

Problem 1:

Multiple linear regression model:

$$\text{Death Rate} = 12.267 + 0.0073 * \text{Doctor_Availability} + 0.00058 * \text{Hospital_Availability} - 0.3302 * \text{Capital_Income} - 0.0094 * \text{Population_Density}$$

Regression Diagnostics:

1. Test for linearity (Lack Of Fit Test): The LOF test did not give any values in the ANOVA table because of the absence of replicated predictor observations in the data. In the ANOVA table, SSLOF = SSE and values for test statistic and p value were not displayed. Hence, using this test is not ideal. Instead when we look at scatter plots for Death rate vs all of its corresponding predictors and observe that we clearly see no pattern in the plot and can conclude that there is a strong non-linear pattern in the data.
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 Hospital Availability 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
Doctor Availability	1.17	$0.2842 > \alpha = 0.05$	Fail to Reject H_0
Hospital Availability	1.52	$0.2239 > \alpha$	Fail to Reject H_0
Capital Income	2.31	$0.1345 > \alpha$	Fail to Reject H_0
Population Density	2.40	$0.1279 > \alpha$	Fail to Reject H_0

Therefore, since we fail to reject H_0 for all individual predictors, we can conclude that there is constant variance in the error terms and hence no transformation is required for any of the 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	3.86	$0.4255 > \alpha = 0.05$	Fail to Reject H_0
Doctor Availability	0.66	$0.4164 > \alpha$	Fail to Reject H_0
Hospital Availability	1.27	$0.2597 > \alpha$	Fail to Reject H_0
Capital Income	2.41	$0.1207 > \alpha$	Fail to Reject H_0
Population Density	0.00	$0.9687 > \alpha$	Fail to Reject H_0

Therefore, since we fail to reject H_0 for the test with the full model and individual predictors, we can conclude that there is constant variance. And our assumption from the residual plot for predictor hospital availability is incorrect and that the points are outliers.

3. Test for normality:

- QQ plot: The QQ plot of residuals looks like it is skewed to the right and hence we might have to transform DeathRate variable. Since it is right - skewed we want to shrink large values by transforming it to $\log(Y)$ or \sqrt{Y} .
- Shapiro-Wilks test:
 H_0 : Death Rate is normally distributed. vs H_A : Death Rate is not normally distributed.
 $W_{obs} = 0.9088$ and $p = 0.0007 < \alpha = 0.05$. Hence, we reject H_0 and conclude that response variable Death Rate is not normally distributed and requires the aforementioned transformation.

Problem 2:

Test for Normality: The only assumption that was not met above was the normality assumption; so we transform our response to $\log(\text{DeathRate})$. The QQ-Plot for this log transformation was heavily right-skewed and since this transformation does not work we use BoxCox transformation to determine the best value for transformation.

The optimal $\lambda = 2.25$ from the boxcox transformation. Hence, we can transform our response to $(\text{DeathRate})^2$ instead to make sure our analysis is easy. Also, the QQ-plots for residuals of $(\text{DeathRate})^2$ and $(\text{DeathRate})^{2.25}$ does not have much difference so we can stick with the square transformation and we will use this squared response in the following analysis.

$$\text{Therefore, } (\text{Death Rate})^2 = 156.834 + 0.1513 * \text{Doctor_Availability} + 0.00784 * \text{Hospital_Availability} - 7.4519 * \text{Capital_Income} - 0.1753 * \text{Population_Density}$$

Diagnostics:

1. Lack of fit test: Same as Problem 1 because there are no replicated observations. But using scatter plot, there is still strong visual evidence on non-linearity. Maybe a transformation on one or more of the predictor variables will solve this issue.
2. Test for homogeneity: The residual plot did not show any patterns just like in problem 1. Also, the results of the Breush-Pagan test only differed in test statistic and p-values but the conclusion was that there is constant variance amongst predictor variables. (SEE OUTPUT FOR TEST STATISTIC AND P-VALUE). Here the Brown Forsythe test was not conducted because it deals with predictor variables' medians and since we did not do any transformation on the predictor variables we make use of the same conclusions as in Problem 1.

Problem 3:

Results table					
Hypotheses	Predictor	Extra SS	F_{obs}	p-value	Conclusion
$H_0 : \beta_3 = 0$	Capital Income	$\text{SSR}(X_3 X_1, X_2) = 3880.98$	5.16	$0.0277 < \alpha$	Reject H_0
$H_0 : \beta_4 = 0$	Population density	$\text{SSR}(X_4 X_1, X_2, X_3) = 3300.65$	4.38	$0.0416 < \alpha$	Reject H_0

Using Type III SS we can include the same predictors as Type I SS which are Capital Income and Population density from the original model:

Results table					
Hypotheses	Predictor	Extra SS	F_{obs}	p-value	Conclusion
$H_0 : \beta_3 = 0$	Capital Income	$\text{SSR}(X_3 X_1, X_2, X_4) = 2588.16$	3.44	$0.0699 < \alpha$	Reject H_0
$H_0 : \beta_4 = 0$	Population density	$\text{SSR}(X_4 X_1, X_2, X_3) = 3300.65$	4.38	$0.0416 < \alpha$	Reject H_0

When we dropped the variable Hospital Availability and Doctor Availability, we observed that Capital Income was insignificant and we drop it for that reason. Then, the tentative model is:

$$(\text{Death Rate})^2 = 110.2534 - 0.1893 * \text{Population Density}.$$

Results table Type I and Type III					
Hypotheses	Predictor	Extra SS	F_{obs}	p-value	Conclusion
$H_0 : \beta_4 = 0$	Population Density	$\text{SSR}(X_4 X_3) = 4148.489$	5.31	$0.0253 < \alpha$	Reject H_0

This result can also be concluded from the Partial F test conducted with assuming

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \text{ vs } H_A : \text{At least one inequality.}$$

$F_{obs} = 1.64$ and $p = 0.1921 > 0.05$. So, we fail to Reject H_0 and can conclude that the predictors dropped (Dr Availability, Hosp Availability and Capital Income) do not significantly improve the model's ability to predict the dependent variable compared to the model without these predictors (tentative model).

Problem 4:

Best Models using:

- Adjusted $R^2 = 0.1210$:
 $\text{DeathRate}^2 = b_0 + b_1 * \text{DrAvail} + b_2 * \text{CapInc} + b_3 * \text{PopD}$
- Mallow's $C_p = 3.4012 \approx p = 3$:
 $\text{DeathRate}^2 = b_0 + b_1 * \text{DrAvail} + b_2 * \text{CapInc} + b_3 * \text{PopD}$
- BIC = 356.9939 is the least:
 $\text{DeathRate}^2 = b_0 + b_1 * \text{DrAvail} + b_2 * \text{CapInc} + b_3 * \text{PopD}$
- Stepwise Selection:
 $\text{DeathRate}^2 = 110.253 - 0.189 * \text{PopD}$
- Forward Selection:
 $\text{DeathRate}^2 = 160.856 + 0.176 * \text{DrAvail} - 7.840 * \text{CapInc} - 0.163 * \text{PopD}$
- Backward Selection:
 $\text{DeathRate}^2 = b_0 + b_1 * \text{HospAvail} + b_2 * \text{DrAvail} + b_3 * \text{CapInc}$

The best models using the Adj. R^2 , C_p and, the BIC method are all the exact same model. Whereas, with the automatic model selection procedure differ. The best model consists of only one predictor in the stepwise selection and in the other two methods there are three predictors but only Forward selection matches the Adj. R^2 , C_p and, the BIC criterion. Hence, the best model is the simplest model:

$$\text{DeathRate}^2 = 160.856 + 0.176 * \text{DrAvail} - 7.840 * \text{CapInc} - 0.163 * \text{PopD} \quad (1)$$

Problem 5:

For the final model:

Coefficient of Multiple determination, $R^2 = 0.1717$

Coefficient of multiple correlation, $r^2 = \sqrt{R^2} = \sqrt{0.1717} = 0.4143$

We use the square of partial correlation coefficient to obtain partial determination coefficient.

<u>Coefficients of Partial Determination</u>	<u>Coefficients of partial correlation</u>
$R_{Y1 3,4}^2 = (0.22149)^2 = 0.04905$	0.22149
$R_{Y3 1,4}^2 = (-0.27294)^2 = 0.07449$	-0.27294
$R_{Y4 1,3}^2 = (-0.27667)^2 = 0.07654$	-0.27667

Problem 6:

Largest coefficient of partial determination = $R_{Y1|3,4}^2 = 0.04905$.

Now, we fit our first model DeathRate^2 on predictors CapInc and PopD and get $R^2 = 0.1290$. Then, we fit predictor DrAvail on CapInc and PopD and get an $R^2 = 0.1936$.

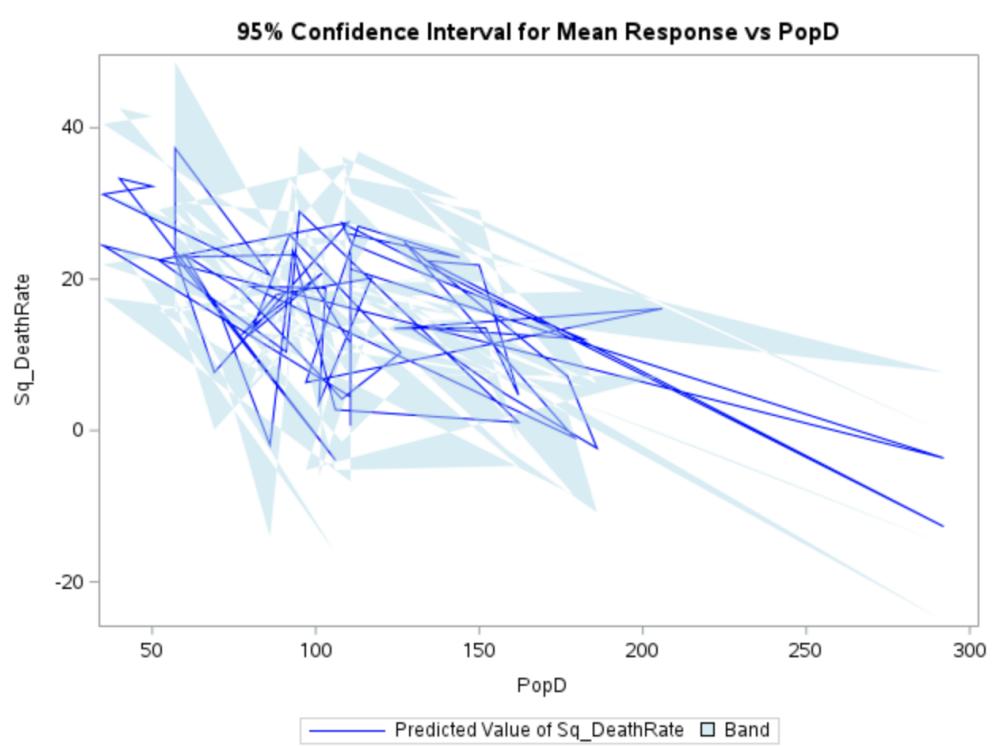
Therefore, $R_{Y1|3,4}^2 = \text{Correlation between } e(Y|\text{CapInc}, \text{PopD}) \text{ and } e(Y|\text{DrAvail}, \text{CapInc}, \text{PopD})$
 $= 0.21851^2 = 0.04774 \approx 0.04905$.

For alternative interpretation of multiple determination:

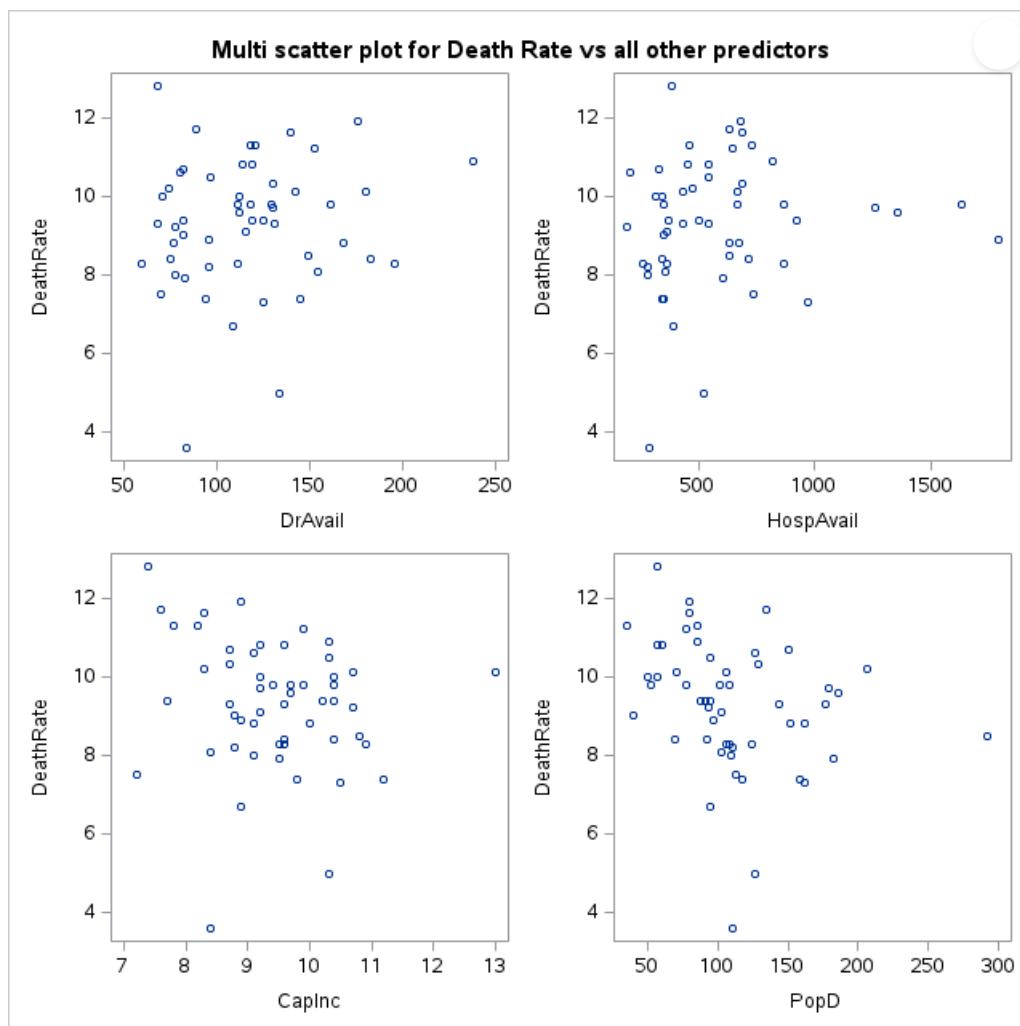
$R^2 = [\text{Corr}(Y, \hat{Y})]^2$. Here, $[\text{Corr}(Y, \hat{Y})]^2 = 0.41441^2 = 0.171735 = R^2 = 0.1717$

Problem 7:

SEE OUTPUT for confidence intervals for mean response and prediction intervals for a death rate of a new city.
(Unfortunately, simultaneous confidence bands not done because output kept giving a scraggly plot after several tries. One of them attached for reference)



Relevant SAS outputs:



Multi scatter plot for Death Rate vs all other predictors

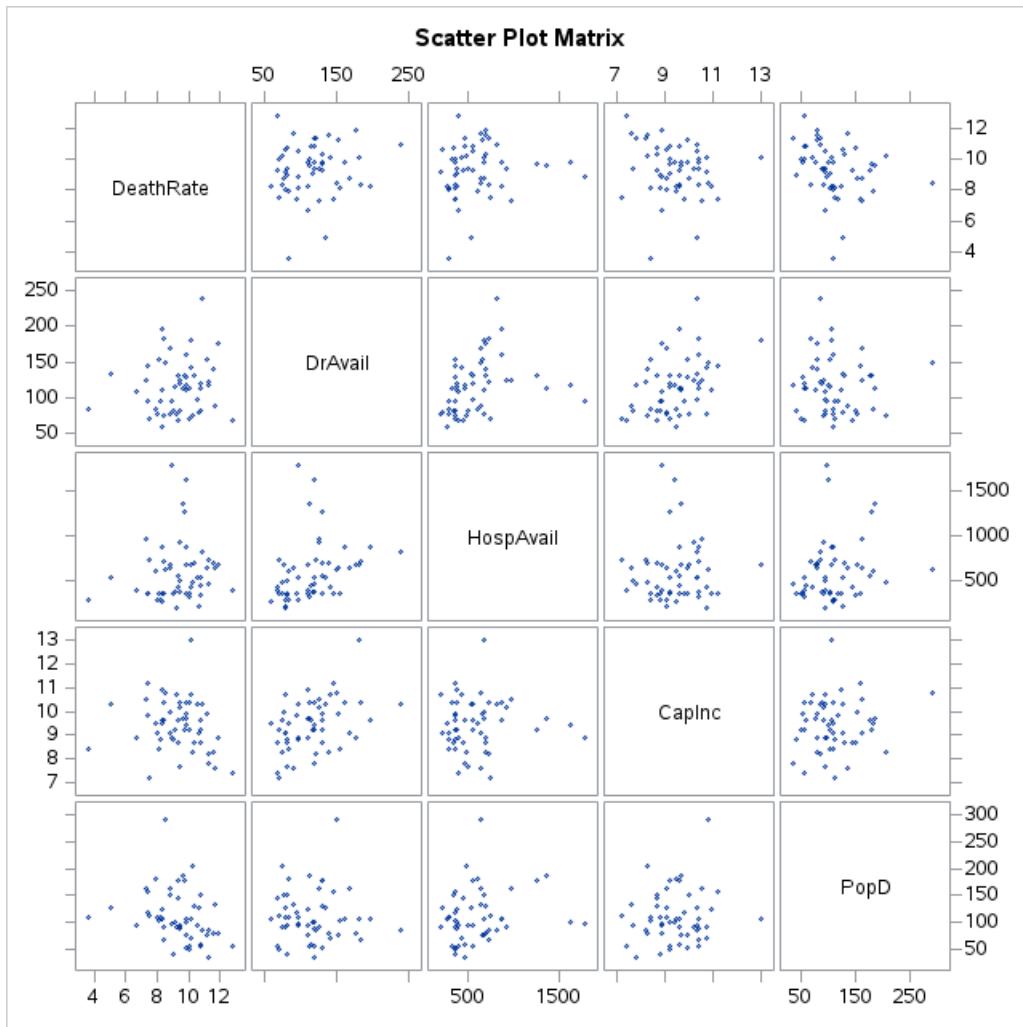
The CORR Procedure

5 Variables: DeathRate DrAvail HospAvail CapInc PopD

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
DeathRate	53	9.30566	1.66253	493.20000	3.60000	12.80000
DrAvail	53	116.09434	37.88660	6153	60.00000	238.00000
HospAvail	53	589.79245	332.61831	31259	190.00000	1792
CapInc	53	9.43585	1.07544	500.10000	7.20000	13.00000
PopD	53	110.64151	47.17973	5864	35.00000	292.00000

Pearson Correlation Coefficients, N = 53 Prob > r under H0: Rho=0					
	DeathRate	DrAvail	HospAvail	CapInc	PopD
DeathRate	1.00000	0.11577 0.4091	0.11059 0.4305	-0.17199 0.2181	-0.27761 0.0442
DrAvail	0.11577 0.4091	1.00000	0.29563 0.0316	0.43329 0.0012	-0.01994 0.8873
HospAvail	0.11059 0.4305	0.29563 0.0316	1.00000	0.02750 0.8450	0.18662 0.1809
CapInc	-0.17199 0.2181	0.43329 0.0012	0.02750 0.8450	1.00000	0.12874 0.3582
PopD	-0.27761 0.0442	-0.01994 0.8873	0.18662 0.1809	0.12874 0.3582	1.00000

1) Correlation Plot b/w Predictors & Response



LOF Test

Multi scatter plot for Death Rate vs all other predictors

The REG Procedure
Model: MODEL1
Dependent Variable: DeathRate

Number of Observations Read	53
Number of Observations Used	53

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	20.65433	5.16358	2.01	0.1075
Error	48	123.07398	2.56404		
Lack of Fit	48	123.07398	2.56404		
Pure Error	0	0	.	.	.
Corrected Total	52	143.72830			

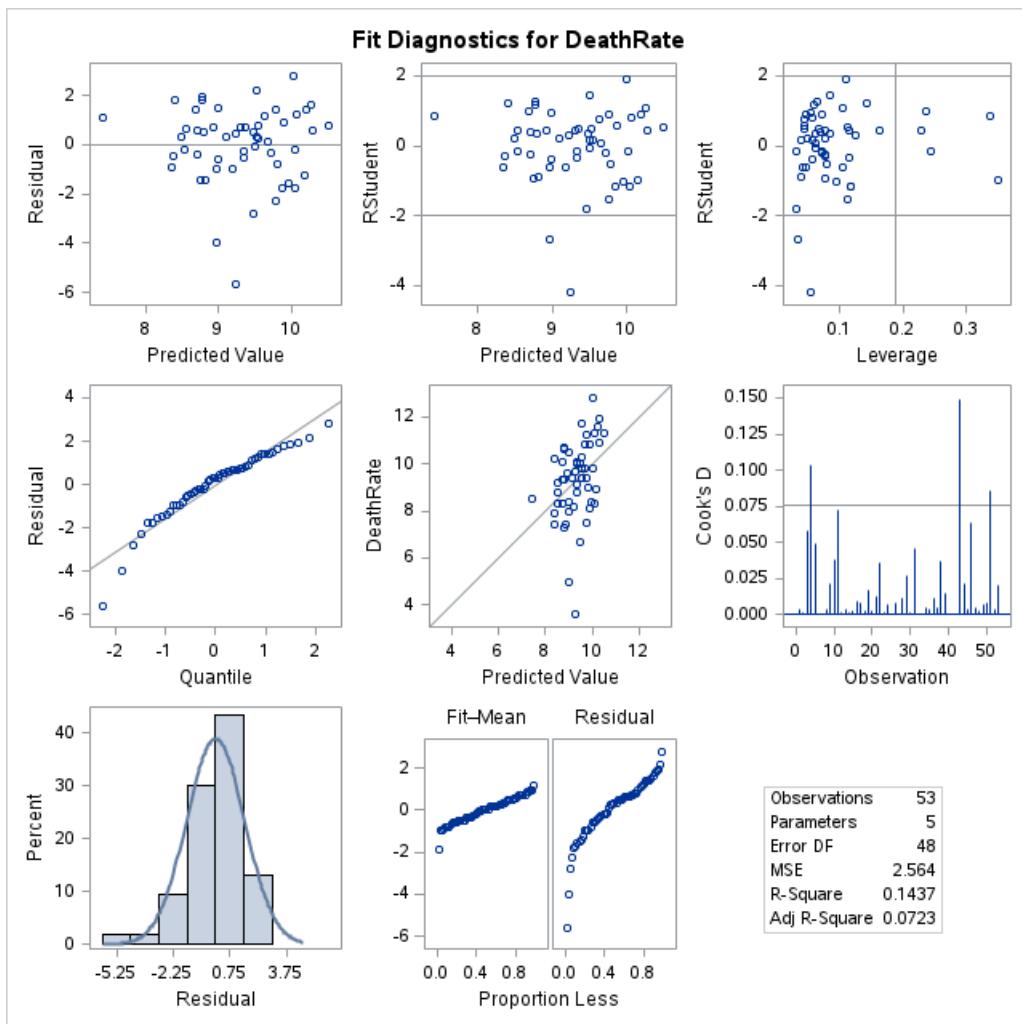
Root MSE	1.60126	R-Square	0.1437
Dependent Mean	9.30566	Adj R-Sq	0.0723
Coeff Var	17.20740		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	12.26626	2.02015	6.07	<.0001
DrAvail	1	0.00739	0.00693	1.07	0.2917
HospAvail	1	0.00058372	0.00072191	0.81	0.4228
CapInc	1	-0.33023	0.23455	-1.41	0.1656
PopD	1	-0.00946	0.00489	-1.94	0.0587

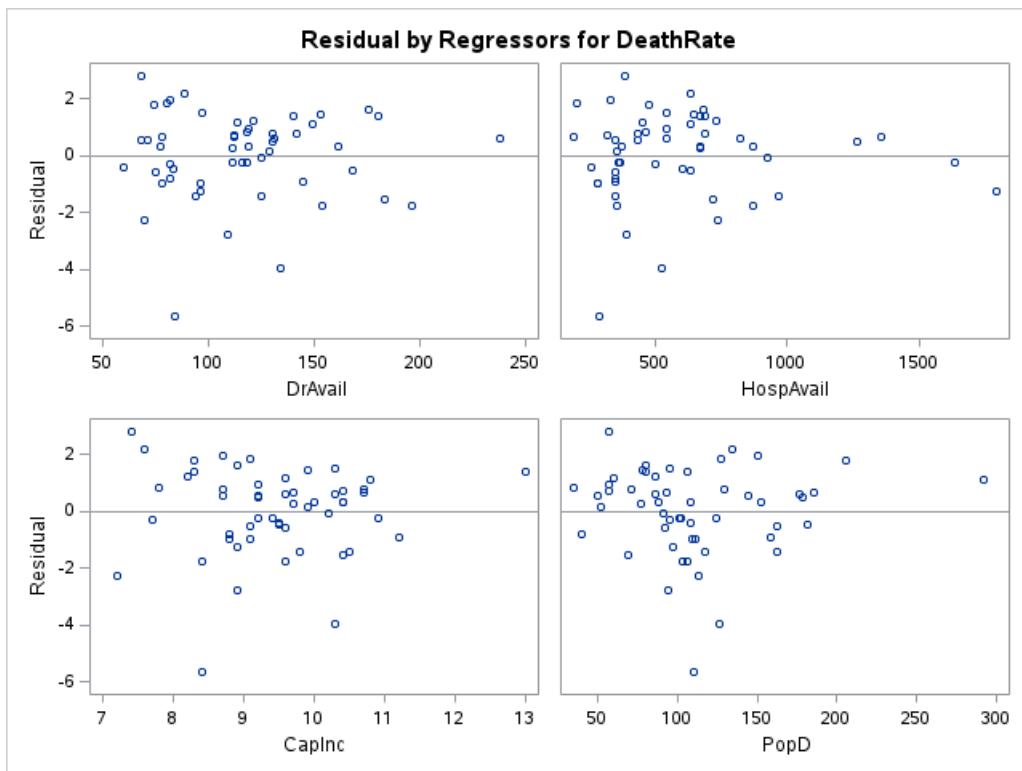
Multi scatter plot for Death Rate vs all other predictors

The REG Procedure

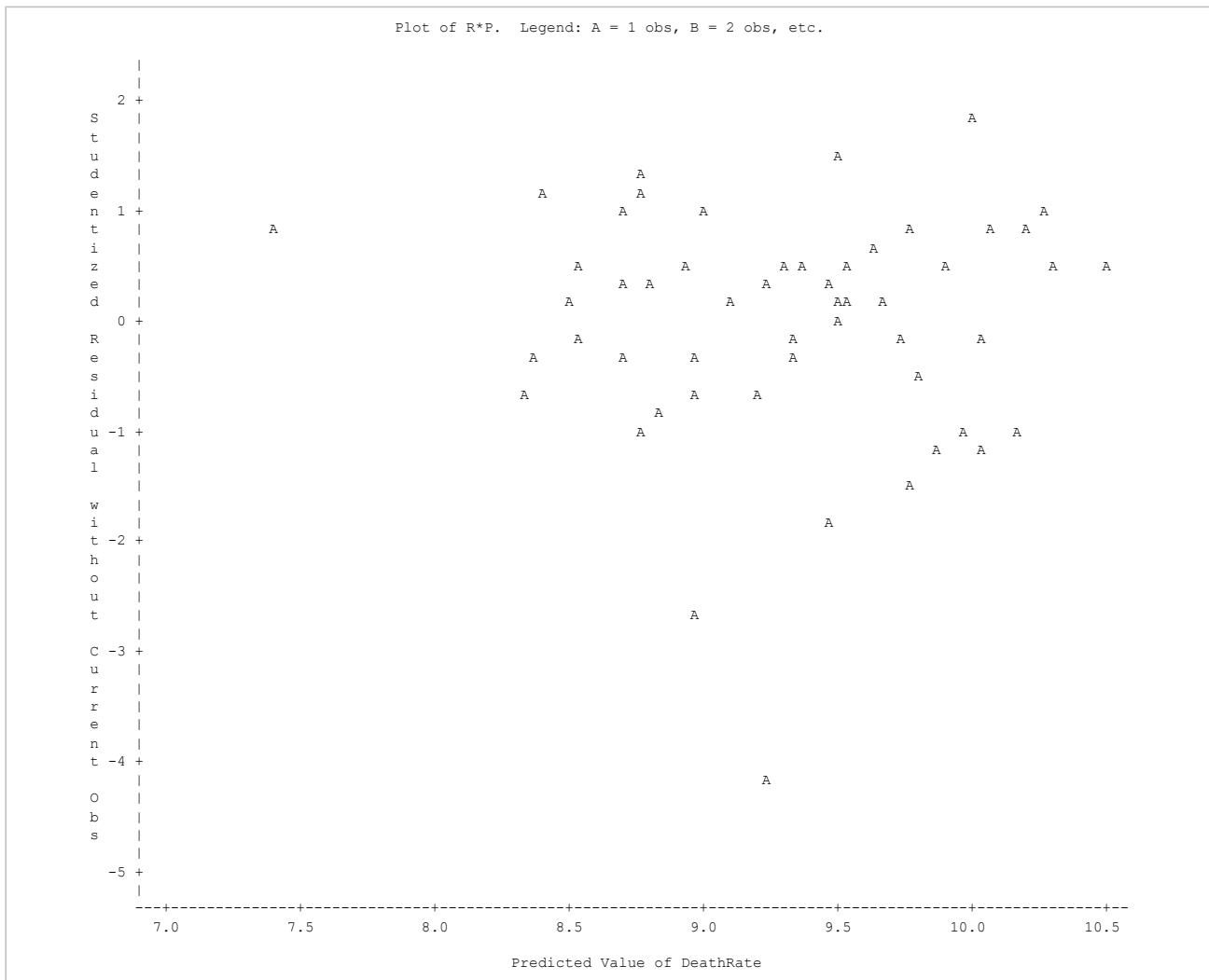
Model: MODEL1
Dependent Variable: DeathRate



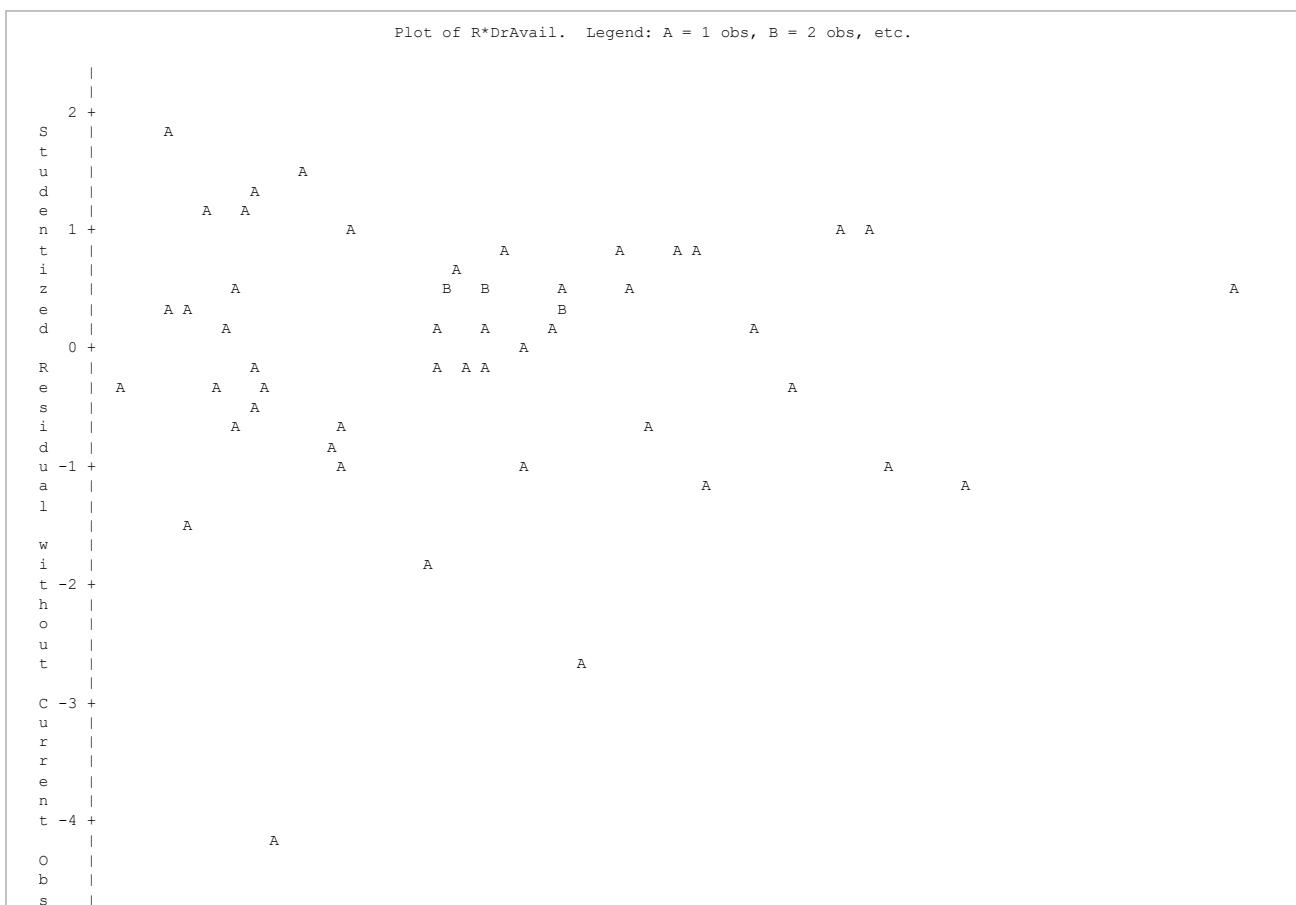
Residual Plot

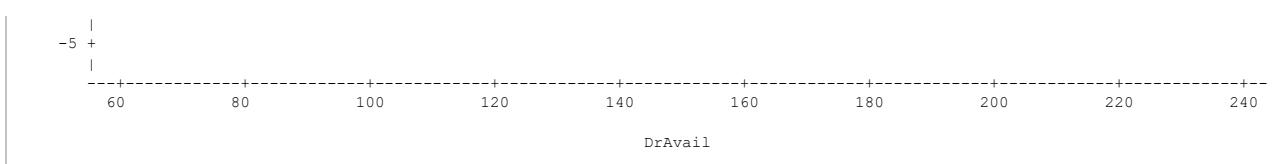


Multi scatter plot for Death Rate vs all other predictors

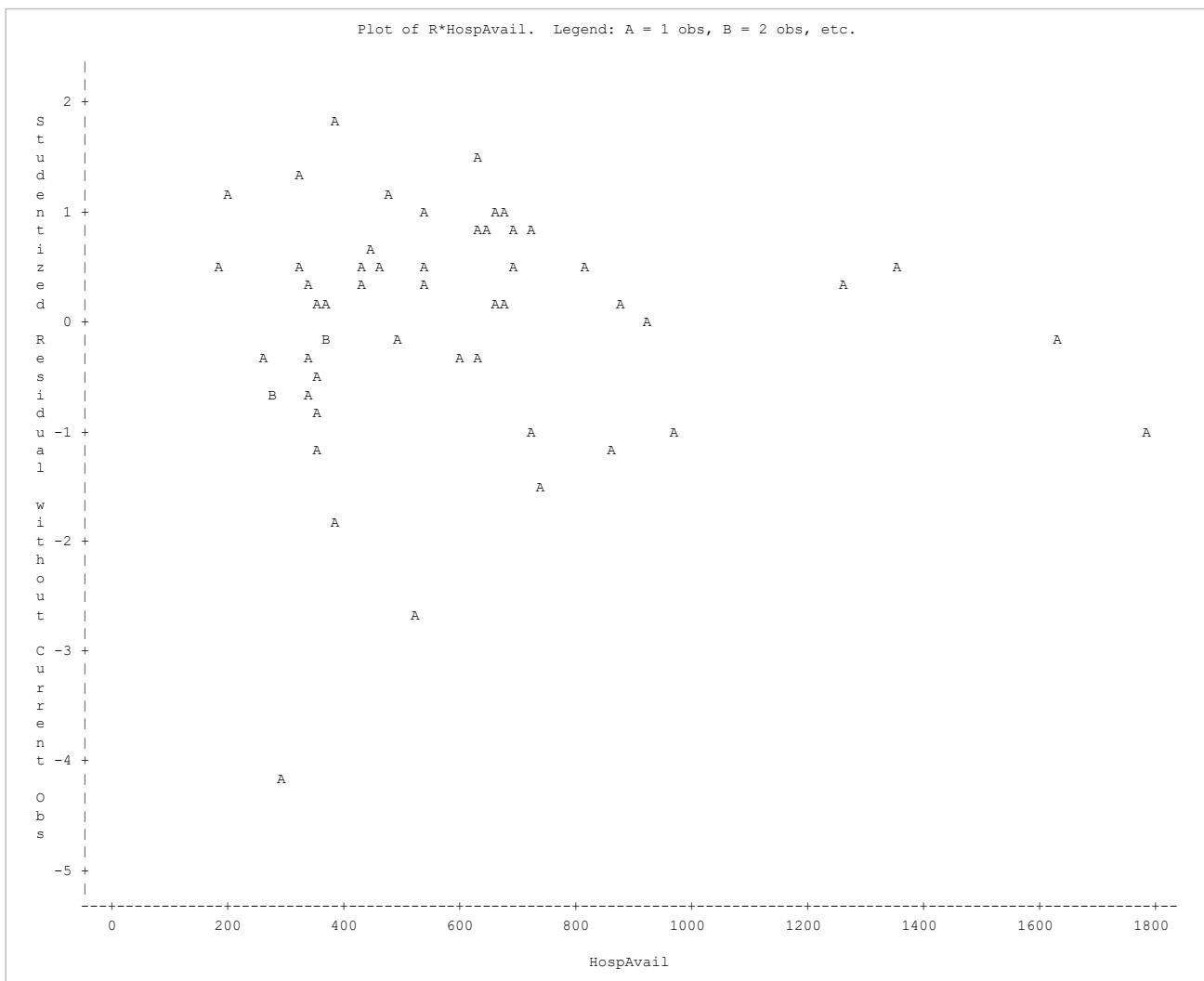


Multi scatter plot for Death Rate vs all other predictors

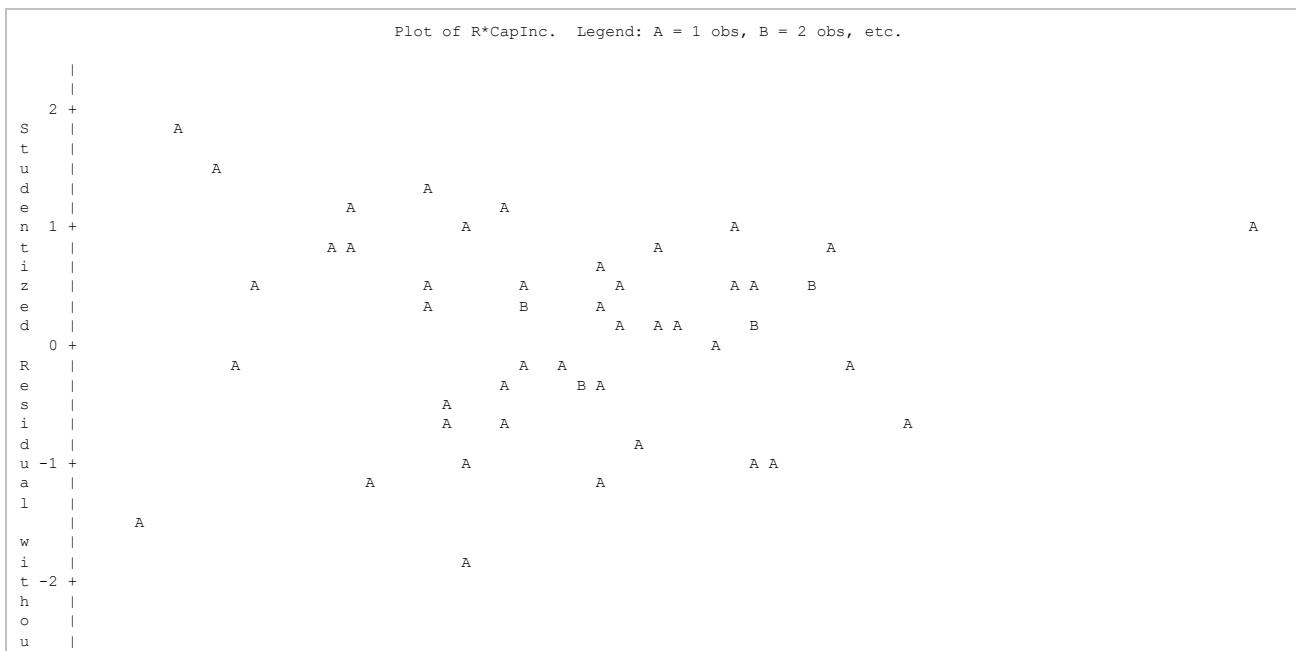


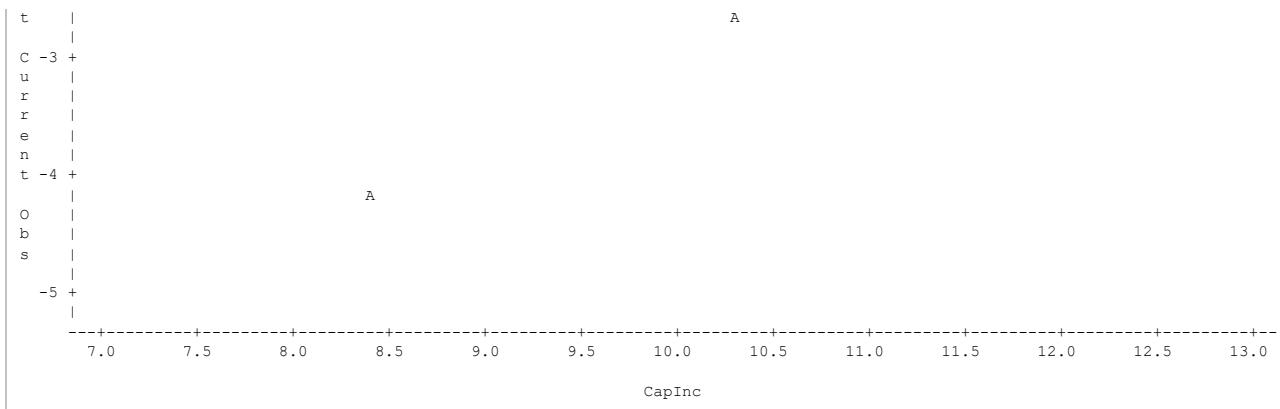


Multi scatter plot for Death Rate vs all other predictors

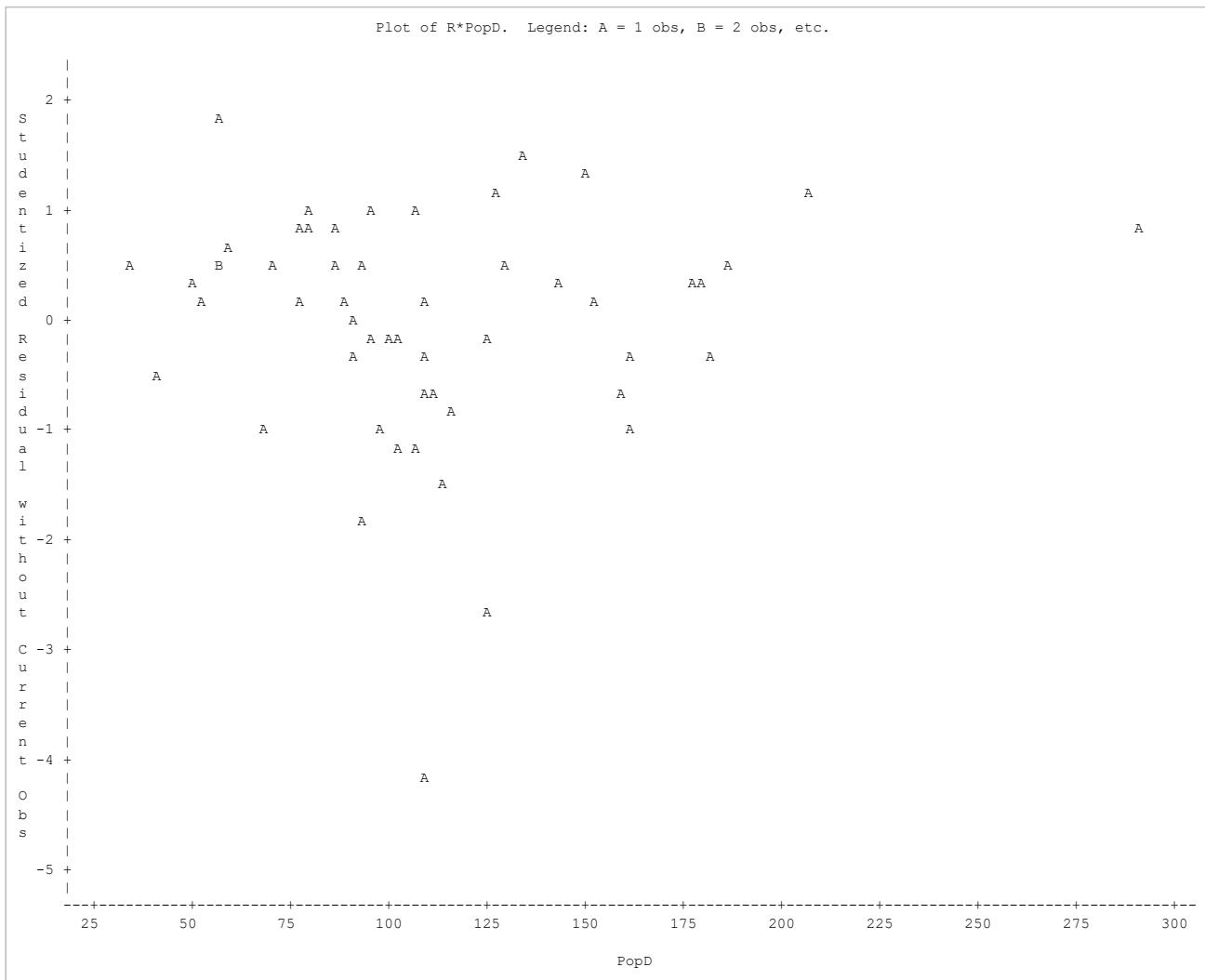


Multi scatter plot for Death Rate vs all other predictors

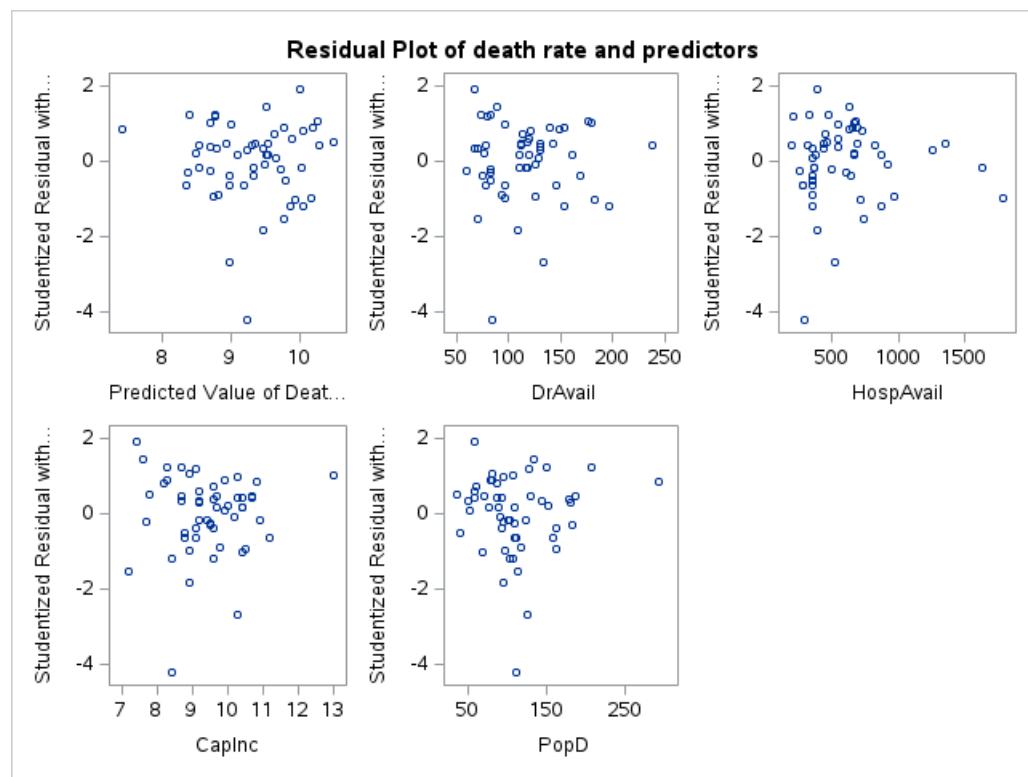




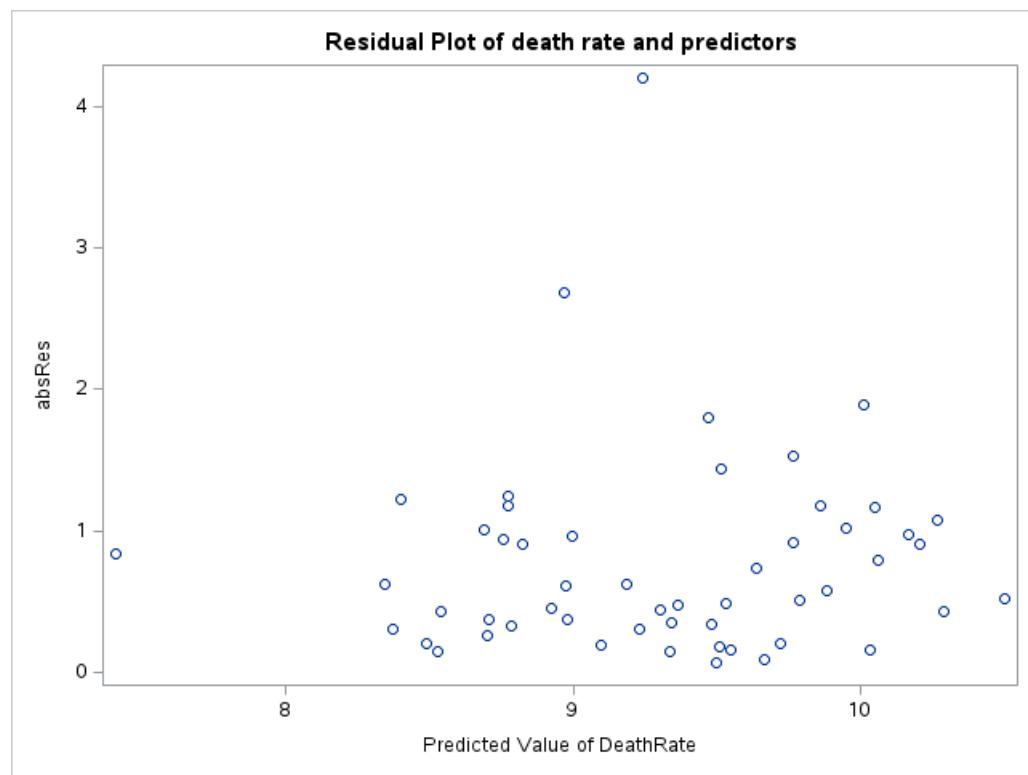
Multi scatter plot for Death Rate vs all other predictors



Plot of Residuals of each Predictor



Plot of Absolute Residuals



Breush-Pagan Test

Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	DeathRate
Parameters	b0 b1 b2 b3 b4
Equations	DeathRate

The Equation to Estimate is

The Equation to Estimate is	
DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 1 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	1

Final Convergence Criteria	
R	0
PPC	0
RPC(b0)	121447.1
Object	0.973542
Trace(S)	2.564041
Objective Value	2.322151

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
DeathRate	5	48	123.1	2.5640	1.6013	0.1437	0.0723

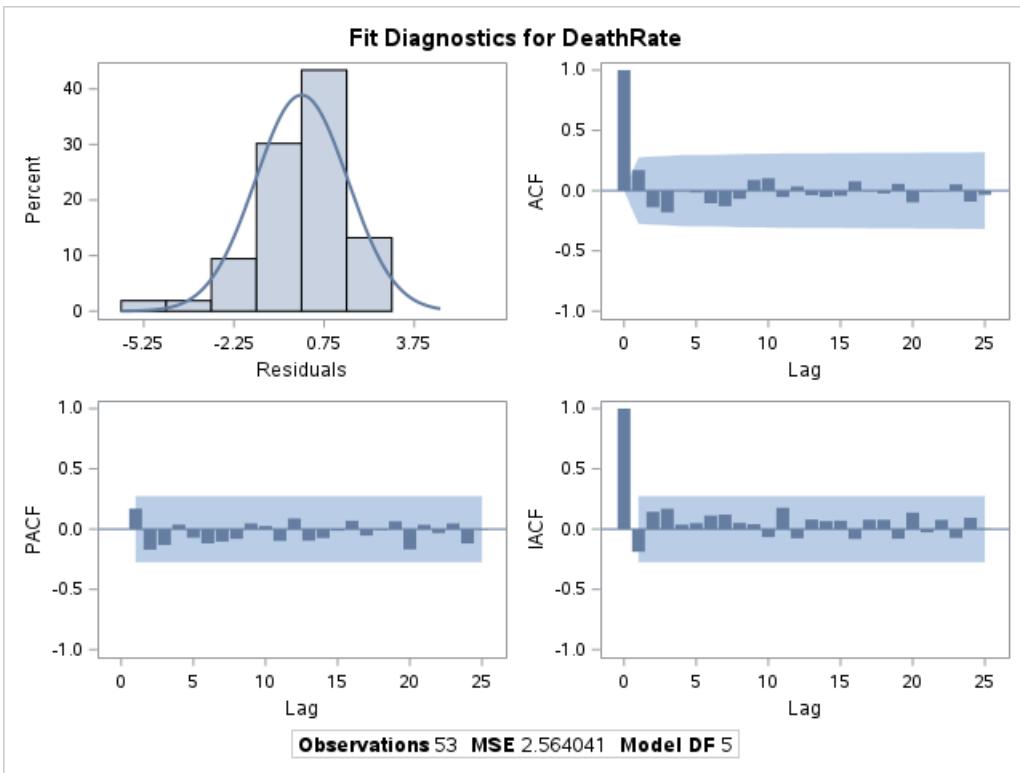
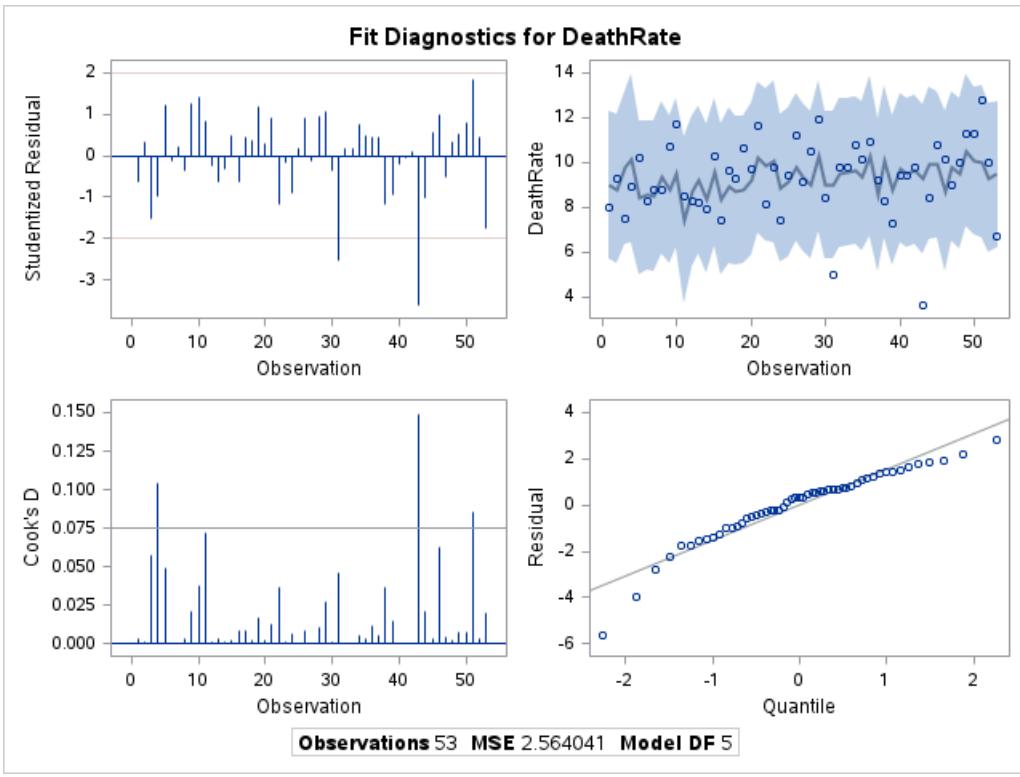
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	12.26626	2.0201	6.07	<.0001	
b1	0.007392	0.00693	1.07	0.2917	
b2	0.000584	0.000722	0.81	0.4228	
b3	-0.33023	0.2346	-1.41	0.1656	
b4	-0.00946	0.00489	-1.94	0.0587	

Number of Observations		Statistics for System	
Used	53	Objective	2.3222
Missing	0	Objective*N	123.0740

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
DeathRate	White's Test	10.75	14	0.7058	Cross of all vars
	Breusch-Pagan	3.86	4	0.4255	DrAvail, HospAvail, CapInc, PopD, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	DeathRate
Parameters(Value(t Value))	b0(12.266255167(6.0719625374)) b1(0.007391615(1.0660501197)) b2(0.0005837157(0.808567984)) b3(-0.330230237(-1.407920563)) b4(-0.009462885(-1.936411417))
Equations	DeathRate

The Equation to Estimate is

The Equation to Estimate is	
DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	2.564041
Objective Value	2.322151

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
DeathRate	5	48	123.1	2.5640	1.6013	0.1437	0.0723

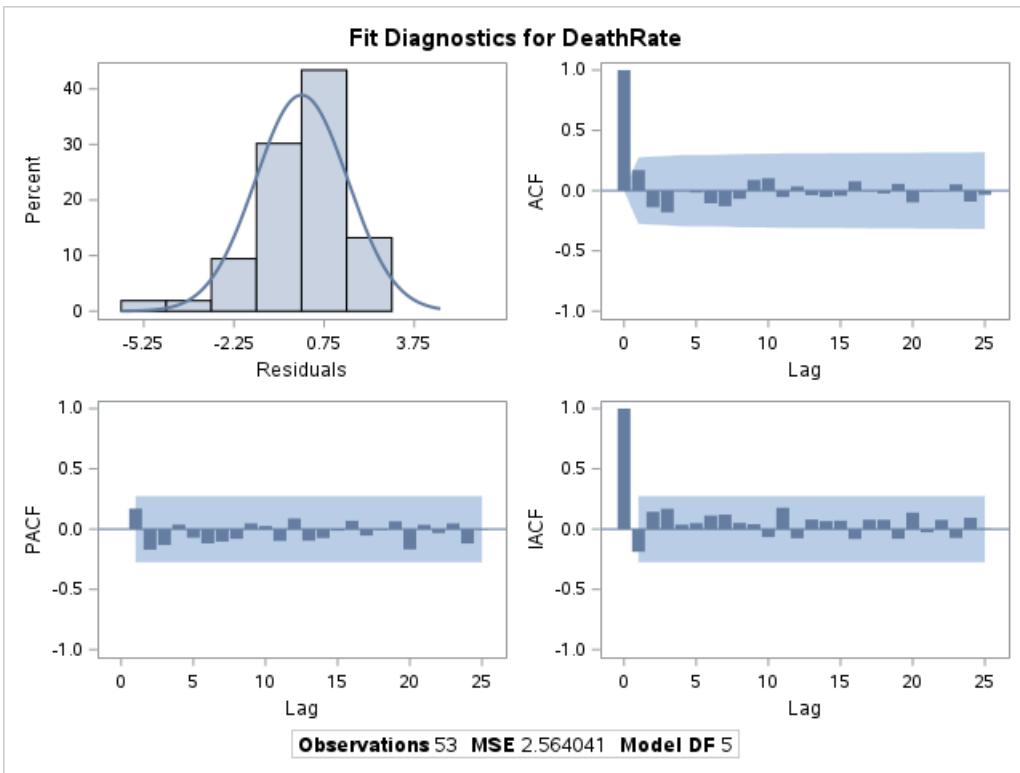
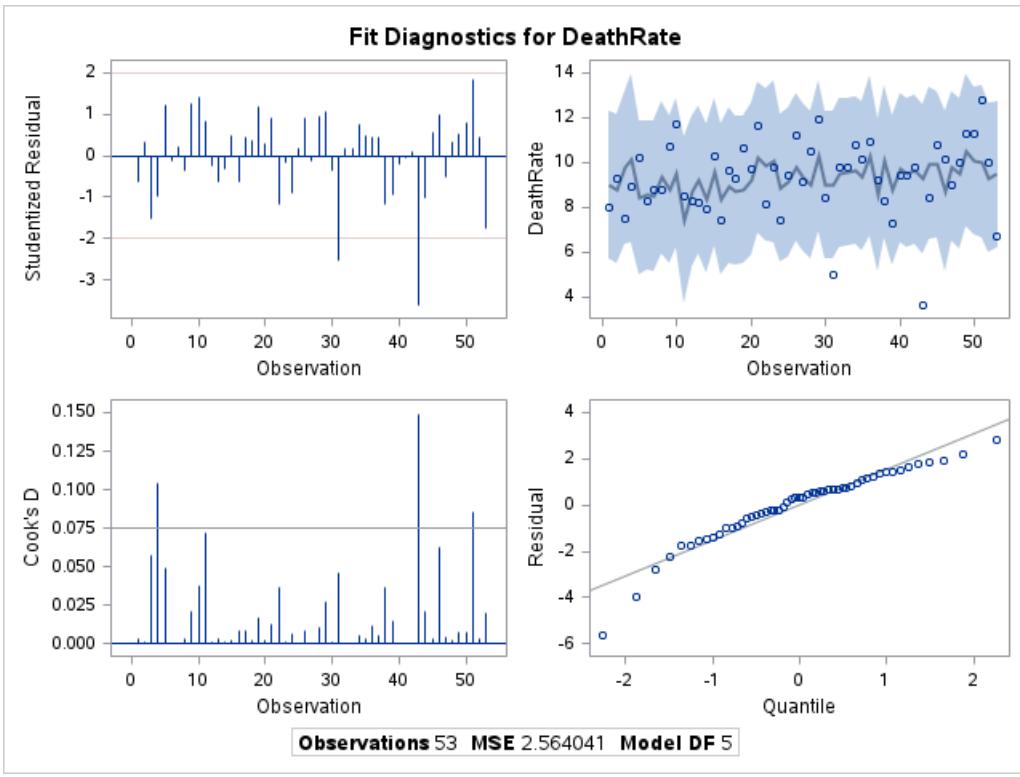
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	12.26626	2.0201	6.07	<.0001	
b1	0.007392	0.00693	1.07	0.2917	
b2	0.000584	0.000722	0.81	0.4228	
b3	-0.33023	0.2346	-1.41	0.1656	
b4	-0.00946	0.00489	-1.94	0.0587	

Number of Observations		Statistics for System	
Used	53	Objective	2.3222
Missing	0	Objective*N	123.0740

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
DeathRate	Breusch-Pagan	0.66	1	0.4164	DrAvail, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	DeathRate
Parameters(Value(t Value))	b0(12.266255167(6.0719625374)) b1(0.007391615(1.0660501197)) b2(0.0005837157(0.808567984)) b3(-0.330230237(-1.407920563)) b4(-0.009462885(-1.936411417))
Equations	DeathRate

The Equation to Estimate is

The Equation to Estimate is	
DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	2.564041
Objective Value	2.322151

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
DeathRate	5	48	123.1	2.5640	1.6013	0.1437	0.0723

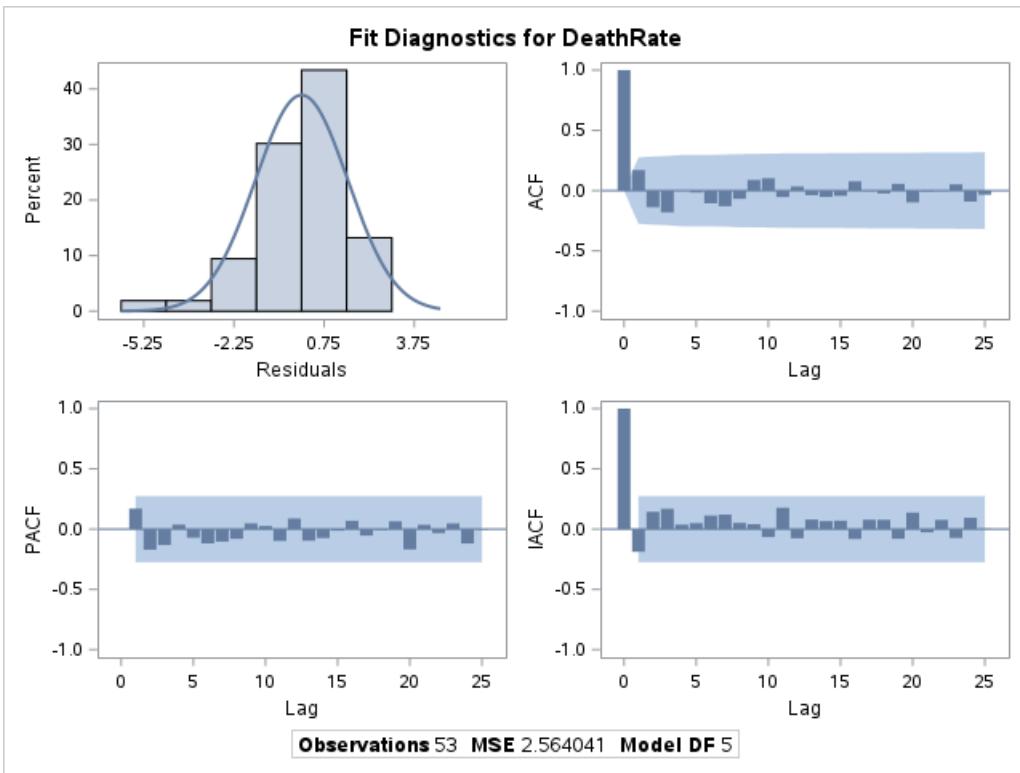
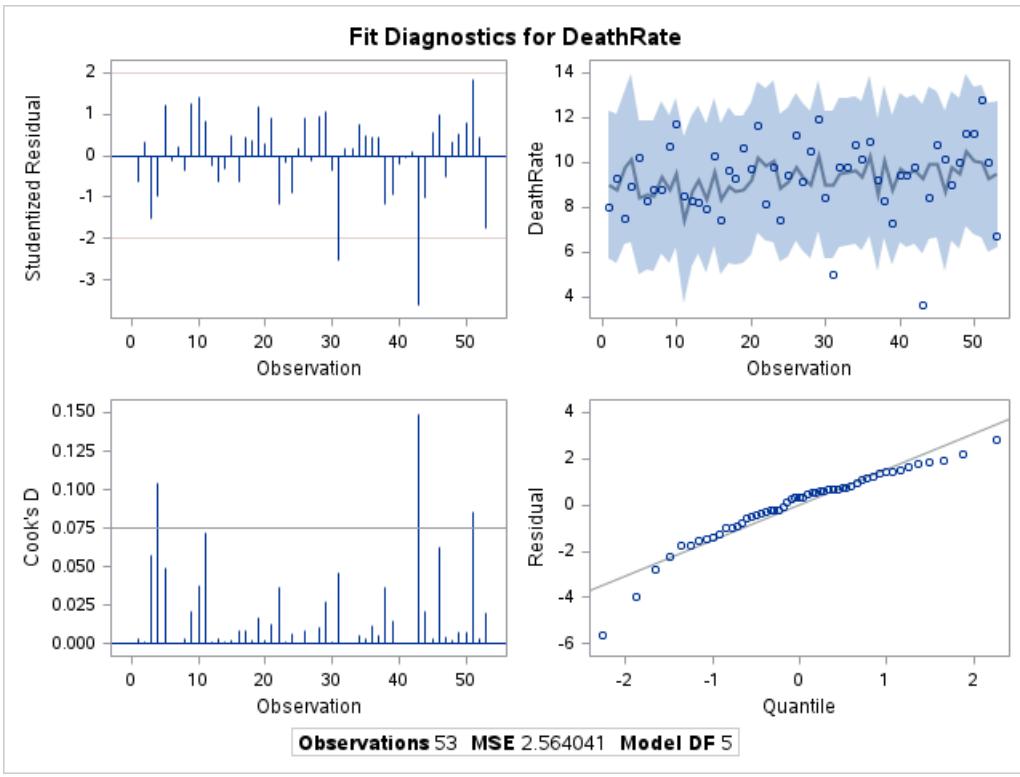
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	12.26626	2.0201	6.07	<.0001	
b1	0.007392	0.00693	1.07	0.2917	
b2	0.000584	0.000722	0.81	0.4228	
b3	-0.33023	0.2346	-1.41	0.1656	
b4	-0.00946	0.00489	-1.94	0.0587	

Number of Observations		Statistics for System	
Used	53	Objective	2.3222
Missing	0	Objective*N	123.0740

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
DeathRate	Breusch-Pagan	1.27	1	0.2597	HospAvail, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	DeathRate
Parameters(Value(t Value))	b0(12.266255167(6.0719625374)) b1(0.007391615(1.0660501197)) b2(0.0005837157(0.808567984)) b3(-0.330230237(-1.407920563)) b4(-0.009462885(-1.936411417))
Equations	DeathRate

The Equation to Estimate is

The Equation to Estimate is	
DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	2.564041
Objective Value	2.322151

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
DeathRate	5	48	123.1	2.5640	1.6013	0.1437	0.0723

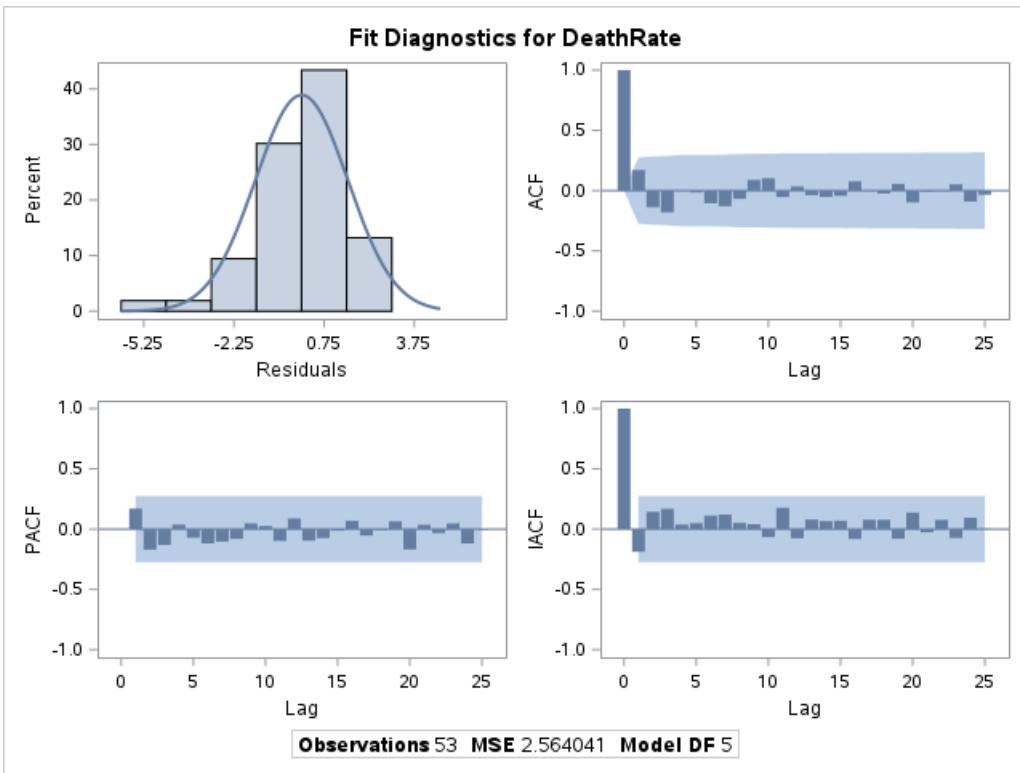
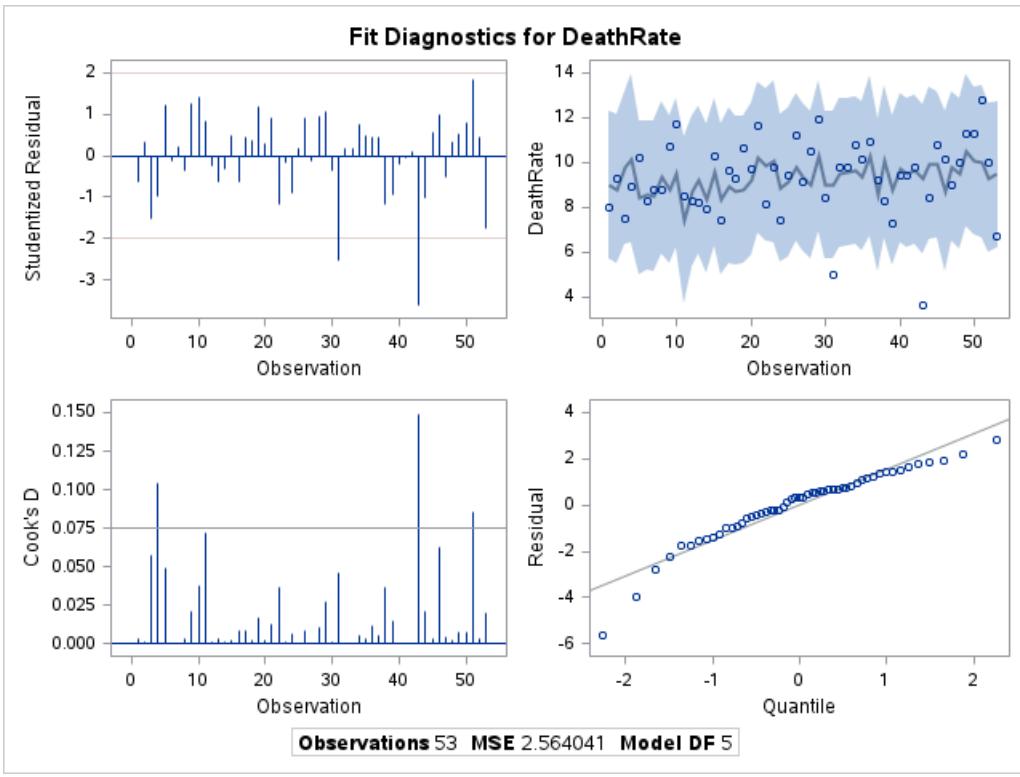
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	12.26626	2.0201	6.07	<.0001	
b1	0.007392	0.00693	1.07	0.2917	
b2	0.000584	0.000722	0.81	0.4228	
b3	-0.33023	0.2346	-1.41	0.1656	
b4	-0.00946	0.00489	-1.94	0.0587	

Number of Observations		Statistics for System	
Used	53	Objective	2.3222
Missing	0	Objective*N	123.0740

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
DeathRate	Breusch-Pagan	2.41	1	0.1207	CapInc, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	DeathRate
Parameters(Value(t Value))	b0(12.266255167(6.0719625374)) b1(0.007391615(1.0660501197)) b2(0.0005837157(0.808567984)) b3(-0.330230237(-1.407920563)) b4(-0.009462885(-1.936411417))
Equations	DeathRate

The Equation to Estimate is

The Equation to Estimate is	
DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	2.564041
Objective Value	2.322151

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
DeathRate	5	48	123.1	2.5640	1.6013	0.1437	0.0723

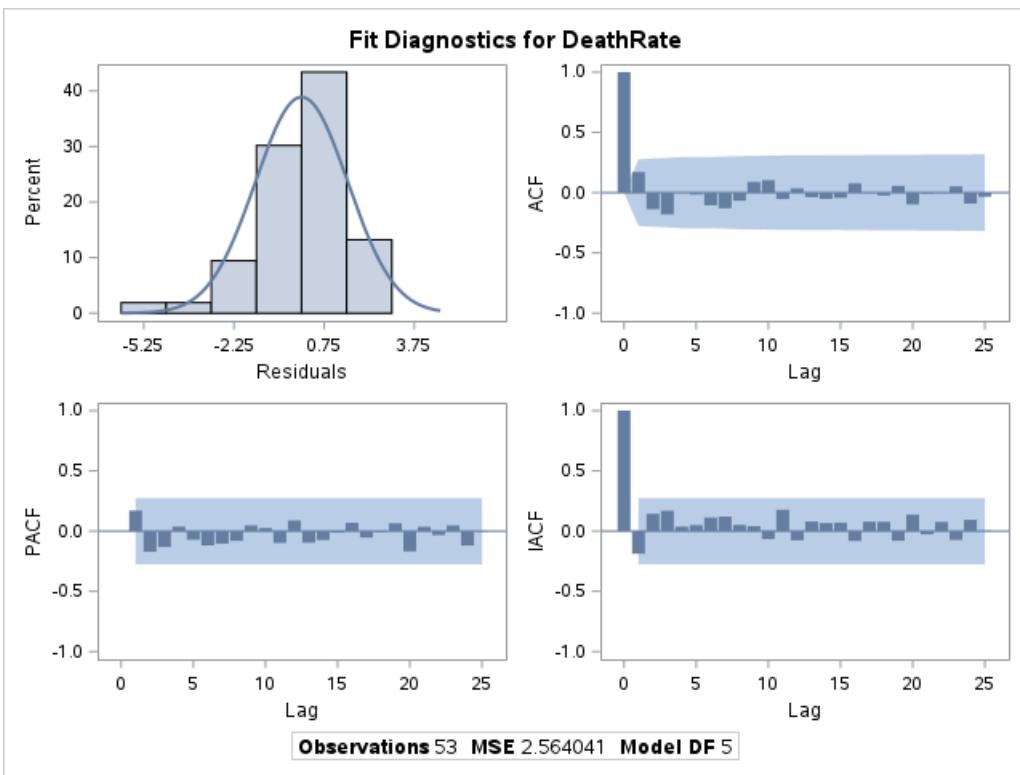
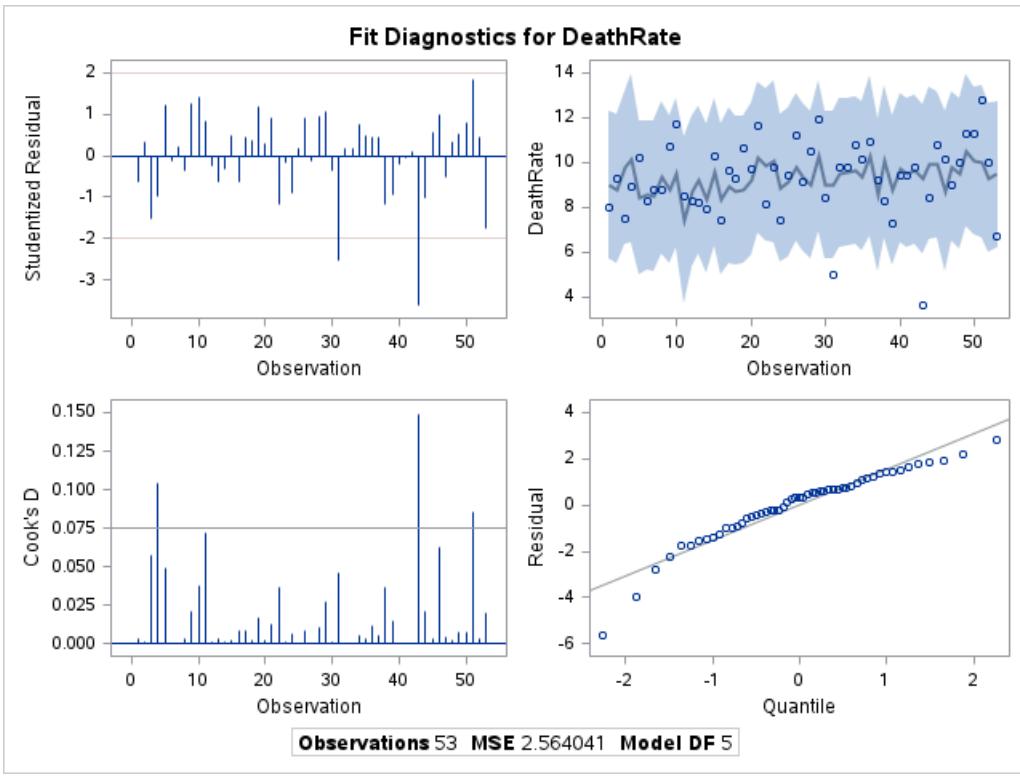
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	12.26626	2.0201	6.07	<.0001	
b1	0.007392	0.00693	1.07	0.2917	
b2	0.000584	0.000722	0.81	0.4228	
b3	-0.33023	0.2346	-1.41	0.1656	
b4	-0.00946	0.00489	-1.94	0.0587	

Number of Observations		Statistics for System	
Used	53	Objective	2.3222
Missing	0	Objective*N	123.0740

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
DeathRate	Breusch-Pagan	0.00	1	0.9687	PopD, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The GLM Procedure

Class Level Information		
Class	Levels	Values
Group	2	0 1

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

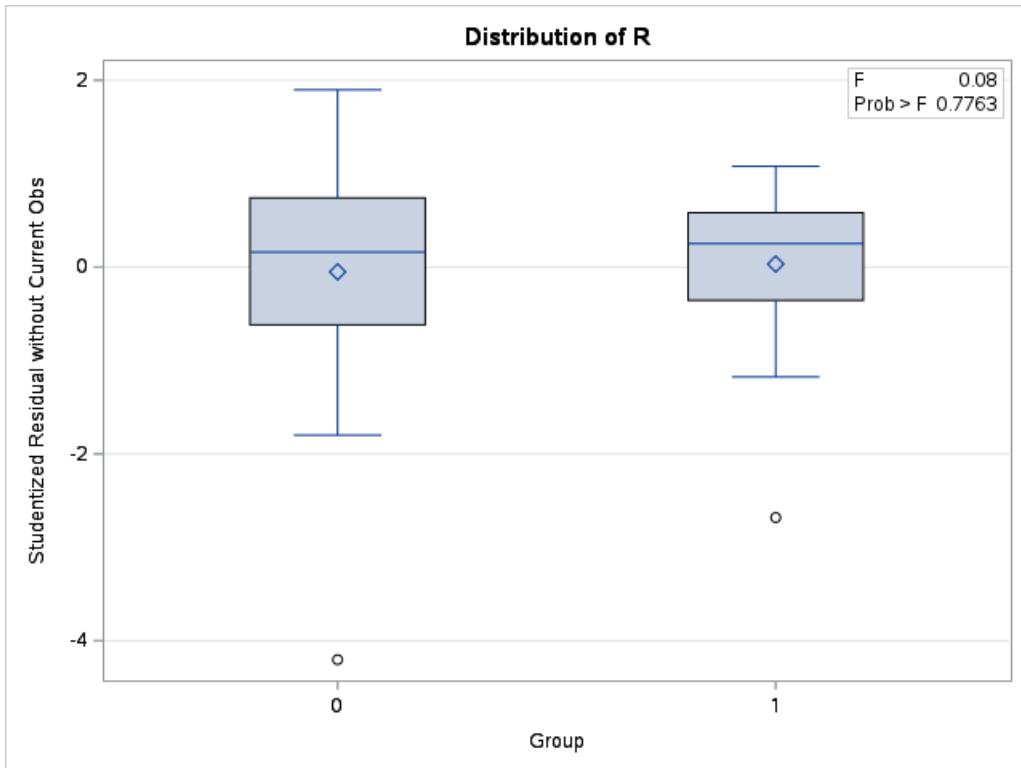
Dependent Variable: R Studentized Residual without Current Obs

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.09272336	0.09272336	0.08	0.7763
Error	51	57.95188106	1.13631139		
Corrected Total	52	58.04460443			

R-Square	Coeff Var	Root MSE	R Mean
0.001597	-10539.78	1.065979	-0.010114

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Group	1	0.09272336	0.09272336	0.08	0.7763

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Group	1	0.09272336	0.09272336	0.08	0.7763



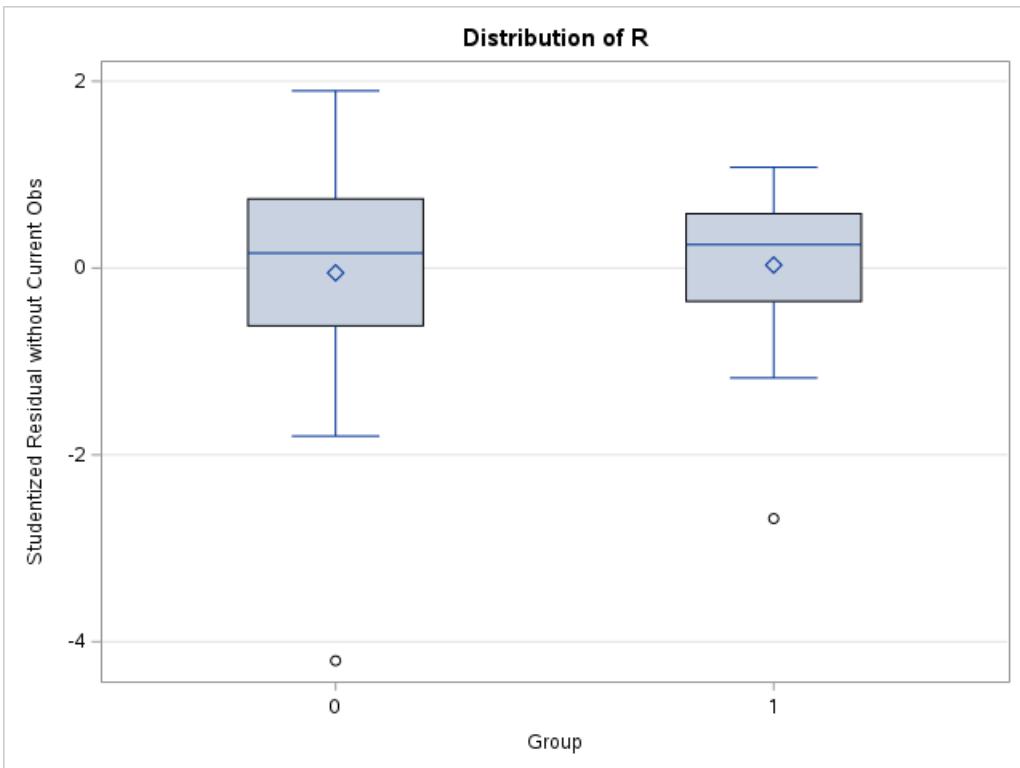
Residual Plot of death rate and predictors

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	0.6806	0.6806	1.17	0.2842
Error	51	29.6245	0.5809		

Residual Plot of death rate and predictors

The GLM Procedure



Level of Group	N	R	
		Mean	Std Dev
0	27	-0.05115898	1.22183891
1	26	0.03250990	0.87491103

Residual Plot of death rate and predictors

The GLM Procedure

Class Level Information		
Class	Levels	Values
Group	2	0 1

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

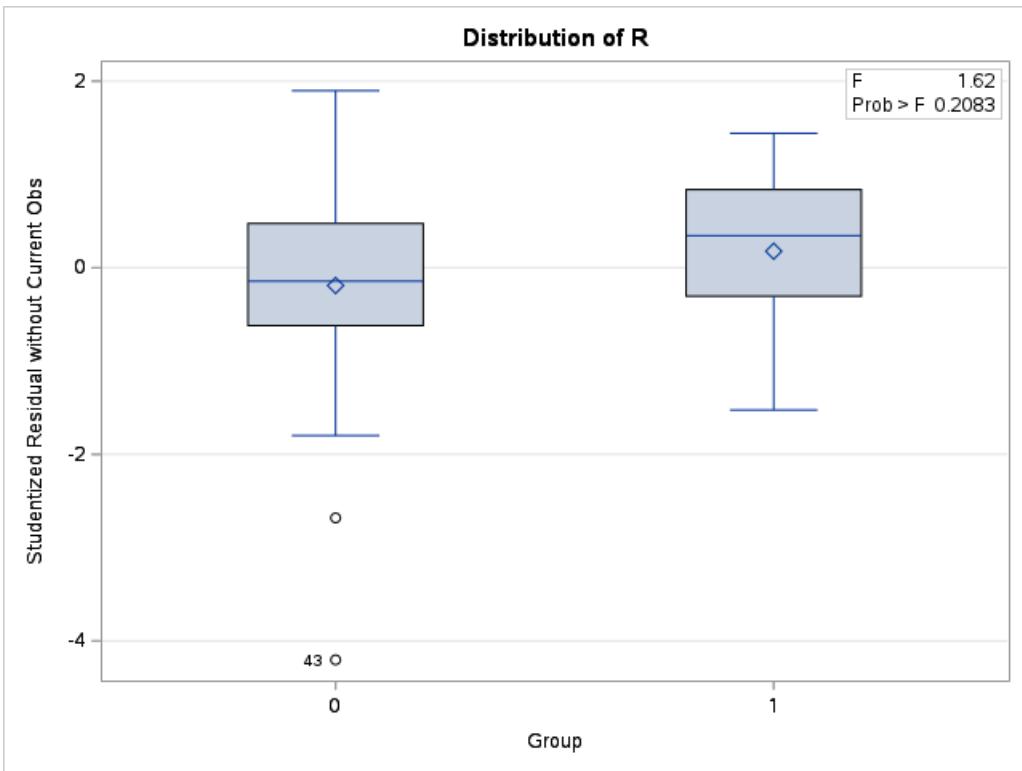
Dependent Variable: R Studentized Residual without Current Obs

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.79108940	1.79108940	1.62	0.2083
Error	51	56.25351502	1.10301010		
Corrected Total	52	58.04460443			

R-Square	Coeff Var	Root MSE	R Mean
0.030857	-10384.19	1.050243	-0.010114

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Group	1	1.79108940	1.79108940	1.62	0.2083

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Group	1	1.79108940	1.79108940	1.62	0.2083



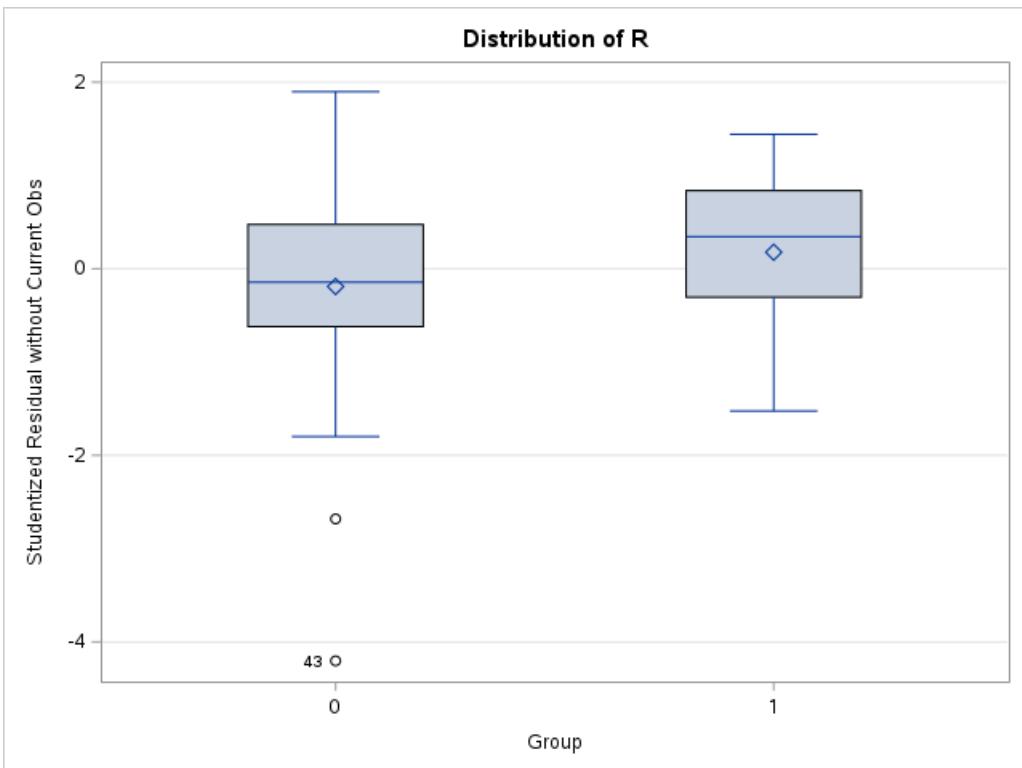
Residual Plot of death rate and predictors

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	0.8107	0.8107	1.52	0.2239
Error	51	27.2724	0.5348		

Residual Plot of death rate and predictors

The GLM Procedure



Level of Group	N	R	
		Mean	Std Dev
0	27	-0.19050928	1.25219156
1	26	0.17721983	0.78704355

Residual Plot of death rate and predictors

The GLM Procedure

Class Level Information		
Class	Levels	Values
Group	2	0 1

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

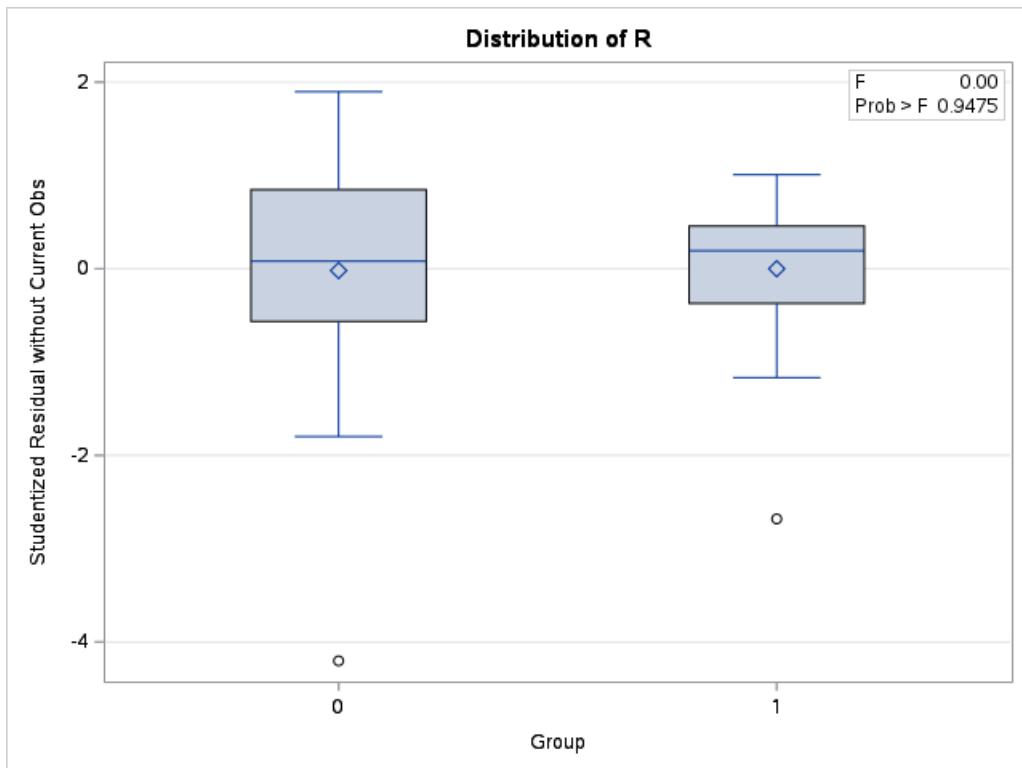
Dependent Variable: R Studentized Residual without Current Obs

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00497750	0.00497750	0.00	0.9475
Error	51	58.03962692	1.13803190		
Corrected Total	52	58.04460443			

R-Square	Coeff Var	Root MSE	R Mean
0.000086	-10547.75	1.066786	-0.010114

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Group	1	0.00497750	0.00497750	0.00	0.9475

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Group	1	0.00497750	0.00497750	0.00	0.9475



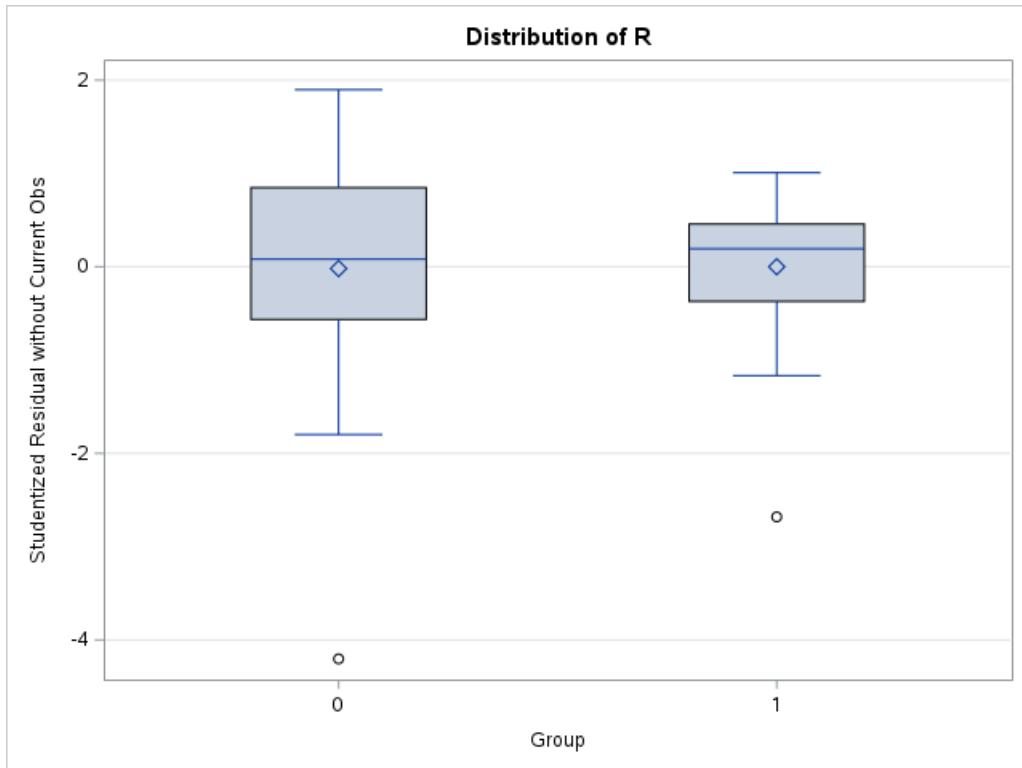
Residual Plot of death rate and predictors

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	1.2624	1.2624	2.31	0.1345
Error	51	27.8446	0.5460		

Residual Plot of death rate and predictors

The GLM Procedure



Residual Plot of death rate and predictors

The GLM Procedure

Class Level Information		
Class	Levels	Values
Group	2	0 1

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

Dependent Variable: R Studentized Residual without Current Obs

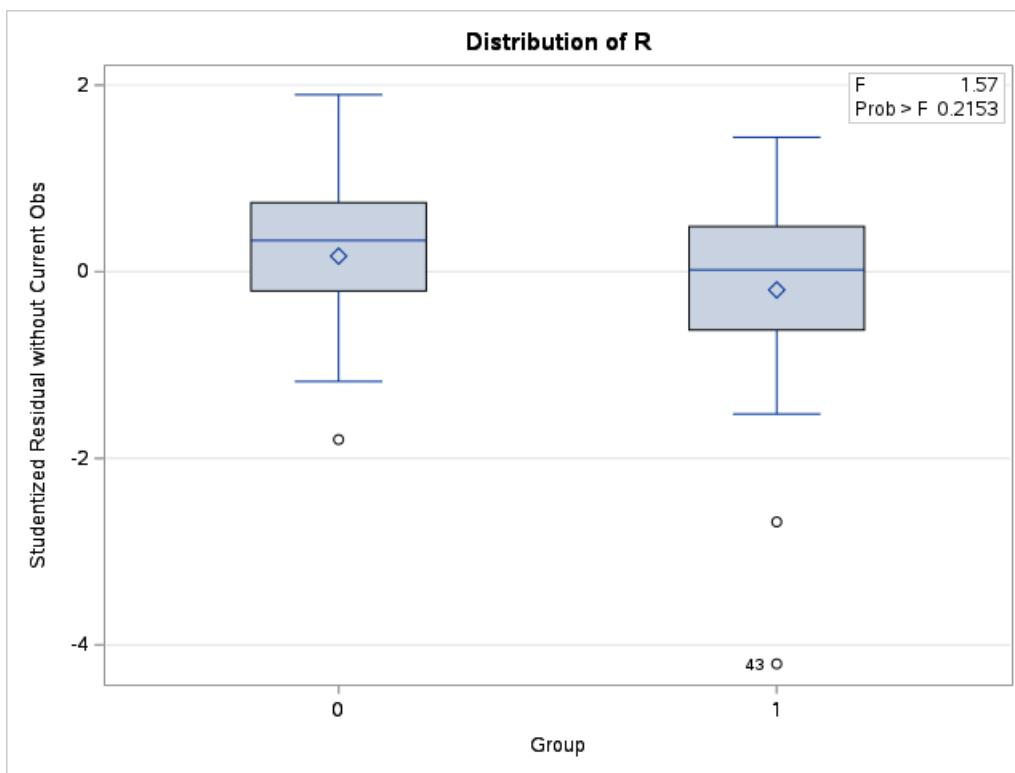
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.73786495	1.73786495	1.57	0.2153
Error	51	56.30673948	1.10405372		
Corrected Total	52	58.04460443			

R-Square	Coeff Var	Root MSE	R Mean
0.029940	-10389.10	1.050740	-0.010114

Source	DF	Type I SS	Mean Square	F Value	Pr > F

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Group	1	1.73786495	1.73786495	1.57	0.2153

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Group	1	1.73786495	1.73786495	1.57	0.2153



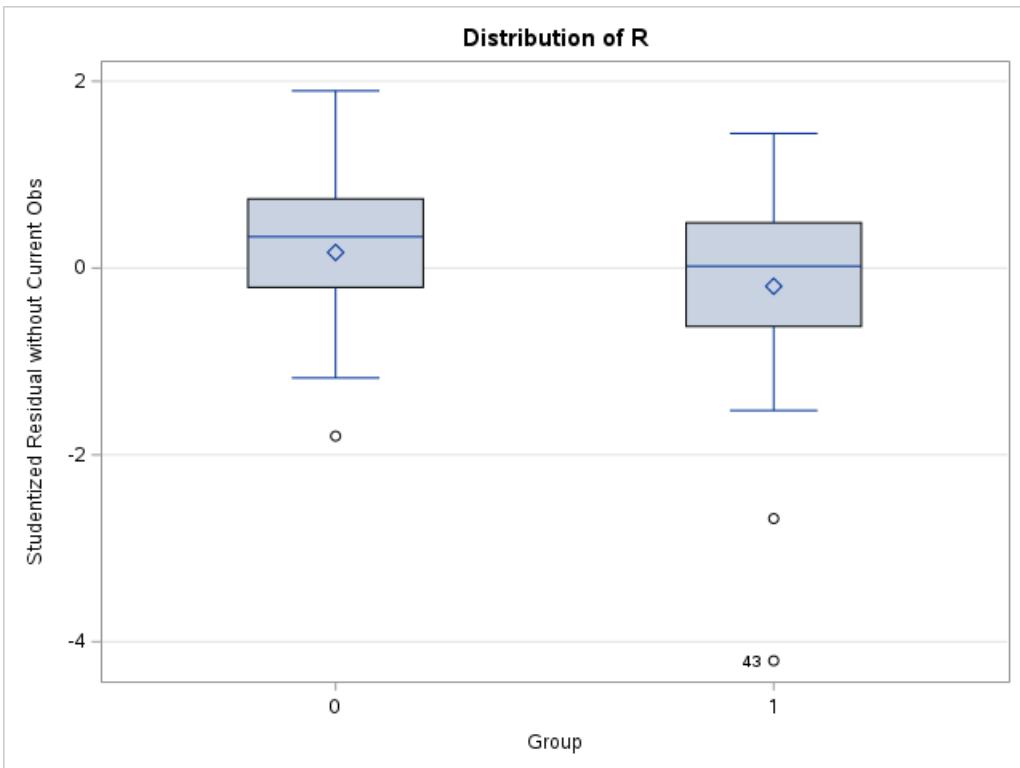
Residual Plot of death rate and predictors

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	1.2725	1.2725	2.40	0.1279
Error	51	27.0902	0.5312		

Residual Plot of death rate and predictors

The GLM Procedure



Level of Group	N	R	
		Mean	Std Dev
0	27	0.16758099	0.79581140
1	26	-0.19464315	1.26238709

Test for Normality

Residual Plot of death rate and predictors

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

Moments			
N	53	Sum Weights	53
Mean	-0.0101139	Sum Observations	-0.536035
Std Deviation	1.05652373	Variance	1.11624239
Skewness	-1.4640972	Kurtosis	3.99060205
Uncorrected SS	58.0500258	Corrected SS	58.0446044
Coeff Variation	-10446.288	Std Error Mean	0.1451247

Basic Statistical Measures			
Location		Variability	
Mean	-0.01011	Std Deviation	1.05652
Median	0.19351	Variance	1.11624
Mode	.	Range	6.10039
		Interquartile Range	1.09173

Tests for Location: Mu0=0				
Test	Statistic	p Value		
Student's t	t	-0.06969	Pr > t	0.9447
Sign	M	3.5	Pr >= M	0.4101
Signed Rank	S	80.5	Pr >= S	0.4813

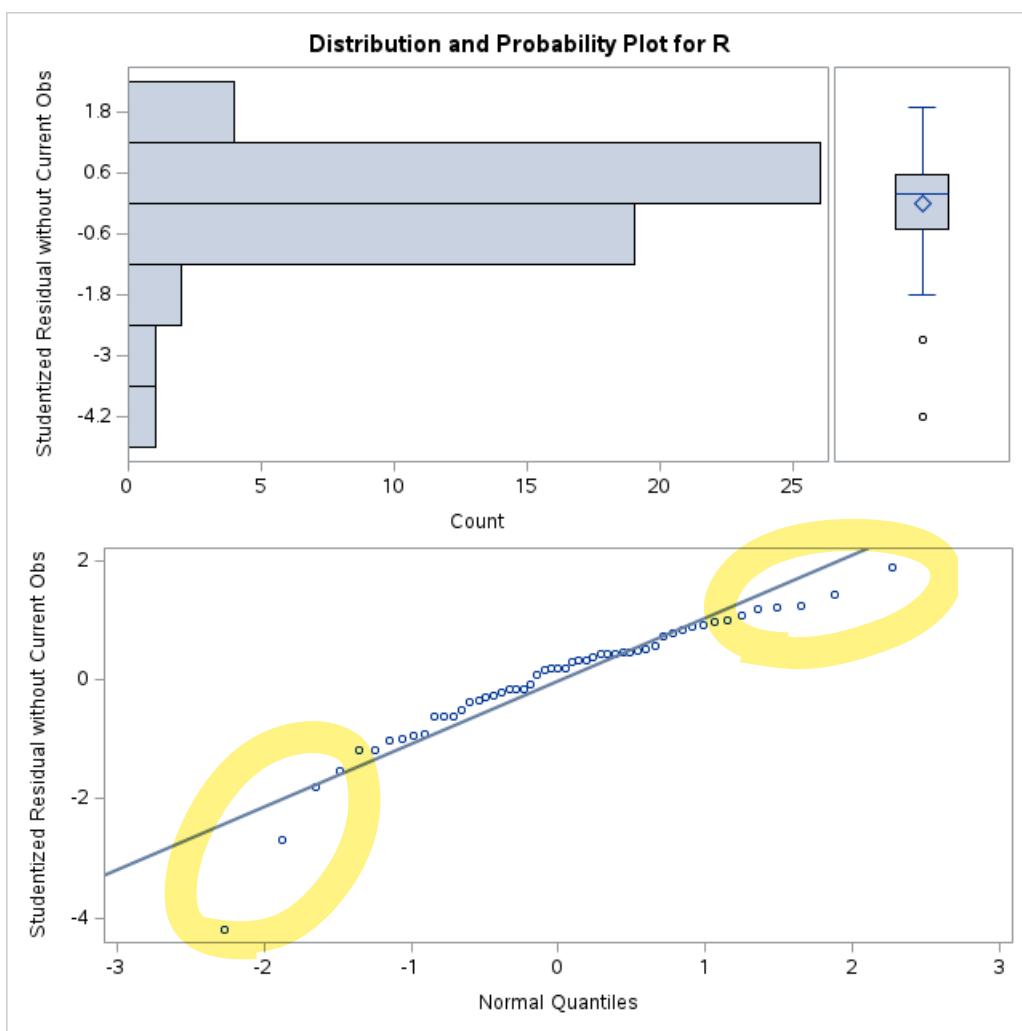
Tests for Normality				
Test	Statistic	p Value		
Shapiro-Wilk	W	0.908813	Pr < W	0.0007
Kolmogorov-Smirnov	D	0.111314	Pr > D	0.0979
Cramer-von Mises	W-Sq	0.156212	Pr > W-Sq	0.0202
Anderson-Darling	A-Sq	1.010857	Pr > A-Sq	0.0108

Quantiles (Definition 5)	
Level	Quantile
100% Max	1.897573

Quantiles (Definition 5)	
Level	Quantile
99%	1.897573
95%	1.251599
90%	1.079139
75% Q3	0.581303
50% Median	0.193510
25% Q1	-0.510427
10%	-1.166922
5%	-1.798496
1%	-4.202817
0% Min	-4.202817

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-4.20282	43	1.18206	19
-2.68113	31	1.21972	5
-1.79850	53	1.25160	9
-1.52445	3	1.44157	10
-1.17590	22	1.89757	51

qq Plot



2) LOF Test for Transformed Response

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: log_DeathRate

Number of Observations Read	53
Number of Observations Used	53

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

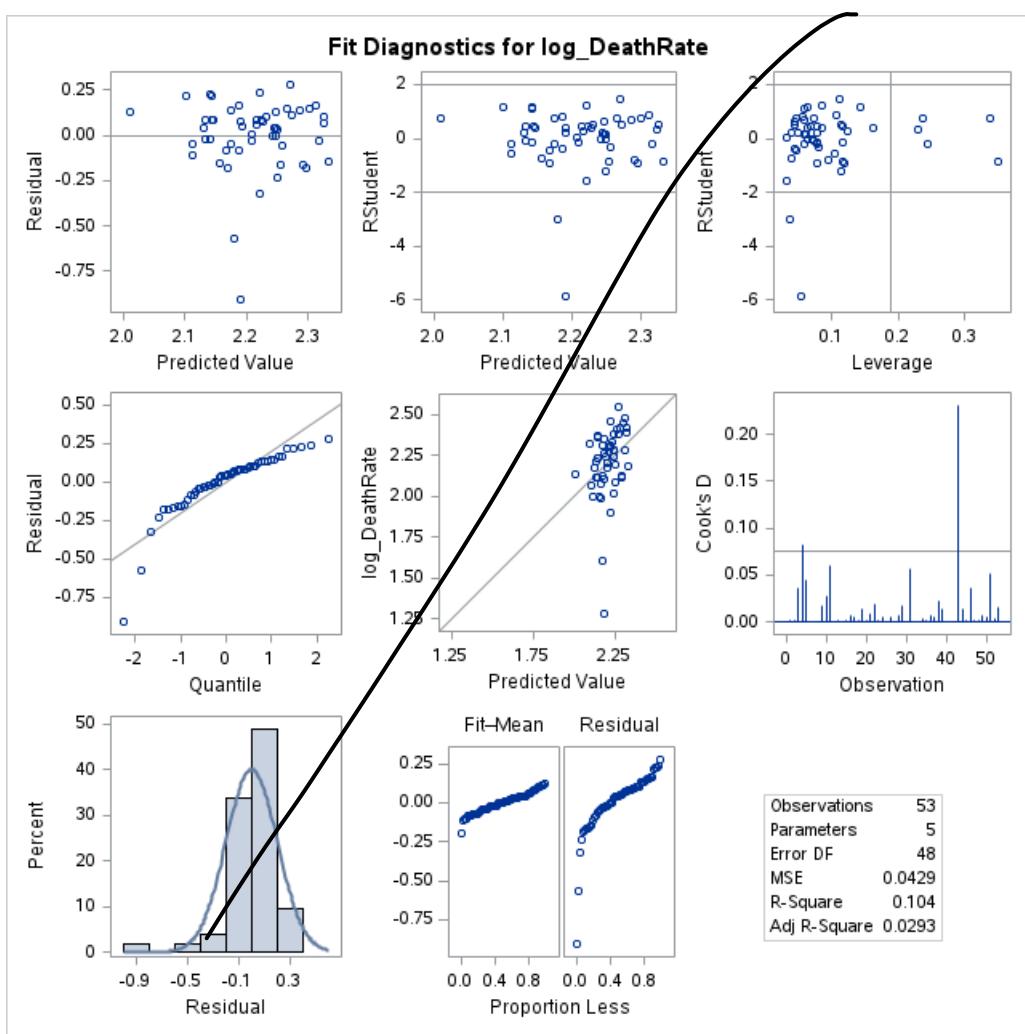
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.23906	0.05977	1.39	0.2508
Error	48	2.06050	0.04293		
Lack of Fit	48	2.06050	0.04293		
Pure Error	0	0			
Corrected Total	52	2.29956			

Root MSE	0.20719	R-Square	0.1040
Dependent Mean	2.21149	Adj R-Sq	0.0293
Coeff Var	9.36874		

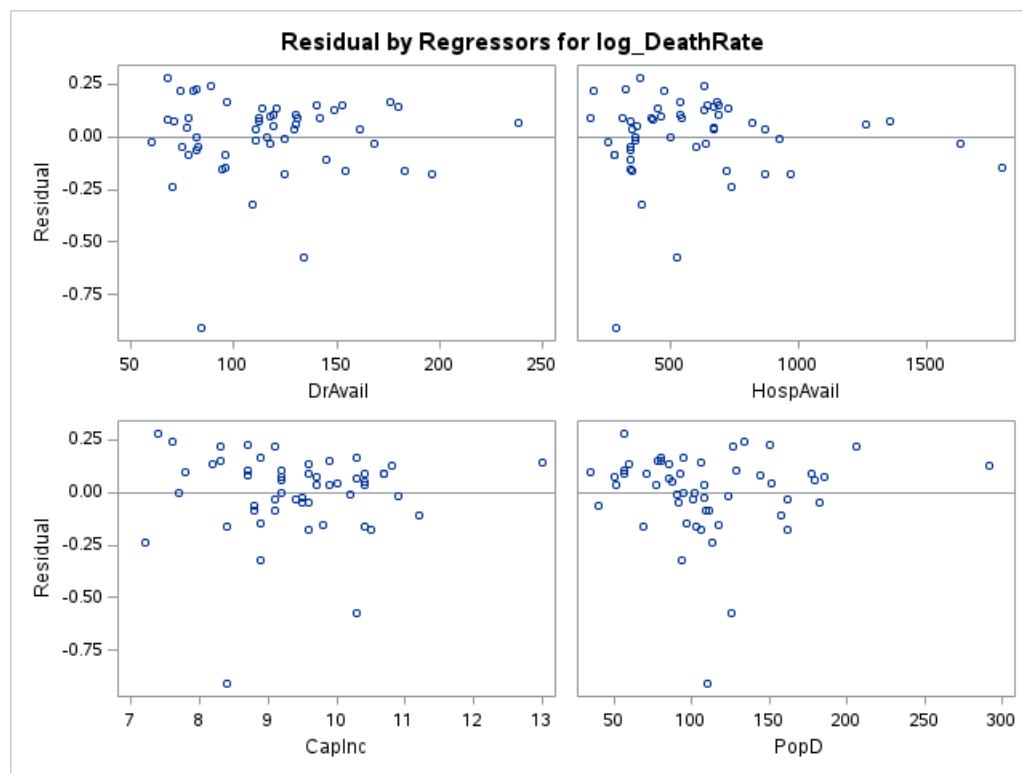
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.44680	0.26139	9.36	<.0001
DrAvail	1	0.00073420	0.00089715	0.82	0.4172
HospAvail	1	0.00008885	0.00009341	0.95	0.3463
CapInc	1	-0.02714	0.03035	-0.89	0.3756
PopD	1	-0.00106	0.00063231	-1.67	0.1014

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: log_DeathRate



2) Similar Regression Diagnostics to 1)



Residual Plot of death rate and predictors

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

Moments			
N	53	Sum Weights	53
Mean	-0.0271079	Sum Observations	-1.4367202
Std Deviation	1.13783681	Variance	1.2946726
Skewness	-3.0077994	Kurtosis	13.3072795
Uncorrected SS	67.3619216	Corrected SS	67.3229751
Coeff Variation	-4197.4318	Std Error Mean	0.15629391

Basic Statistical Measures			
Location		Variability	
Mean	-0.02711	Std Deviation	1.13784
Median	0.21760	Variance	1.29467
Mode	.	Range	7.32674
		Interquartile Range	0.82219

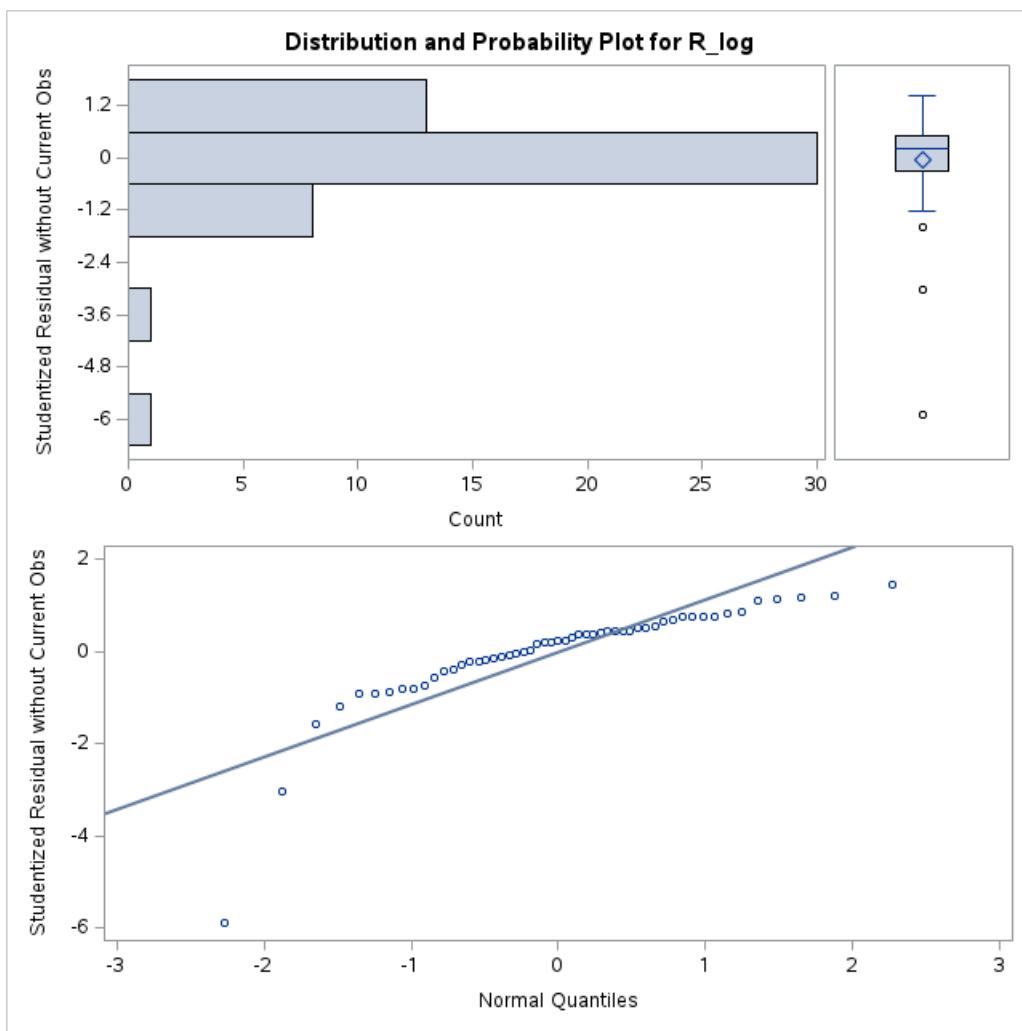
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	-0.17344	Pr > t	0.8630
Sign	M	4.5	Pr >= M	0.2717
Signed Rank	S	125.5	Pr >= S	0.2706

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.742269	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.16537	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.456002	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.850326	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	1.448054
99%	1.448054
95%	1.160342
90%	0.845110
75% Q3	0.525638

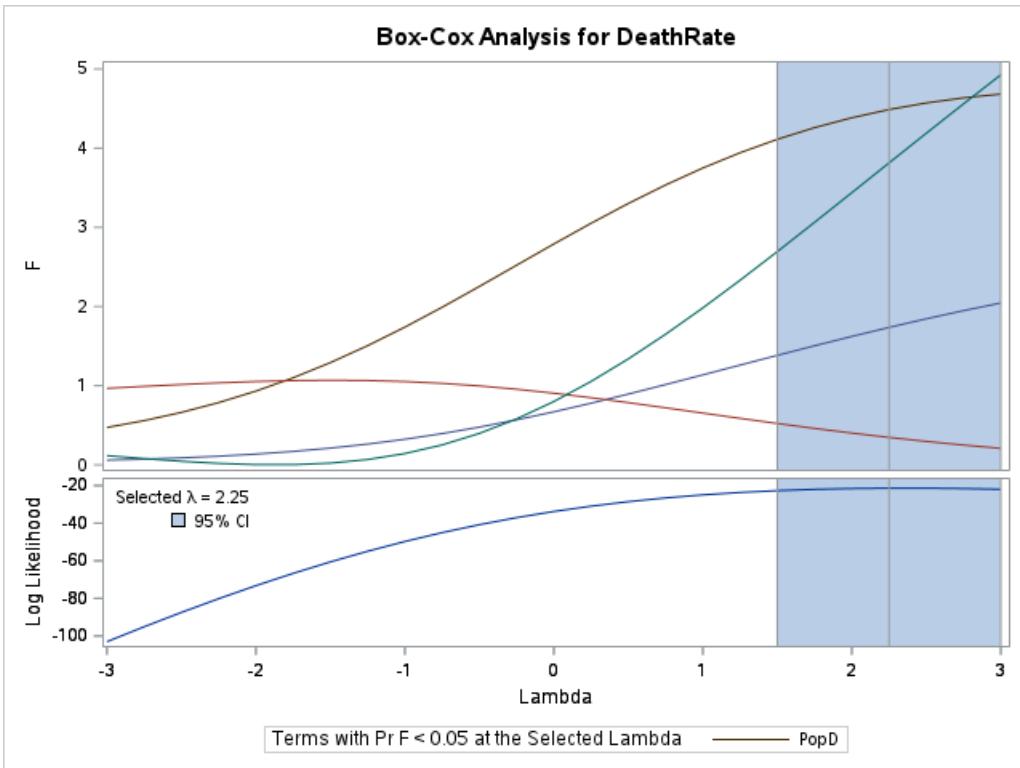
Quantiles (Definition 5)	
Level	Quantile
50% Median	0.217600
25% Q1	-0.296554
10%	-0.906523
5%	-1.586130
1%	-5.878681
0% Min	-5.878681

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-5.878681	43	1.08717	19
-3.028113	31	1.14363	9
-1.586130	53	1.16034	5
-1.205757	3	1.21137	10
-0.918546	38	1.44805	51



Residual Plot of death rate and predictors

The TRANSREG Procedure



2) Regression Diagnostics for transformed data

Residual Plot of death rate and predictors

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

Moments			
N	53	Sum Weights	53
Mean	0.00018785	Sum Observations	0.00995604
Std Deviation	1.03204041	Variance	1.0651074
Skewness	-0.4031479	Kurtosis	0.50944986
Uncorrected SS	55.3855867	Corrected SS	55.3855848
Coeff Variation	549396.669	Std Error Mean	0.14176165

Basic Statistical Measures			
Location		Variability	
Mean	0.000188	Std Deviation	1.03204
Median	0.157877	Variance	1.06511
Mode	.	Range	5.44818
		Interquartile Range	1.17410

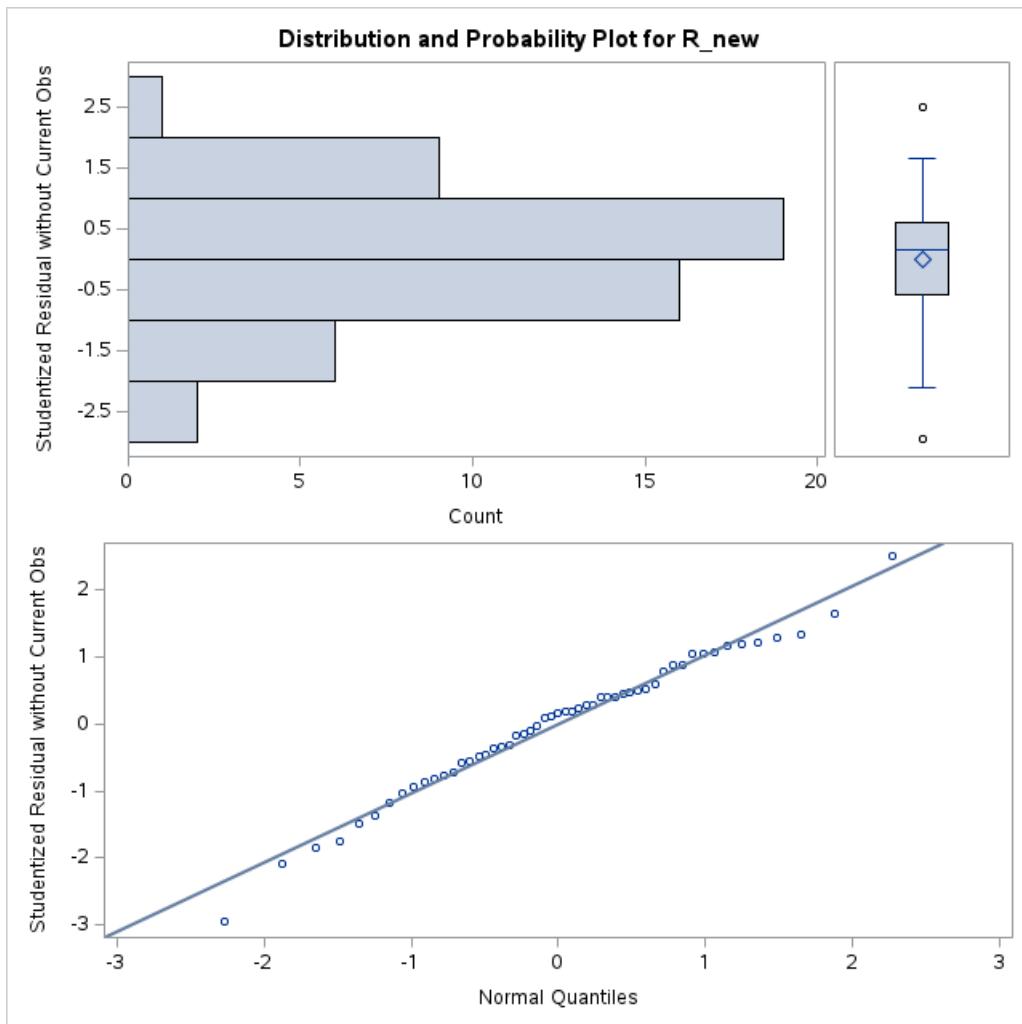
Tests for Location: Mu0=0				
Test	Statistic	p Value		
Student's t	t	0.001325	$\Pr > t $	0.9989
Sign	M	2.5	$\Pr \geq M $	0.5831
Signed Rank	S	38.5	$\Pr \geq S $	0.7367

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.984465	$\Pr < W$	0.7168
Kolmogorov-Smirnov	D	0.083261	$\Pr > D$	>0.1500
Cramer-von Mises	W-Sq	0.04655	$\Pr > W\text{-Sq}$	>0.2500
Anderson-Darling	A-Sq	0.303175	$\Pr > A\text{-Sq}$	>0.2500

Quantiles (Definition 5)	
Level	Quantile
100% Max	2.509145
99%	2.509145
95%	1.333165
90%	1.198785
75% Q3	0.595184
50% Median	0.157877

Quantiles (Definition 5)	
Level	Quantile
25% Q1	-0.578911
10%	-1.372759
5%	-1.841024
1%	-2.939035
0% Min	-2.939035

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-2.93904	43	1.21279	46
-2.09757	31	1.27869	9
-1.84102	53	1.33316	29
-1.74744	3	1.65240	10
-1.48784	22	2.50915	51



Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7857.09329	1964.27332	2.61	0.0470
Error	48	36135	752.81581		
Lack of Fit	48	36135	752.81581		
Pure Error	0	0	.		

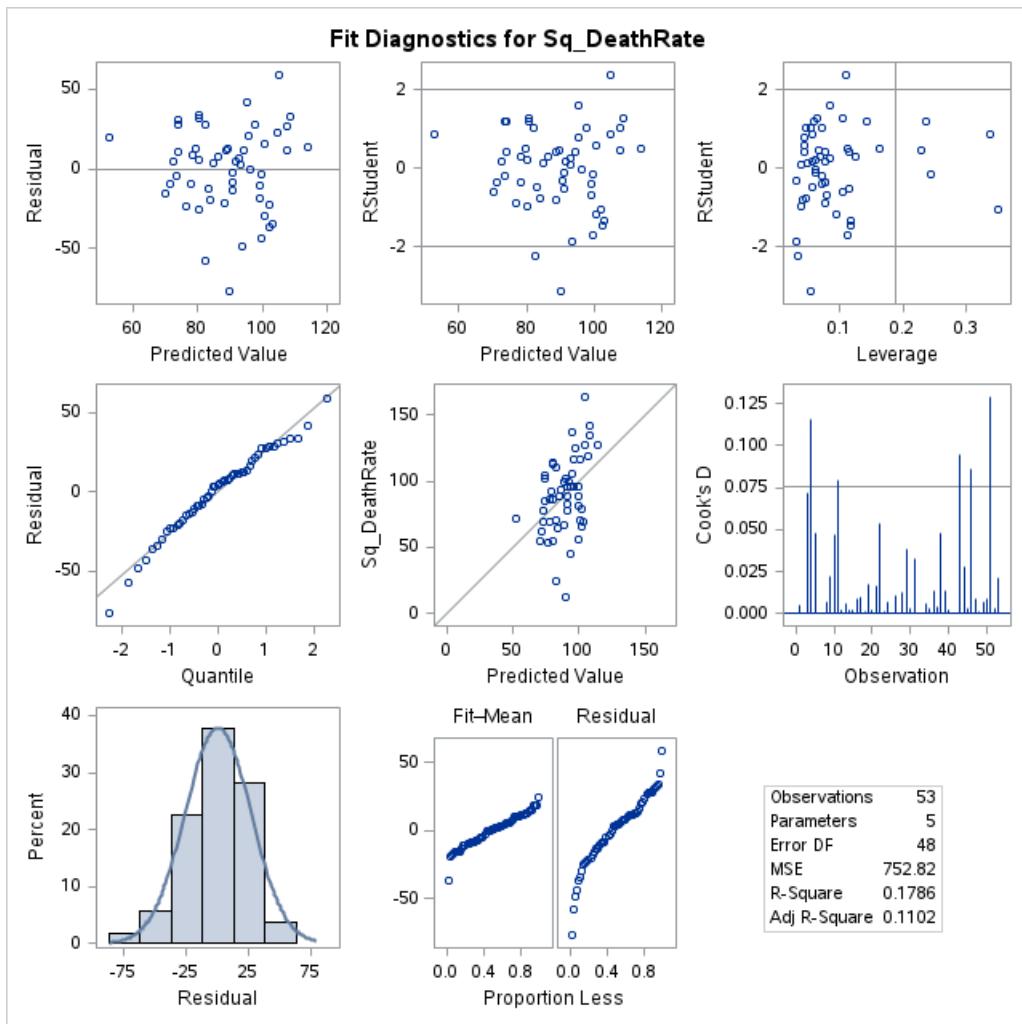
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Corrected Total	52	43992			

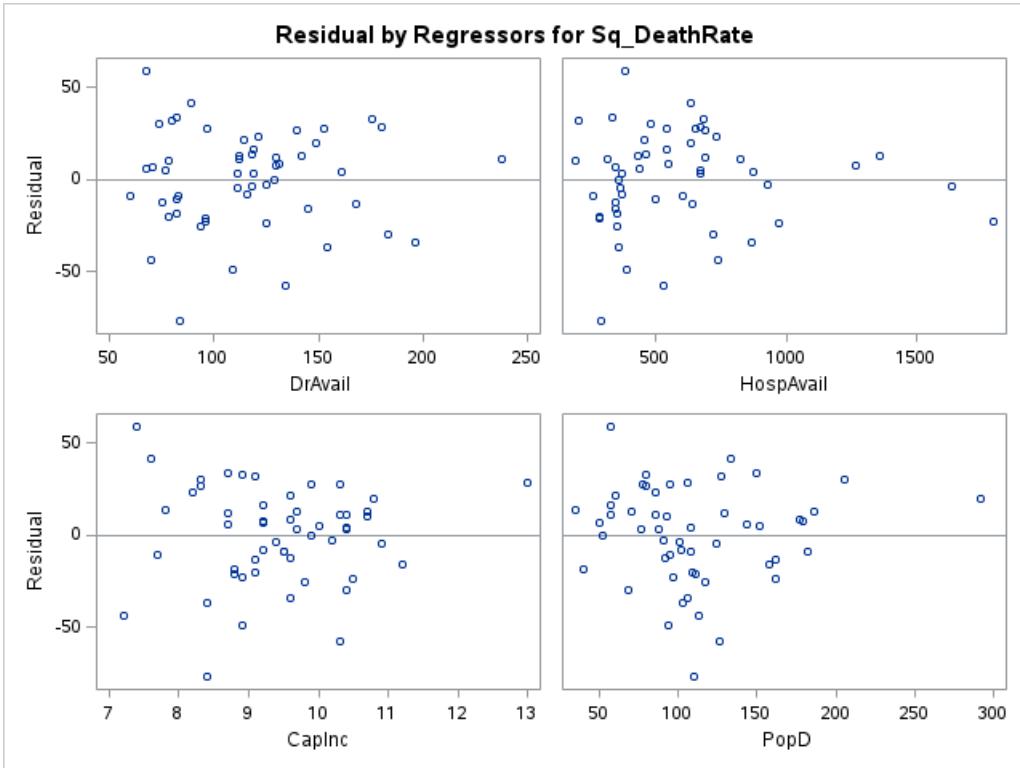
Root MSE	27.43749	R-Square	0.1786
Dependent Mean	89.30717	Adj R-Sq	0.1102
Coeff Var	30.72260		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	156.83437	34.61504	4.53	<.0001
DrAvail	1	0.15131	0.11881	1.27	0.2089
HospAvail	1	0.00784	0.01237	0.63	0.5295
CapInc	1	-7.45199	4.01902	-1.85	0.0699
PopD	1	-0.17533	0.08374	-2.09	0.0416

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

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

Moments			
N	53	Sum Weights	53
Mean	-0.0013466	Sum Observations	-0.0713675
Std Deviation	1.03358498	Variance	1.06829792
Skewness	-0.5625504	Kurtosis	0.79520839
Uncorrected SS	55.5515878	Corrected SS	55.5514917
Coeff Variation	-76757.648	Std Error Mean	0.14197382

Basic Statistical Measures			
Location		Variability	
Mean	-0.00135	Std Deviation	1.03358
Median	0.16581	Variance	1.06830
Mode	.	Range	5.52078
		Interquartile Range	1.19181

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	-0.00948	Pr > t	0.9925
Sign	M	2.5	Pr >= M	0.5831
Signed Rank	S	42.5	Pr >= S	0.7105

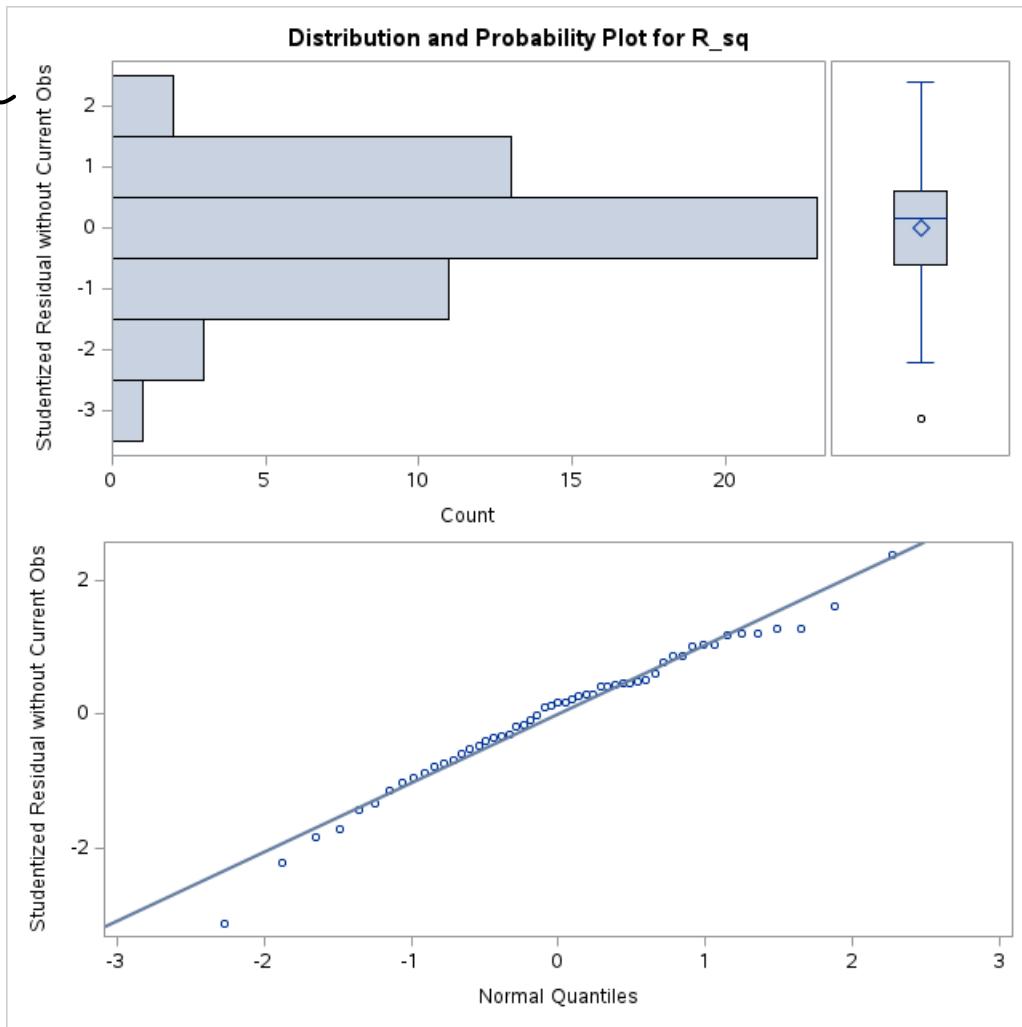
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.977976	Pr < W	0.4310
Kolmogorov-Smirnov	D	0.08953	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.057374	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.368832	Pr > A-Sq	>0.2500

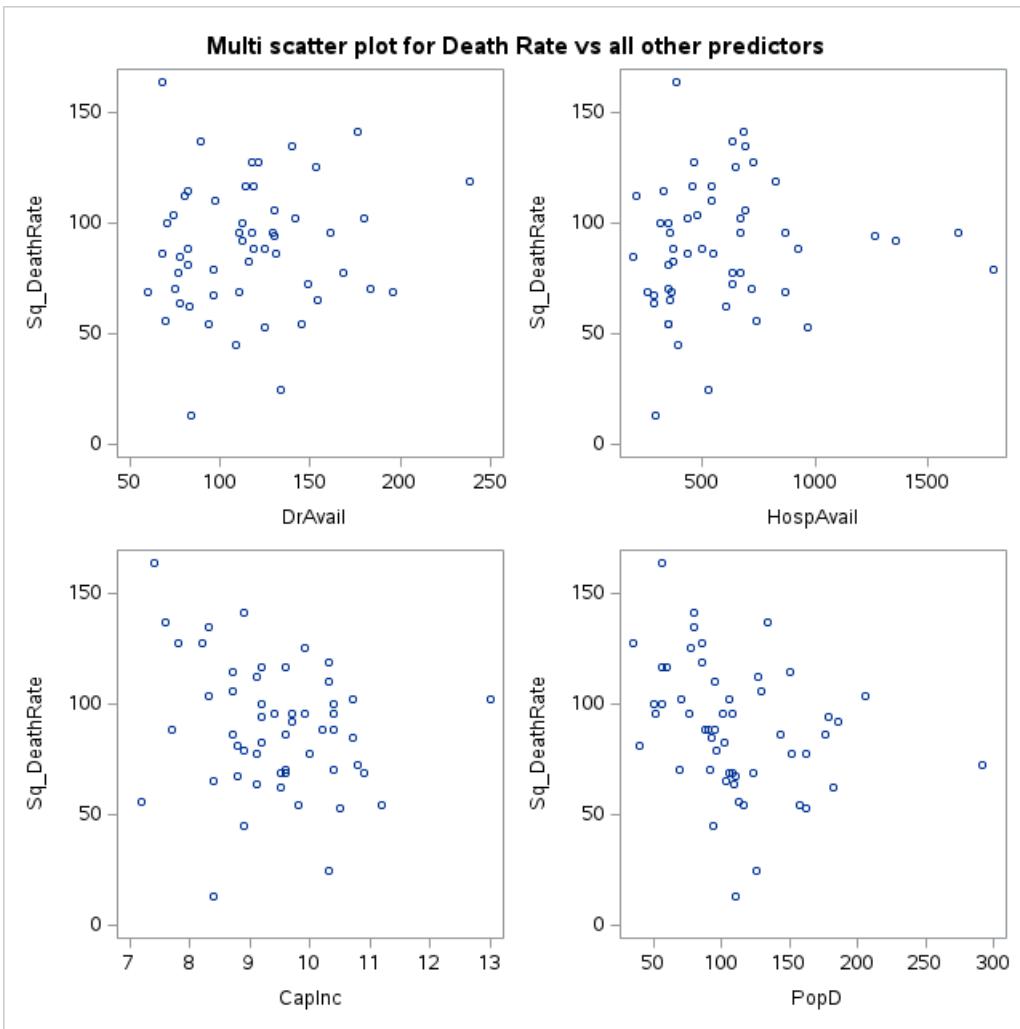
Quantiles (Definition 5)	
Level	Quantile
100% Max	2.383029
99%	2.383029
95%	1.286392
90%	1.196158
75% Q3	0.596876

Quantiles (Definition 5)	
Level	Quantile
50% Median	0.165810
25% Q1	-0.594934
10%	-1.341343
5%	-1.847724
1%	-3.137755
0% Min	-3.137755

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-3.13776	43	1.20321	19
-2.21341	31	1.28162	9
-1.84772	53	1.28639	29
-1.71709	3	1.61739	10
-1.43675	22	2.38303	51

qq Plot
of transformed
Var





Multi scatter plot for Death Rate vs all other predictors

The CORR Procedure

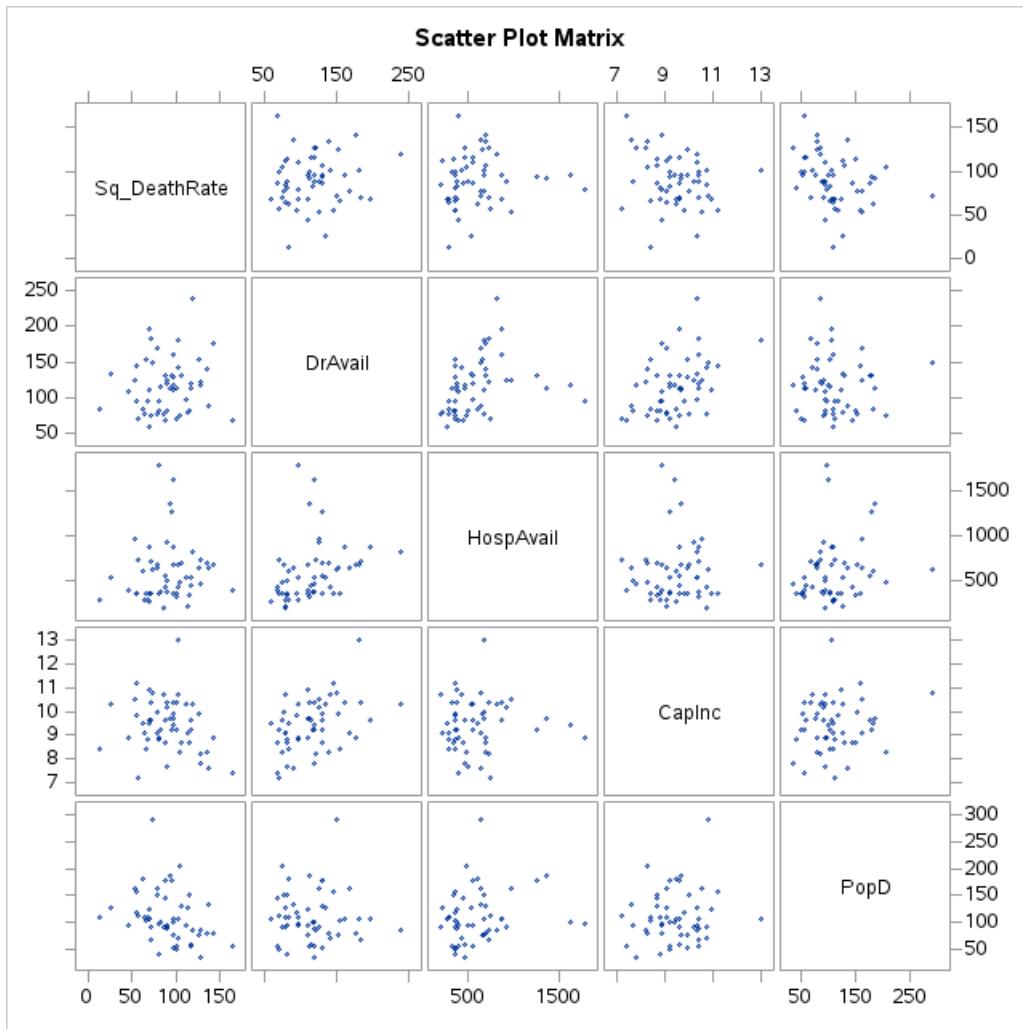
5 Variables:	Sq_DeathRate DrAvail HospAvail CapInc PopD
--------------	--

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Sq_DeathRate	53	89.30717	29.08616	4733	12.96000	163.84000
DrAvail	53	116.09434	37.88660	6153	60.00000	238.00000
HospAvail	53	589.79245	332.61831	31259	190.00000	1792
CapInc	53	9.43585	1.07544	500.10000	7.20000	13.00000
PopD	53	110.64151	47.17973	5864	35.00000	292.00000

Pearson Correlation Coefficients, N = 53

Prob > |r| under H0: Rho=0

	Sq_DeathRate	DrAvail	HospAvail	CapInc	PopD
Sq_DeathRate	1.00000	0.10987 0.4335	0.08721 0.5346	-0.22428 0.1064	-0.30708 0.0253
DrAvail	0.10987 0.4335	1.00000	0.29563 0.0316	0.43329 0.0012	-0.01994 0.8873
HospAvail	0.08721 0.5346	0.29563 0.0316	1.00000	0.02750 0.8450	0.18662 0.1809
CapInc	-0.22428 0.1064	0.43329 0.0012	0.02750 0.8450	1.00000	0.12874 0.3582
PopD	-0.30708 0.0253	-0.01994 0.8873	0.18662 0.1809	0.12874 0.3582	1.00000



LOF Test for Response²

Multi scatter plot for Death Rate vs all other predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7857.09329	1964.27332	2.61	0.0470
Error	48	36135	752.81581		
Lack of Fit	48	36135	752.81581		
Pure Error	0	0			
Corrected Total	52	43992			

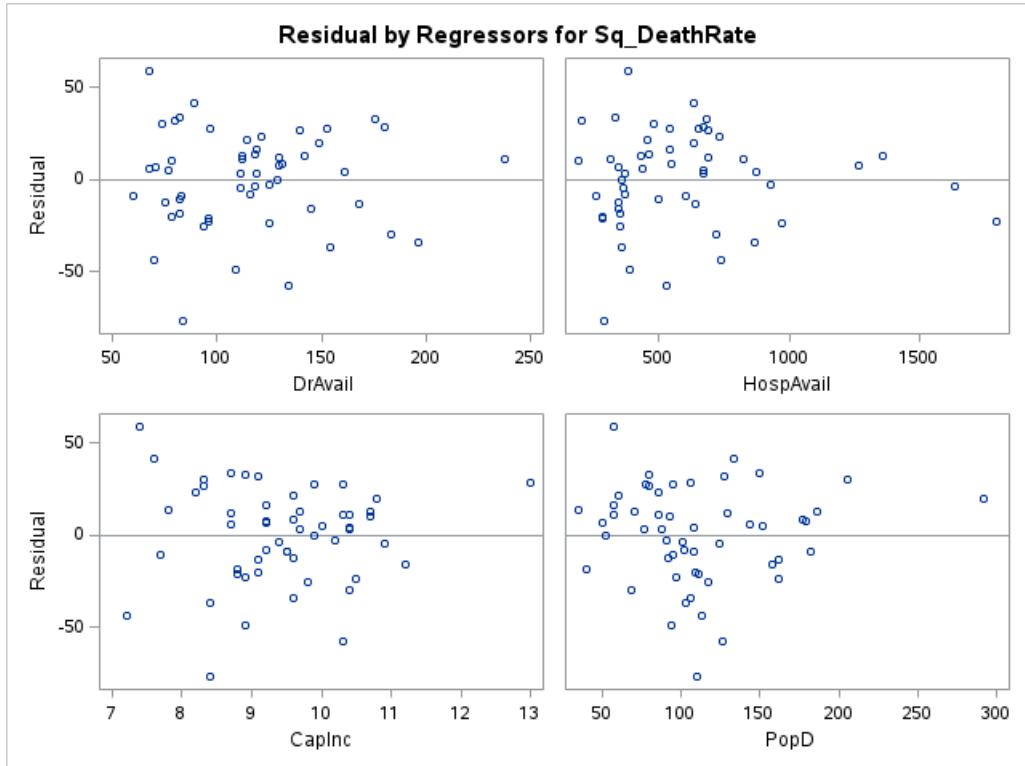
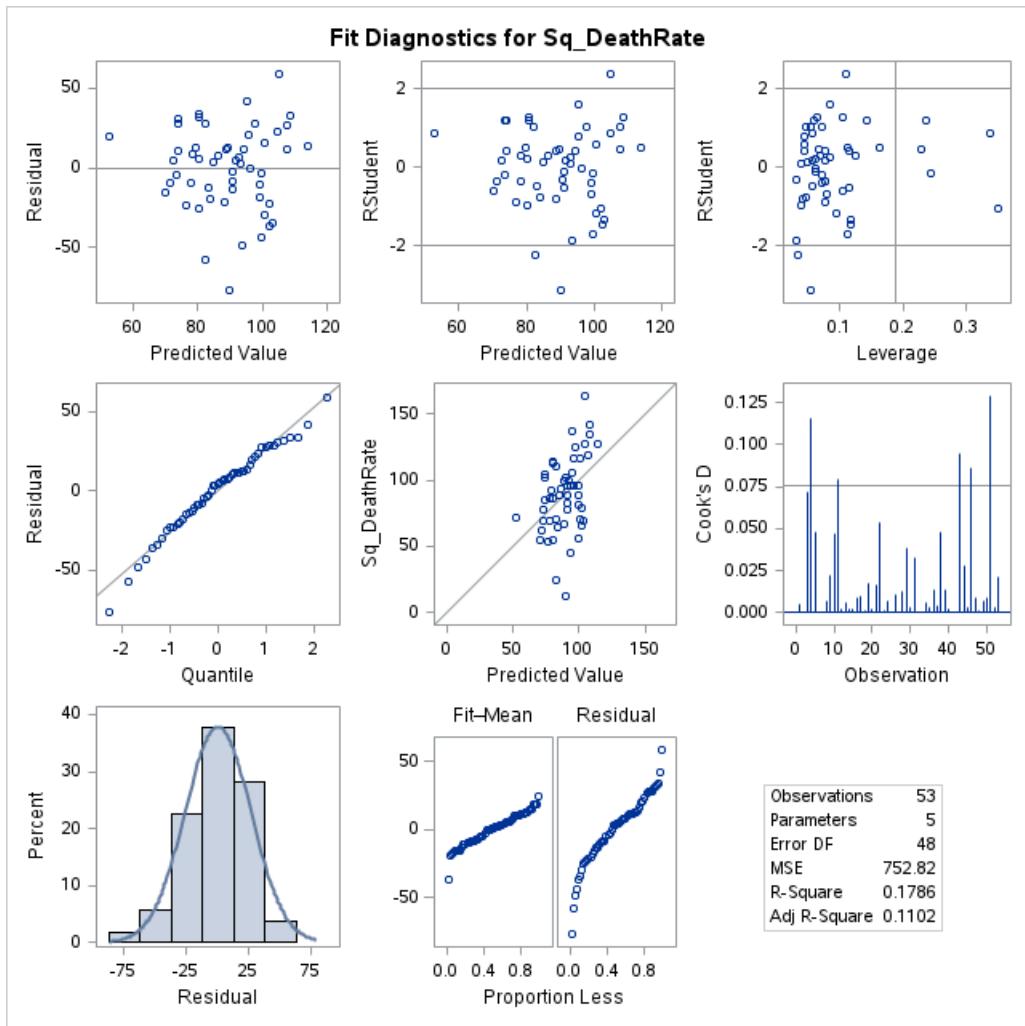
Root MSE	27.43749	R-Square	0.1786
Dependent Mean	89.30717	Adj R-Sq	0.1102
Coeff Var	30.72260		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	156.83437	34.61504	4.53	<.0001
DrAvail	1	0.15131	0.11881	1.27	0.2089
HospAvail	1	0.00784	0.01237	0.63	0.5295
CapInc	1	-7.45199	4.01902	-1.85	0.0699
PopD	1	-0.17533	0.08374	-2.09	0.0416

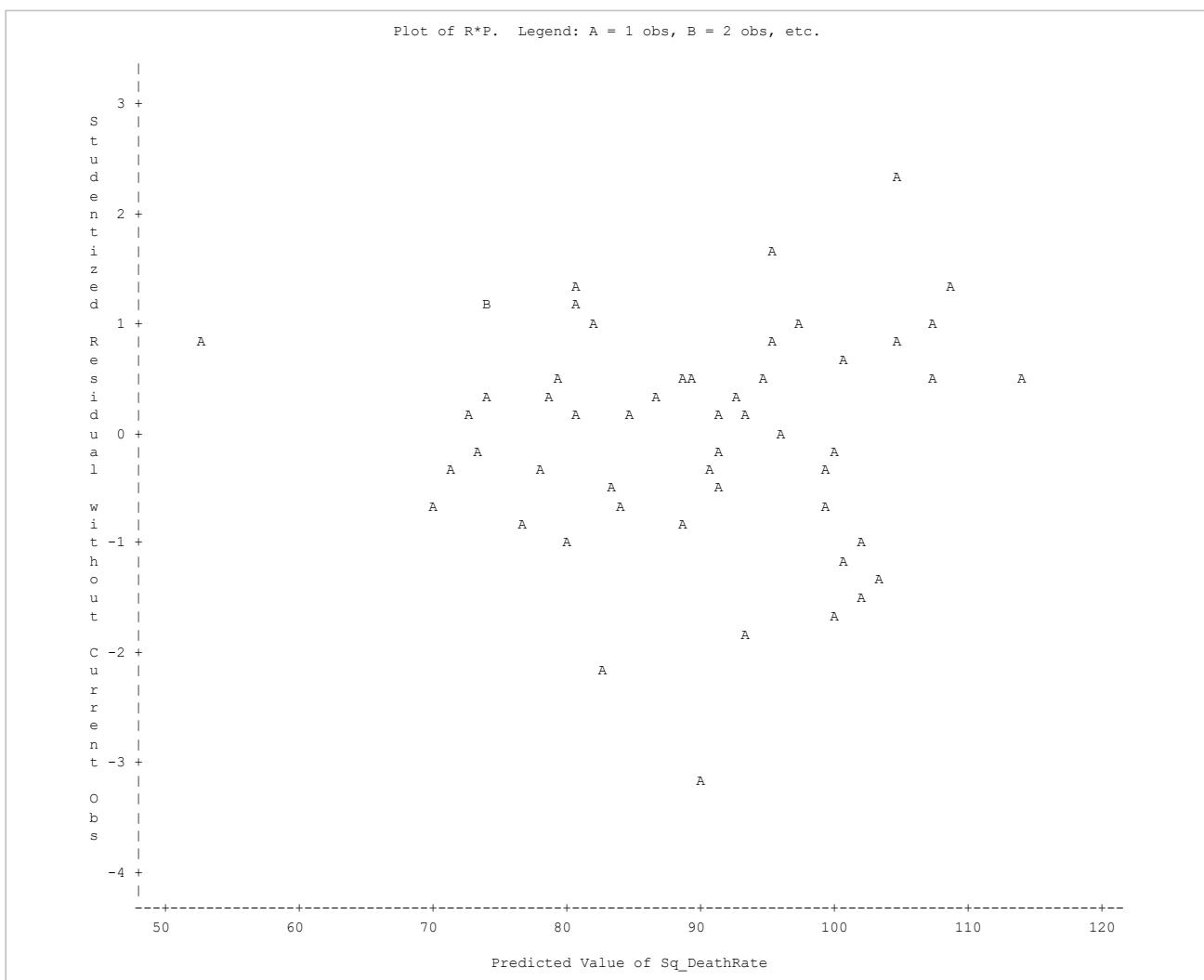
Multi scatter plot for Death Rate vs all other predictors

The REG Procedure

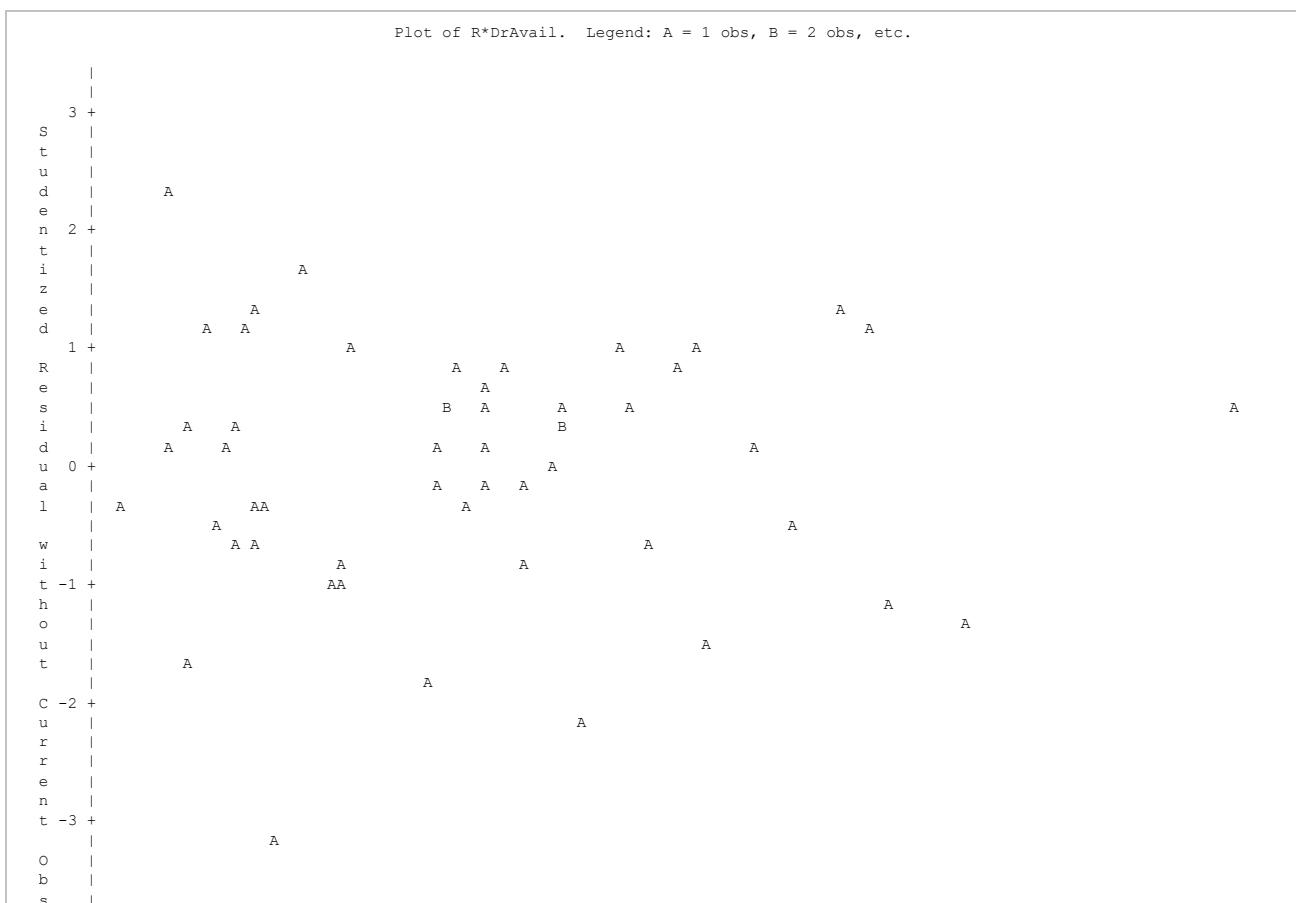
Model: MODEL1
Dependent Variable: Sq_DeathRate

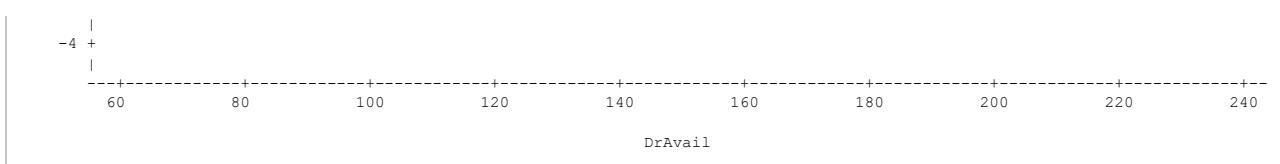


Multi scatter plot for Death Rate vs all other predictors

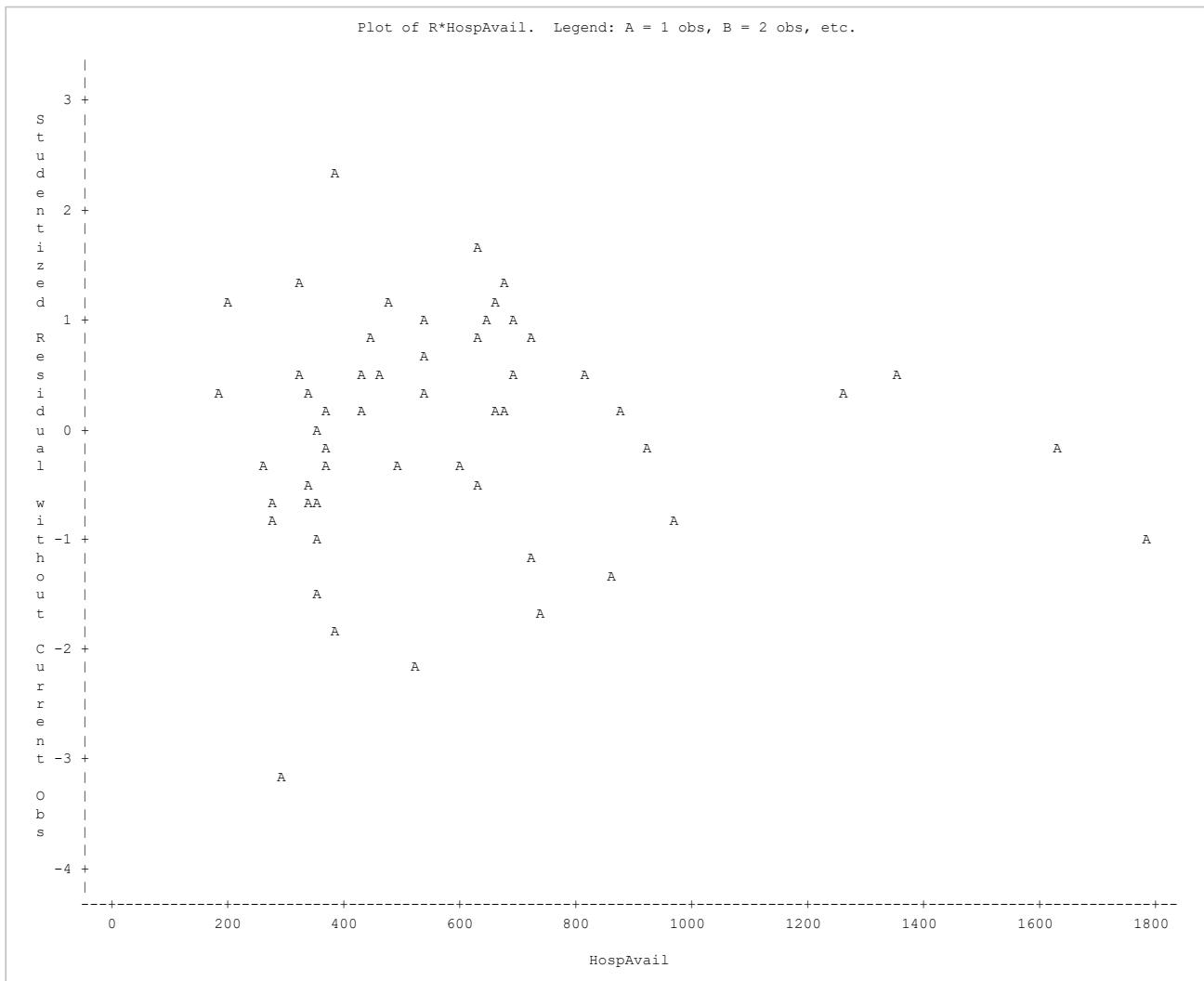


Multi scatter plot for Death Rate vs all other predictors

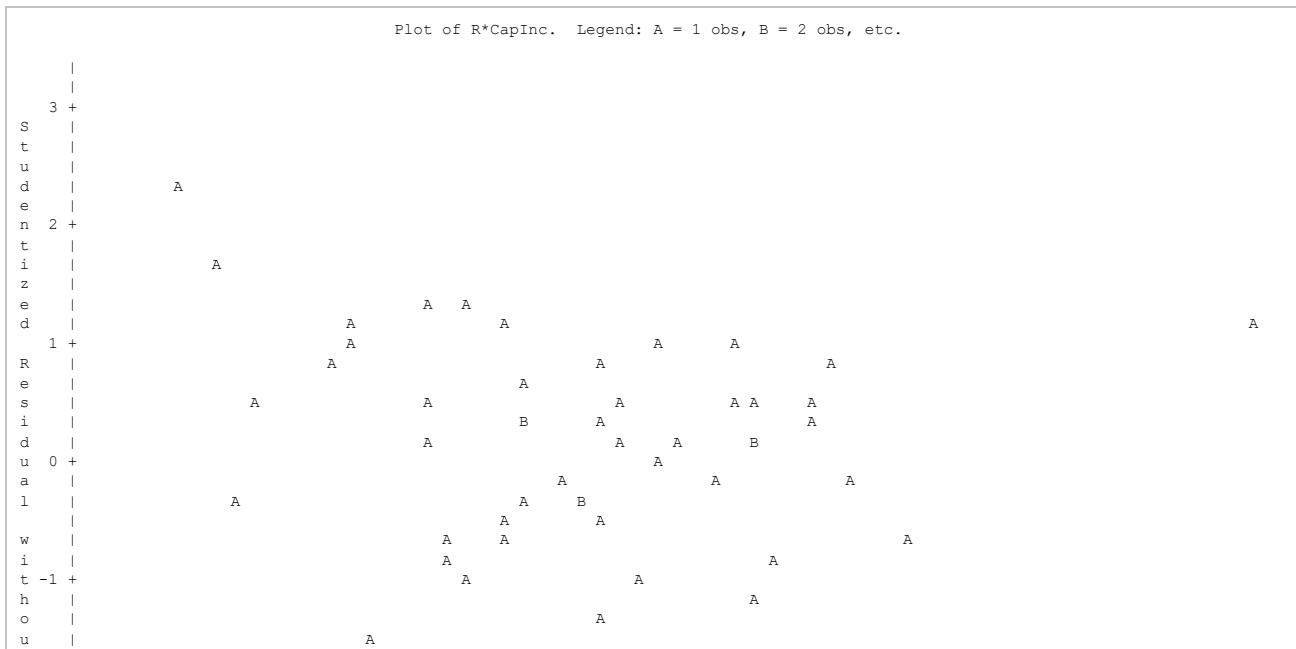


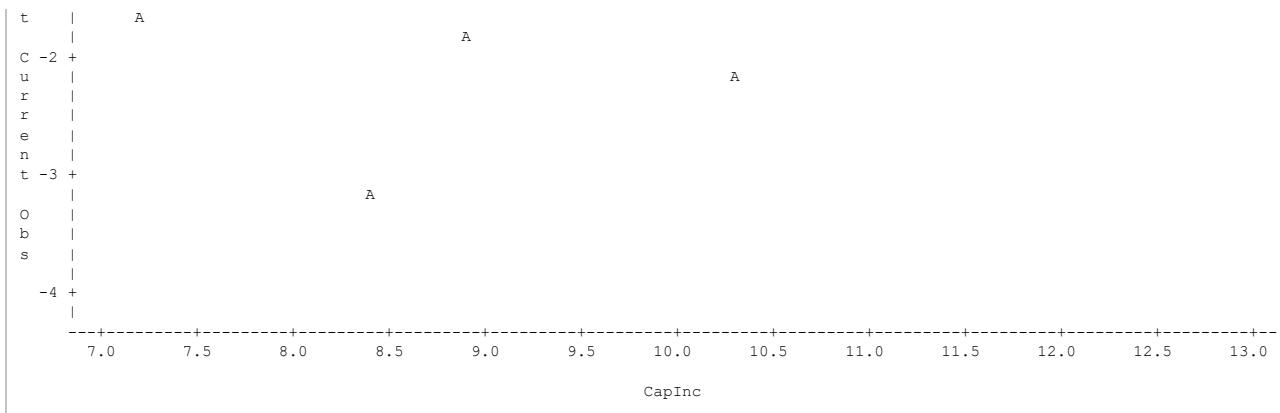


Multi scatter plot for Death Rate vs all other predictors

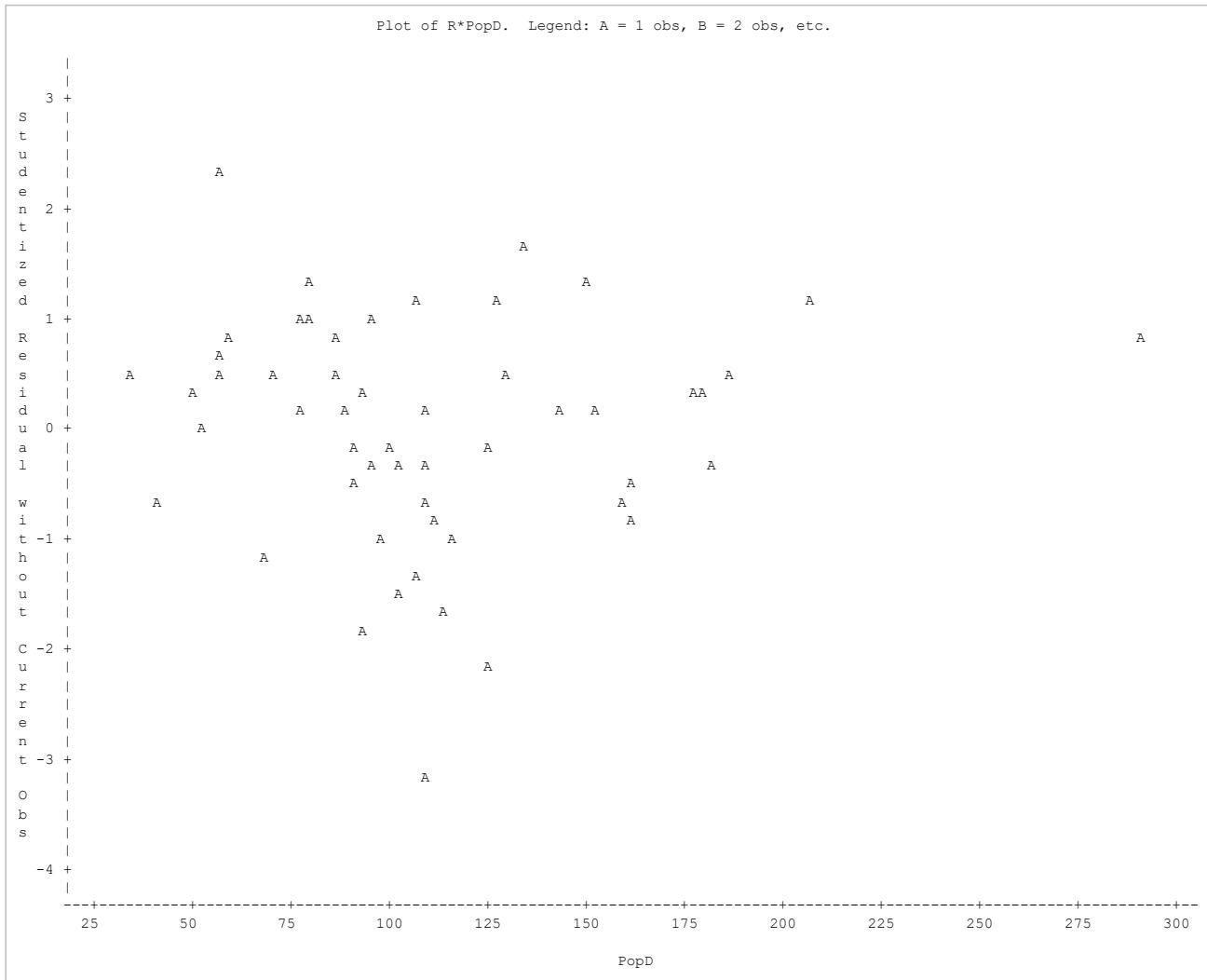


Multi scatter plot for Death Rate vs all other predictors

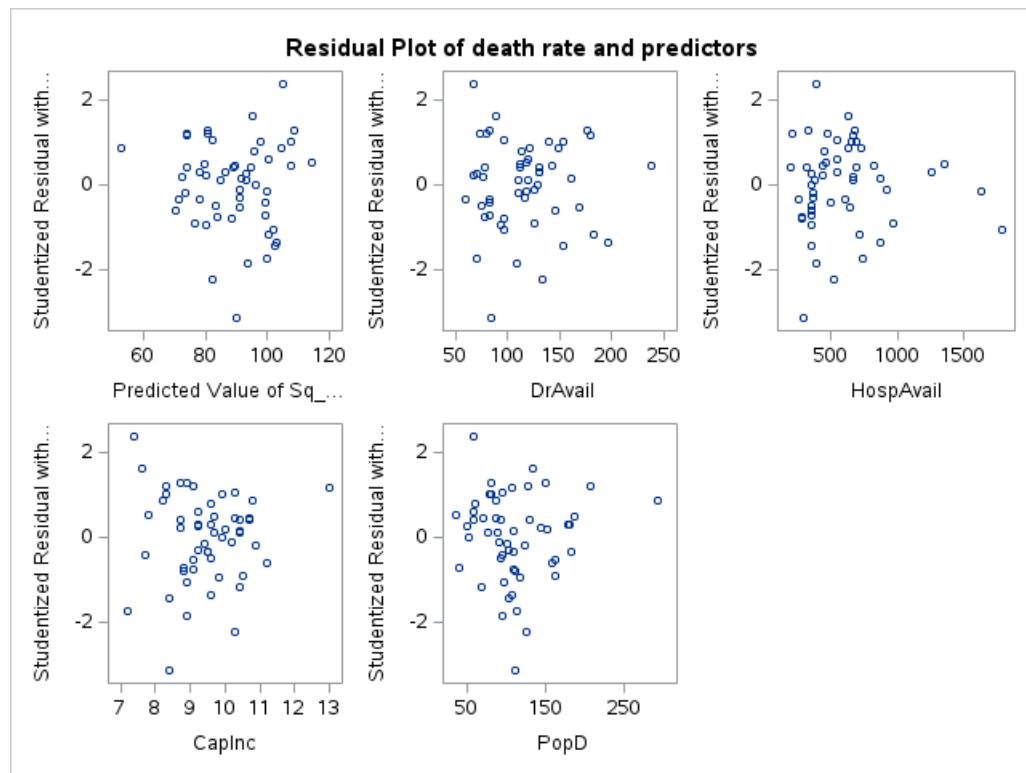




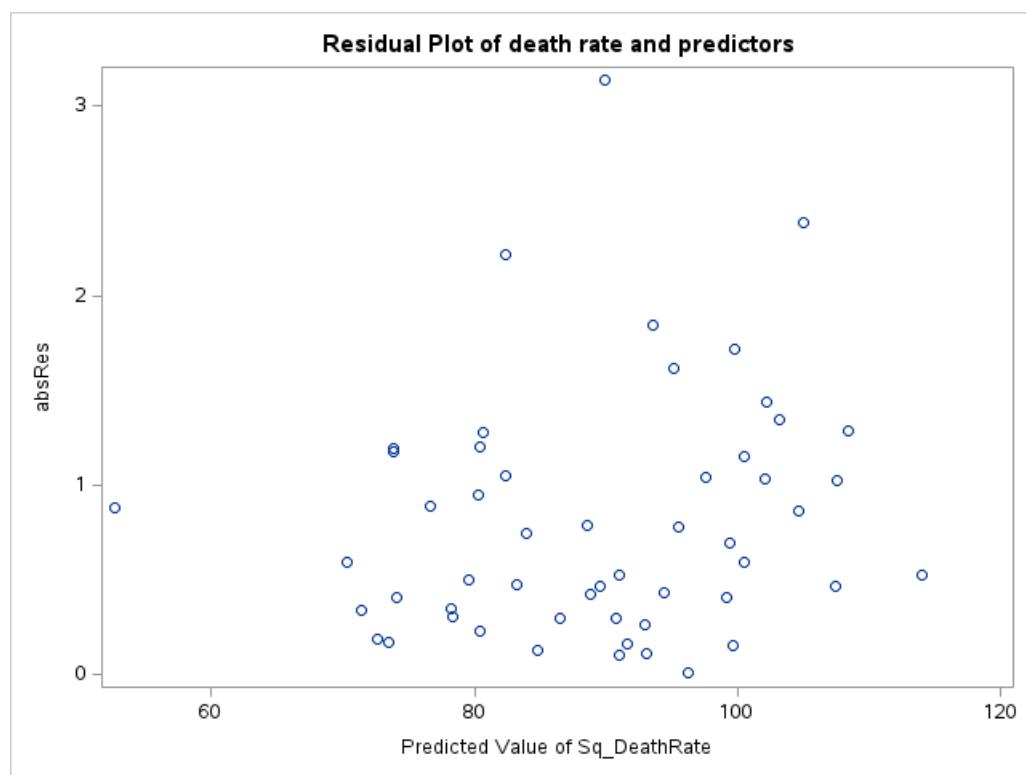
Multi scatter plot for Death Rate vs all other predictors



Residual PLOTS



Absolute
Residual
plot



Breush-Pagan Test

Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	Sq_DeathRate
Parameters	b0 b1 b2 b3 b4
Equations	Sq_DeathRate

The Equation to Estimate is

The Equation to Estimate is	
Sq_DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 1 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	1

Final Convergence Criteria	
R	0
PPC	0
RPC(b0)	1552815
Object	0.922443
Trace(S)	752.8158
Objective Value	681.7955

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
Sq_DeathRate	5	48	36135.2	752.8	27.4375	0.1786	0.1102

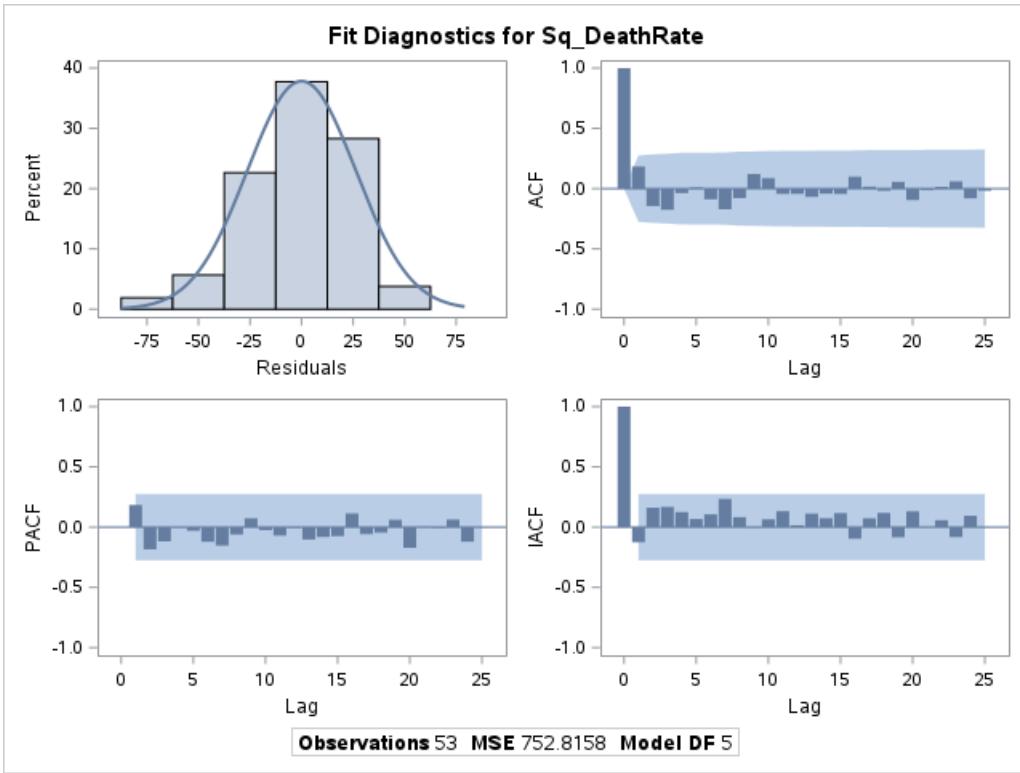
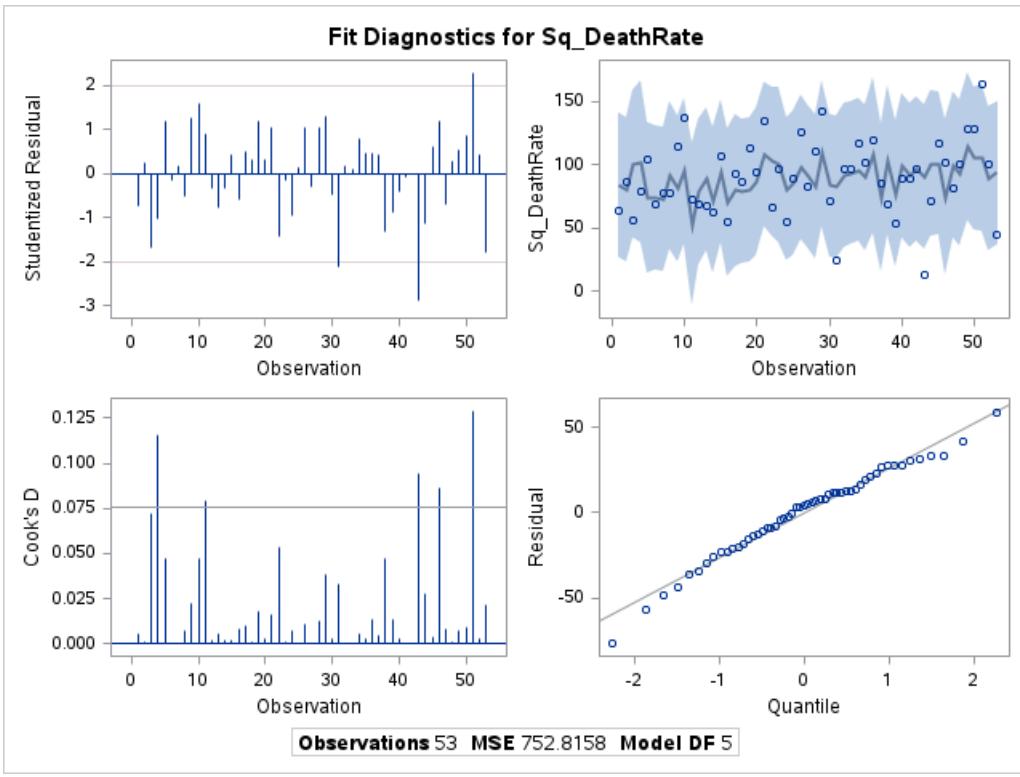
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	156.8344	34.6150	4.53	<.0001	
b1	0.151314	0.1188	1.27	0.2089	
b2	0.007835	0.0124	0.63	0.5295	
b3	-7.45199	4.0190	-1.85	0.0699	
b4	-0.17533	0.0837	-2.09	0.0416	

Number of Observations		Statistics for System	
Used	53	Objective	681.7955
Missing	0	Objective*N	36135

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Sq_DeathRate	White's Test	16.04	14	0.3112	Cross of all vars
	Breusch-Pagan	7.06	4	0.1329	DrAvail, HospAvail, CapInc, PopD, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	Sq_DeathRate
Parameters(Value(t Value))	b0(156.83437099(4.530816474)) b1(0.1513135787(1.2736037289)) b2(0.0078351104(0.6334005813)) b3(-7.451989641(-1.854179156)) b4(-0.175332844(-2.093897858))
Equations	Sq_DeathRate

The Equation to Estimate is

The Equation to Estimate is	
Sq_DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	752.8158
Objective Value	681.7955

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
Sq_DeathRate	5	48	36135.2	752.8	27.4375	0.1786	0.1102

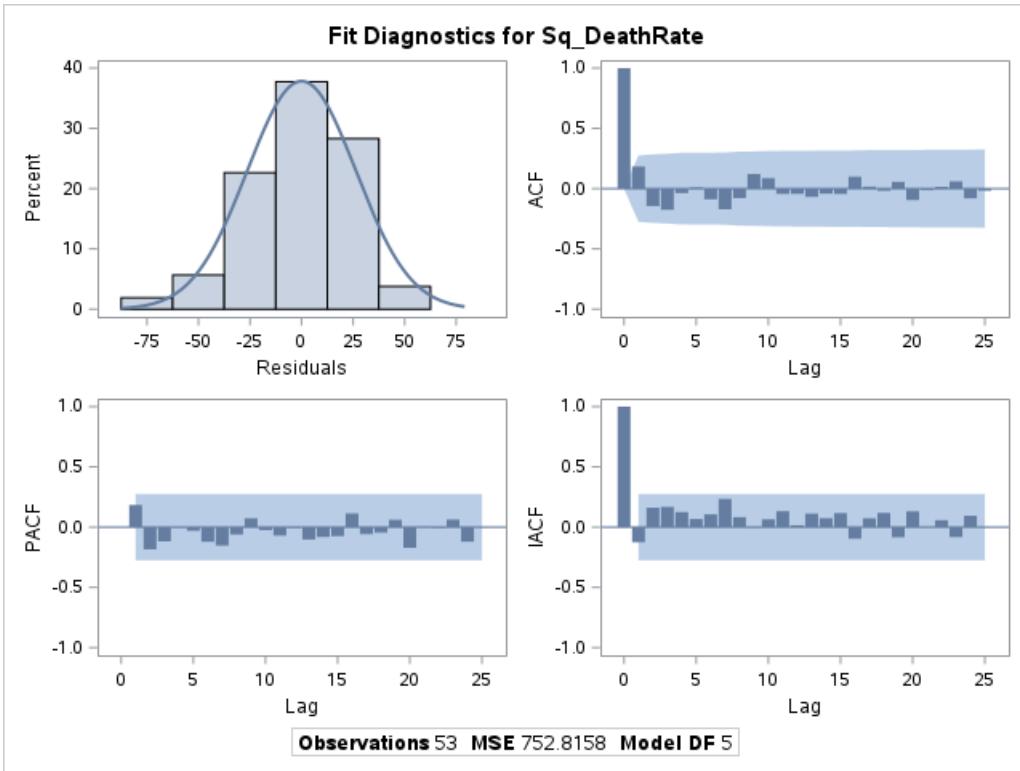
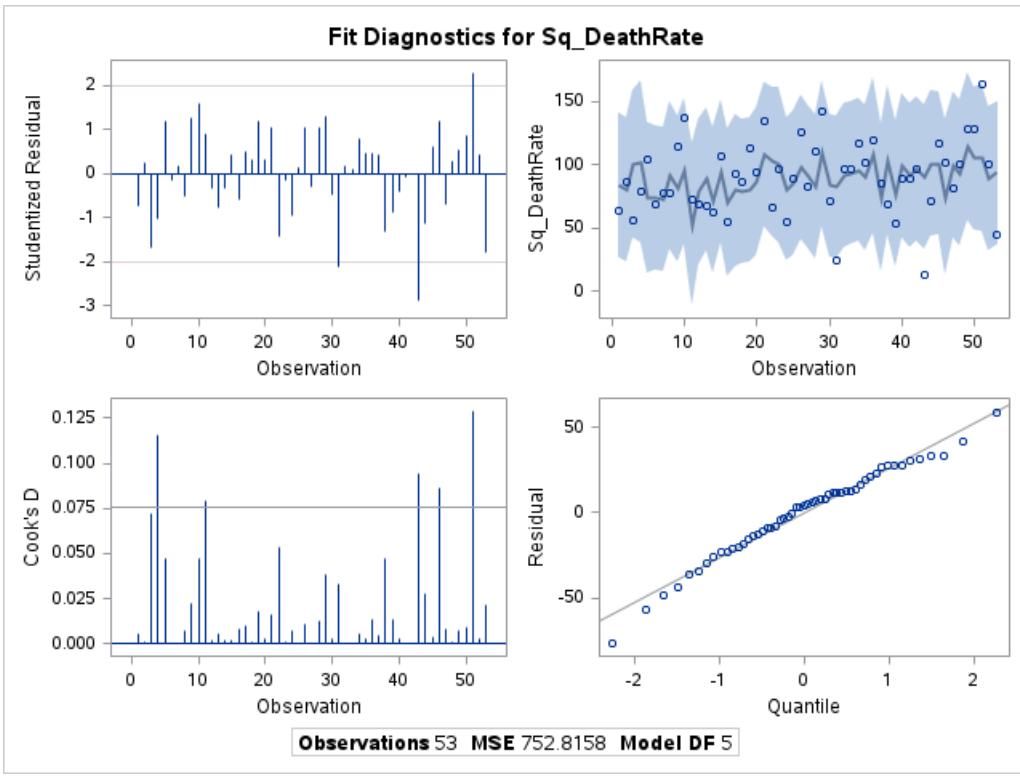
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	156.8344	34.6150	4.53	<.0001	
b1	0.151314	0.1188	1.27	0.2089	
b2	0.007835	0.0124	0.63	0.5295	
b3	-7.45199	4.0190	-1.85	0.0699	
b4	-0.17533	0.0837	-2.09	0.0416	

Number of Observations		Statistics for System	
Used	53	Objective	681.7955
Missing	0	Objective*N	36135

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Sq_DeathRate	Breusch-Pagan	0.69	1	0.4073	DrAvail, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	Sq_DeathRate
Parameters(Value(t Value))	b0(156.83437099(4.530816474)) b1(0.1513135787(1.2736037289)) b2(0.0078351104(0.6334005813)) b3(-7.451989641(-1.854179156)) b4(-0.175332844(-2.093897858))
Equations	Sq_DeathRate

The Equation to Estimate is

The Equation to Estimate is	
Sq_DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	752.8158
Objective Value	681.7955

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
Sq_DeathRate	5	48	36135.2	752.8	27.4375	0.1786	0.1102

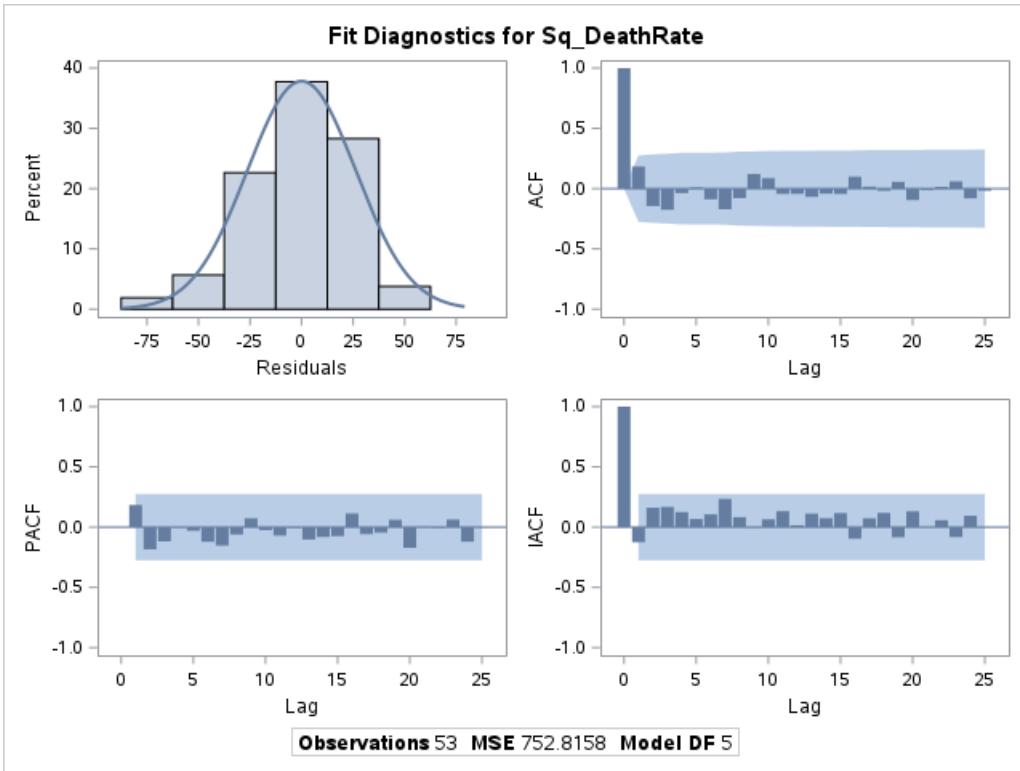
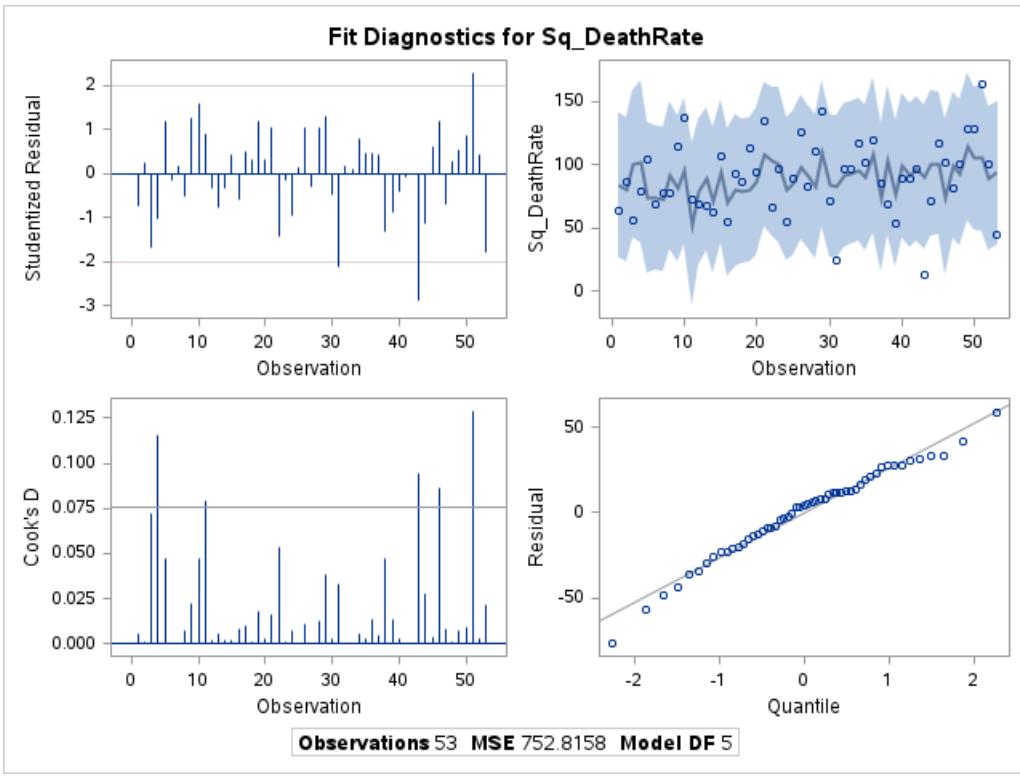
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	156.8344	34.6150	4.53	<.0001	
b1	0.151314	0.1188	1.27	0.2089	
b2	0.007835	0.0124	0.63	0.5295	
b3	-7.45199	4.0190	-1.85	0.0699	
b4	-0.17533	0.0837	-2.09	0.0416	

Number of Observations		Statistics for System	
Used	53	Objective	681.7955
Missing	0	Objective*N	36135

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Sq_DeathRate	Breusch-Pagan	1.39	1	0.2376	HospAvail, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	Sq_DeathRate
Parameters(Value(t Value))	b0(156.83437099(4.530816474)) b1(0.1513135787(1.2736037289)) b2(0.0078351104(0.6334005813)) b3(-7.451989641(-1.854179156)) b4(-0.175332844(-2.093897858))
Equations	Sq_DeathRate

The Equation to Estimate is

The Equation to Estimate is	
Sq_DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	752.8158
Objective Value	681.7955

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
Sq_DeathRate	5	48	36135.2	752.8	27.4375	0.1786	0.1102

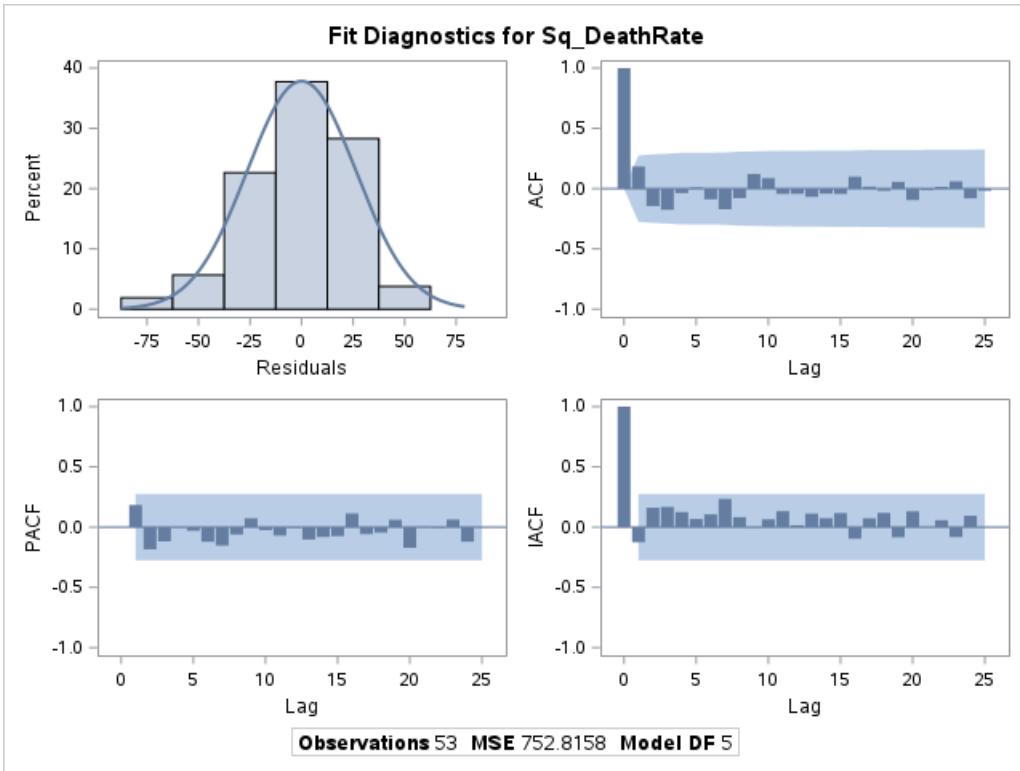
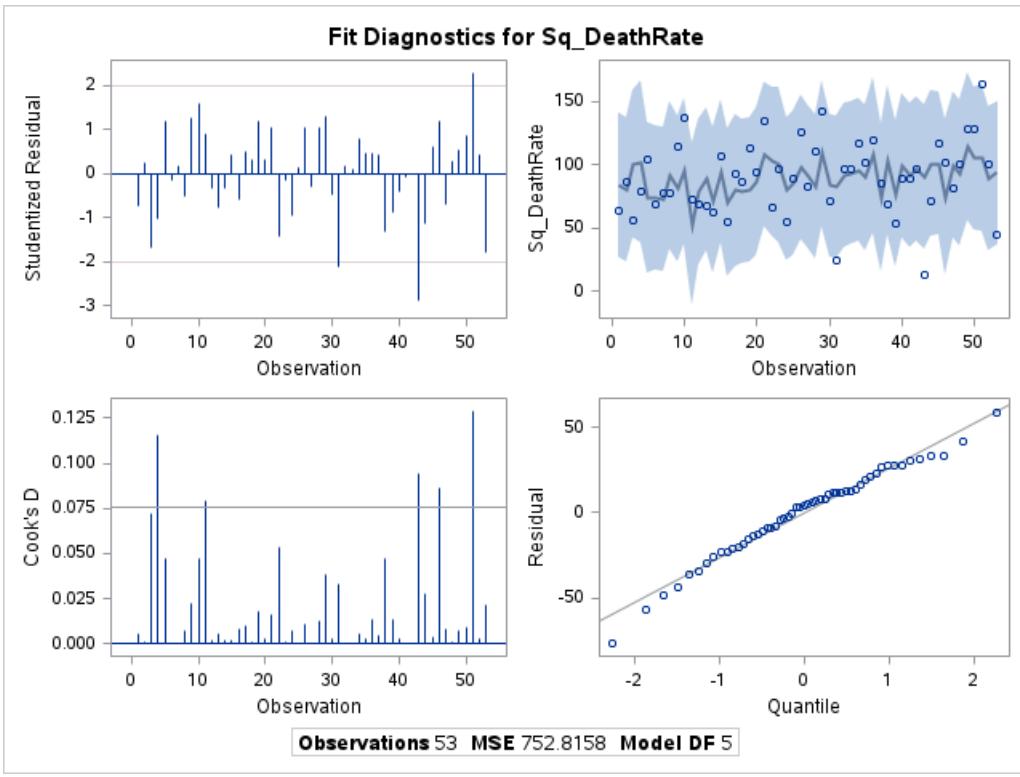
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	156.8344	34.6150	4.53	<.0001	
b1	0.151314	0.1188	1.27	0.2089	
b2	0.007835	0.0124	0.63	0.5295	
b3	-7.45199	4.0190	-1.85	0.0699	
b4	-0.17533	0.0837	-2.09	0.0416	

Number of Observations		Statistics for System	
Used	53	Objective	681.7955
Missing	0	Objective*N	36135

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Sq_DeathRate	Breusch-Pagan	5.38	1	0.0204	CapInc, 1

Residual Plot of death rate and predictors

The MODEL Procedure



Residual Plot of death rate and predictors

The MODEL Procedure

Model Summary	
Model Variables	1
Parameters	5
Equations	1
Number of Statements	1

Model Variables	Sq_DeathRate
Parameters(Value(t Value))	b0(156.83437099(4.530816474)) b1(0.1513135787(1.2736037289)) b2(0.0078351104(0.6334005813)) b3(-7.451989641(-1.854179156)) b4(-0.175332844(-2.093897858))
Equations	Sq_DeathRate

The Equation to Estimate is

The Equation to Estimate is	
Sq_DeathRate =	F(b0(1), b1(DrAvail), b2(HospAvail), b3(CapInc), b4(PopD))

NOTE: At OLS Iteration 0 CONVERGE=0.001 Criteria Met.

Residual Plot of death rate and predictors

The MODEL Procedure OLS Estimation Summary

Data Set Options	
DATA=	D

Minimization Summary	
Parameters Estimated	5
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	0
PPC	0
RPC	.
Object	.
Trace(S)	752.8158
Objective Value	681.7955

Observations Processed	
Read	53
Solved	53

Residual Plot of death rate and predictors

The MODEL Procedure

Nonlinear OLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
Sq_DeathRate	5	48	36135.2	752.8	27.4375	0.1786	0.1102

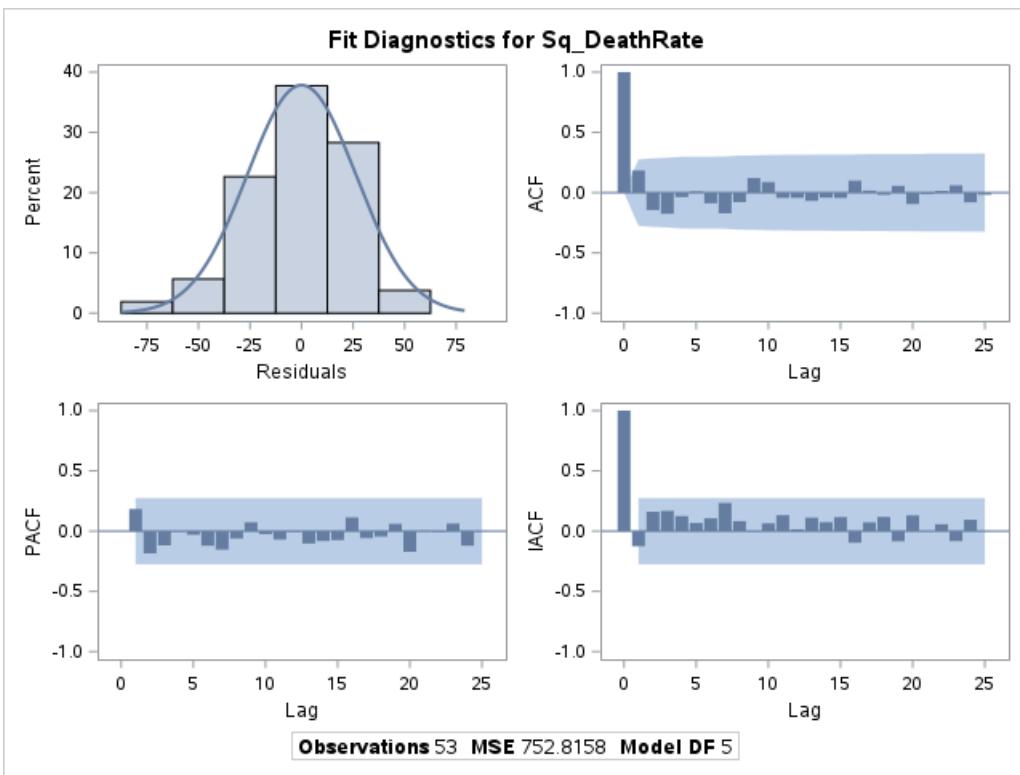
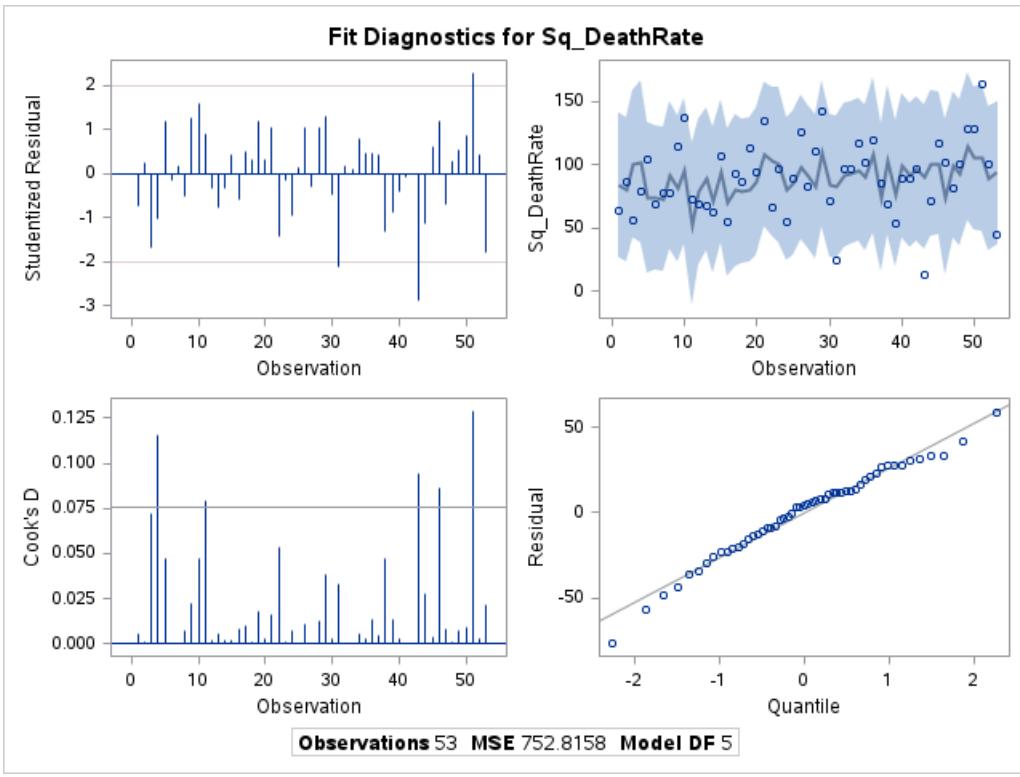
Nonlinear OLS Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > t	
b0	156.8344	34.6150	4.53	<.0001	
b1	0.151314	0.1188	1.27	0.2089	
b2	0.007835	0.0124	0.63	0.5295	
b3	-7.45199	4.0190	-1.85	0.0699	
b4	-0.17533	0.0837	-2.09	0.0416	

Number of Observations		Statistics for System	
Used	53	Objective	681.7955
Missing	0	Objective*N	36135

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Sq_DeathRate	Breusch-Pagan	0.08	1	0.7727	PopD, 1

Residual Plot of death rate and predictors

The MODEL Procedure



3)

Residual Plot of death rate and predictors

The GLM Procedure

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

Dependent Variable: Sq_DeathRate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7857.09329	1964.27332	2.61	0.0470
Error	48	36135.15894	752.81581		

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Corrected Total	52	43992.25224			

R-Square	Coeff Var	Root MSE	Sq_DeathRate Mean
0.178602	30.72260	27.43749	89.30717

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DrAvail	1	531.042251	531.042251	0.71	0.4051
HospAvail	1	144.412739	144.412739	0.19	0.6634
CapInc	1	3880.986458	3880.986458	5.16	0.0277
PopD	1	3300.651845	3300.651845	4.38	0.0416

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DrAvail	1	1221.117277	1221.117277	1.62	0.2089
HospAvail	1	302.026915	302.026915	0.40	0.5295
CapInc	1	2588.165960	2588.165960	3.44	0.0699
PopD	1	3300.651845	3300.651845	4.38	0.0416

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	156.8343710	34.61503504	4.53	<.0001
DrAvail	0.1513136	0.11880742	1.27	0.2089
HospAvail	0.0078351	0.01236991	0.63	0.5295
CapInc	-7.4519896	4.01902352	-1.85	0.0699
PopD	-0.1753328	0.08373515	-2.09	0.0416

Residual Plot of death rate and predictors

The GLM Procedure

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

Dependent Variable: Sq_DeathRate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6635.97602	2211.99201	2.90	0.0441
Error	49	37356.27622	762.37298		
Corrected Total	52	43992.25224			

R-Square	Coeff Var	Root MSE	Sq_DeathRate Mean
0.150844	30.91701	27.61110	89.30717

Source	DF	Type I SS	Mean Square	F Value	Pr > F
HospAvail	1	334.619874	334.619874	0.44	0.5107
CapInc	1	2262.251908	2262.251908	2.97	0.0913
PopD	1	4039.104234	4039.104234	5.30	0.0256

Source	DF	Type III SS	Mean Square	F Value	Pr > F
HospAvail	1	960.631204	960.631204	1.26	0.2671
CapInc	1	1535.494540	1535.494540	2.01	0.1622
PopD	1	4039.104234	4039.104234	5.30	0.0256

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	150.8342624	34.50993489	4.37	<.0001
HospAvail	0.0131532	0.01171752	1.12	0.2671
CapInc	-5.0952764	3.59027208	-1.42	0.1622
PopD	-0.1916683	0.08327058	-2.30	0.0256

Residual Plot of death rate and predictors

The GLM Procedure

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

Dependent Variable: Sq_DeathRate

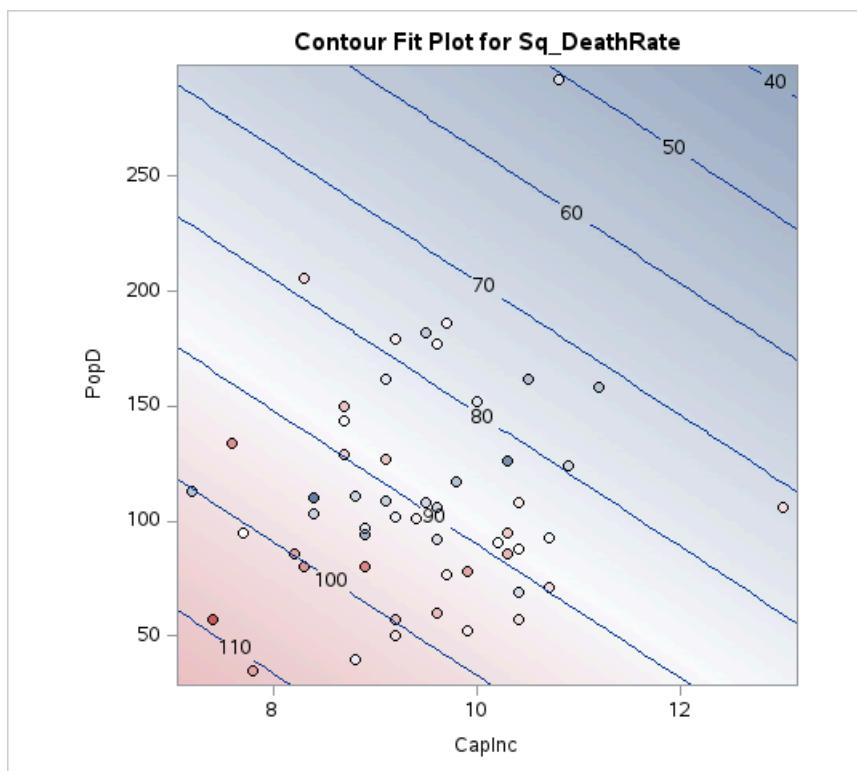
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5675.34481	2837.67241	3.70	0.0316
Error	50	38316.90742	766.33815		
Corrected Total	52	43992.25224			

R-Square	Coeff Var	Root MSE	Sq_DeathRate Mean
0.129008	30.99730	27.68281	89.30717

Source	DF	Type I SS	Mean Square	F Value	Pr > F
CapInc	1	2212.952724	2212.952724	2.89	0.0955
PopD	1	3462.392088	3462.392088	4.52	0.0385

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CapInc	1	1526.855179	1526.855179	1.99	0.1643
PopD	1	3462.392088	3462.392088	4.52	0.0385

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	156.5461757	34.22137620	4.57	<.0001
CapInc	-5.0808897	3.59957368	-1.41	0.1643
PopD	-0.1744056	0.08205074	-2.13	0.0385



Residual Plot of death rate and predictors

The GLM Procedure

Number of Observations Read	53
Number of Observations Used	53

Residual Plot of death rate and predictors

The GLM Procedure

Dependent Variable: Sq_DeathRate

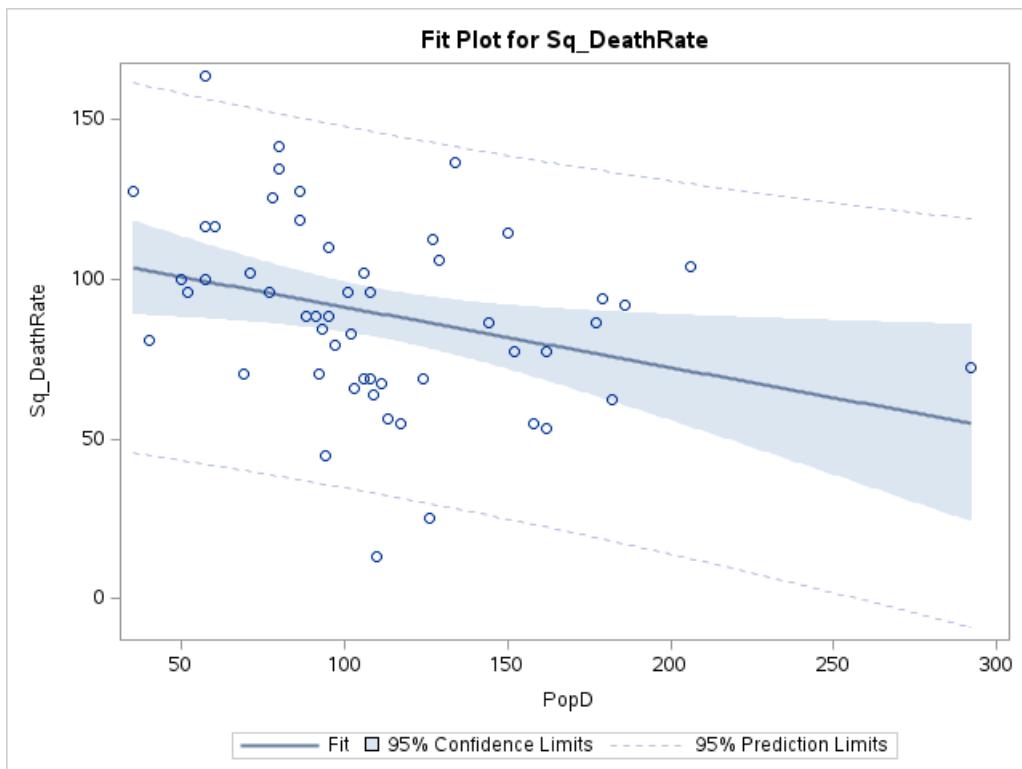
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4148.48963	4148.48963	5.31	0.0253
Error	51	39843.76260	781.25025		
Corrected Total	52	43992.25224			

R-Square	Coeff Var	Root MSE	Sq_DeathRate Mean
0.094300	31.29744	27.95085	89.30717

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PopD	1	4148.489634	4148.489634	5.31	0.0253

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PopD	1	4148.489634	4148.489634	5.31	0.0253

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	110.2534066	9.86740590	11.17	<.0001
PopD	-0.1893163	0.08215575	-2.30	0.0253



Residual Plot of death rate and predictors

The REG Procedure

Model: MODEL1

Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

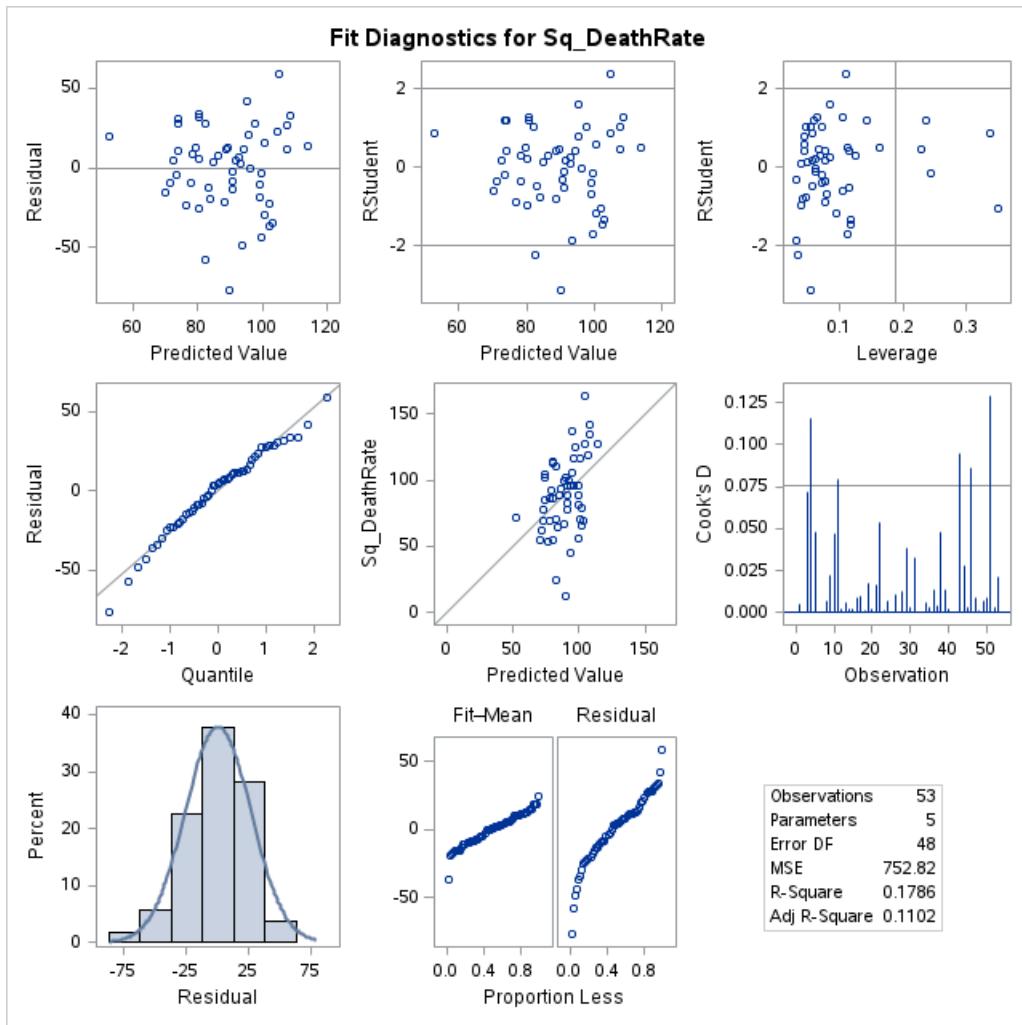
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7857.09329	1964.27332	2.61	0.0470
Error	48	36135	752.81581		
Corrected Total	52	43992			

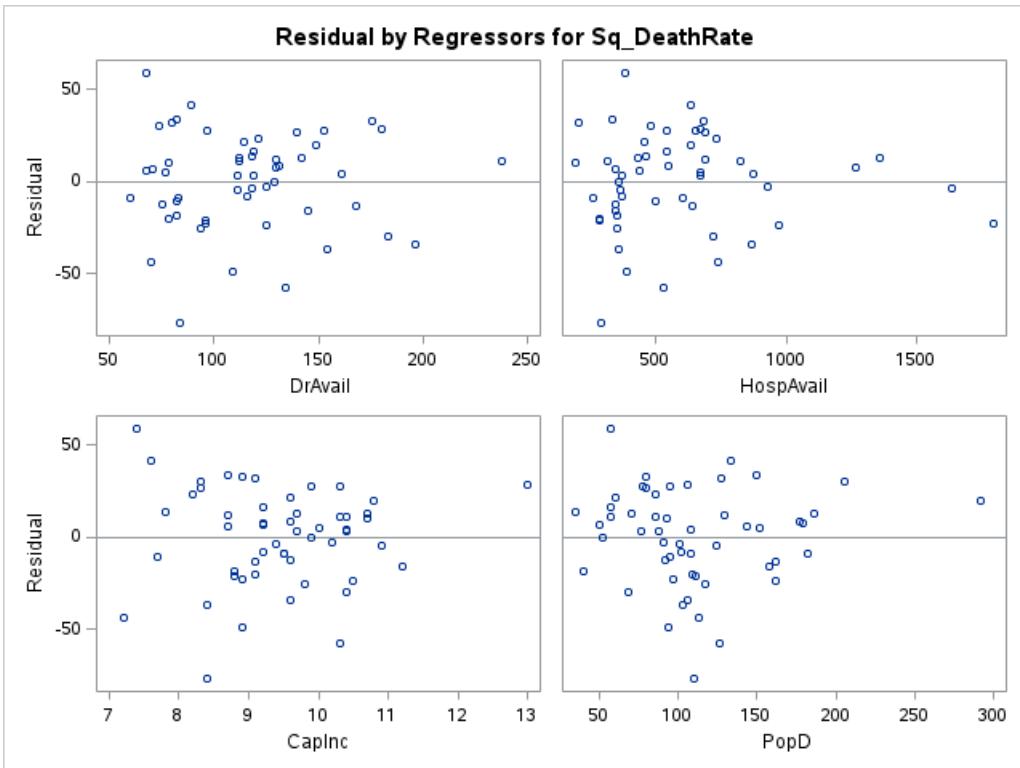
Root MSE	27.43749	R-Square	0.1786
Dependent Mean	89.30717	Adj R-Sq	0.1102
Coeff Var	30.72260		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	156.83437	34.61504	4.53	<.0001
DrAvail	1	0.15131	0.11881	1.27	0.2089
HospAvail	1	0.00784	0.01237	0.63	0.5295
CapInc	1	-7.45199	4.01902	-1.85	0.0699
PopD	1	-0.17533	0.08374	-2.09	0.0416

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1

Test x1x2x3 Results for Dependent Variable Sq_DeathRate				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	1236.20122	1.64	0.1921
Denominator	48	752.81581		

Residual Plot of death rate and predictors

The CORR Procedure

3 Partial Variables:	DrAvail HospAvail CapInc
2 Variables:	Sq_DeathRate PopD

Simple Statistics								
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
DrAvail	53	116.09434	37.88660	6153	60.00000	238.00000		
HospAvail	53	589.79245	332.61831	31259	190.00000	1792		
CapInc	53	9.43585	1.07544	500.10000	7.20000	13.00000		
Sq_DeathRate	53	89.30717	29.08616	4733	12.96000	163.84000	804.81247	28.36922
PopD	53	110.64151	47.17973	5864	35.00000	292.00000	2191	46.80999

Pearson Partial Correlation Coefficients, N = 53 Prob > r under H0: Partial Rho=0		
	Sq_DeathRate	PopD
Sq_DeathRate	1.00000	-0.28930 0.0416
PopD	-0.28930 0.0416	1.00000

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Adjusted R-Square Selection Method

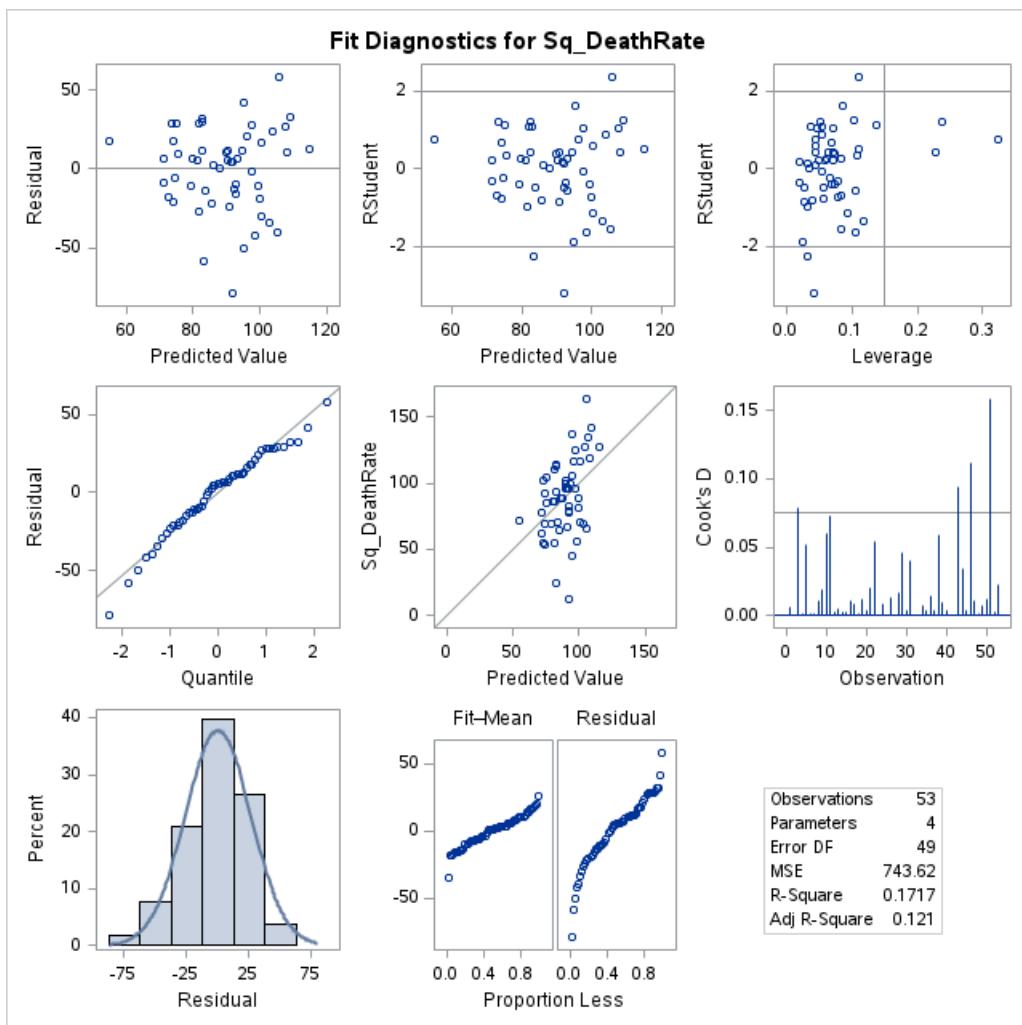
4)

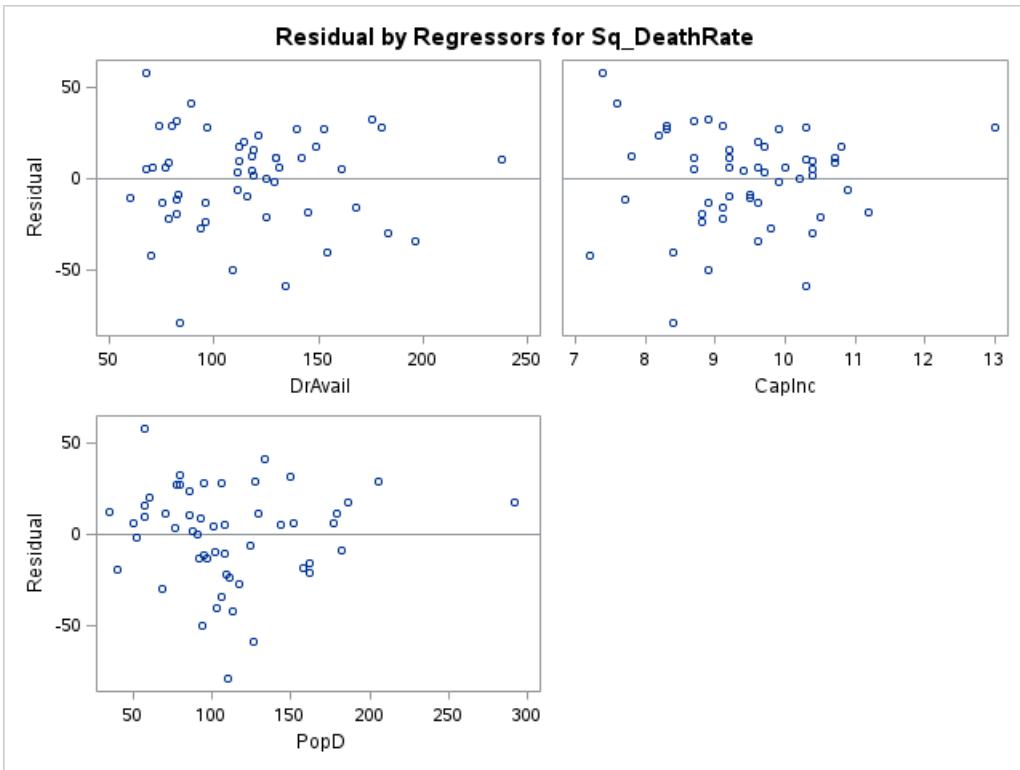
Number of Observations Read	53
Number of Observations Used	53

Number in Model	Adjusted R-Square	R-Square	C(p)	BIC	Variables in Model
3	0.1210	0.1717	3.4012	356.9939	DrAvail CapInc PopD
4	0.1102	0.1786	5.0000	358.8306	DrAvail HospAvail CapInc PopD
3	0.0989	0.1508	4.6221	358.1074	HospAvail CapInc PopD
2	0.0942	0.1290	3.8981	357.1622	CapInc PopD
2	0.0806	0.1159	4.6617	357.8611	HospAvail PopD
1	0.0765	0.0943	3.9263	356.9943	PopD
2	0.0693	0.1051	5.2971	358.4352	DrAvail PopD
2	0.0672	0.1031	5.4132	358.5394	DrAvail CapInc
3	0.0659	0.1198	6.4380	359.7181	DrAvail HospAvail PopD
3	0.0487	0.1036	7.3844	360.5370	DrAvail HospAvail CapInc
1	0.0317	0.0503	6.4974	359.3188	CapInc
2	0.0214	0.0590	7.9874	360.7941	HospAvail CapInc
1	-0.0073	0.0121	8.7315	361.2533	DrAvail
1	-0.0119	0.0076	8.9924	361.4744	HospAvail
2	-0.0240	0.0154	10.5397	362.9324	DrAvail HospAvail

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

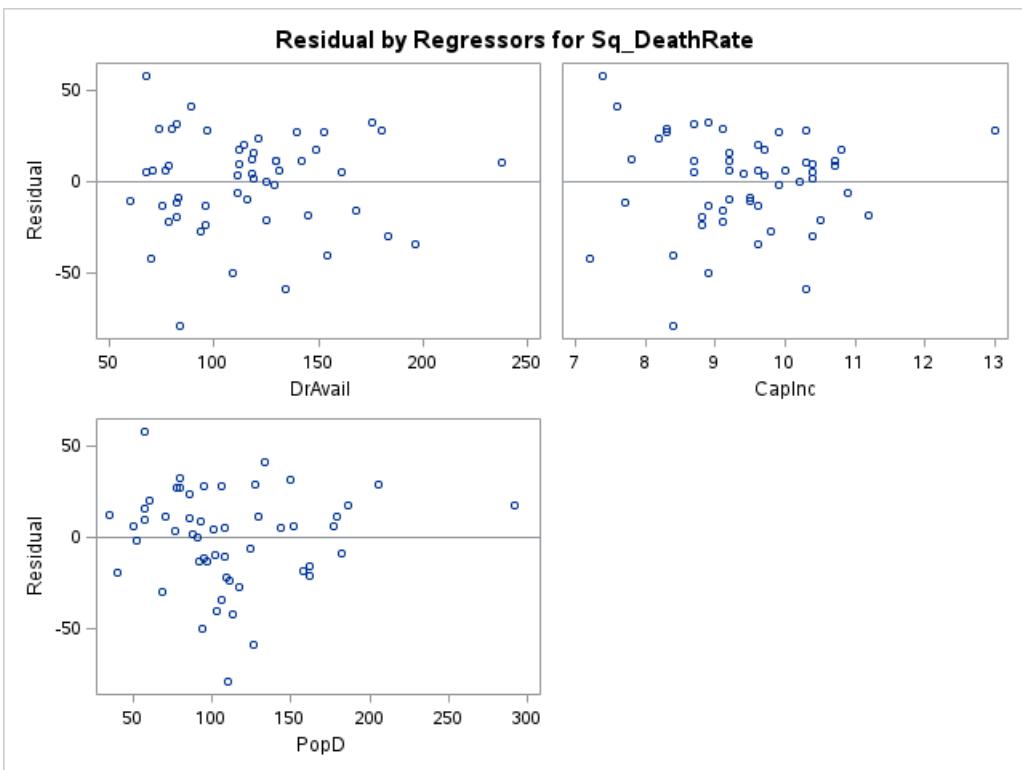
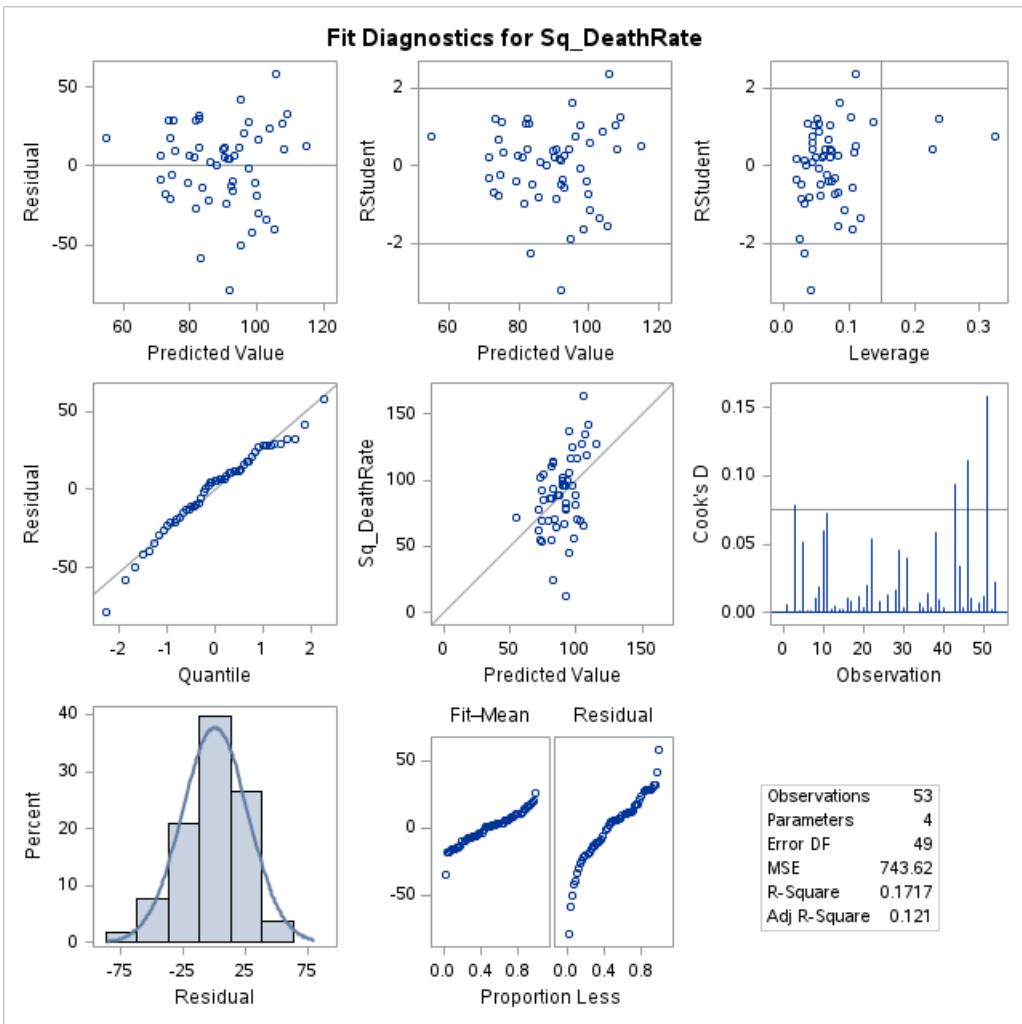
C(p) Selection Method

Number of Observations Read	53
Number of Observations Used	53

Number in Model	C(p)	R-Square	Adjusted R-Square	BIC	Variables in Model
3	3.4012	0.1717	0.1210	356.9939	DrAvail CapInc PopD
2	3.8981	0.1290	0.0942	357.1622	CapInc PopD
1	3.9263	0.0943	0.0765	356.9943	PopD
3	4.6221	0.1508	0.0989	358.1074	HospAvail CapInc PopD
2	4.6617	0.1159	0.0806	357.8611	HospAvail PopD
4	5.0000	0.1786	0.1102	358.8306	DrAvail HospAvail CapInc PopD
2	5.2971	0.1051	0.0693	358.4352	DrAvail PopD
2	5.4132	0.1031	0.0672	358.5394	DrAvail CapInc
3	6.4380	0.1198	0.0659	359.7181	DrAvail HospAvail PopD
1	6.4974	0.0503	0.0317	359.3188	CapInc
3	7.3844	0.1036	0.0487	360.5370	DrAvail HospAvail CapInc
2	7.9874	0.0590	0.0214	360.7941	HospAvail CapInc
1	8.7315	0.0121	-0.0073	361.2533	DrAvail
1	8.9924	0.0076	-0.0119	361.4744	HospAvail
2	10.5397	0.0154	-0.0240	362.9324	DrAvail HospAvail

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate



Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read | 53 |

Number of Observations Used 53

Stepwise Selection: Step 1

Variable PopD Entered: R-Square = 0.0943 and C(p) = 3.9263

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4148.48963	4148.48963	5.31	0.0253
Error	51	39844	781.25025		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	110.25341	9.86741	97537	124.85	<.0001
PopD	-0.18932	0.08216	4148.48963	5.31	0.0253

Bounds on condition number: 1, 1

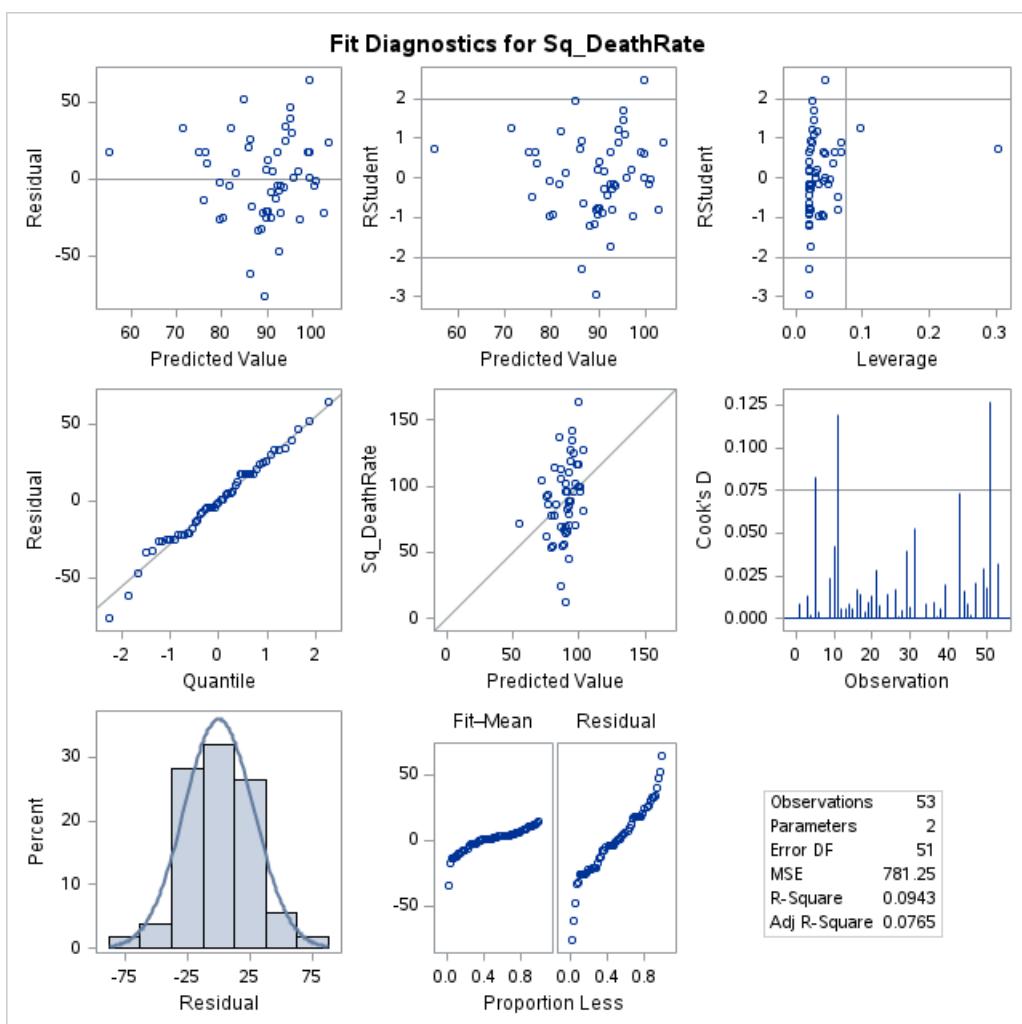
All variables left in the model are significant at the 0.1500 level.

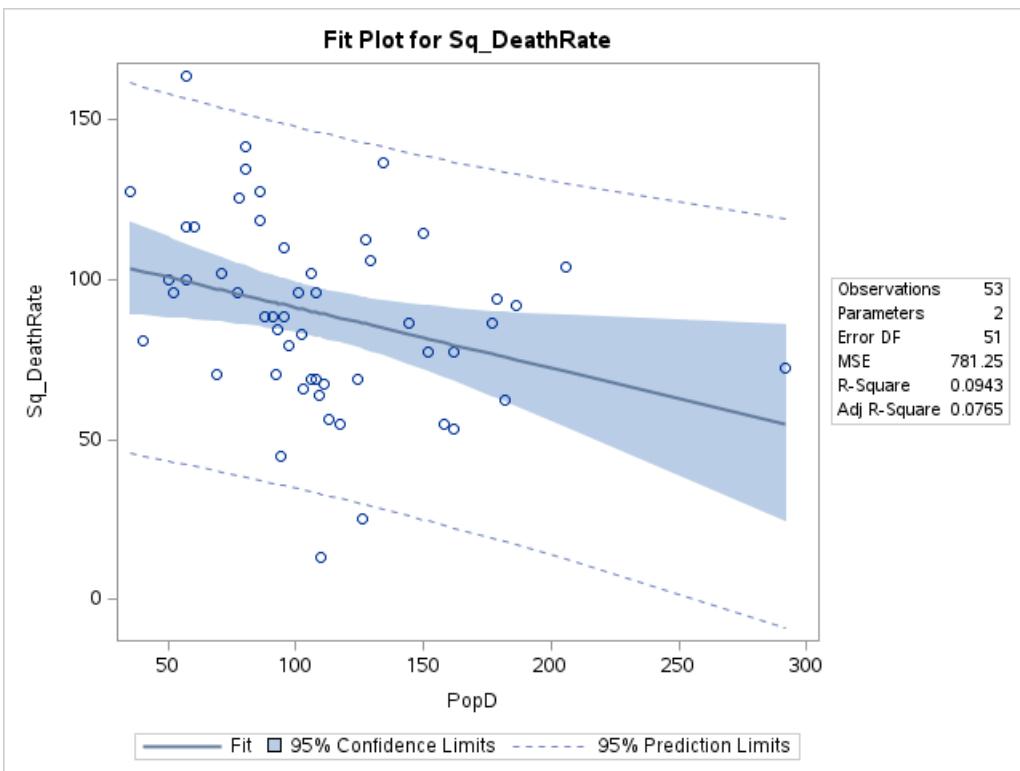
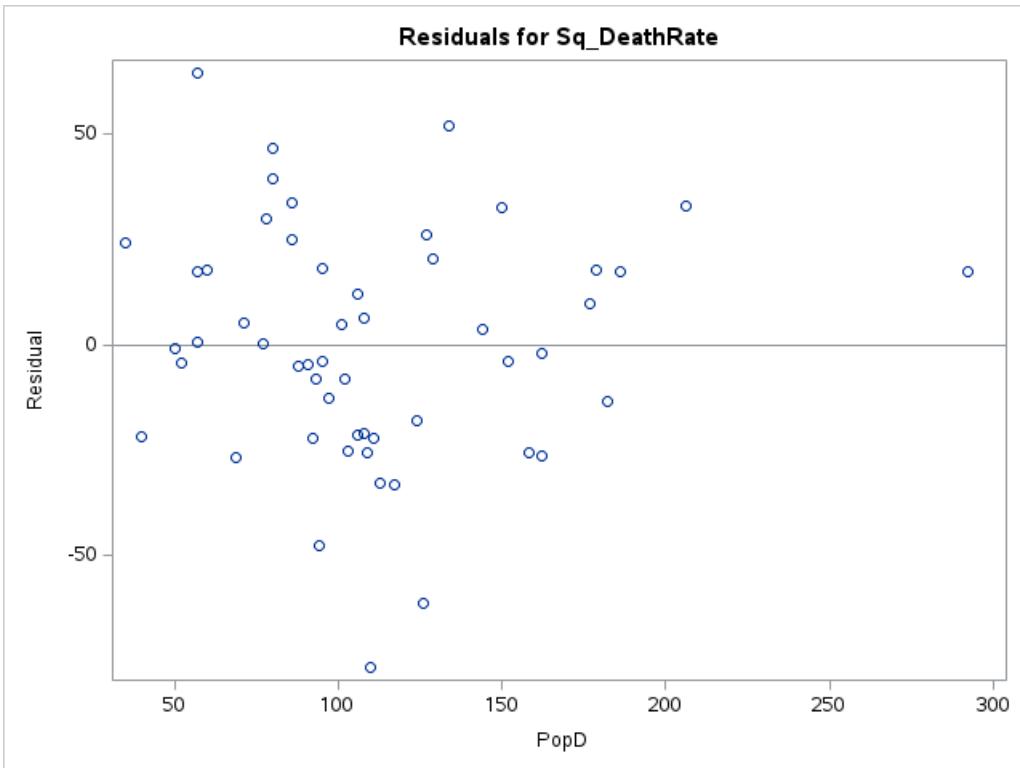
No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	PopD		1	0.0943	0.0943	3.9263	5.31	0.0253

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

The REG Procedure
 Model: MODEL1
 Dependent Variable: `Sq_DeathRate`

Number of Observations Read	53
Number of Observations Used	53

Forward Selection: Step 1

Variable `PopD` Entered: R-Square = 0.0943 and C(p) = 3.9263

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4148.48963	4148.48963	5.31	0.0253

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Error	51	39844	781.25025		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	110.25341	9.86741	97537	124.85	<.0001
PopD	-0.18932	0.08216	4148.48963	5.31	0.0253

Bounds on condition number: 1, 1

Forward Selection: Step 2

Variable CapInc Entered: R-Square = 0.1290 and C(p) = 3.8981

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5675.34481	2837.67241	3.70	0.0316
Error	50	38317	766.33815		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	156.54618	34.22138	16037	20.93	<.0001
CapInc	-5.08089	3.59957	1526.85518	1.99	0.1643
PopD	-0.17441	0.08205	3462.39209	4.52	0.0385

Bounds on condition number: 1.0169, 4.0674

Forward Selection: Step 3

Variable DrAvail Entered: R-Square = 0.1717 and C(p) = 3.4012

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7555.06638	2518.35546	3.39	0.0253
Error	49	36437	743.61604		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	160.85644	33.81906	16823	22.62	<.0001
DrAvail	0.17672	0.11115	1879.72157	2.53	0.1183
CapInc	-7.84004	3.94771	2932.88310	3.94	0.0526
PopD	-0.16348	0.08112	3020.30858	4.06	0.0494

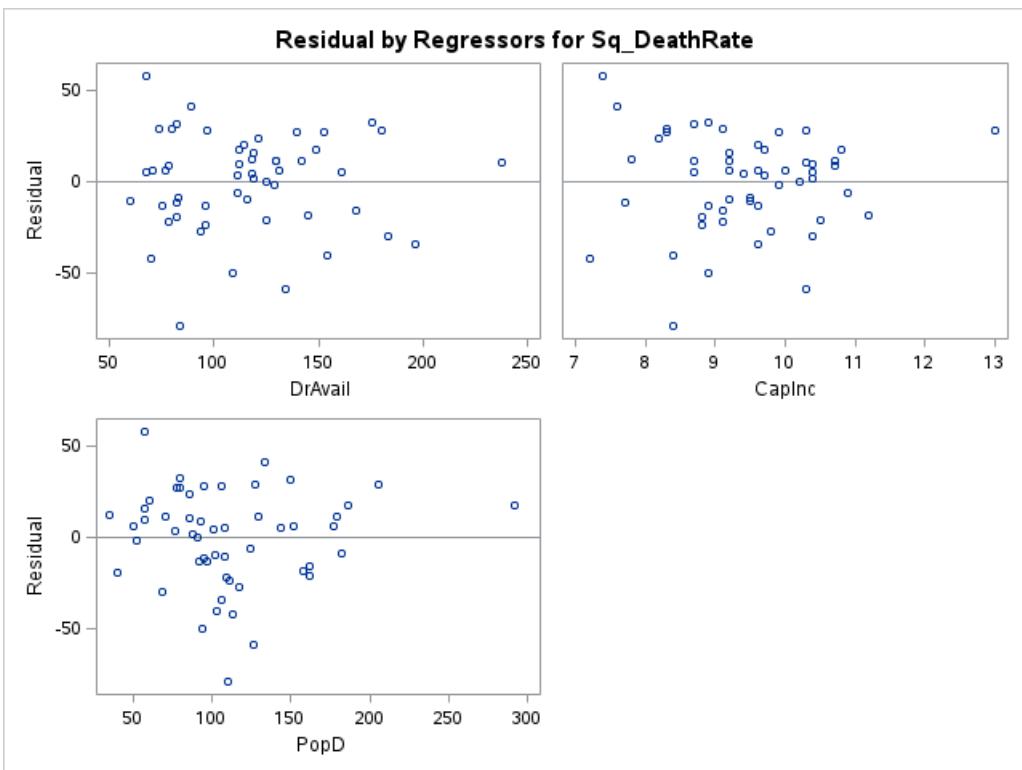
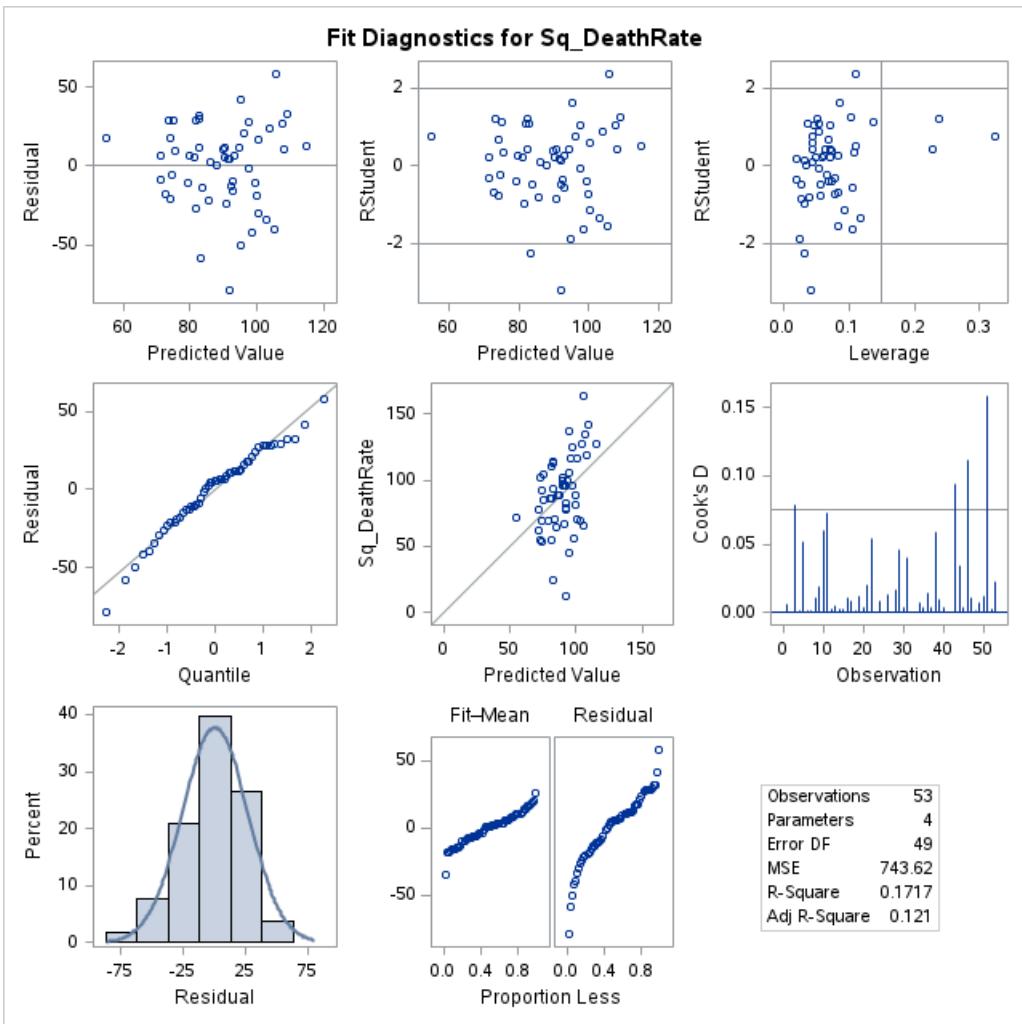
Bounds on condition number: 1.2604, 10.574

No other variable met the 0.5000 significance level for entry into the model.

Summary of Forward Selection							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	PopD	1	0.0943	0.0943	3.9263	5.31	0.0253
2	CapInc	2	0.0347	0.1290	3.8981	1.99	0.1643
3	DrAvail	3	0.0427	0.1717	3.4012	2.53	0.1183

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate



Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read | 53 |

Backward Elimination: Step 0

All Variables Entered: R-Square = 0.1786 and C(p) = 5.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7857.09329	1964.27332	2.61	0.0470
Error	48	36135	752.81581		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	156.83437	34.61504	15454	20.53	<.0001
DrAvail	0.15131	0.11881	1221.11728	1.62	0.2089
HospAvail	0.00784	0.01237	302.02692	0.40	0.5295
CapInc	-7.45199	4.01902	2588.16596	3.44	0.0699
PopD	-0.17533	0.08374	3300.65185	4.38	0.0416

Bounds on condition number: 1.3995, 19.749

Backward Elimination: Step 1

Variable HospAvail Removed: R-Square = 0.1717 and C(p) = 3.4012

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7555.06638	2518.35546	3.39	0.0253
Error	49	36437	743.61604		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	160.85644	33.81906	16823	22.62	<.0001
DrAvail	0.17672	0.11115	1879.72157	2.53	0.1183
CapInc	-7.84004	3.94771	2932.88310	3.94	0.0526
PopD	-0.16348	0.08112	3020.30858	4.06	0.0494

Bounds on condition number: 1.2604, 10.574

Backward Elimination: Step 2

Variable DrAvail Removed: R-Square = 0.1290 and C(p) = 3.8981

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5675.34481	2837.67241	3.70	0.0316
Error	50	38317	766.33815		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	156.54618	34.22138	16037	20.93	<.0001
CapInc	-5.08089	3.59957	1526.85518	1.99	0.1643
PopD	-0.17441	0.08205	3462.39209	4.52	0.0385

Bounds on condition number: 1.0169, 4.0674

Backward Elimination: Step 3

Variable CapInc Removed: R-Square = 0.0943 and C(p) = 3.9263

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4148.48963	4148.48963	5.31	0.0253
Error	51	39844	781.25025		
Corrected Total	52	43992			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	110.25341	9.86741	97537	124.85	<.0001

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
PopD	-0.18932	0.08216	4148.48963	5.31	0.0253

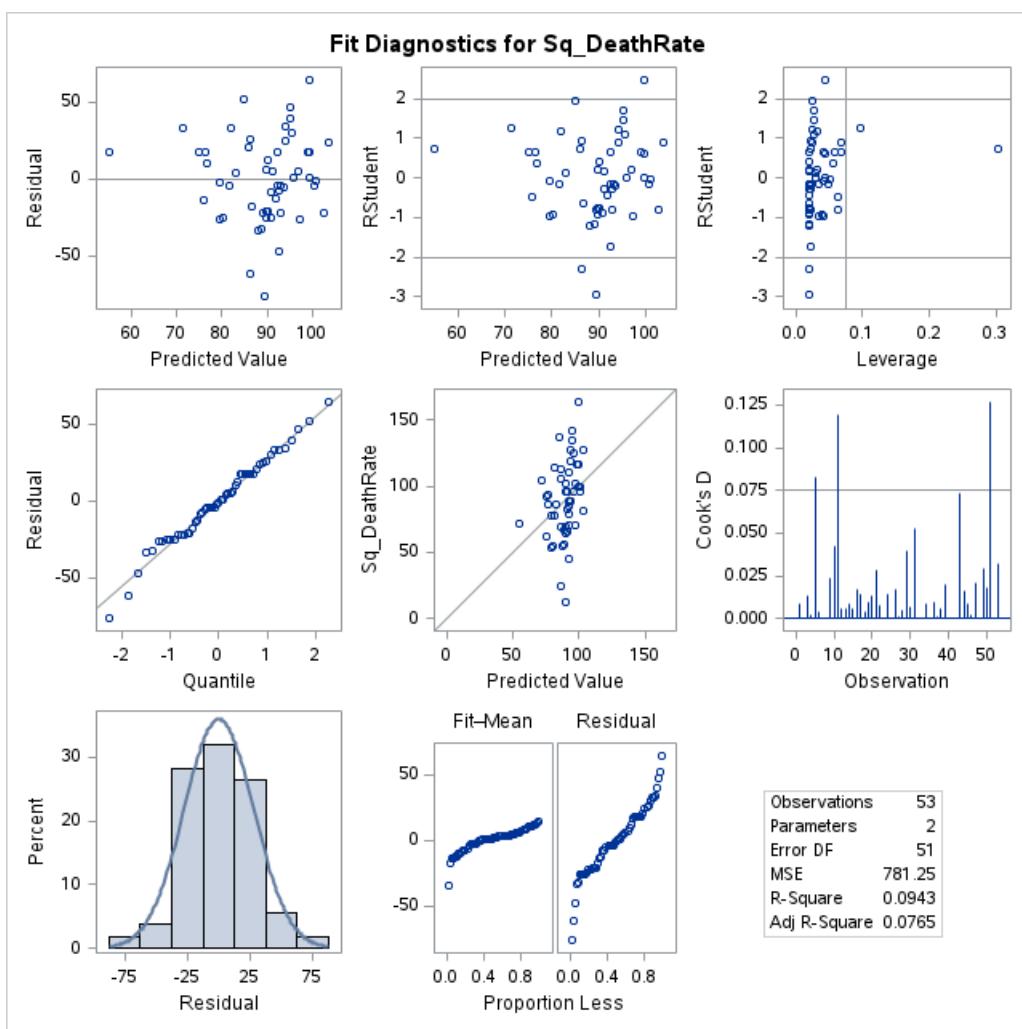
Bounds on condition number: 1, 1

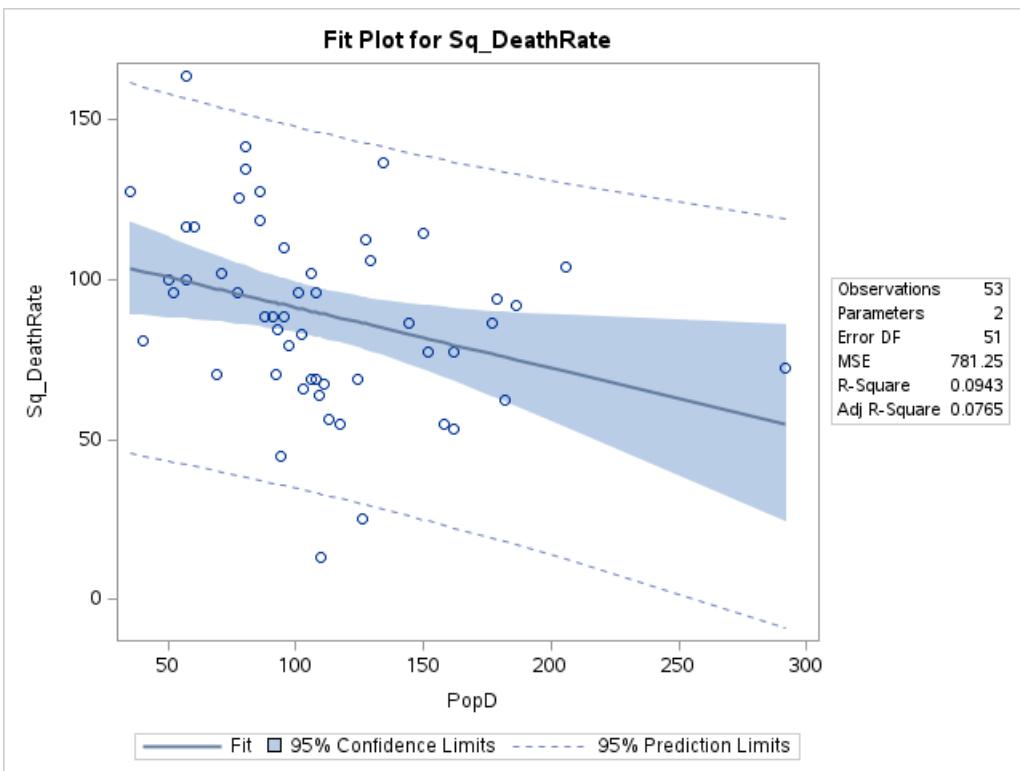
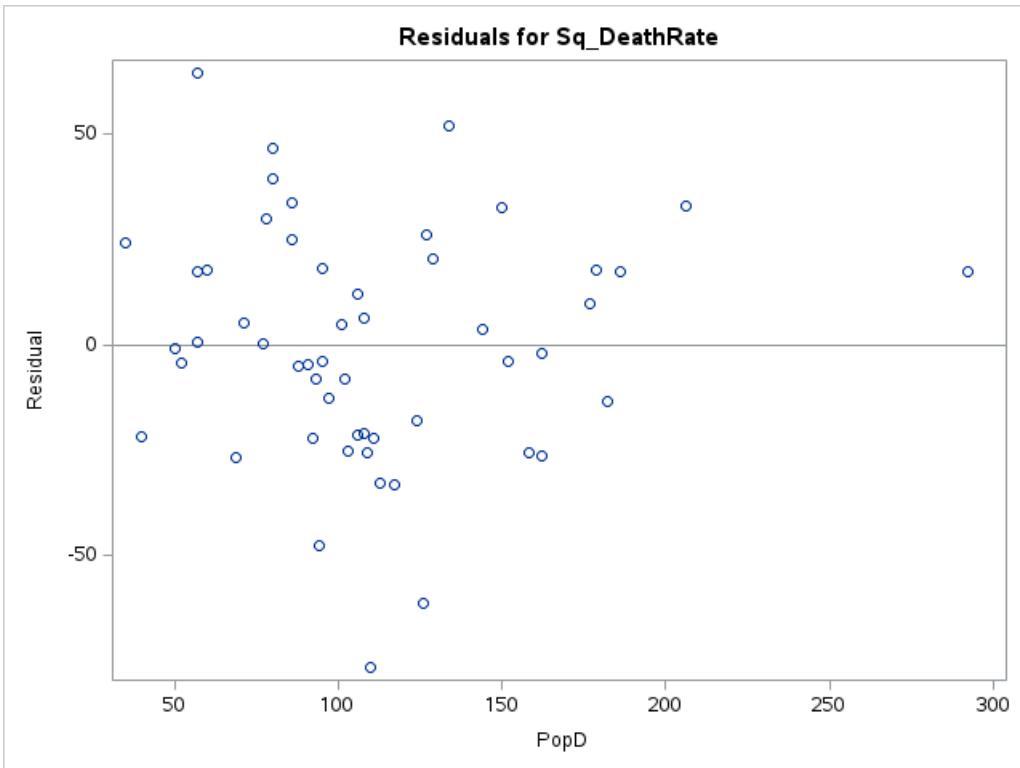
All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	HospAvail	3	0.0069	0.1717	3.4012	0.40	0.5295
2	DrAvail	2	0.0427	0.1290	3.8981	2.53	0.1183
3	CapInc	1	0.0347	0.0943	3.9263	1.99	0.1643

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

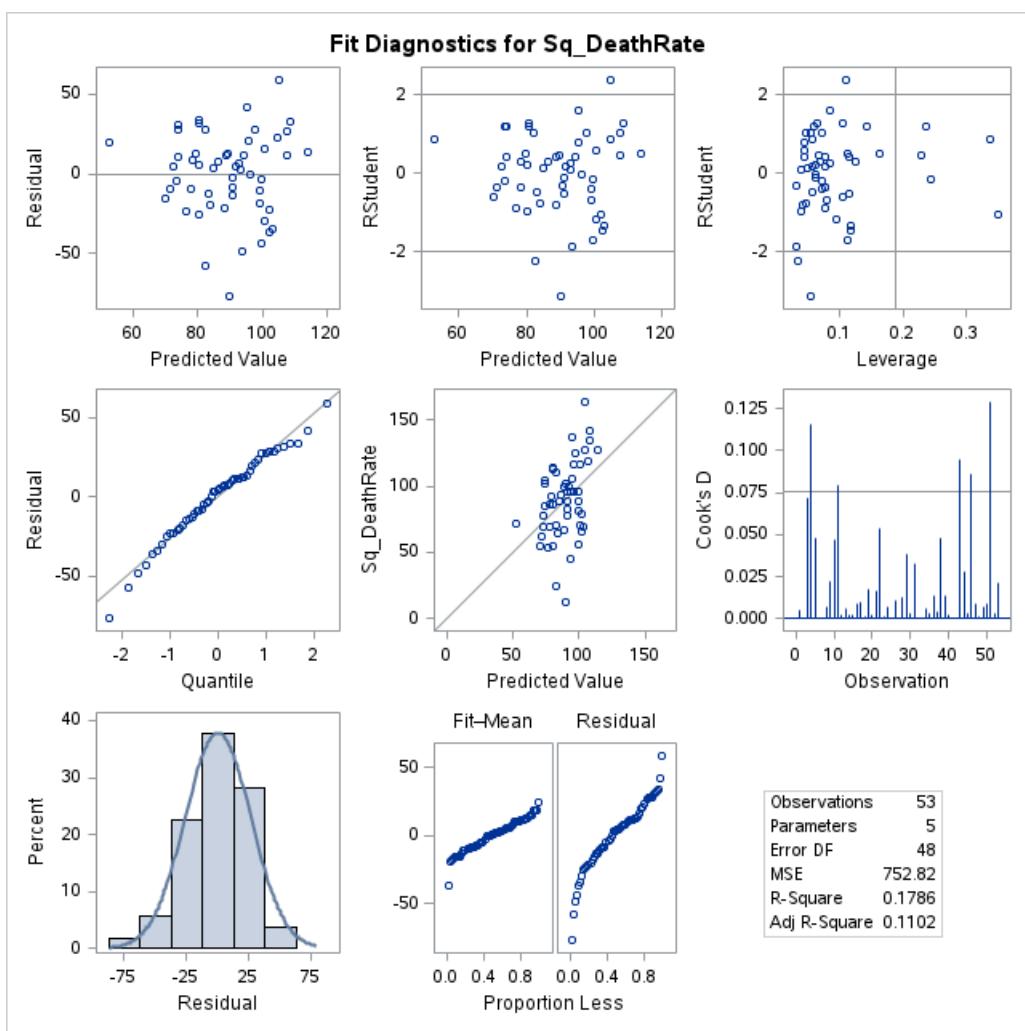
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7857.09329	1964.27332	2.61	0.0470
Error	48	36135	752.81581		
Corrected Total	52	43992			

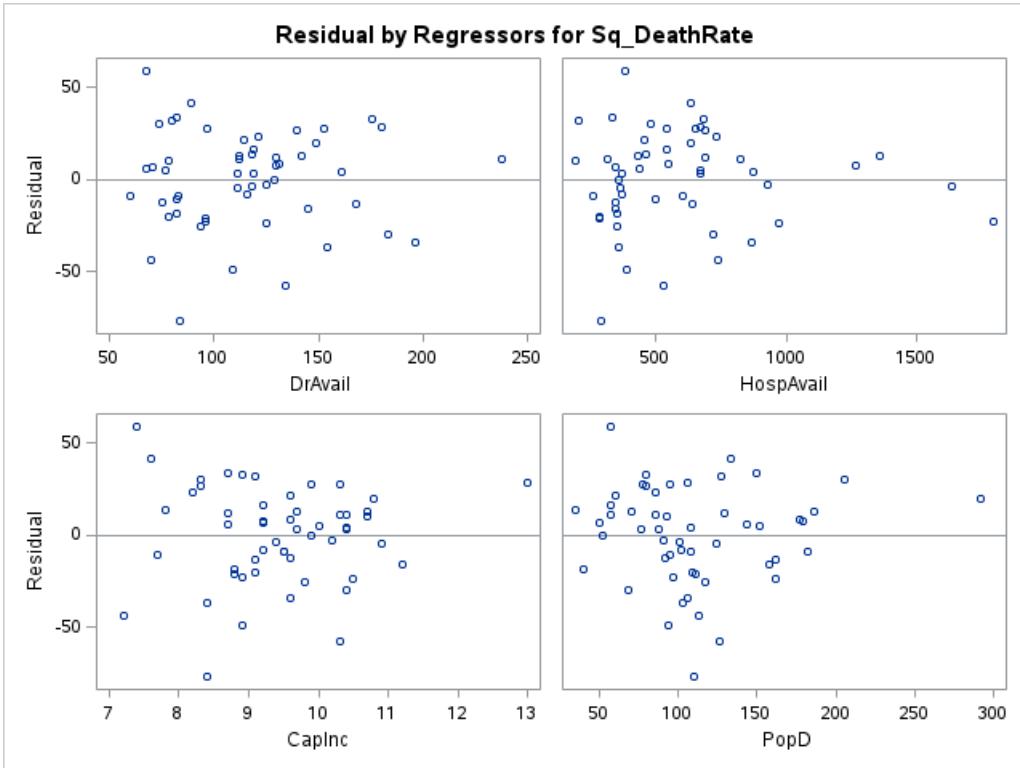
Root MSE	27.43749	R-Square	0.1786
Dependent Mean	89.30717	Adj R-Sq	0.1102
Coeff Var	30.72260		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	156.83437	34.61504	4.53	<.0001
DrAvail	1	0.15131	0.11881	1.27	0.2089
HospAvail	1	0.00784	0.01237	0.63	0.5295
CapInc	1	-7.45199	4.01902	-1.85	0.0699
PopD	1	-0.17533	0.08374	-2.09	0.0416

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1

Test x2 Results for Dependent Variable Sq_DeathRate				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	302.02692	0.40	0.5295
Denominator	48	752.81581		

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

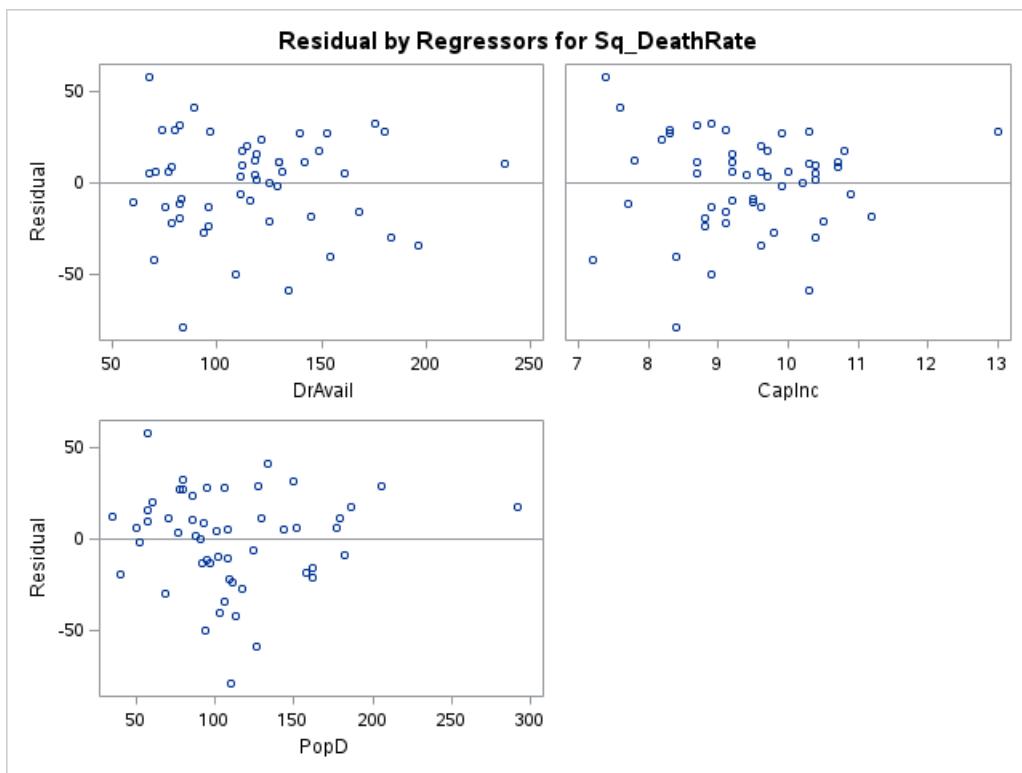
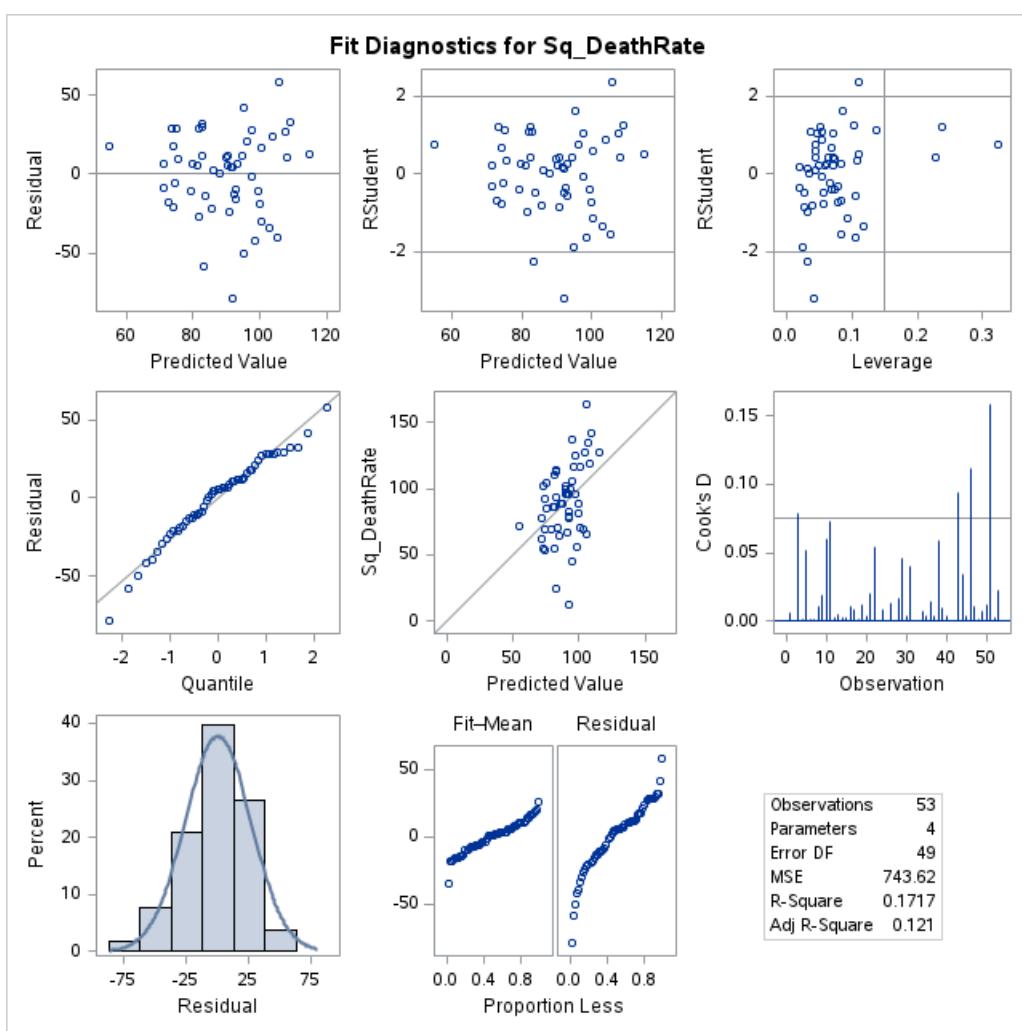
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7555.06638	2518.35546	3.39	0.0253
Error	49	36437	743.61604		
Corrected Total	52	43992			

Root MSE	27.26932	R-Square	0.1717
Dependent Mean	89.30717	Adj R-Sq	0.1210
Coeff Var	30.53431		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	160.85644	33.81906	4.76	<.0001
DrAvail	1	0.17672	0.11115	1.59	0.1183
CapInc	1	-7.84004	3.94771	-1.99	0.0526
PopD	1	-0.16348	0.08112	-2.02	0.0494

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate



Residual Plot of death rate and predictors

The CORR Procedure

2 Partial Variables:	CapInc PopD
2 Variables:	Sq_DeathRate DrAvail

Simple Statistics								
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
CapInc	53	9.43585	1.07544	500.10000	7.20000	13.00000		
PopD	53	110.64151	47.17973	5864	35.00000	292.00000		
Sq_DeathRate	53	89.30717	29.08616	4733	12.96000	163.84000	766.33815	27.68281
DrAvail	53	116.09434	37.88660	6153	60.00000	238.00000	1204	34.69653

Pearson Partial Correlation Coefficients, N = 53 Prob > r under H0: Partial Rho=0		
	Sq_DeathRate	DrAvail
Sq_DeathRate	1.00000	0.22149 0.1183
DrAvail	0.22149 0.1183	1.00000

Residual Plot of death rate and predictors

The CORR Procedure

2 Partial Variables:	DrAvail PopD
2 Variables:	Sq_DeathRate CapInc

Simple Statistics								
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
DrAvail	53	116.09434	37.88660	6153	60.00000	238.00000		
PopD	53	110.64151	47.17973	5864	35.00000	292.00000		
Sq_DeathRate	53	89.30717	29.08616	4733	12.96000	163.84000	787.40138	28.06067
CapInc	53	9.43585	1.07544	500.10000	7.20000	13.00000	0.95431	0.97689

Pearson Partial Correlation Coefficients, N = 53 Prob > r under H0: Partial Rho=0		
	Sq_DeathRate	CapInc
Sq_DeathRate	1.00000	-0.27294 0.0526
CapInc	-0.27294 0.0526	1.00000

Residual Plot of death rate and predictors

The CORR Procedure

2 Partial Variables:	DrAvail CapInc
2 Variables:	Sq_DeathRate PopD

Simple Statistics								
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
DrAvail	53	116.09434	37.88660	6153	60.00000	238.00000		
CapInc	53	9.43585	1.07544	500.10000	7.20000	13.00000		
Sq_DeathRate	53	89.30717	29.08616	4733	12.96000	163.84000	789.14989	28.09181
PopD	53	110.64151	47.17973	5864	35.00000	292.00000	2260	47.54211

Pearson Partial Correlation Coefficients, N = 53 Prob > r under H0: Partial Rho=0		
	Sq_DeathRate	PopD
Sq_DeathRate	1.00000	-0.27667 0.0494
PopD	-0.27667 0.0494	1.00000

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read 53

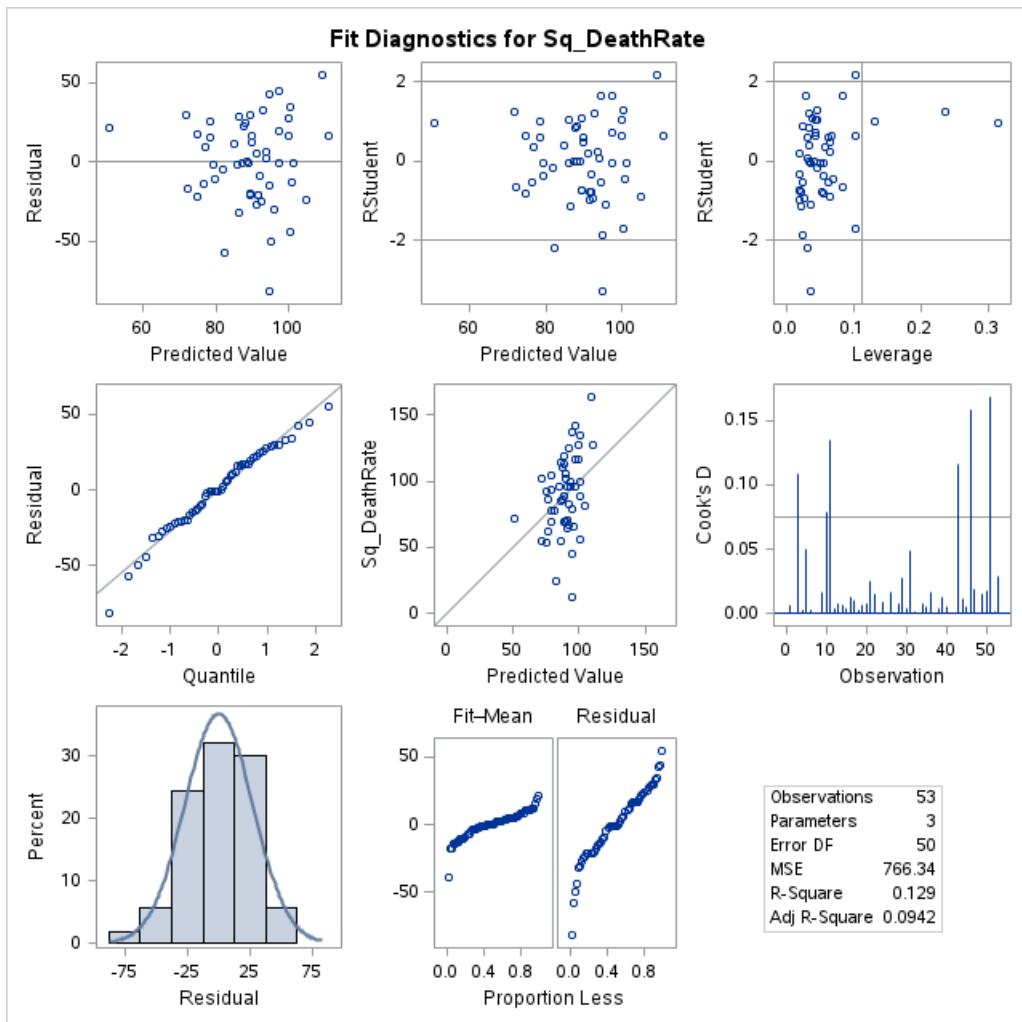
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5675.34481	2837.67241	3.70	0.0316
Error	50	38317	766.33815		
Corrected Total	52	43992			

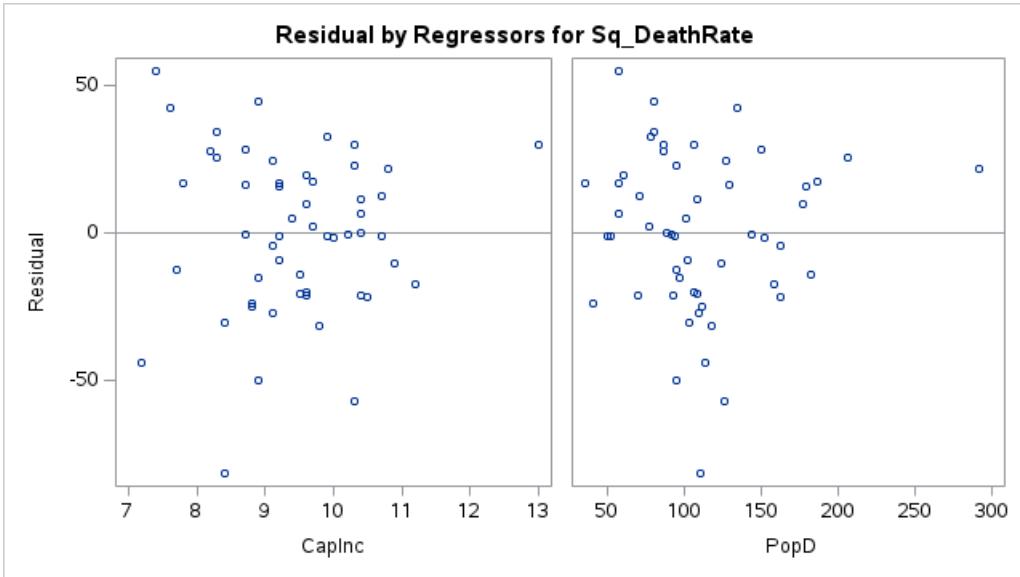
Root MSE	27.68281	R-Square	0.1290
Dependent Mean	89.30717	Adj R-Sq	0.0942
Coeff Var	30.99730		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	156.54618	34.22138	4.57	<.0001
CapInc	1	-5.08089	3.59957	-1.41	0.1643
PopD	1	-0.17441	0.08205	-2.13	0.0385

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate





Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: DrAvail

Number of Observations Read	53
Number of Observations Used	53

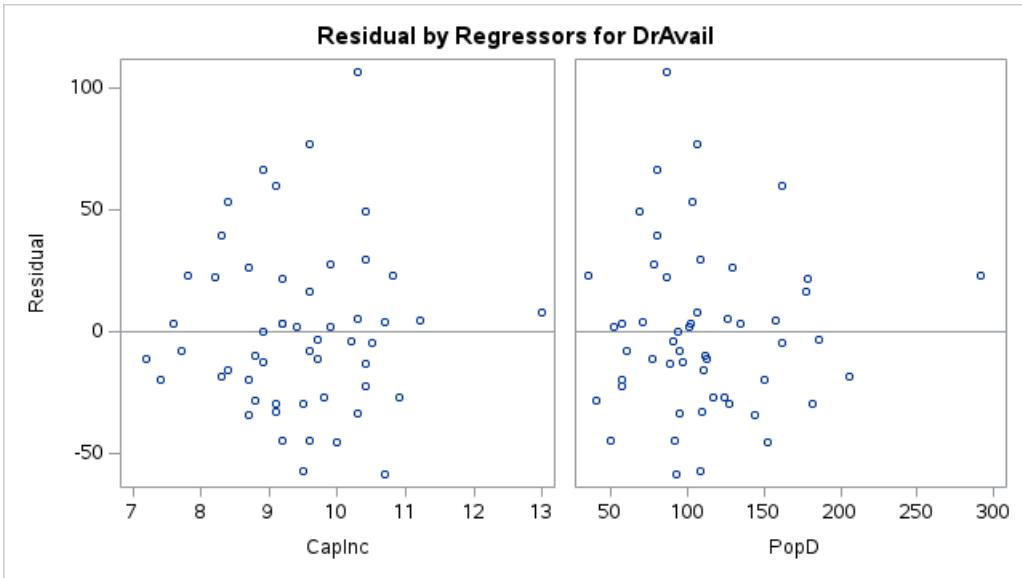
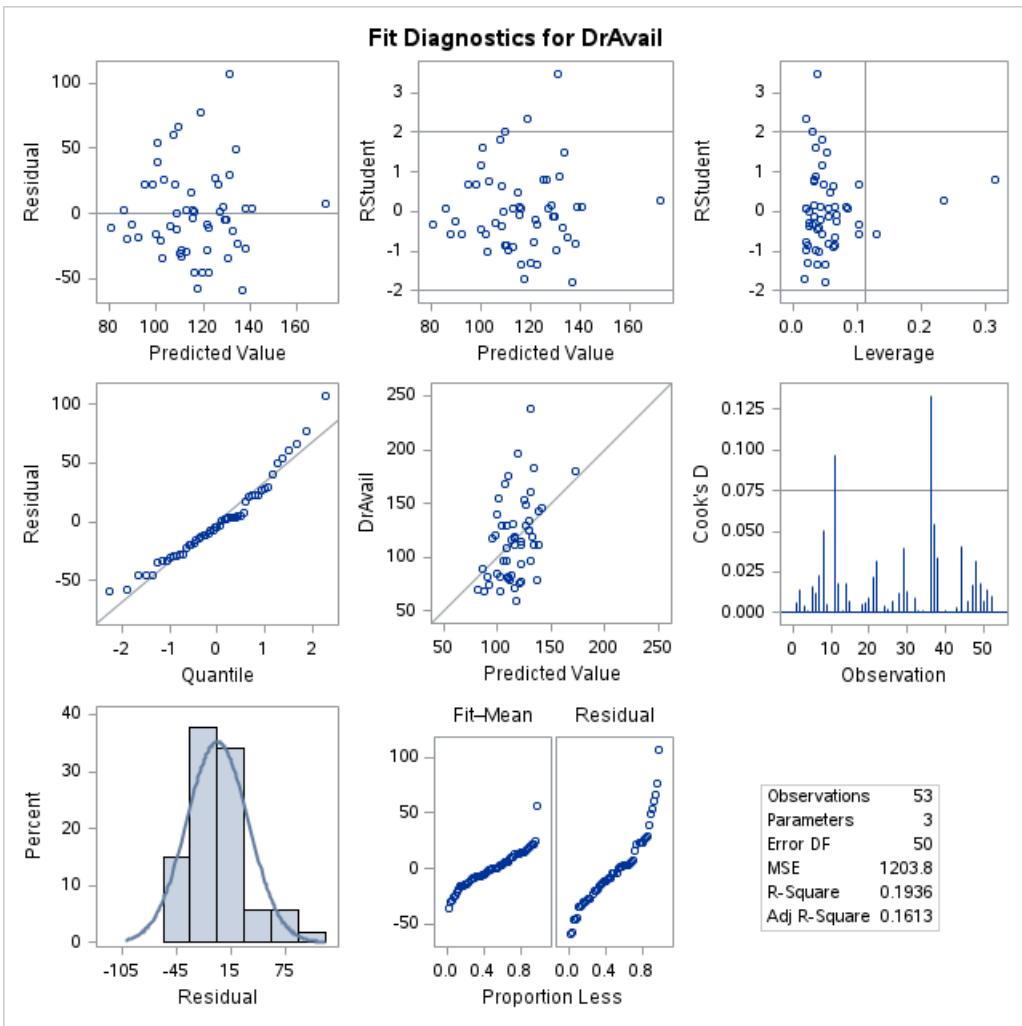
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	14448	7224.03725	6.00	0.0046
Error	50	60192	1203.84908		
Corrected Total	52	74641			

Root MSE	34.69653	R-Square	0.1936
Dependent Mean	116.09434	Adj R-Sq	0.1613
Coeff Var	29.88649		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-24.39094	42.89170	-0.57	0.5721
CapInc	1	15.61347	4.51156	3.46	0.0011
PopD	1	-0.06183	0.10284	-0.60	0.5504

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: DrAvail



Residual Plot of death rate and predictors

The CORR Procedure

2 Variables: R_X3X4 R_X1gX3X4

Simple Statistics								
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label	
R_X3X4	53	0.00290	1.03841	0.15378	-3.29057	2.16787	Studentized Residual without Current Obs	
R_X1gX3X4	53	0.01140	1.03401	0.60425	-1.78055	3.47270	Studentized Residual without Current Obs	

Pearson Correlation Coefficients, N = 53		
		Prob > r under H0: Rho=0
	R_X3X4	R_X1gX3X4
R_X3X4 Studentized Residual without Current Obs	1.00000	0.21851 0.1160
R_X1gX3X4 Studentized Residual without Current Obs	0.21851 0.1160	1.00000

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

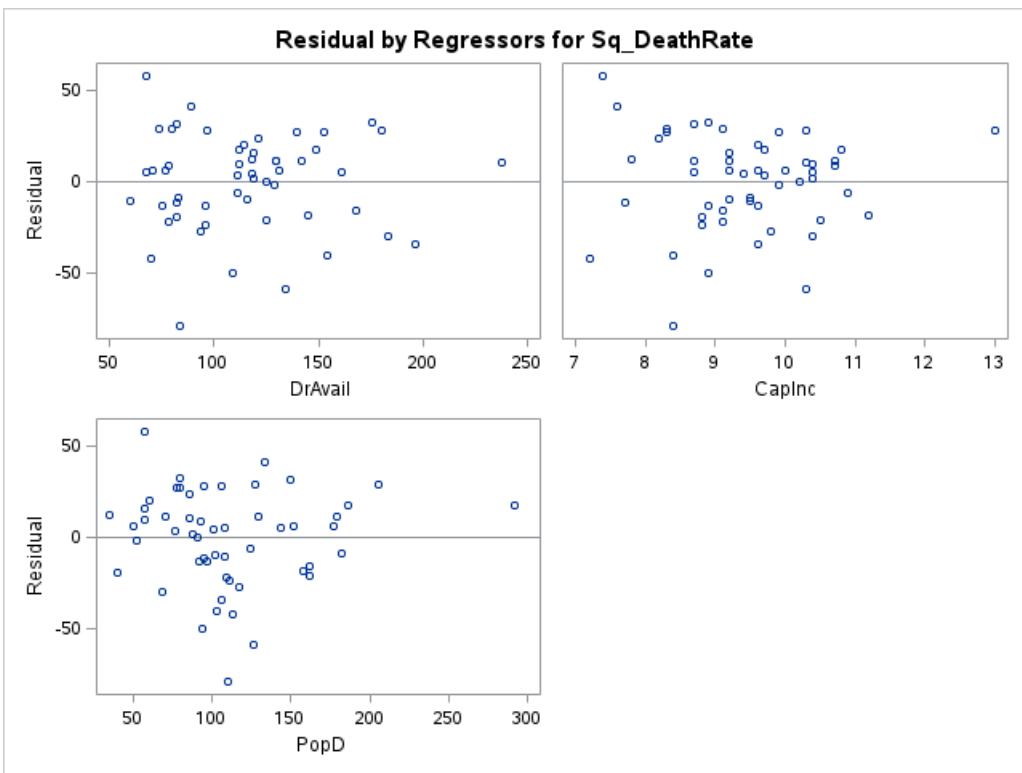
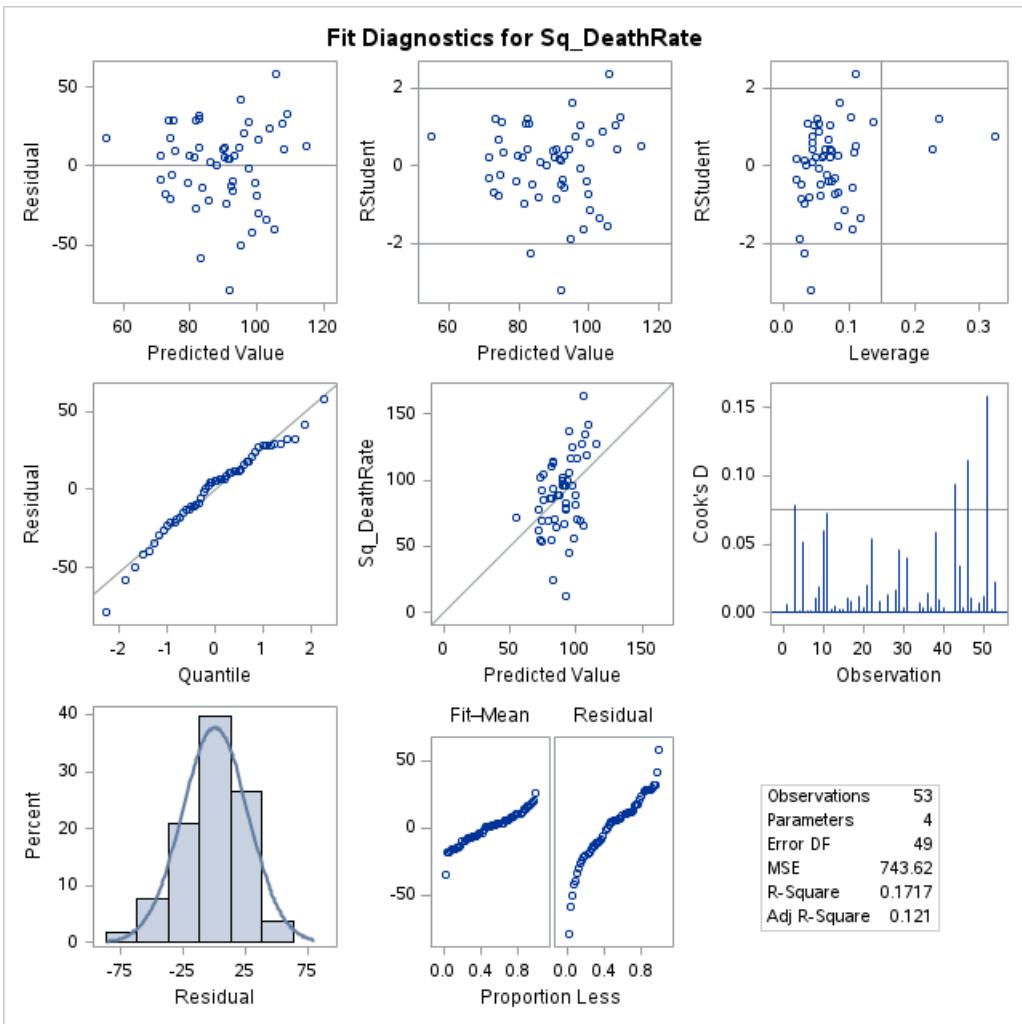
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7555.06638	2518.35546	3.39	0.0253
Error	49	36437	743.61604		
Corrected Total	52	43992			

Root MSE	27.26932	R-Square	0.1717
Dependent Mean	89.30717	Adj R-Sq	0.1210
Coeff Var	30.53431		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	160.85644	33.81906	4.76	<.0001
DrAvail	1	0.17672	0.11115	1.59	0.1183
CapInc	1	-7.84004	3.94771	-1.99	0.0526
PopD	1	-0.16348	0.08112	-2.02	0.0494

Residual Plot of death rate and predictors

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate



Residual Plot of death rate and predictors

The CORR Procedure

2 Variables: Sq_DeathRate P

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Sq_DeathRate	53	89.30717	29.08616	4733	12.96000	163.84000	
P	53	89.30717	12.05362	4733	54.77881	114.83485	Predicted Value of Sq_DeathRate

Pearson Correlation Coefficients, N = 53 Prob > r under H0: Rho=0		
	Sq_DeathRate	P
Sq_DeathRate	1.00000	0.41441 0.0020
P	0.41441 0.0020	1.00000

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y						
Variable	Intercept	DrAvail	CapInc	PopD	Sq_DeathRate	
Intercept	53	6153	500.09999657	5864	4733.280035	
DrAvail	6153	788969	58976.799578	678924	555802.83393	
CapInc	500.09999657	58976.799578	4779.0099338	55671.499755	44297.699051	
PopD	5864	678924	55671.499755	764550	501784.23414	
Sq_DeathRate	4733.280035	555802.83393	44297.699051	501784.23414	466708.09921	

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Number of Observations Read	53
Number of Observations Used	53

X'X Inverse, Parameter Estimates, and SSE					
Variable	Intercept	DrAvail	CapInc	PopD	Sq_DeathRate
Intercept	1.5380635449	0.000405216	-0.160374479	-0.000478749	160.85644449
DrAvail	0.000405216	0.000166134	-0.000259392	1.0272196E-6	0.1767159452
CapInc	-0.160374479	-0.000259392	0.020957606	-0.000065656	-7.840038815
PopD	-0.000478749	1.0272196E-6	-0.000065656	8.8485697E-6	-0.163479084
Sq_DeathRate	160.85644449	0.1767159452	-7.840038815	-0.163479084	36437.185858

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7555.06638	2518.35546	3.39	0.0253
Error	49	36437	743.61604		
Corrected Total	52	43992			

Root MSE	27.26932	R-Square	0.1717
Dependent Mean	89.30717	Adj R-Sq	0.1210
Coeff Var	30.53431		

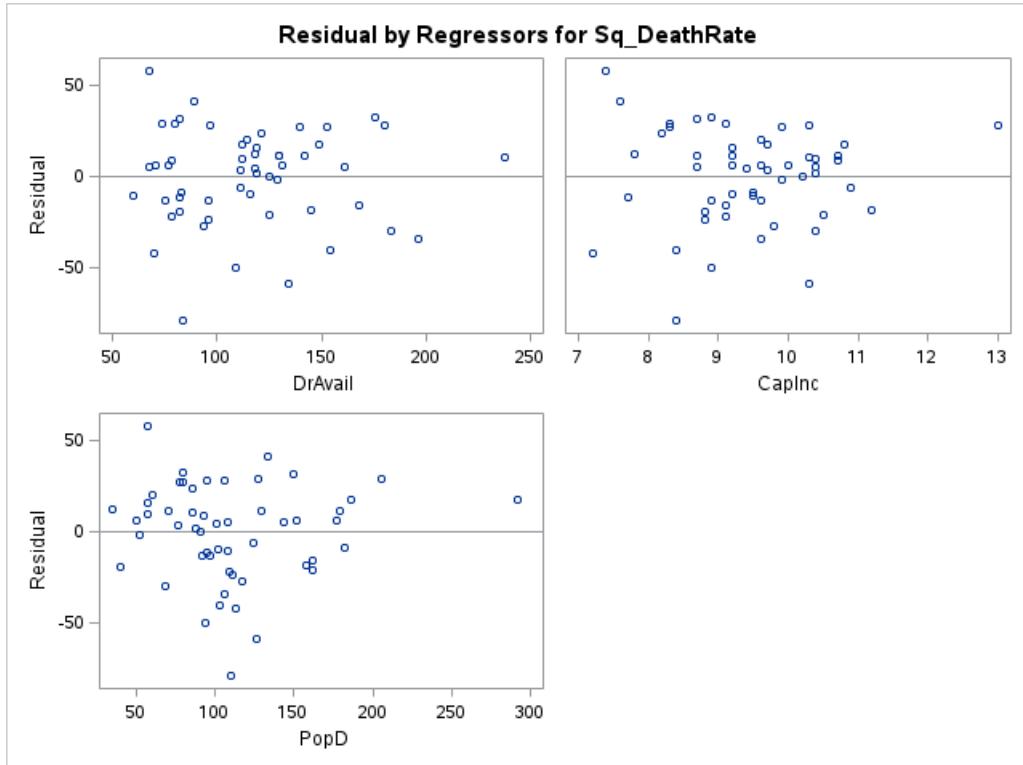
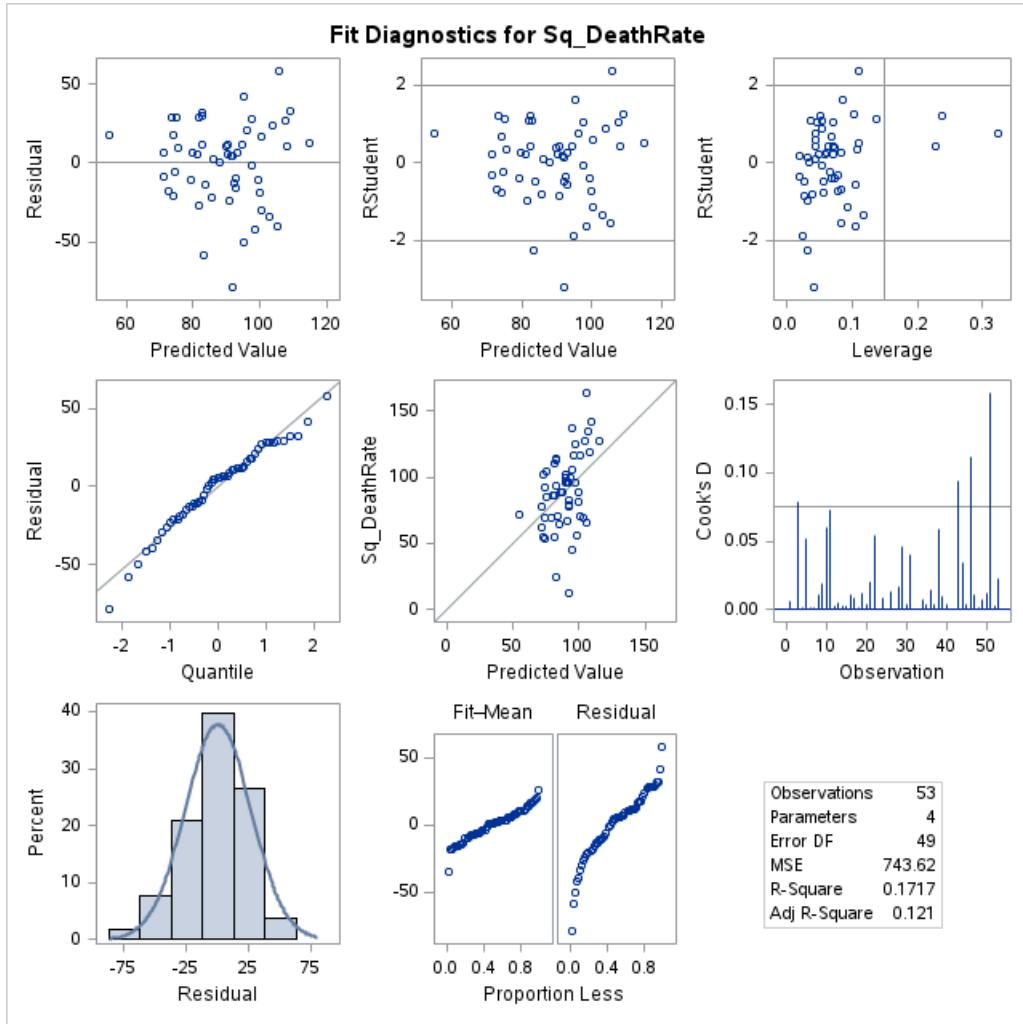
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	160.85644	33.81906	4.76	<.0001
DrAvail	1	0.17672	0.11115	1.59	0.1183
CapInc	1	-7.84004	3.94771	-1.99	0.0526
PopD	1	-0.16348	0.08112	-2.02	0.0494

The REG Procedure
Model: MODEL1
Dependent Variable: Sq_DeathRate

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Output Statistics		Residual		
				95% CL Mean				
				95% CL Predict				
1	64.0	85.4767	5.3703	74.6847	96.2687	29.6244	141.3290	-21.4767
2	86.5	81.1238	6.6828	67.6942	94.5534	24.7025	137.5451	5.3662
3	56.3	98.3051	8.8777	80.4648	116.1455	40.6745	155.9358	-42.0551
4	79.2	92.1874	4.5021	83.1400	101.2347	36.6458	147.7290	-12.9774
5	104.0	75.1844	10.0925	54.9027	95.4661	16.7519	133.6169	28.8556
6	68.9	74.7441	7.0568	60.5630	88.9252	18.1392	131.3490	-5.8541
7	77.4	71.2144	7.2723	56.6000	85.8287	14.4994	127.9294	6.2256
8	77.4	92.7168	8.8831	74.8654	110.5681	35.0827	150.3508	-15.2768
9	114.5	82.6170	6.1743	70.2092	95.0247	26.4301	138.8038	31.8730
10	136.9	95.0937	7.9527	79.1120	111.0753	38.0110	152.1763	41.7963
11	72.3	54.7788	15.5071	23.6161	85.9415	-8.2619	117.8195	17.4712

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
				95% CL Lower	95% CL Upper	95% CL Lower	95% CL Upper	
12	68.9	79.3233	7.3922	64.4681	94.1784	22.5458	136.1008	-10.4333
13	67.2	90.6827	4.5169	81.6055	99.7598	35.1362	146.2291	-23.4427
14	62.4	71.2903	7.6089	55.9997	86.5809	14.3973	128.1833	-8.8803
15	106.1	94.5324	5.7212	83.0352	106.0296	38.5395	150.5252	11.5576
16	54.8	72.8421	7.8660	57.0349	88.6494	15.8081	129.8762	-18.0821
17	92.2	74.1931	7.1200	59.8850	88.5013	17.5563	130.8300	17.9669
18	86.5	79.8061	6.7582	66.2250	93.3872	23.3485	136.2637	6.6839
19	112.4	82.8875	5.3481	72.1402	93.6349	27.0438	138.7312	29.4725
20	94.1	82.4384	7.2336	67.9019	96.9749	25.7434	139.1334	11.6516
21	134.6	107.4460	7.3029	92.7703	122.1217	50.7152	164.1769	27.1140
22	65.6	105.3760	7.9259	89.4484	121.3036	48.3085	162.4436	-39.7660
23	96.0	91.5012	3.8304	83.8038	99.1986	36.1635	146.8389	4.5388
24	54.8	81.5083	5.0057	71.4490	91.5676	25.7929	137.2237	-26.7483
25	88.4	85.9631	5.7398	74.4285	97.4976	29.9625	141.9636	2.3969
26	125.4	97.5262	5.8529	85.7645	109.2880	41.4785	153.5740	27.9138
27	82.8	92.5523	3.8965	84.7220	100.3825	37.1959	147.9086	-9.7423
28	110.3	81.7150	6.3185	69.0174	94.4125	25.4634	137.9666	28.5350
29	141.6	109.1038	8.7746	91.4705	126.7370	51.5369	166.6707	32.5062
30	70.6	83.8057	6.4553	70.8332	96.7781	27.4914	140.1200	-13.2457
31	25.0	83.1856	4.9332	73.2720	93.0992	27.4964	138.8749	-58.1856
32	96.0	90.1156	6.0696	77.9182	102.3130	33.9748	146.2564	5.9244
33	96.0	91.8356	4.9532	81.8819	101.7894	36.1392	147.5321	4.2044
34	116.6	95.9289	5.6996	84.4751	107.3828	39.9450	151.9129	20.7111
35	102.0	90.4547	6.9467	76.4947	104.4147	33.9047	147.0046	11.5553
36	118.8	108.1032	13.0425	81.8934	134.3131	47.3581	168.8484	10.7068
37	84.6	75.5483	8.9830	57.4964	93.6002	17.8518	133.2448	9.0917
38	68.9	102.8996	9.3765	84.0568	121.7424	44.9508	160.8484	-34.0096
39	53.3	74.1419	6.4591	61.1618	87.1220	17.8259	130.4580	-20.8519
40	88.4	99.4483	7.2313	84.9165	113.9802	42.7545	156.1421	-11.0883
41	88.4	88.1009	5.0221	78.0086	98.1933	32.3796	143.8223	0.2590
42	96.0	97.5355	6.4237	84.6267	110.4444	41.2358	153.8352	-1.4955
43	13.0	91.8616	5.5337	80.7412	102.9819	35.9449	147.7782	-78.9016
44	70.6	100.3790	8.3670	83.5649	117.1931	43.0577	157.7003	-29.8190
45	116.6	100.4390	5.7206	88.9429	111.9350	44.4464	156.4316	16.2010
46	102.0	73.4160	13.2624	46.7643	100.0678	12.4789	134.3531	28.5940
47	81.0	99.8156	7.6458	84.4509	115.1804	42.9027	156.7286	-18.8156
48	100.0	93.1010	7.9306	77.1638	109.0381	36.0308	150.1711	6.8990
49	127.7	114.8349	9.0734	96.6011	133.0686	57.0812	172.5885	12.8552
50	127.7	103.8916	6.4270	90.9761	116.8071	47.5904	160.1927	23.7984
51	163.8	105.5385	9.0353	87.3814	123.6957	47.8090	163.2680	58.3015
52	100.0	89.7939	7.3869	74.9494	104.6385	33.0192	146.5687	10.2061
53	44.9	94.9751	4.3350	86.2636	103.6867	39.4872	150.4630	-50.0851

Sum of Residuals	0
Sum of Squared Residuals	36437
Predicted Residual SS (PRESS)	42693



```

* Create a pointer named HD to the data file;
filename HD "/home/u63986019/health.csv";

DATA c; /* Assign name c to data */
INFILE HD DSD FIRSTOBS = 2; /* Since the data is a CSV, use DSD FIRSTOBS = 2*/
INPUT DeathRate DrAvail HospAvail CapInc PopD; /*Input names of columns*/
RUN;

/* 1 */

/* make a multi-celled scatter plot*/
PROC SGSCATTER DATA = c;
TITLE "Multi scatter plot for Death Rate vs all other predictors";
PLOT (DeathRate)*(DrAvail HospAvail CapInc PopD);
RUN;

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

/* Check correlation coefficients*/
PROC CORR DATA = c plots = matrix;
VAR DeathRate DrAvail HospAvail CapInc PopD;
RUN;

/* Lack of fit test to check linearity*/
PROC REG DATA = c;
MODEL DeathRate = DrAvail HospAvail CapInc PopD / lackfit;
OUTPUT OUT=D RSTUDENT=R PREDICTED=P;
RUN;

/* Residual plot*/
PROC PLOT Data = D;
plot R*(P DrAvail HospAvail CapInc PopD);
RUN;

PROC SGSCATTER Data = D;
TITLE "Residual Plot of death rate and predictors ";
plot R*(P DrAvail HospAvail CapInc PopD);
RUN;

*hosp avail 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 SGSCATTER DATA = D;
SCATTER x = P Y = absRes;
RUN;

*No noticeable patterns in Residual plot but presence of outliers;

/* Breusch Pagan Test*/
PROC MODEL DATA = D;
PARMS b0 b1 b2 b3 b4;
DeathRate = b0 + b1*DrAvail + b2*HospAvail + b3*CapInc + b4*PopD;
/*Breusch-Pagan test for heteroscedasticity (BREUSCH) wrt the specified predictors*/
fit DeathRate /WHITE BREUSCH=(DrAvail HospAvail CapInc PopD);
*Breusch-Pagan test for heteroscedasticity individually for each of the predictors;
fit DeathRate /BREUSCH = (DrAvail);
fit DeathRate /BREUSCH = (HospAvail);
fit DeathRate /BREUSCH = (CapInc);
fit DeathRate /BREUSCH = (PopD);
RUN;

*P value is very high so dont RH0 => there is constant variance;

/* Brown Forsythe Test for homogeneity of variance*/
/* Get the medians of the predictors*/
PROC UNIVARIATE DATA = D NOPRINT;
VAR DrAvail HospAvail CapInc PopD;
OUTPUT OUT = Medians Median = MedDrA Median = MedHosp Median = MedInc Median = MedPopD N=N;
RUN;

```

```

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

/*Test for homogeneity of residuals grouped by DR availability*/
DATA DrABF;
MERGE D Medians;
Group = (DrAvail > MedDrA);
RUN;

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

/*Test for homogeneity of residuals grouped by hosp avail*/
DATA HoABF;
MERGE D Medians;
Group = (HospAvail > MedHosp);
RUN;

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

/*Test for homogeneity of residuals grouped by cap inc*/
DATA IncBF;
MERGE D Medians;
Group = (CapInc > MedInc);
RUN;

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

/*Test for homogeneity of residuals grouped by pop den*/
DATA PopDBF;
MERGE D Medians;
Group = (PopD > MedPopD);
RUN;

PROC GLM Data = PopDBF;
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;

*p val < 0.0001 implies RH0 => Normality assumption is violated so do transformation on Y.
QQ plot is right tailed so SHRINK values by doing log or sqrt transfromation;
*-----;
/* 2 */

*log transformation on Y;
Data c_log;
SET c;
log_DeathRate = log(DeathRate);
RUN;

PROC REG DATA = c_log;
MODEL log_DeathRate = DrAvail HospAvail CapInc PopD/ lackfit;
/*save the studentized residuals and the predicted values */
OUTPUT OUT=D_log RSTUDENT=R_log PREDICTED=P_log;

```

```

RUN;

/*Residual QQ plot*/
PROC UNIVARIATE DATA = D_log NORMAL PLOT; /* Check normality of the studentized residuals */
VAR R_log;
RUN;

*QQ plot is still heavily skewed to the right so use BOXCOX transformation to figure out best lambda;

*Find best transformation lambda using boxcox;
PROC TRANSREG DATA=c; /* Find Box-Cox transformation power */
MODEL BoxCox(DeathRate)=identity(DrAvail HospAvail CapInc PopD);
RUN;

DATA c; SET c;
DeathRate225 = (DeathRate**2.25-1)/2.25;
RUN;

PROC REG Data = c NOPRINT;
MODEL DeathRate225 = DrAvail HospAvail CapInc PopD;
OUTPUT OUT=F RSTUDENT=R_new;
RUN;

PROC UNIVARIATE DATA=F NORMAL PLOT; /* Check normality of Studentized residuals */
VAR R_new;
RUN;

*boxcox lambda = 2.25 implies square transformation is apt;

/* Square Transformation on y */
Data c_sq;
SET c;
Sq_DeathRate = DeathRate**2;
RUN;

PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD/ lackfit;
/*save the studentized residuals and the predicted values */
OUTPUT OUT=D_sq RSTUDENT=R_sq PREDICTED=P_sq;
RUN;

/*Residual QQ plot does not differ as much from QQplot with (deathrate)^2.25 */
PROC UNIVARIATE DATA = D_sq NORMAL PLOT; /* Check normality of the studentized residuals */
VAR R_sq;
RUN;

-----Diagnostics for transformed data-----

/* make a multi-celled scatter plot*/
PROC SGSCATTER DATA = c_sq;
TITLE "Multi scatter plot for Death Rate vs all other predictors";
PLOT (Sq_DeathRate)*(DrAvail HospAvail CapInc PopD);
RUN;

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

/* Check correlation coefficients*/
PROC CORR DATA = c_sq plots = matrix;
VAR Sq_DeathRate DrAvail HospAvail CapInc PopD;
RUN;

/* Lack of fit test to check linearity*/
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD / lackfit;
OUTPUT OUT=D RSTUDENT=R PREDICTED=P;
RUN;

/* Residual plot*/
PROC PLOT Data = D;
plot R*(P DrAvail HospAvail CapInc PopD);
RUN;

PROC SGSCATTER Data = D;
TITLE "Residual Plot of death rate and predictors ";
plot R*(P DrAvail HospAvail CapInc PopD);
RUN;

```

```

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

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

*No noticeable patterns in Residual plot but presence of outliers;

/* Breusch Pagan Test*/
PROC MODEL DATA = D;
PARMS b0 b1 b2 b3 b4;
Sq_DeathRate = b0 + b1*DrAvail + b2*HospAvail + b3*CapInc + b4*PopD;
/*Breusch-Pagan test for heteroscedasticity (BREUSCH) wrt the specified predictors*/
fit Sq_DeathRate /WHITE BREUSCH=(DrAvail HospAvail CapInc PopD);
*Breusch-Pagan test for heteroscedasticity individually for each of the predictors;
fit Sq_DeathRate /BREUSCH = (DrAvail);
fit Sq_DeathRate /BREUSCH = (HospAvail);
fit Sq_DeathRate /BREUSCH = (CapInc);
fit Sq_DeathRate /BREUSCH = (PopD);
RUN;

*p value is very high so dont RH0 => there is constant variance;

*-----;

/* 3 */

/* Prints both Type I and Type III SS along with partial F tests */
PROC GLM DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD;
RUN;

*Drop one variable at a time. In the above code, CapInc and PopD are significant;
*Again CapInc and PopD are significant;
PROC GLM DATA = c_sq;
MODEL Sq_DeathRate = HospAvail CapInc PopD;
RUN;

*Here, only PopD is significant. So, we drop CapInc;
PROC GLM DATA = c_sq;
MODEL Sq_DeathRate = CapInc PopD;
RUN;

*Tentative model Extra SS;
PROC GLM DATA = c_sq;
MODEL Sq_DeathRate = PopD;
RUN;

*Partial F test;
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD;
x1x2x3: test DrAvail = HospAvail = CapInc = 0;
RUN;

PROC CORR DATA = c_sq;
VAR Sq_DeathRate PopD;
PARTIAL DrAvail HospAvail CapInc;
RUN;

*-----;

/* 4 */

/* Model selection */
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD / selection = adjrsq cp bic; /* selection is based on Adj R^2 */
RUN;

PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD / selection = cp adjrsq bic; /* selection is based on cp */
RUN;

```

```
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD / selection = stepwise; /* stepwise selection */
RUN;
```

```
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD / selection = f; /* forward selection*/
RUN;
```

```
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD/ selection = b; /* backward selection*/
RUN;
```

```
*-----;
```

```
/* 5 */
```

```
/* Note: Coefficient of partial determination is square of partial correlation coefficient */
```

```
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail HospAvail CapInc PopD;
x2: test HospAvail = 0;
RUN;
```

```
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail CapInc PopD;
output out=predicted predicted=P;
RUN;
```

```
PROC CORR DATA = c_sq;
VAR Sq_DeathRate DrAvail;
* gives partial correlation coefficient between deathrate, popD after adjusting for hospavail capinc and drAvail;
PARTIAL CapInc PopD;
RUN;
```

```
PROC CORR DATA = c_sq;
VAR Sq_DeathRate CapInc;
* gives partial correlation coefficient between deathrate, popD after adjusting for hospavail capinc and drAvail;
PARTIAL DrAvail PopD;
RUN;
```

```
PROC CORR DATA = c_sq;
VAR Sq_DeathRate PopD;
* gives partial correlation coefficient between deathrate, popD after adjusting for hospavail capinc and drAvail;
PARTIAL DrAvail CapInc;
RUN;
```

```
*-----;
```

```
/* 6 */
```

```
*Alternative interpretation in terms of simple determination;
```

```
*Y on X3 and X4;
proc reg data = c_sq;
model Sq_Deathrate = CapInc PopD;
Output out =X3X4 PREDICTED=P1 RSTUDENT=R_X3X4;
RUN;
```

```
*X1 on X3 and X4;
proc reg data = c_sq;
model DrAvail = CapInc PopD;
Output OUT = X1GX3X4 PREDICTED=P2 RSTUDENT=R_X1gX3X4;
RUN;
```

```
/* Merge the datasets to calculate correlation between variables of two different datasets */
data merged;
merge X3X4 X1GX3X4;
run;
```

```
/* Correlation between residuals of the above regression models */
proc corr data=merged;
var R_X3X4 R_X1gX3X4;
run;
```

```
*Alternative interpretation of coefficient of multiple determination;
```

```
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail CapInc PopD;
output out=predicted predicted=P;
```

```
RUN;

proc corr data = predicted;
var Sq_DeathRate P;
run;

*-----;

/* 7 */

*95% interval estimates for mean response and prediction intervals for the death rate of a new city;
PROC REG DATA = c_sq;
MODEL Sq_DeathRate = DrAvail CapInc PopD /CLM CLI XPX I ALPHA = 0.05; /* BEST and FINAL model obtained by forward selection*/
OUTPUT OUT=D RSTUDENT=R PREDICTED=P LCLM=lowerCI UCLM=upperCI LCL=lowerPI UCL=upperPI;
RUN;
```