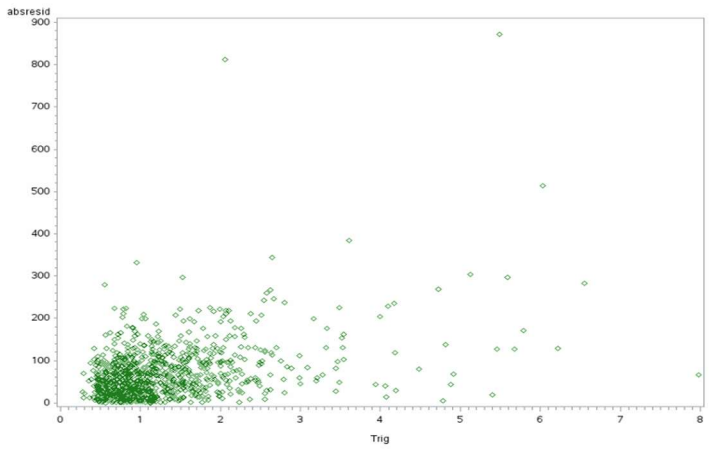


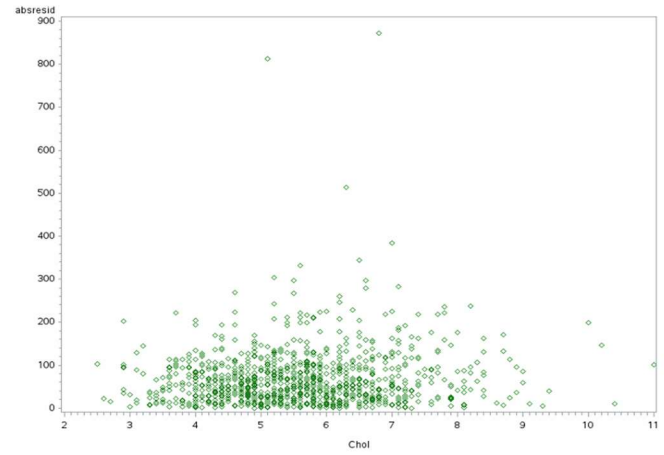


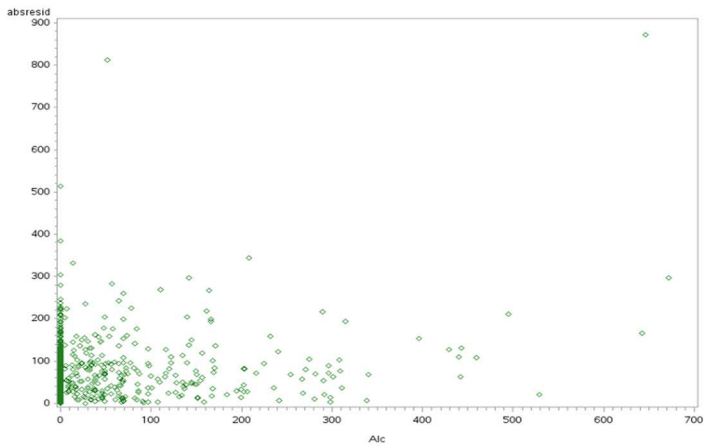


Plot of absolute residuals of predictors of best model:









Ordinary Least Squares

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10114906	2528727	271.77	<.0001
Error	993	9239669	9304.80238		
Corrected Total	997	19354575			

Root MSE	96.46140	R-Square	0.5226
Dependent Mean	330.10120	Adj R-Sq	0.5207
Coeff Var	29.22177		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	98.23003	22.02861	4.46	<.0001	55.00205	141.45801
DBP	1	1.43222	0.22587	6.34	<.0001	0.98898	1.87546
Chol	1	-6.19569	2.58585	-2.40	0.0168	-11.27005	-1.12134
Trig	1	98.58245	3.71421	26.54	<.0001	91.29384	105.87106
Alc	1	0.43157	0.04003	10.78	<.0001	0.35301	0.51013

Weighted Least Squares Iteration 1

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Weight: wt

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	977.00831	244.25208	143.81	<.0001
Error	993	1686.53871	1.69843		
Corrected Total	997	2663.54702			

Root MSE	1.30324	R-Square	0.3668
Dependent Mean	289.34824	Adj R-Sq	0.3643
Coeff Var	0.45040		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	77.79161	17.32962	4.49	<.0001	43.78473	111.79850
DBP	1	1.68099	0.18583	9.05	<.0001	1.31633	2.04564
Chol	1	-3.85856	2.18272	-1.77	0.0774	-8.14184	0.42472
Trig	1	86.59896	5.12979	16.88	<.0001	76.53249	96.66542
Alc	1	0.43564	0.05536	7.87	<.0001	0.32701	0.54428

Weighted Least Squares Iteration 2

Weighted Least Squares Iteration 3

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Weight: wt2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	972.61449	243.15362	141.82	<.0001
Error	993	1702.50366	1.71451		
Corrected Total	997	2675.11815			

Root MSE	1.30939	R-Square	0.3636
Dependent Mean	287.80505	Adj R-Sq	0.3610
Coeff Var	0.45496		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	75.26328	17.26337	4.36	<.0001	41.38640	109.14017
DBP	1	1.69389	0.18665	9.08	<.0001	1.32761	2.06017
Chol	1	-3.48379	2.16356	-1.61	0.1077	-7.72947	0.76189
Trig	1	85.84075	5.15626	16.65	<.0001	75.72233	95.95917
Alc	1	0.44443	0.05730	7.76	<.0001	0.33198	0.55688

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Weight: wt3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	972.27243	243.06811	141.67	<.0001
Error	993	1703.71358	1.71572		
Corrected Total	997	2675.98601			

Root MSE	1.30986	R-Square	0.3633
Dependent Mean	287.61327	Adj R-Sq	0.3608
Coeff Var	0.45542		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	75.07199	17.25729	4.35	<.0001	41.20704	108.93694
DBP	1	1.69546	0.18677	9.08	<.0001	1.32894	2.06197
Chol	1	-3.46718	2.16264	-1.60	0.1092	-7.71104	0.77669
Trig	1	85.76030	5.15525	16.64	<.0001	75.64386	95.87673
Alc	1	0.44709	0.05785	7.73	<.0001	0.33357	0.56060

Weighted Least Squares Iteration 4

The REG Procedure
Model: MODEL1
Dependent Variable: UricA

Number of Observations Read	998
Number of Observations Used	998

Weight: wt4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	972.32074	243.08019	141.66	<.0001
Error	993	1703.98295	1.71599		
Corrected Total	997	2676.30369			

Root MSE	1.30996	R-Square	0.3633
Dependent Mean	287.57451	Adj R-Sq	0.3607
Coeff Var	0.45552		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	75.04400	17.25648	4.35	<.0001	41.18064	108.90735
DBP	1	1.69580	0.18680	9.08	<.0001	1.32922	2.06237
Chol	1	-3.46712	2.16262	-1.60	0.1092	-7.71095	0.77671
Trig	1	85.74604	5.15447	16.64	<.0001	75.63114	95.86094
Alc	1	0.44781	0.05799	7.72	<.0001	0.33401	0.56161

4:

Iteration 0(Parameter estimates and first 10 observations of residuals. This iteration has no weight because it's OLS)

Obs	Model	Dependent	Variable	DF	Estimate	StdErr	tValue	Probt
1	MODEL1	UricA	Intercept	1	98.23003	22.02861	4.46	<.0001
2	MODEL1	UricA	DBP	1	1.43222	0.22587	6.34	<.0001
3	MODEL1	UricA	Chol	1	-6.19569	2.58585	-2.40	0.0168
4	MODEL1	UricA	Trig	1	98.58245	3.71421	26.54	<.0001
5	MODEL1	UricA	Alc	1	0.43157	0.04003	10.78	<.0001

OLS Residuals and Weights

Obs	residual
1	-81.973
2	-25.119
3	-92.576
4	-95.347
5	132.925
6	-48.991
7	344.511
8	217.638
9	-117.935
10	111.362

Iteration 1:

Obs	Model	Dependent	Variable	DF	Estimate	StdErr	tValue	Probt	LowerCL	UpperCL
1	MODEL1	UricA	Intercept	1	86.58116	16.80564	5.15	<.0001	53.60234	119.55997
2	MODEL1	UricA	DBP	1	1.64756	0.17291	9.53	<.0001	1.30824	1.98688
3	MODEL1	UricA	Chol	1	-7.06153	1.99899	-3.53	0.0004	-10.98427	-3.13878
4	MODEL1	UricA	Trig	1	94.75207	3.15925	29.99	<.0001	88.55246	100.95168
5	MODEL1	UricA	Alc	1	0.37386	0.03377	11.07	<.0001	0.30759	0.44013

Iteration 1 Residuals and Weights

Obs	residual1	wt
1	-69.906	0.90522
2	-15.492	0.99090
3	-88.190	0.87994
4	-94.836	0.87290
5	138.368	0.76088
6	-41.087	0.96560
7	361.310	0.02020
8	230.825	0.43248
9	-106.386	0.80904
10	120.483	0.82876

Iteration 2:

Obs	Model	Dependent	Variable	DF	Estimate	StdErr	tValue	Probt	LowerCL	UpperCL
1	MODEL1	UricA	Intercept	1	85.75681	16.80326	5.10	<.0001	52.78266	118.73095
2	MODEL1	UricA	DBP	1	1.68939	0.17283	9.77	<.0001	1.35023	2.02854
3	MODEL1	UricA	Chol	1	-7.16299	1.99801	-3.59	0.0004	-11.08381	-3.24216
4	MODEL1	UricA	Trig	1	92.55702	3.16049	29.29	<.0001	86.35498	98.75905
5	MODEL1	UricA	Alc	1	0.35124	0.03302	10.64	<.0001	0.28644	0.41604

Iteration 2 Residuals and Weights

Obs	residual2	wt2
1	-64.502	0.93120
2	-12.337	0.99656
3	-86.703	0.89165
4	-93.509	0.87527
5	140.686	0.74445
6	-38.672	0.97596
7	368.908	0.00418
8	236.283	0.38222
9	-102.258	0.84439
10	124.885	0.80280

Iteration 3:

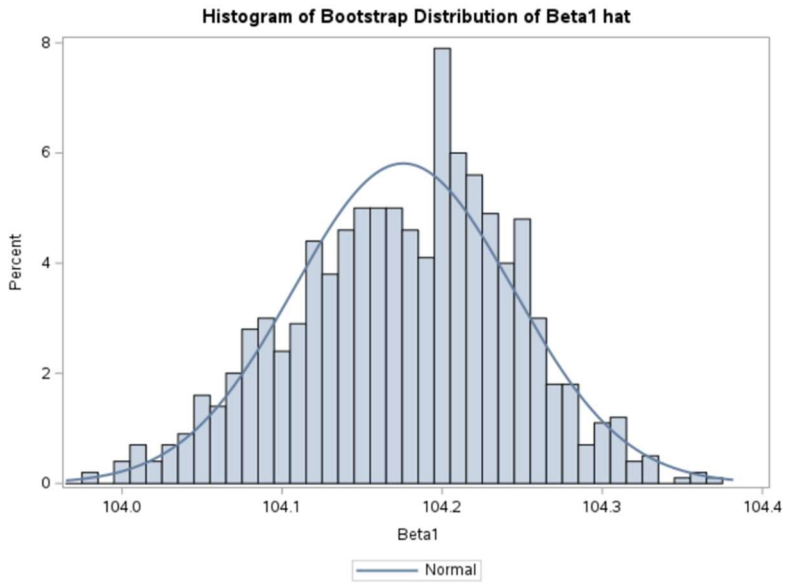
Obs	Model	Dependent	Variable	DF	Estimate	StdErr	tValue	Probt	LowerCL	UpperCL
1	MODEL1	UricA	Intercept	1	85.85647	16.77745	5.12	<.0001	52.93298	118.77996
2	MODEL1	UricA	DBP	1	1.69944	0.17255	9.85	<.0001	1.36083	2.03806
3	MODEL1	UricA	Chol	1	-7.11892	1.99520	-3.57	0.0004	-11.03424	-3.20361
4	MODEL1	UricA	Trig	1	91.40628	3.16122	28.91	<.0001	85.20280	97.60976
5	MODEL1	UricA	Alc	1	0.34320	0.03275	10.48	<.0001	0.27893	0.40746

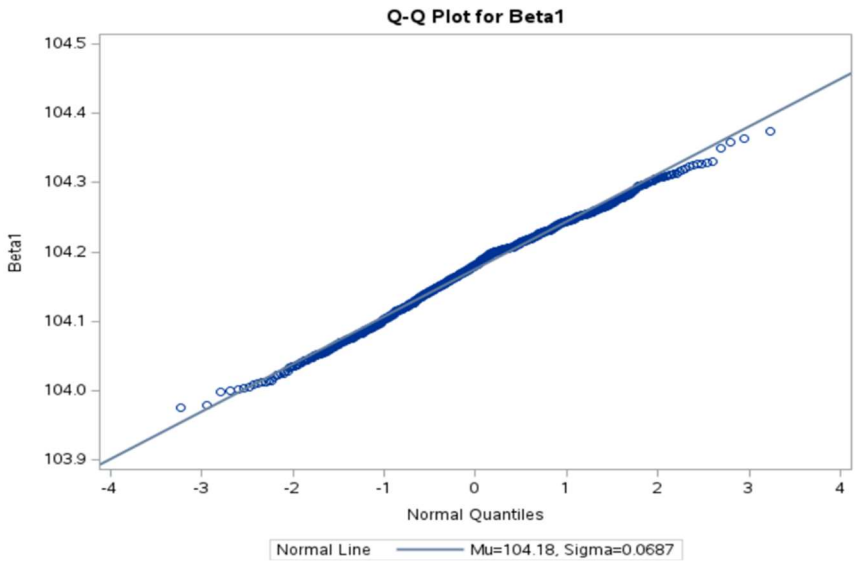
Iteration 3 Residuals and Weights

Obs	final_residual	wt3
1	-61.865	0.94087
2	-11.022	0.99781
3	-86.033	0.89447
4	-92.570	0.87782
5	141.758	0.73481
6	-37.716	0.97854
7	372.180	0.00033
8	238.482	0.35668
9	-100.432	0.85481
10	127.089	0.78763

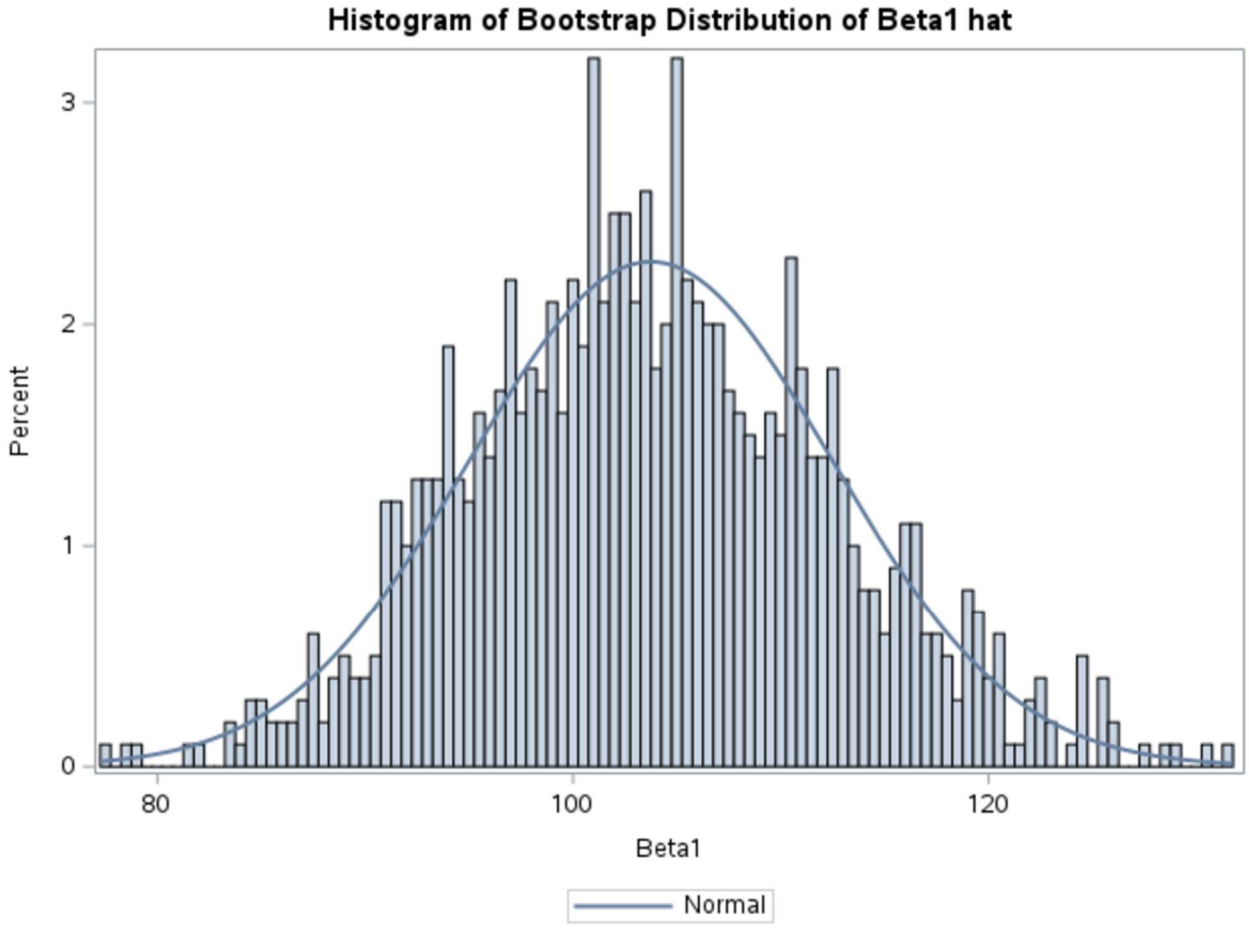
5:

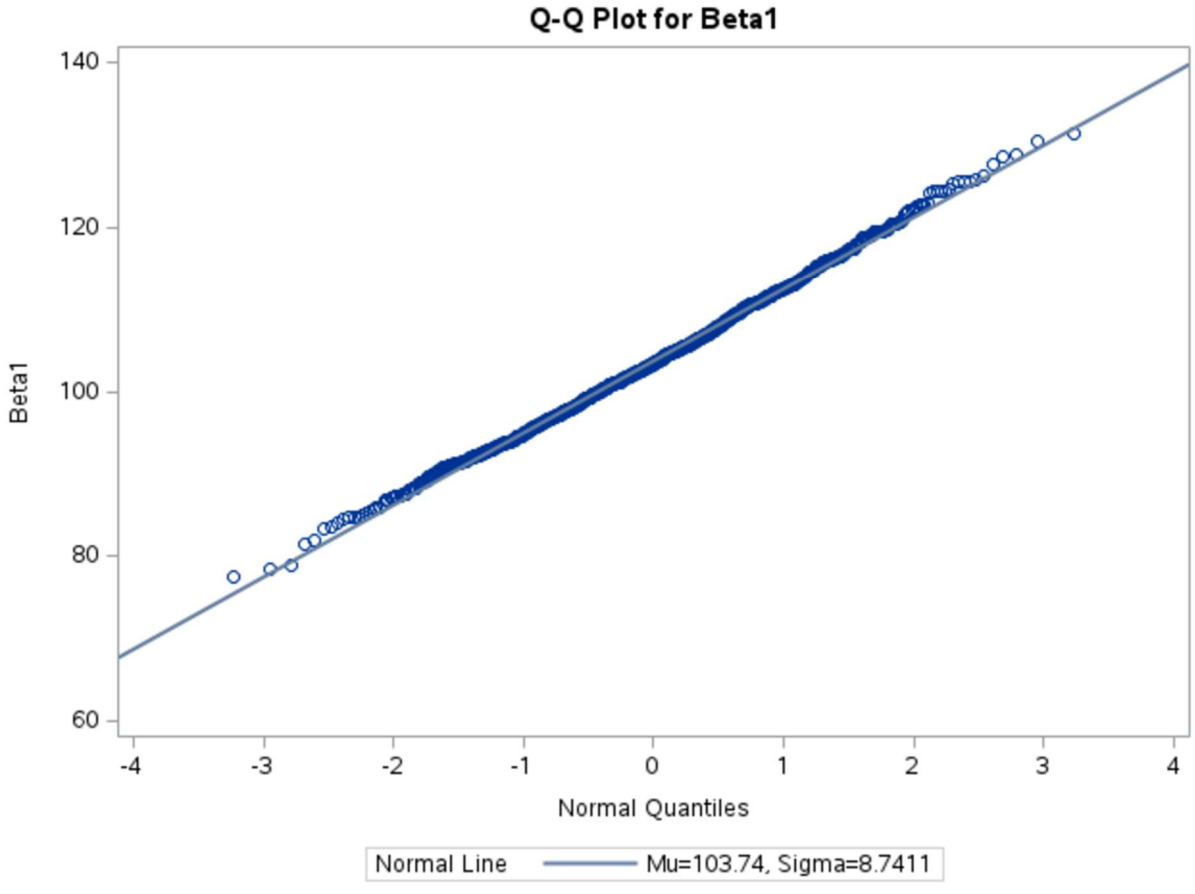
Resampling residuals (fixed X):





Resampling (X, Y):





6:

Univariate logistic regression model fits

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	464.279
SC	888.772	473.329
-2 Log L	882.247	460.279

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	421.9681	1	<.0001
Score	346.9107	1	<.0001
Wald	158.2321	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.0637	0.3742	183.1372	<.0001
clump_thickness	1	-0.9191	0.0731	158.2321	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	0.399	0.346	0.460

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	258.759
SC	888.772	267.809
-2 Log L	882.247	254.759

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	627.4888	1	<.0001
Score	458.8699	1	<.0001
Wald	143.3401	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.1743	0.3879	177.9157	<.0001
size_uniformity	1	-1.5980	0.1335	143.3401	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
size_uniformity	0.202	0.156	0.263

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	271.584
SC	888.772	280.634
-2 Log L	882.247	267.584

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	614.6638	1	<.0001
Score	460.0919	1	<.0001
Wald	149.3395	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.1642	0.3865	178.5237	<.0001
shape_uniformity	1	-1.4726	0.1205	149.3395	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
shape_uniformity	0.229	0.181	0.290

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	467.078
SC	888.772	476.128
-2 Log L	882.247	463.078

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	419.1697	1	<.0001
Score	339.9705	1	<.0001
Wald	137.8512	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.2732	0.2218	217.8694	<.0001
marginal_adhesion	1	-1.0439	0.0889	137.8512	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
marginal_adhesion	0.352	0.296	0.419

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	456.146
SC	888.772	465.196
-2 Log L	882.247	452.146

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	430.1016	1	<.0001
Score	325.8980	1	<.0001
Wald	147.0733	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.0321	0.3497	207.1146	<.0001
epithelial_size	1	-1.4604	0.1204	147.0733	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
epithelial_size	0.232	0.183	0.294

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	343.865
SC	888.772	352.915
-2 Log L	882.247	339.865

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	542.3829	1	<.0001
Score	462.0677	1	<.0001
Wald	146.6061	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	3.5188	0.2316	230.7504	<.0001
bare_nucleoli	1	-0.8554	0.0706	146.6061	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
bare_nucleoli	0.425	0.370	0.488

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	392.216
SC	888.772	401.266
-2 Log L	882.247	388.216

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	494.0310	1	<.0001
Score	391.7474	1	<.0001
Wald	135.4307	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.2797	0.3918	181.5462	<.0001
bland_chromatin	1	-1.3652	0.1173	135.4307	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
bland_chromatin	0.255	0.203	0.321

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	467.515
SC	888.772	476.565
-2 Log L	882.247	463.515

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	418.7323	1	<.0001
Score	352.5565	1	<.0001
Wald	125.0152	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.9770	0.1978	226.4033	<.0001
normal_nucleoli	1	-0.9267	0.0829	125.0152	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
normal_nucleoli	0.396	0.337	0.466

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	718.739
SC	888.772	727.789
-2 Log L	882.247	714.739

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	167.5084	1	<.0001
Score	123.0477	1	<.0001
Wald	56.1508	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.4479	0.2287	114.5611	<.0001
mitoses	1	-1.3425	0.1792	56.1508	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mitoses	0.261	0.184	0.371

Multiple logistic regression model fit

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	123.062
SC	888.772	168.312
-2 Log L	882.247	103.062

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	779.1853	9	<.0001
Score	574.9216	9	<.0001
Wald	97.6466	9	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	10.0670	1.1660	74.5437	<.0001
clump_thickness	1	-0.5264	0.1393	14.2804	0.0002
size_uniformity	1	-0.00015	0.2086	0.0000	0.9994
shape_uniformity	1	-0.3336	0.2288	2.1260	0.1448
marginal_adhesion	1	-0.3293	0.1235	7.1105	0.0077
epithelial_size	1	-0.0928	0.1562	0.3525	0.5527
bare_nucleoli	1	-0.3818	0.0939	16.5455	<.0001
bland_chromatin	1	-0.4421	0.1724	6.5773	0.0103
normal_nucleoli	1	-0.2113	0.1128	3.5070	0.0611
mitoses	1	-0.5341	0.3274	2.6621	0.1028

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	0.591	0.450	0.776
size_uniformity	1.000	0.664	1.505
shape_uniformity	0.716	0.457	1.122
marginal_adhesion	0.719	0.565	0.916
epithelial_size	0.911	0.671	1.238
bare_nucleoli	0.683	0.568	0.820
bland_chromatin	0.643	0.458	0.901
normal_nucleoli	0.810	0.649	1.010
mitoses	0.586	0.309	1.113

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	99.6	Somers' D	0.993
Percent Discordant	0.4	Gamma	0.993
Percent Tied	0.0	Tau-a	0.452
Pairs	105672	c	0.996

Removing one variable at a time:

(1) Size uniformity:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	121.062
SC	888.772	161.787
-2 Log L	882.247	103.062

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	779.1853	8	<.0001
Score	573.0414	8	<.0001
Wald	97.6466	8	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	10.0671	1.1487	76.8097	<.0001
clump_thickness	1	-0.5264	0.1387	14.3991	0.0001
shape_uniformity	1	-0.3337	0.1687	3.9148	0.0479
marginal_adhesion	1	-0.3293	0.1212	7.3832	0.0066
epithelial_size	1	-0.0928	0.1554	0.3565	0.5504
bare_nucleoli	1	-0.3818	0.0939	16.5526	<.0001
bland_chromatin	1	-0.4421	0.1709	6.6887	0.0097
normal_nucleoli	1	-0.2113	0.1117	3.5782	0.0585
mitoses	1	-0.5341	0.3263	2.6792	0.1017

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	0.591	0.450	0.775
shape_uniformity	0.716	0.515	0.997
marginal_adhesion	0.719	0.567	0.912
epithelial_size	0.911	0.672	1.236
bare_nucleoli	0.683	0.568	0.820
bland_chromatin	0.643	0.460	0.898
normal_nucleoli	0.810	0.650	1.008
mitoses	0.586	0.309	1.111

(2) Epithelial size:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	119.415
SC	888.772	155.616
-2 Log L	882.247	103.415

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	778.8320	7	<.0001
Score	571.7070	7	<.0001
Wald	97.2616	7	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	9.9623	1.1229	78.7122	<.0001
clump_thickness	1	-0.5270	0.1386	14.4597	0.0001
shape_uniformity	1	-0.3584	0.1668	4.6174	0.0316
marginal_adhesion	1	-0.3416	0.1192	8.2134	0.0042
bare_nucleoli	1	-0.3870	0.0935	17.1198	<.0001
bland_chromatin	1	-0.4581	0.1689	7.3556	0.0067
normal_nucleoli	1	-0.2245	0.1108	4.1046	0.0428
mitoses	1	-0.5315	0.3236	2.6982	0.1005

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	0.590	0.450	0.775
shape_uniformity	0.699	0.504	0.969
marginal_adhesion	0.711	0.563	0.898
bare_nucleoli	0.679	0.565	0.816
bland_chromatin	0.632	0.454	0.881
normal_nucleoli	0.799	0.643	0.993
mitoses	0.588	0.312	1.108

(3) Shape uniformity:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	122.866
SC	888.772	154.541
-2 Log L	882.247	108.866

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	773.3813	6	<.0001
Score	563.6052	6	<.0001
Wald	95.4599	6	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	10.3483	1.1294	83.9578	<.0001
clump_thickness	1	-0.6392	0.1298	24.2657	<.0001
marginal_adhesion	1	-0.4069	0.1196	11.5704	0.0007
bare_nucleoli	1	-0.4546	0.0878	26.7832	<.0001
bland_chromatin	1	-0.5513	0.1520	13.1559	0.0003
normal_nucleoli	1	-0.3176	0.1024	9.6126	0.0019
mitoses	1	-0.5699	0.3095	3.3903	0.0656

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	0.528	0.409	0.681
marginal_adhesion	0.666	0.527	0.842
bare_nucleoli	0.635	0.534	0.754
bland_chromatin	0.576	0.428	0.776
normal_nucleoli	0.728	0.595	0.890
mitoses	0.566	0.308	1.037

(4) Mitoses:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	884.247	121.310
SC	888.772	152.985
-2 Log L	882.247	107.310

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	774.9373	6	<.0001
Score	571.5418	6	<.0001
Wald	98.7664	6	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	9.7497	1.0829	81.0643	<.0001
clump_thickness	1	-0.6160	0.1352	20.7485	<.0001
shape_uniformity	1	-0.3626	0.1605	5.1079	0.0238
marginal_adhesion	1	-0.3369	0.1156	8.4931	0.0036
bare_nucleoli	1	-0.3770	0.0938	16.1645	<.0001
bland_chromatin	1	-0.4684	0.1669	7.8794	0.0050
normal_nucleoli	1	-0.2414	0.1085	4.9506	0.0261

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	0.540	0.414	0.704
shape_uniformity	0.696	0.508	0.953
marginal_adhesion	0.714	0.569	0.896
bare_nucleoli	0.686	0.571	0.824
bland_chromatin	0.626	0.451	0.868
normal_nucleoli	0.786	0.635	0.972

Model obtained using stepwise selection:

Summary of Stepwise Selection								
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	bare_nucleoli		1	1	462.0677		<.0001	bare_nucleoli
2	shape_uniformity		1	2	179.7679		<.0001	shape_uniformity
3	clump_thickness		1	3	30.0308		<.0001	clump_thickness
4	bland_chromatin		1	4	16.5639		<.0001	bland_chromatin
5	marginal_adhesion		1	5	10.1935		0.0014	marginal_adhesion
6	normal_nucleoli		1	6	5.2259		0.0223	normal_nucleoli

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.7497	1.0829	81.0643	<.0001
clump_thickness	1	0.6160	0.1352	20.7485	<.0001
shape_uniformity	1	0.3626	0.1605	5.1079	0.0238
marginal_adhesion	1	0.3369	0.1156	8.4931	0.0036
bare_nucleoli	1	0.3770	0.0938	16.1645	<.0001
bland_chromatin	1	0.4684	0.1669	7.8794	0.0050
normal_nucleoli	1	0.2414	0.1085	4.9506	0.0261

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
clump_thickness	1.852	1.420	2.413
shape_uniformity	1.437	1.049	1.968
marginal_adhesion	1.401	1.117	1.757
bare_nucleoli	1.458	1.213	1.752
bland_chromatin	1.597	1.152	2.215
normal_nucleoli	1.273	1.029	1.575

```
* Create a pointer named HD to the data file;
filename CA "/home/u63986019/sasuser.v94/Cardio.csv";

DATA c; /* Assign name c to data */
INFILE CA DSD FIRSTOBS = 2; /* Since the data is a CSV, use DSD FIRSTOBS = 2*/
INPUT UricA DBP HDL Chol Trig Alc; /*Input names of columns*/
RUN;

/* 1 */

*Full model with all predictors;
PROC REG DATA = c;
MODEL UricA = DBP HDL Chol Trig Alc;
OUTPUT OUT=D RSTUDENT=R PREDICTED=P;
RUN;

/* Model selection */
PROC REG DATA = c;
MODEL UricA = DBP HDL Chol Trig Alc / selection = adjrsq cp bic; /* selection is based on Adj R^2 */
RUN;

*-------------------------------------------------------------------------------------------------------;

/* 2 */

*Fit the best model and test for linearity using LOF test;
PROC REG DATA = c;
MODEL UricA = DBP Chol Trig Alc / lackfit;
OUTPUT OUT=D RSTUDENT=R PREDICTED=P;
RUN;

PROC SGSCATTER DATA = c;
TITLE "Multi scatter plot for Uric Acid levels vs other predictors (Best model)";
PLOT (UricA)*(DBP Chol Trig Alc);
RUN;

*From the scatter plot we can observe that there may be very few outliers for each predictor and there is
clearly a non - linear pattern in the data. Same with the correlation plots below ;

/* Residual plot*/
PROC SGSCATTER Data = D;
TITLE "Residual Plot of uric acid levels and predictors ";
plot R*(P UricA DBP Chol Trig Alc);
RUN;

*Alcohol intake probably has non const var;

/* Save the absolute value of residuals */
DATA D;
SET D;
absRes = abs(R);
RUN;

/* absolute residuals vs fitted values to check homogeneity assumption */
PROC SGPLOT DATA = D;
SCATTER x = P Y = absRes;
RUN;

*No noticable patterns in Residual plot but indicates presence of outliers;

/* Breusch Pagan Test*/
PROC MODEL DATA = D;
PARMS b0 b1 b2 b3 b4;
UricA = b0 + b1*DBP + b2*Chol + b3*Trig + b4*Alc;
/*Breusch-Pagan test for heteroscedasticity (BREUSCH) wrt the specified predictors*/
fit UricA /WHITE BREUSCH=(DBP Chol Trig Alc);
*Breusch-Pagan test for heteroscedasticity individually for each of the predictors;
fit UricA /BREUSCH = (DBP);
fit UricA /BREUSCH = (Chol);
fit UricA /BREUSCH = (Trig);
fit UricA /BREUSCH = (Alc);
RUN;
```

```
*P value is very high for DBP and chol so dont RH0 => there is constant variance. Whereas for Trig
and Alc, p-val is << 0.05 so Reject H0 and conclude that they have non-constant variance;

/* Brown Forsythe Test for homogeneity of variance*/
/* Get the medians of the predictors*/
PROC UNIVARIATE DATA = D NOPRINT;
VAR DBP Chol Trig Alc;
OUTPUT OUT = Medians Median = MedDBP Median = MedChol Median = MedTrig Median = MedAlc N=N;
RUN;

DATA Medians;
SET medians;
DO i = 1 TO N;
OUTPUT;
END;
RUN;

/*Test for homogeneity of residuals grouped by Diastolic BP*/
DATA DBPBF;
MERGE D Medians;
Group = (DBP > MedDBP);
RUN;

PROC GLM Data = DBPBF;
class Group;
model R = Group;
means Group / hovtest = BF;
run;

/*Test for homogeneity of residuals grouped by total cholestrol*/
DATA CholBF;
MERGE D Medians;
Group = (Chol > MedChol);
RUN;

PROC GLM Data = CholBF;
class Group;
model R = Group;
means Group / hovtest = BF;
run;

/*Test for homogeneity of residuals grouped by triglyceride levels*/
DATA TrigBF;
MERGE D Medians;
Group = (Trig > MedTrig);
RUN;

PROC GLM Data = TrigBF;
class Group;
model R = Group;
means Group / hovtest = BF;
run;

/*Test for homogeneity of residuals grouped by alcohol intake*/
DATA AlcBF;
MERGE D Medians;
Group = (Alc > MedAlc);
RUN;

PROC GLM Data = AlcBF;
class Group;
model R = Group;
means Group / hovtest = BF;
run;

/*Residual QQ Plot*/
PROC UNIVARIATE DATA = D NORMAL PLOT; /* Check normality of the studentized residuals */
VAR R;
RUN;

*All points along the normal line implies normality mostly;
PROC TRANSREG DATA=c; /* Find Box-Cox transformation power */
```

```
MODEL BoxCox(UricA)=identity(DBP Chol Trig Alc);
RUN;

*Determine INFLUENTIAL POINTS using hii and dffits(taken from "diagnostics.sas");
PROC REG DATA=c;
MODEL UricA = DBP Chol Trig Alc / INFLUENCE R;
ods output outputstatistics=results;
RUN;

DATA results; set results;
if HatDiagonal > 2*(4/998) then hilev=1; /* check if hii > 2*p/n */
else hilev=0;
if (abs(DFFITS) > 1) then dfflag=1;
else dfflag=0;
Fpercent = 100*probf(CooksD, 4, 994); /* calculate percentile for each Cook's D value using F(p, n-p) dist*/
if (abs(dfb_DBP) > 1) then b1flag=1; /* check if each DFBETAS value > 1 */
else b1flag=0;
if (abs(dfb_Chol) > 1) then b2flag=1;
else b2flag=0;
if (abs(dfb_Trig) > 1) then b3flag=1;
else b3flag=0;
if (abs(dfb_Alc) > 1) then b4flag=1;
else b4flag=0;
RUN;

PROC PRINT DATA = results (obs = 122);
where hilev=1 or dfflag=1 or Fpercent>20 or b1flag=1 or b2flag=1 or b3flag=1 or b4flag=1;
var HatDiagonal hilev DFFITS dfflag CooksD Fpercent dfb_DBP b1flag dfb_Chol b2flag dfb_Trig b3flag dfb_Alc b4flag;
RUN;
*LOT OF INFLUENTIAL POINTS- 122 of them -- with all hilev = 1;

*Determine OUTLIERS (taken from "diagnostics.sas");
DATA results; set results; /* Test for outliers using Bonferroni method */
tvalue = tinv(0.999974949, 992); *alpha/2n = 0.05/2*998 = 0.00002505 and n-p-1 = 998 - 5 - 1;
if (abs(RStudent)) > tvalue then outlier=1;
else outlier=0;
RUN;

PROC PRINT data=results;
where outlier=1;
var RStudent;
RUN;

*outliers are obs: 267, 477, 483;

*Determine if there is COLLINEARITY (taken from "Multicollinearity.SAS");
PROC MEANS data = c;
VAR DBP Chol Trig Alc;
RUN;

/* Check correlation coefficients*/
PROC CORR DATA = c plots = matrix;
VAR UricA DBP Chol Trig Alc;
RUN;

PROC REG DATA = c;
MODEL UricA = DBP Chol Trig Alc / collin tol vif; /* Collinearity diagnostics */
RUN;

*NO MULTICOLLINEARITY IN uncentered VAR;

*-----------------------------------------------------------------------------------------------------;

/* 3: */

PROC GPLOT Data = c;
symbol color=red value = circle;
plot (UricA)*(DBP Chol Trig Alc);
RUN;

PROC REG Data = c;
MODEL UricA = DBP Chol Trig Alc/R clb;
output out=results r =residual;
```

```
RUN;

DATA Step2;
SET results;
absresid = abs(residual);
RUN;

PROC GPLOT Data = Step2;
symbol color=green value = diamond;
plot (absresid)*(DBP Chol Trig Alc);
RUN;

PROC REG Data = Step2;
MODEL absresid = DBP Chol Trig Alc/p; /* option p requests fitted values */
output out = Step3 p =yhat;
RUN;

DATA STEP3;
SET Step3;
wt = 1/(yhat**2);
RUN;

PROC REG Data=Step3; /* weighted least squares regression */
MODEL UricA = DBP Chol Trig Alc/R clb;
WEIGHT wt;
output out=iteration2 r =residual2;
RUN;

/* Reiterate the process - Iteratively reweighted least squares */

DATA iteration2;
SET iteration2;
absresid2 = abs(residual2);
RUN;

PROC REG Data = iteration2;
MODEL absresid2 = DBP Chol Trig Alc/p; /* option p requests fitted values */
output out = results2 p =yhat2;
RUN;

DATA results2;
SET results2;
wt2 = 1/(yhat2**2);
RUN;

PROC REG Data=results2; /* weighted least squares regression */
MODEL UricA = DBP Chol Trig Alc/R clb;
WEIGHT wt2;
output out=iteration3 r =residual3;
RUN;

/* The same process can be iterated using saved residuals in iteration3, if needed */

DATA iteration3;
SET iteration3;
absresid3 = abs(residual3);
RUN;

PROC REG Data = iteration3;
MODEL absresid3 = DBP Chol Trig Alc/p; /* option p requests fitted values */
output out = results3 p =yhat3;
RUN;

DATA results3;
SET results3;
wt3 = 1/(yhat3**2);
RUN;

PROC REG Data=results3; /* weighted least squares regression */
MODEL UricA = DBP Chol Trig Alc/R clb;
WEIGHT wt3;
output out=iteration4 r =residual4;
RUN;
```

```
/* Repeat one more time to check convergence*/
DATA iteration4;
SET iteration4;
absresid4 = abs(residual4);
RUN;

PROC REG Data = iteration4;
MODEL absresid4 = DBP Chol Trig Alc/p; /* option p requests fitted values */
output out = results4 p =yhat4;
RUN;

DATA results4;
SET results4;
wt4 = 1/(yhat4**2);
RUN;

PROC REG Data=results4; /* weighted least squares regression */
MODEL UricA = DBP Chol Trig Alc/R clb;
WEIGHT wt4;
output out=iteration5 r =residual5;
RUN;

/* Converged at third iteration. */

*-------------------------------------------------------------------------------------------------------;

/* 4: */

/* Initial OLS Regression: iteration 0 */
PROC REG DATA=c;
MODEL UricA = DBP Chol Trig Alc / R;
OUTPUT OUT=results RESIDUAL=residual;
ODS OUTPUT ParameterEstimates=ols_coefficients;  /* Capture regression coefficients to compare */
RUN;

/* Display OLS Coefficients */
PROC PRINT DATA=ols_coefficients;
RUN;

PROC PRINT DATA=results (obs=10);
TITLE "OLS Residuals and Weights";
VAR residual;  /* Since weights = 1 in OLS */
RUN;

/* First Iteration - Calculate MAD = (1/0.6745) * median{|ei - median(ei)|} */
PROC UNIVARIATE DATA=results NOPRINT;
VAR residual;
OUTPUT OUT=med_stats MEDIAN=med_resid;
RUN;

DATA results_mad;
SET results;
IF _N_ = 1 THEN SET med_stats;
abs_dev = ABS(residual - med_resid);
RUN;

PROC UNIVARIATE DATA=results_mad NOPRINT;
VAR abs_dev;
OUTPUT OUT=mad_stats MEDIAN=median_abs_dev;
RUN;

DATA Step1;
SET results_mad;
IF _N_ = 1 THEN SET mad_stats;
mad = (1 / 0.6745) * median_abs_dev;
weight = ABS(residual / mad);
RUN;

/* Bisquare Weight Function */
DATA Step2;
SET Step1;
c = 4.685;
```

```
IF weight <= c THEN wt = (1 - (weight/c)**2)**2;
ELSE wt = 0;
RUN;

/* First Iteration of Weighted Regression */
PROC REG DATA=Step2;
MODEL UricA = DBP Chol Trig Alc / R CLB;
WEIGHT wt;
OUTPUT OUT=Iteration1 RESIDUAL=residual1;
ODS OUTPUT ParameterEstimates=coeffs1;  /* Capture coefficients */
RUN;

/* Output Comparison after First Iteration */
PROC PRINT DATA=coeffs1;
RUN;

PROC PRINT DATA=Iteration1 (obs=10);
TITLE "Iteration 1 Residuals and Weights";
VAR residual1 wt;
RUN;

/* Repeat for Second Iteration */
PROC UNIVARIATE DATA=Iteration1 NOPRINT;
VAR residual1;
OUTPUT OUT=med_stats2 MEDIAN=med_resid2;
RUN;

DATA results_mad2;
SET Iteration1;
IF _N_ = 1 THEN SET med_stats2;
abs_dev2 = ABS(residual1 - med_resid2);
RUN;

PROC UNIVARIATE DATA=results_mad2 NOPRINT;
VAR abs_dev2;
OUTPUT OUT=mad_stats2 MEDIAN=median_abs_dev2;
RUN;

DATA Step3;
SET results_mad2;
IF _N_ = 1 THEN SET mad_stats2;
mad2 = (1 / 0.6745) * median_abs_dev2;
weight2 = ABS(residual1 / mad2);
RUN;

/* Bisquare Weight Function for Second Iteration */
DATA Step4;
SET Step3;
c = 4.685;
IF weight2 <= c THEN wt2 = (1 - (weight2/c)**2)**2;
ELSE wt2 = 0;
RUN;

/* Second Iteration of Weighted Regression */
PROC REG DATA=Step4;
MODEL UricA = DBP Chol Trig Alc / R CLB;
WEIGHT wt2;
OUTPUT OUT=Iteration2 RESIDUAL=residual2;
ODS OUTPUT ParameterEstimates=coeffs2;  /* Capture coefficients */
RUN;

/* Output Comparison after Second Iteration */
PROC PRINT DATA=coeffs2;
RUN;

PROC PRINT DATA=Iteration2 (obs=10);
TITLE "Iteration 2 Residuals and Weights";
VAR residual2 wt2;
RUN;

/* Repeat for Third Iteration */
PROC UNIVARIATE DATA=Iteration2 NOPRINT;
VAR residual2;
```

```
OUTPUT OUT=med_stats3 MEDIAN=med_resid3;
RUN;

DATA results_mad3;
SET Iteration2;
IF _N_ = 1 THEN SET med_stats3;
abs_dev3 = ABS(residual2 - med_resid3);
RUN;

PROC UNIVARIATE DATA=results_mad3 NOPRINT;
VAR abs_dev3;
OUTPUT OUT=mad_stats3 MEDIAN=median_abs_dev3;
RUN;

DATA Step5;
SET results_mad3;
IF _N_ = 1 THEN SET mad_stats3;
mad3 = (1 / 0.6745) * median_abs_dev3;
weight3 = ABS(residual2 / mad3);
RUN;

/* Bisquare Weight Function for Third Iteration */
DATA Step6;
SET Step5;
c = 4.685;
IF weight3 <= c THEN wt3 = (1 - (weight3/c)**2)**2;
ELSE wt3 = 0;
RUN;

/* Third Iteration (final iteration) of Weighted Regression */
PROC REG DATA=Step6;
MODEL UricA = DBP Chol Trig Alc / R CLB;
WEIGHT wt3;
OUTPUT OUT=FinalIteration RESIDUAL=final_residual;
ODS OUTPUT ParameterEstimates=final_coeffs;  /* Capture coefficients */
RUN;

/* Output Comparison after Third Iteration */
PROC PRINT DATA=final_coeffs;
RUN;

PROC PRINT DATA=FinalIteration (obs=10);
TITLE "Iteration 3 Residuals and Weights";
VAR final_residual wt3;
RUN;
```

```
/* ----------------------------------------5 (method 1)-------------------------------------------------------------------------*/

* Create a pointer named HD to the data file;
filename CA "/home/u63986019/sasuser.v94/Cardio.csv";

DATA c; /* Assign name c to data */
INFILE CA DSD FIRSTOBS=2; /* Since the data is a CSV, use DSD FIRSTOBS=2 */
INPUT UricA DBP HDL Chol Trig Alc; /* Input names of columns */
RUN;

* Fit the regression model and save residuals;
PROC REG DATA=c OUTEST=OrigEst;
MODEL UricA = Trig / CLB;
OUTPUT OUT=D RSTUDENT=R PREDICTED=P;
RUN;

* Bootstrap: Resample residuals with fixed X (Trig);
PROC SURVEYSELECT DATA=D OUT=BootRes
METHOD=URS /* Unrestricted random sampling */
SAMPRATE=1
REPS= 1000/* Number of bootstrap samples */
SEED=123; /* Random seed for reproducibility */
RUN;

* Generate new bootstrap datasets and fit regression models;
DATA Bootstrap;
SET BootRes;
BY Replicate; /* Group by bootstrap replicate */
YBoot = P + R; /* Recompute response using predicted values and resampled residuals */
RUN;

PROC REG DATA=Bootstrap OUTEST=BootEst NOPRINT;
BY Replicate; /* Fit separate regression for each replicate */
MODEL YBoot = Trig;
RUN;

* Extract beta_1 estimates for all bootstrap samples;
DATA Beta1;
SET BootEst;
WHERE _TYPE_ = "PARMS"; /* Filter rows containing parameter estimates */
KEEP Replicate Trig;
RENAME Trig=Beta1; /* Trig contains the coefficient value */
RUN;

* Plot the histogram and Q-Q plot of the bootstrap distribution;
PROC SGPLOT DATA=Beta1;
HISTOGRAM Beta1 / BINWIDTH=0.01; /* Adjust BINWIDTH if necessary */
DENSITY Beta1 / TYPE=NORMAL; /* Overlay normal density for comparison*/
TITLE "Histogram of Bootstrap Distribution of Beta1 hat";
RUN;

PROC UNIVARIATE DATA=Beta1 NOPRINT;
VAR Beta1;
QQPLOT Beta1 / NORMAL(MU=EST SIGMA=EST);
TITLE "Q-Q Plot of Bootstrap Distribution of Beta1 hat";
RUN;

/* Save the original Beta_1 = 104.173 */
DATA OrigBeta;
SET OrigEst;
WHERE _TYPE_ = "PARMS"; /* Only keep parameter estimates */
KEEP Trig; /* Keep the coefficient for Trig */
RENAME Trig = OriginalBeta;
RUN;

/*Compute Bootstrap Mean and SE */
PROC MEANS DATA=Beta1 NOPRINT;
VAR Beta1;
OUTPUT OUT=SummaryStats MEAN=BootMean STD=BootSE;
RUN;

/*Combine Original Beta with Bootstrap Summary */
DATA BiasAndSE;
MERGE SummaryStats OrigBeta; /* Merge datasets */
Bias = BootMean - OriginalBeta; /* Calculate Bias */
KEEP BootMean BootSE Bias OriginalBeta;
RUN;

/*Display Results */
PROC PRINT DATA=BiasAndSE LABEL;
```

```
LABEL BootMean = "Boot Mean of Beta_1"
BootSE = "Boot SE of Beta_1"
Bias = "Bootstrap Bias of Beta_1"
OriginalBeta = "Original Beta_1";
TITLE "Bias and Standard Error of Bootstrap Estimates";
RUN;

/* Percentiles for Beta_1 bootstrap estimates */
PROC UNIVARIATE DATA=Beta1 NOPRINT;
VAR Beta1;
OUTPUT OUT=Beta1Percentiles PCTLPTS=2.5 97.5 PCTLPRE=Beta1_;
RUN;

/* Add Original Beta and Compute Bias */
DATA BiasData;
MERGE Beta1 OrigBeta; /* Add OriginalBeta to Beta1 dataset */
Bias = Beta1 - OriginalBeta; /* Compute Bias */
RUN;

/* Calculate Percentiles of Bias */
PROC UNIVARIATE DATA=BiasData NOPRINT;
VAR Bias;
OUTPUT OUT=BiasPercentiles PCTLPTS=2.5 97.5 PCTLPRE=Bias_;
RUN;

DATA FinalPercentiles;
MERGE Beta1Percentiles BiasPercentiles;
RUN;

PROC PRINT DATA=FinalPercentiles LABEL;
TITLE "Percentiles of Bootstrap Estimates and Bias";
RUN;

/* Calculate mean and standard error of bootstrap beta_1 estimates */
PROC MEANS DATA=Beta1 NOPRINT;
VAR Beta1;
OUTPUT OUT=BootstrapStats MEAN=MeanBeta1 STD=SEBeta1;
RUN;

/* Compute Normal Approximation CI */
DATA NormalCI;
SET BootstrapStats;
OriginalBeta = 104.173; /* Replace with the original beta_1 value */
Z = 1.96; /* Z value for 95% CI */
LNorm = OriginalBeta - Z * SEBeta1;
UNorm = OriginalBeta + Z * SEBeta1;
RUN;

PROC PRINT DATA=NormalCI LABEL;
VAR OriginalBeta LNorm UNorm;
TITLE "95% Confidence Interval Using Normal Approximation";
RUN;

/* Compute Basic Bootstrap CI */
DATA BasicCI;
SET Beta1Percentiles; /* This dataset already contains the 2.5th and 97.5th percentiles */
OriginalBeta = 104.173; /* Replace with the original beta_1 value */
LBasic = 2 * OriginalBeta - Beta1_97_5;
UBasic = 2 * OriginalBeta - Beta1_2_5;
RUN;

PROC PRINT DATA=BasicCI LABEL;
VAR OriginalBeta LBasic UBasic;
TITLE "95% Confidence Interval Using Basic Bootstrap";
RUN;

/* Compute Percentile Bootstrap CI */
DATA PercentileCI;
SET Beta1Percentiles; /* This dataset already contains the 2.5th and 97.5th percentiles */
LPer = Beta1_2_5;
UPer = Beta1_97_5;
RUN;

PROC PRINT DATA=PercentileCI LABEL;
VAR LPer UPer;
TITLE "95% Confidence Interval Using Percentile Bootstrap";
RUN;
```

```
/* ----------------------------------------5 (method 2)-------------------------------------------------------------------------*/

* Create a pointer named HD to the data file;
filename CA "/home/u63986019/sasuser.v94/Cardio.csv";

DATA c; /* Assign name c to data */
INFILE CA DSD FIRSTOBS = 2; /* Since the data is a CSV, use DSD FIRSTOBS = 2*/
INPUT UricA DBP HDL Chol Trig Alc; /*Input names of columns*/
RUN;

/* X,Y BOTH */
* Fit the initial regression model and output residuals and predicted values;
PROC REG DATA=c OUTEST=Coefficients NOPRINT;
MODEL UricA = Trig;
OUTPUT OUT=D RSTUDENT=R PREDICTED=P;
RUN;

* Set up bootstrap sampling using SURVEYSELECT;
PROC SURVEYSELECT DATA=c
OUT=Bootstrap
METHOD=URS
SEED=812
SAMPRATE=1
OUTHITS
REPS=1000;
RUN;

* Perform regression on each bootstrap sample;
DATA BootstrapResults;
SET Bootstrap;
BY Replicate;
RUN;

PROC REG DATA=Bootstrap OUTEST=BootCoefficients NOPRINT;
BY Replicate;
MODEL UricA = Trig;
RUN;

* Extract regression coefficient estimates for Trig;
DATA Beta1Estimates;
SET BootCoefficients;
KEEP Replicate Trig;
RENAME Trig=Beta1;
RUN;

* Generate histogram and Q-Q plot for the bootstrap distribution of Beta1;
PROC SGPLOT DATA=Beta1Estimates;
HISTOGRAM Beta1 / BINWIDTH=0.5; /* Adjust BINWIDTH if necessary */
DENSITY Beta1 / TYPE=NORMAL; /* Overlay normal density for comparison*/
TITLE "Histogram of Bootstrap Distribution of Beta1 hat";
RUN;

PROC UNIVARIATE DATA=Beta1Estimates;
VAR Beta1;
QQPLOT Beta1 / NORMAL(MU=EST SIGMA=EST); * Overlay a normal Q-Q plot;
TITLE "Bootstrap Distribution of Beta1 (Trig)";
RUN;

* Compute the original estimate of Beta1 from the full dataset;
PROC REG DATA=c OUTEST=OrigEstimate NOPRINT;
MODEL UricA = Trig;
RUN;

* Extract the original Beta1 estimate;
DATA OrigEstimate;
SET OrigEstimate;
KEEP Trig;
RENAME Trig=Beta1_Original;
RUN;

* Compute bias and standard error from bootstrap estimates;
PROC MEANS DATA=Beta1Estimates NOPRINT;
VAR Beta1;
OUTPUT OUT=BootstrapStats MEAN=Beta1_Mean STD=Beta1_SE;
RUN;

* Merge the original estimate with bootstrap statistics;
DATA BiasAndSE;
MERGE OrigEstimate BootstrapStats;
Bias = Beta1_Mean - Beta1_Original; /* Bias formula */
```

```
KEEP Beta1_Original Beta1_Mean Beta1_SE Bias;
RUN;

* Display results;
PROC PRINT DATA=BiasAndSE;
TITLE "Bias and Standard Error of Bootstrap Estimates for Beta1";
RUN;

* Compute the 2.5th and 97.5th percentiles of the bootstrap Beta1 estimates;
PROC UNIVARIATE DATA=Beta1Estimates NOPRINT;
VAR Beta1;
OUTPUT OUT=Percentiles PCTLPTS=2.5 97.5 PCTLPRE=Beta1_;
RUN;

* Display the results;
PROC PRINT DATA=Percentiles;
TITLE "2.5th and 97.5th Percentiles of Bootstrap Beta1 Estimates";
RUN;

* Calculate the original estimate of Beta1 from the full dataset;
PROC REG DATA=c OUTEST=OrigEstimate NOPRINT;
MODEL UricA = Trig;
RUN;

* Extract the original Beta1 estimate;
DATA OrigEstimate;
SET OrigEstimate;
KEEP Trig;
RENAME Trig=Beta1_Original;
RUN;

* Merge original estimate with bootstrap estimates;
DATA BiasEstimates;
SET Beta1Estimates;
IF _N_ = 1 THEN SET OrigEstimate;
Bias = Beta1 - Beta1_Original; /* Compute bias for each bootstrap sample */
RUN;

* Compute the 2.5th and 97.5th percentiles of the bootstrap bias distribution;
PROC UNIVARIATE DATA=BiasEstimates NOPRINT;
VAR Bias;
OUTPUT OUT=BiasPercentiles PCTLPTS=2.5 97.5 PCTLPRE=Bias_;
RUN;

* Display the results;
PROC PRINT DATA=BiasPercentiles;
TITLE "2.5th and 97.5th Percentiles of Bootstrap Bias Estimates";
RUN;

* Calculate the original estimate of Beta1 from the full dataset;
PROC REG DATA=c OUTEST=OrigEstimate NOPRINT;
MODEL UricA = Trig;
RUN;

* Extract the original Beta1 estimate;
DATA OrigEstimate;
SET OrigEstimate;
KEEP Trig;
RENAME Trig=Beta1_Original;
RUN;

*Calculate mean and standard error of bootstrap estimates;
PROC MEANS DATA=Beta1Estimates NOPRINT;
VAR Beta1;
OUTPUT OUT=BootstrapStats MEAN=Beta1_Mean STD=Beta1_SE;
RUN;

*Calculate the percentiles for the percentile bootstrap method;
PROC UNIVARIATE DATA=Beta1Estimates NOPRINT;
VAR Beta1;
OUTPUT OUT=PercentileCI PCTLPTS=2.5 97.5 PCTLPRE=Beta1_;
RUN;

* Merge datasets to compute all CIs;
DATA BootstrapCIs;
MERGE OrigEstimate BootstrapStats PercentileCI;
    * Normal Approximation CI;
    Beta1_Lower_Normal = Beta1_Mean - 1.96 * Beta1_SE;
    Beta1_Upper_Normal = Beta1_Mean + 1.96 * Beta1_SE;

    * Basic Bootstrap CI;
```

```
    Beta1_Lower_Basic = 2 * Beta1_Original - Beta1_97_5;
    Beta1_Upper_Basic = 2 * Beta1_Original - Beta1_2_5;

RUN;

*Display the CIs;
PROC PRINT DATA=BootstrapCIs;
TITLE "95% Confidence Intervals for Beta1 Bootstrap Estimates";
RUN;
```

```
/*import data from Excel file called my_data.xlsx*/
proc import out=c
datafile="/home/u63986019/sasuser.v94/breast_tumor.xlsx"
dbms=xlsx
replace;
getnames=YES;
run;

PROC Logistic Data = c descending; /* descending to model P(Y=1) instead of P(Y=0) */
MODEL class = clump_thickness;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = size_uniformity;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = shape_uniformity;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = marginal_adhesion;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = epithelial_size;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = bare_nucleoli;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = bland_chromatin;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = normal_nucleoli;
RUN;
*SIGNIFICANT;

PROC Logistic Data = c descending;
MODEL class = mitoses;
RUN;
*SIGNIFICANT;

/* ALL PREDICTORS ARE SIGNIFICANT */

*--------------------------Multiple LRM-------------------------------------------------;
PROC LOGISTIC Data = c descending;
MODEL class = clump_thickness size_uniformity shape_uniformity marginal_adhesion epithelial_size
bare_nucleoli bland_chromatin normal_nucleoli mitoses;
RUN;

/* ALL PREDICTORS ARE SIGNIFICANT JOINTLY USING LRT but drop ones with high p-value */

*size uniformity has high p value, so remove it ;
PROC LOGISTIC Data = c descending;
MODEL class = clump_thickness shape_uniformity marginal_adhesion epithelial_size bare_nucleoli
bland_chromatin normal_nucleoli mitoses;
RUN;

*epithelial size has high p value, so remove it next;
PROC LOGISTIC Data = c descending;
MODEL class = clump_thickness shape_uniformity marginal_adhesion bare_nucleoli bland_chromatin
```

```
normal_nucleoli mitoses;
RUN;

*-2 log l changed from 103.062 to 103.415 => difference = 0.353 so, removing this variable did not
substantially change the model;

*Next is shape uniformity;
PROC LOGISTIC Data = c descending;
MODEL class = clump_thickness marginal_adhesion bare_nucleoli bland_chromatin normal_nucleoli mitoses;
RUN;

*Next high p value is mitoses, remove it;
PROC LOGISTIC Data = c descending;
MODEL class = clump_thickness shape_uniformity marginal_adhesion bare_nucleoli bland_chromatin
normal_nucleoli;
RUN;

*-2 log l changed from 103.415 to 107.310 => difference = 3.895 > p = 0.1. so, removing this variable
did substantially change the model and we must keep it because it helps model pred performance;

*FINAL MODEL;
PROC LOGISTIC Data = c descending;
MODEL class = clump_thickness shape_uniformity marginal_adhesion bare_nucleoli bland_chromatin
normal_nucleoli mitoses;
RUN;

*--------------------STEPWISE SELECTION--------------------------------------;
PROC LOGISTIC DATA = c;
MODEL class = clump_thickness size_uniformity shape_uniformity marginal_adhesion epithelial_size
bare_nucleoli bland_chromatin normal_nucleoli mitoses / selection = stepwise; /* stepwise selection */
RUN;
```