Joe Griffin
Assignment 4
April 11, 2021

Question 1 attached. Calculations were done on excel.

Question 2:

a. Y hat = **-1.26902 + 0.36499 (x6) - 0.43635 (x7) + 0.22577 (x9) + 0.17655 (x11) + 0.78167 (x12) + 0.15911 (x16)**
b. R^2 = 1-(SSE/SST) = 1 - (28.50808 / 140.62760) = **0.79727962363**

   Standard error estimate = sqrt(SSE/N) = sqrt(28.50808/100) = **0.53392958337**

   The model is pretty accurate. The R^2 value is not quite 1, but it is still very high and definitely shows high correlation. The standard error estimate also supports this idea.

c. I would remove x7 and x16 as these variables have the two lowest parameters.
d. R^2 = 1-(SSE/SST) = 1 - (32.55145 /140.62760) = **0.7685273019**

   Standard error estimate = sqrt(SSE/N) = sqrt(32.55145/100) = **0.57053878045**

   The standard error estimate and R^2 go up slightly in this model, but that will happen when you remove variables. Even though these values go up slightly, there are now no negative parameters, and even more significantly the F value increases significantly from 60.96 to 78.85. The F score is another important factor in determining the accuracy of a model, and higher F scores signify more accurate models.

LEWIS UNIVERSITY

Multivariate Data Analysis: Multiple Regression Analysis

Note: Submit your solutions in one single PDF file.

1. **(By Hand)** For the dependent variable Y and the independent variables X1 and X2, the linear regression model is given by:

   Y=0.08059*X1-0.16109*X2+5.26570. Complete the following table:

Actual Y	x1	X2	Predicted Y	Residuals (Predication Error)
6	6.8	4.7	5.057	0.943
3.1	5.3	5.5	4.807	1.707
5.8	4.5	6.2	4.630	-1.170
4.5	8.8	7	4.847	0.347
4.5	6.8	6.1	4.831	0.331
3.7	8.5	5.1	5.130	1.429
5.4	8.9	4.8	5.210	-0.190
5.1	6.9	5.4	4.952	-0.148
5.8	9.3	5.9	5.065	-0.735
5.7	8.4	5.4	5.073	-0.627

$SST = \sum (y-\bar{y})^2 = 8.724$ $\quad \bar{y} = 4.96$

Is this a good model? Why? Why not?

$SSE = \sum (y-\hat{y})^2 = \sum (\text{residuals}^2) = 8.438$

$R^2 = 1 - SSE/SST = 1 - 8.438/8.724 = 0.0328$

2. For the data set associated with this homework (HBAT). Using X19 as the dependent variable and (X6, X7, X9, X11, X12 and X16) as the independent variables:

   a. Find the parameters (coefficients) for the Linear Regression Model, then write down the equation of the model.
   b. Find the coefficient of determination and the standard error of the estimate. How accurate is the model?
   c. If you are asked to remove two independent variables, which two variables would you choose and Why?
   d. After removing the two variables found in part c, re-run parts a and b. Compare the results. Which model is more accurate and why?

This is not a good model, as the R^2 value is very low. You want the R^2 value to be as close to 1 as possible. Some of the residuals are very large, which explain the large SSE value, and as a result produce a small R^2 value.

Model: MODEL1
Dependent Variable: x19 x19

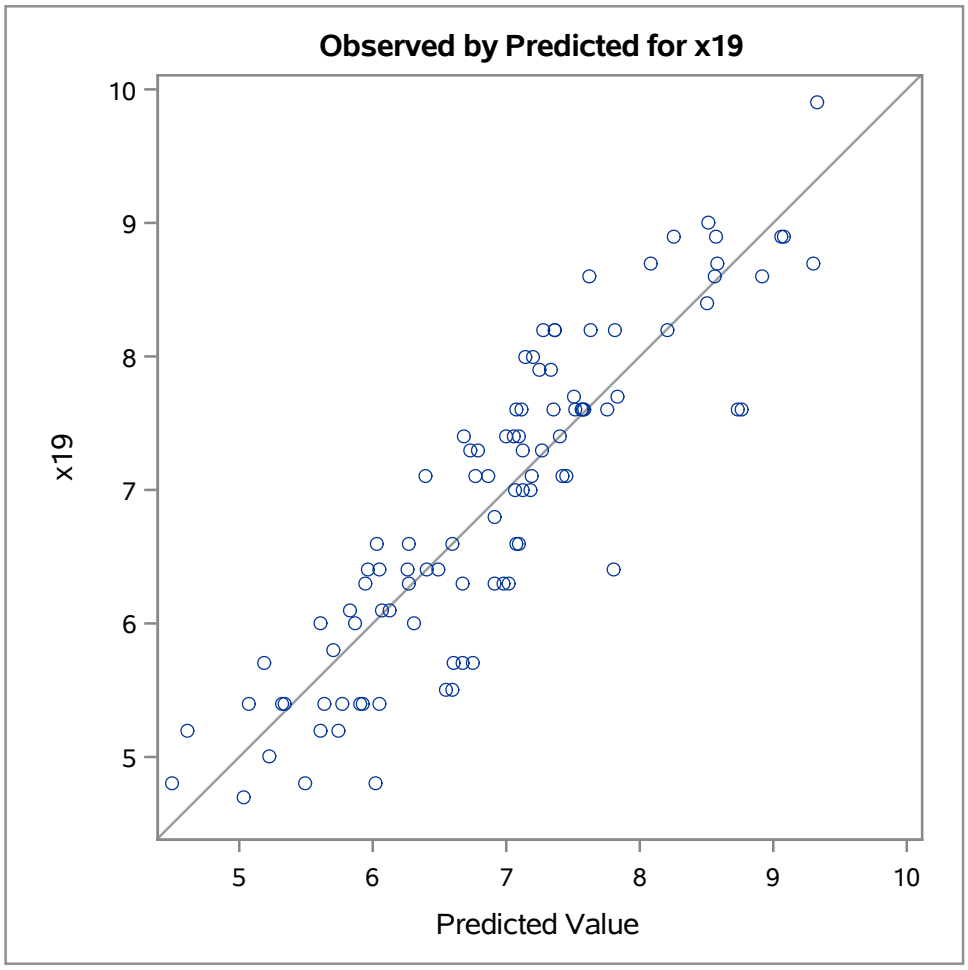
Number of Observations Read	100
Number of Observations Used	100

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	6	112.11952	18.68659	60.96	<.0001
Error	93	28.50808	0.30654		
Corrected Total	99	140.62760			

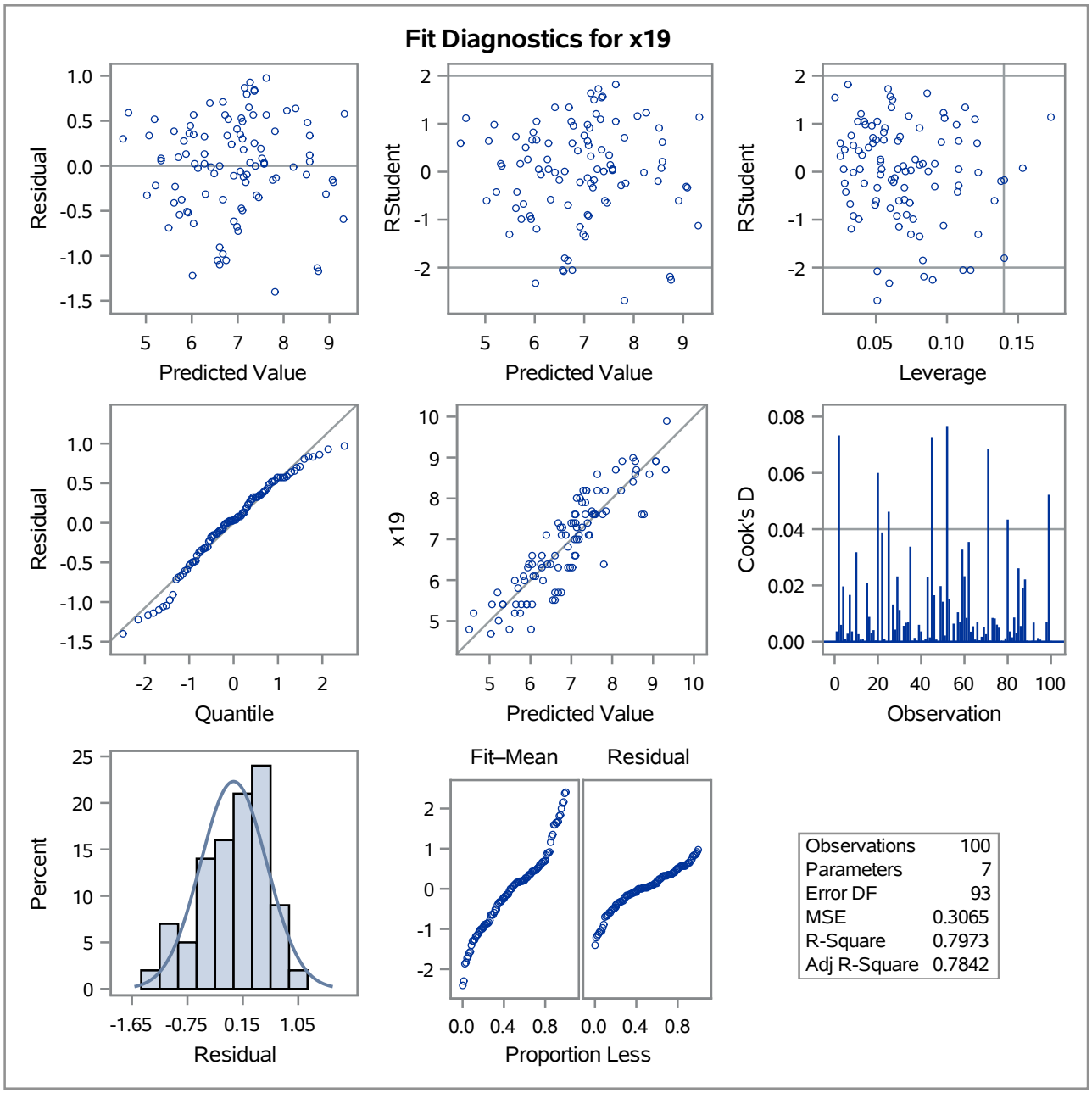
Root MSE	0.55366	**R-Square**	0.7973
Dependent Mean	6.91800	**Adj R-Sq**	0.7842
Coeff Var	8.00317		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	-1.26902	0.49935	-2.54	0.0127
x6	x6	1	0.36499	0.04676	7.81	<.0001
x7	x7	1	-0.43635	0.13103	-3.33	0.0012
x9	x9	1	0.22577	0.08074	2.80	0.0063
x11	x11	1	0.17655	0.06034	2.93	0.0043
x12	x12	1	0.78167	0.08814	8.87	<.0001
x16	x16	1	0.15911	0.09215	1.73	0.0875

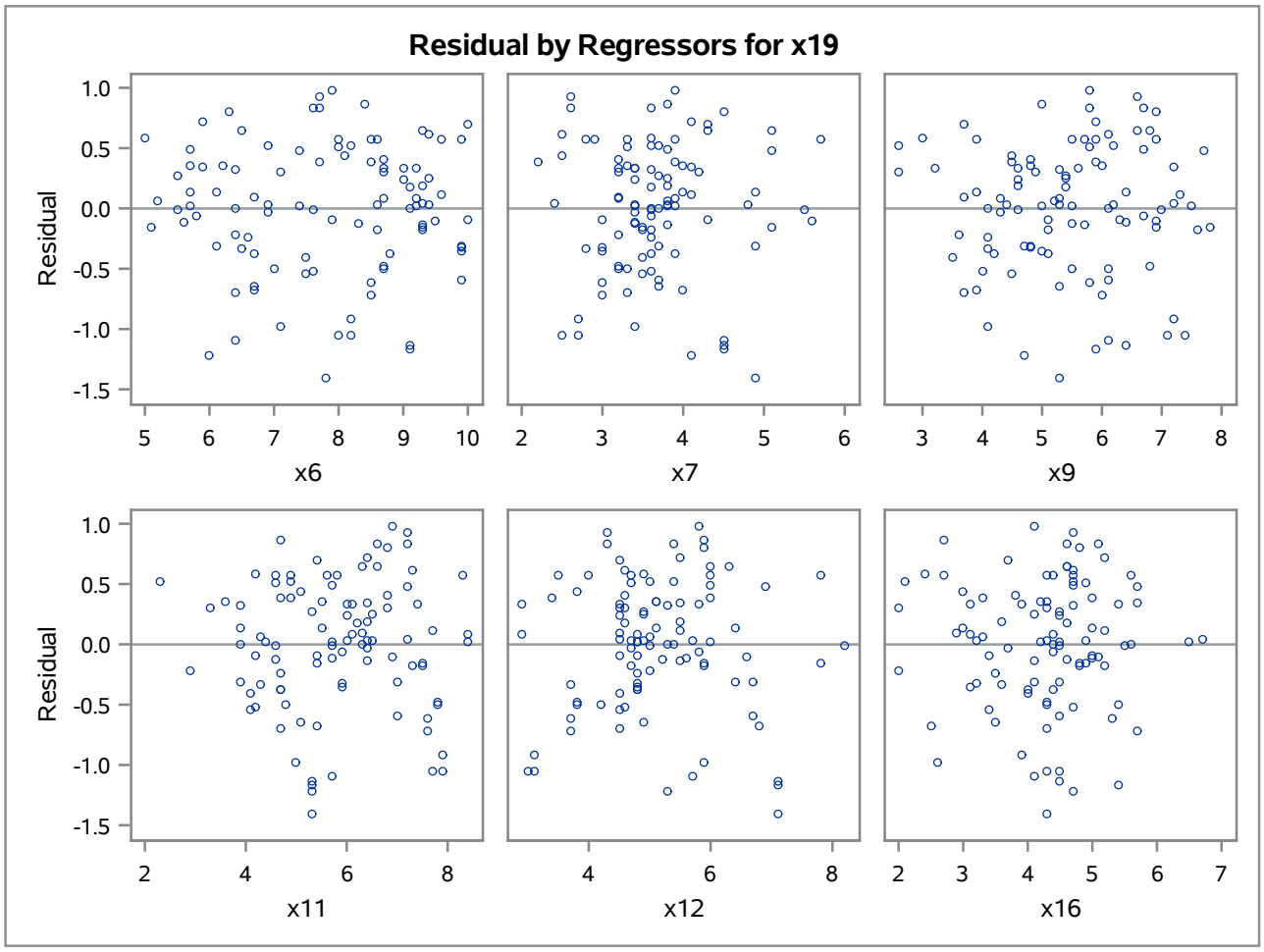
Model: MODEL1
Dependent Variable: x19 x19



Model: MODEL1
Dependent Variable: x19 x19



Model: MODEL1
Dependent Variable: x19 x19



Model: MODEL1
Dependent Variable: x19 x19

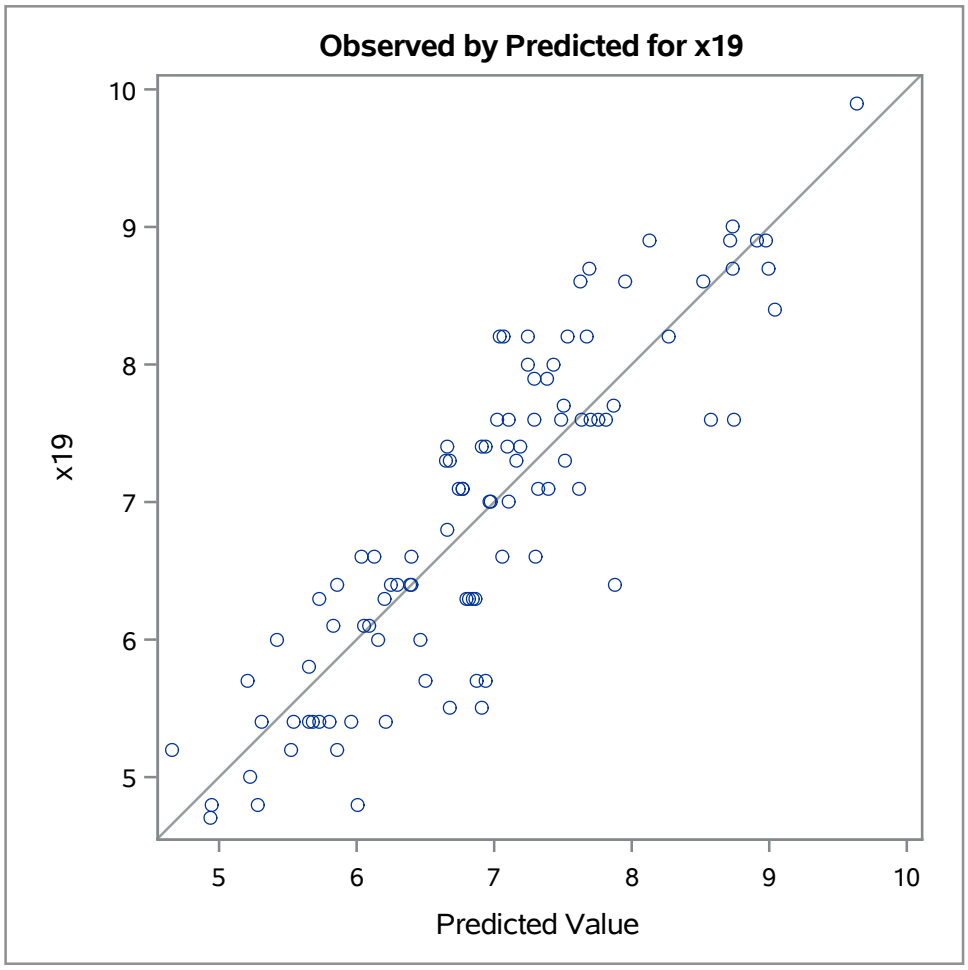
Number of Observations Read	100
Number of Observations Used	100

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	4	108.07615	27.01904	78.85	<.0001
Error	95	32.55145	0.34265		
Corrected Total	99	140.62760			

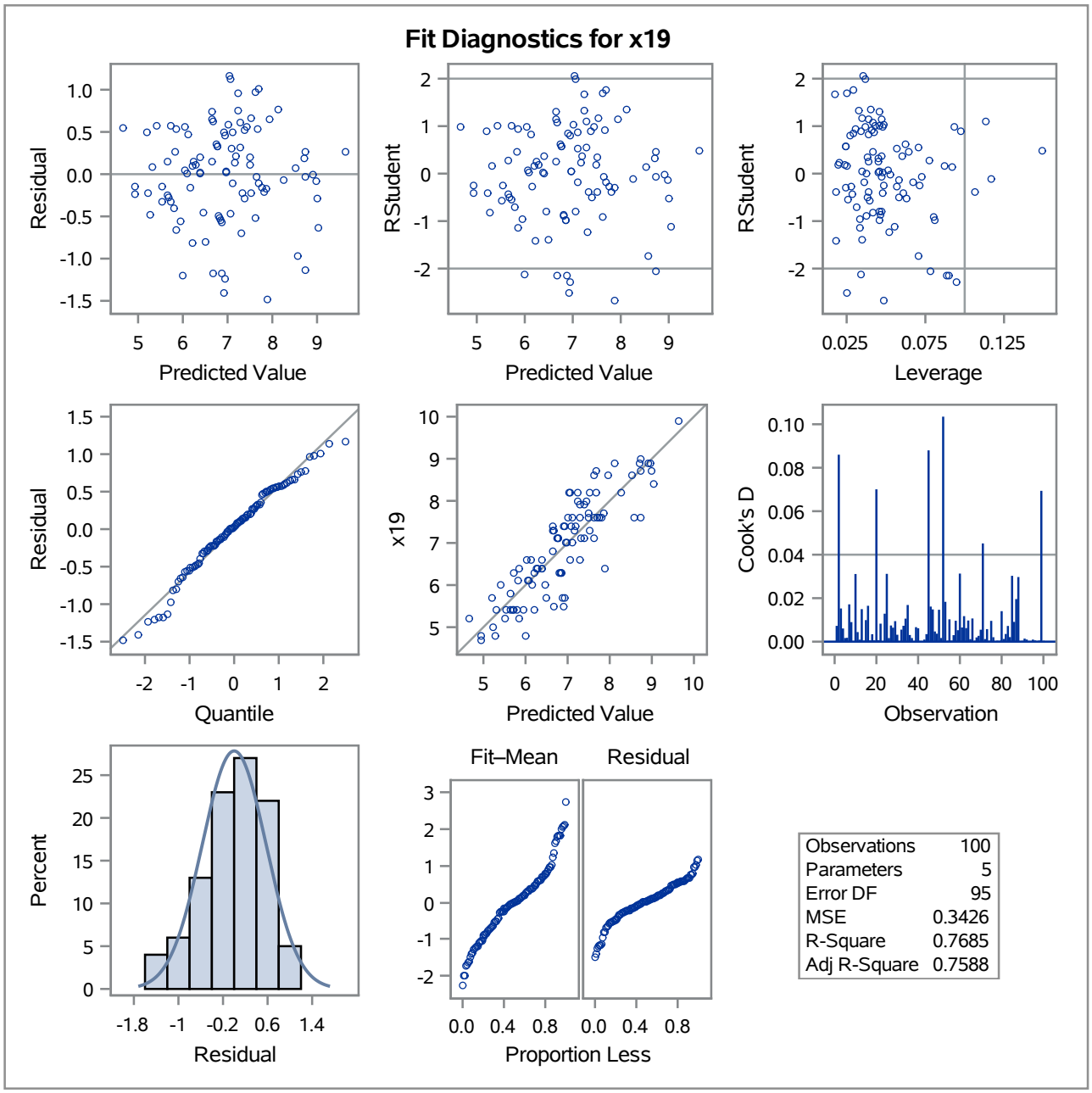
Root MSE	0.58536	**R-Square**	0.7685
Dependent Mean	6.91800	**Adj R-Sq**	0.7588
Coeff Var	8.46141		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	-1.63599	0.49775	-3.29	0.0014
x6	x6	1	0.37568	0.04932	7.62	<.0001
x9	x9	1	0.33604	0.06324	5.31	<.0001
x11	x11	1	0.16288	0.06366	2.56	0.0121
x12	x12	1	0.55547	0.05817	9.55	<.0001

Model: MODEL1
Dependent Variable: x19 x19



Model: MODEL1
Dependent Variable: x19 x19



Model: MODEL1
Dependent Variable: x19 x19