Q1 (10 points)

Q1(10 points). Consider the linear regression model
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$.

1(5). The residuals are listed below:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
0.2, 0.3, -0.8, -0.8, -0.3, 0.4, 0.1,-0.1, -0.4, -0.7, 0.6, -0.1, -0.1, 0.3,0.2.

Do a Durbin-Watson test H_0 : the error terms are not (first-order) autocorrelated. H_a: the error terms are negatively or positively (first-order) autocorrelated. Let $\alpha = 0.1$.

2(5). The residuals are: 0.2, 0.3, -0.8, -0.8, -0.3, calculate $e_{(i)}$, $z_{(i)}$ and construct the normal plot.

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

① H_0: No A.C. (autocorrelated)

H_a: + A.C. or − A.C.

$n = t = 15$, $\alpha = 0.1$, $K-1 = 3-1 = 2$, two Indep. variables

$$T.S.:\ d = \frac{\sum_{t=2}^{15}(e_t - e_{t-1})^2}{\sum_{t=1}^{15} e_t^2}$$

$$= \frac{(0.3-0.2)^2 + (-0.8-0.3)^2 + \dots + (0.2-0.3)^2}{(0.2)^2 + (0.3)^2 + \dots + (0.2)^2}$$

$$= 1.6268$$

$d_{L,\alpha/2} = d_{L,0.05} = 0.95$, $d_{U,\alpha/2} = d_{U,0.05} = 1.54$

$4 - d_{U,0.05} = 4 - 1.54 = 2.46 > d_{U,0.05}$

Since $d = 1.6268 > d_{U,0.05}$ and $4 - d_{U,\alpha/2} > d_{U,0.05}$, we do not reject H_0.

②

i	$e_{(i)}$	$(3i-1)/(3n+1)$	$z_{(i)}$
1	−0.8	$[3(1)-1]/[3(5)+1] = 0.125$	$z_{(1)} = -1.15$
2	−0.8	$[3(2)-1]/[3(5)+1] = 0.3125$	−0.489
3	−0.3	$[3(3)-1]/[3(5)+1] = 0.5$	0
4	0.2	$[3(4)-1]/[3(5)+1] = 0.6875$	0.489
5	0.3	$[3(5)-1]/[3(5)+1] = 0.875$	1.15

$P(Z \le z_{(1)}) = 0.125$

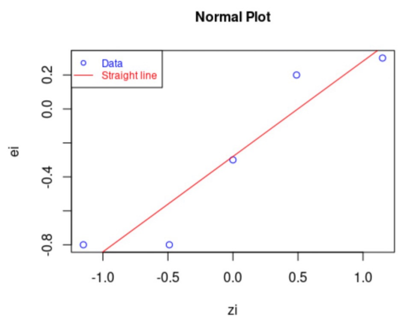
$P(Z \le z_{(2)}) = 0.3125$

$P(Z \le z_{(3)}) = 0.5$

$P(Z \le z_{(4)}) = 0.6875$

$P(Z \le z_{(5)}) = 0.875$

$n = 5$



Q2 (20 points)

Q2(20 points). The sample data set is show as following

x: 1, 0, 3, 2

y: 4, 3, 1, 3

Apply the matrix form to calculate

1(9 points) The least squares estimates of β_0, β_1.

2(4 points) SS_{Res}, SS_R, SS_T, multiple coefficient of determination R^2.

3(2 points) Predict y_0 given $x_0 = [1, 2]$.

4(5 points) 90% prediction interval for y_0 given $x_0 = [1, 2]$.

1

$$\tilde{X} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 3 \\ 1 & 2 \end{bmatrix}, \quad \tilde{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 3 & 2 \end{bmatrix}, \quad \tilde{y} = \begin{bmatrix} 4 \\ 3 \\ 1 \\ 3 \end{bmatrix}$$

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

$$(\tilde{X}'\tilde{X}) = \underset{2\times4}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 3 & 2 \end{bmatrix}} \cdot \underset{4\times2}{\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 3 \\ 1 & 2 \end{bmatrix}}$$

$$= \underset{2\times2}{\begin{bmatrix} 1(1)+1(1)+1(1)+1(1) & 1(1)+1(0)+1(3)+1(2) \\ 1(1)+0(1)+3(1)+2(1) & 1(1)+0(0)+3(3)+2(2) \end{bmatrix}}$$

$$= \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}$$

$$(\tilde{X}'\tilde{X})^{-1} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}^{-1}$$

$$= \frac{1}{4(14) - 6(6)} \begin{bmatrix} 14 & -6 \\ -6 & 4 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 14 & -6 \\ -6 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 14/20 & -6/20 \\ -6/20 & 4/20 \end{bmatrix}$$

$$= \begin{bmatrix} 7/10 & -3/10 \\ -3/10 & 1/5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.7 & -0.3 \\ -0.3 & 0.2 \end{bmatrix}$$

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

$$= \begin{bmatrix} 7/10 & -3/10 \\ -3/10 & 1/5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \\ 1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 3.8 \\ -0.7 \end{bmatrix}$$

$\therefore \beta_0 = 3.8 \quad , \quad \beta_1 = -0.7$

② $SS_{Res} = \tilde{y}'\tilde{y} - \hat{\beta}'\tilde{X}'\tilde{y}$

$$= \begin{bmatrix} 4 & 3 & 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 3.8 & -0.7 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \\ 1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 4(4)+3(3)+1(1)+3(3) \end{bmatrix} - \begin{bmatrix} 3.1 & 3.8 & 1.7 & 2.4 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \\ 1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 35 \end{bmatrix} - \begin{bmatrix} 3.1 & 3.8 & 1.7 & 2.4 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \\ 1 \\ 3 \end{bmatrix}$$

$$= 35 - 32.7$$

$$= 2.3$$

$SS_R = \hat{\beta}\tilde{X}'\tilde{y} - n\bar{y}^2$

$$= 32.7 - 4\left(\frac{4+3+1+3}{4}\right)^2$$

$$= 32.7 - 4(2.75)^2$$

$$= 2.45$$

$SS_T = \tilde{y}'\tilde{y} - n\bar{y}^2$

$$= 35 - 4(2.75)^2$$

$$= 4.75$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{2.45}{4.75} = 0.5158$$

③ $y_0 = \tilde{x}_0 \beta = \begin{bmatrix} 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3.8 \\ -0.7 \end{bmatrix} = 2.4$

④ $\sigma = \sqrt{\frac{SS_{Res}}{n-k}} = \sqrt{\frac{2.3}{2}} = 1.0724$

$h_{00} = \tilde{x}_0' (\tilde{X}'\tilde{X})^{-1} \tilde{x}_0 = \begin{bmatrix} 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.7 & -0.3 \\ -0.3 & 0.2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 0.3$

$t_{0.05}(2) = 2.92$

$\therefore y_0 \pm 2.92 (1.0724) \sqrt{1+0.3} = (-1.1704, 5.9704)$

Q3 (20 points)

Q3(20 points). Apply SAS to work on the following sample data.

i	x_1	x_2	y
1	1.43	2.79	1.23
2	7.90	5.59	6.12
3	-3.40	3.58	-1.90
4	52.87	9.73	44.61
5	54.39	9.32	41.28
6	4.70	0.13	10.56
7	39.68	8.91	34.78
8	21.75	7.03	16.57
9	12.62	6.46	10.09
10	-1.85	2.97	2.01

$\alpha = 0.01$ Find the following information from the SAS output.

1.Fit the data to linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \epsilon$.

a. (3 points)Specify the least squares point estimators on your output.

b. (3 points)Implement an F-test for a portion of the model: $H_0 : \beta_2 = \beta_3 = 0$, H_a : at least one of β_2, β_3 is not 0. Specify the test statistic $F(x_2, x_2^2|x_1)$ value and P-value on your output, and make conclusion.

c. (4 points)Code to display $\tilde{X}'\tilde{X}, (\tilde{X}'\tilde{X})^{-1}, \tilde{X}'\tilde{y}, \tilde{y}'\tilde{y}$, specify these matrices/vectors/numbers on your output.

d. (3 points)Implement a Durbin-Watson test for first-order autocorrelation test (H_0: no autocorrelation, H_a: positive autocorrelation). Specify the value of the test statistic and P-value on your output and give the conclusion.

e. (2 points)Use the "Residual by Regressors for y" panel of the regression model output to validate the constant variance assumption.

f. (2 points)Calculate the prediction of y_0 for $x_{01} = 10, x_{02} = 5$.

2 (3 points) Fit the data to linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \epsilon$, compare this with the above model (in sub-question 1.) using multiple coefficients of determination R^2.

1.Fit the data to linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \epsilon$.

a. (3 points)Specify the least squares point estimators on your output.

Parameter Estimate
$\beta_0 =$ 9.91079
$\beta_1 =$ 0.30646
$\beta_2 =$ -5.24023
$\beta_3 =$ 0.73677

b. (3 points)Implement an F-test for a portion of the model: $H_0 : \beta_2 = \beta_3 = 0$, H_a : at least one of β_2, β_3 is not 0. Specify the test statistic $F(x_2, x_2^2|x_1)$ value and P-value on your output, and make conclusion.

The REG Procedure
Model: MODEL1

Test mytest Results for Dependent Variable y

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	25.63895	22.63	0.0016
Denominator	6	1.13318		

$\rightarrow$ p-value

T.S. $F(x_2, x_2^2 | x_1)$

$\boxed{\text{conclusion}}$: since p-value $= 0.0016 < \alpha = 0.01$, we reject H_0, and conclude that at least one of x_2 and x_2^2 is significant.

c. (4 points)Code to display $\tilde{X}'\tilde{X}$, $(\tilde{X}'\tilde{X})^{-1}$, $\tilde{X}'\tilde{y}$, $\tilde{y}'\tilde{y}$, specify these matrices/vectors/numbers on your output.

$\tilde{X}'\tilde{X}$ $\tilde{X}'\tilde{y}$ $\tilde{y}'\tilde{y}$

Model Crossproducts X'X X'Y Y'Y					
Variable	Intercept	x1	x2	x2sq	y
Intercept	10	190.09	56.51	412.7623	165.35
x1	190.09	8061.8657	1640.4116	14679.655592	6574.034
x2	56.51	1640.4116	412.7623	3323.565419	1348.5262
x2sq	412.7623	14679.655592	3323.565419	28273.561383	12004.502344
y	165.35	6574.034	1348.5262	12004.502344	5438.2429

$(\tilde{X}'\tilde{X})^{-1}$

X'X Inverse, Parameter Estimates, and SSE					
Variable	Intercept	x1	x2	x2sq	y
Intercept	1.4489080779	-0.085966558	-0.856100865	0.1241163792	9.9107930382
x1	-0.085966558	0.0140719141	0.0854121885	-0.016091361	0.306459682
x2	-0.856100865	0.0854121885	0.684714617	-0.112336367	-5.24022871
x2sq	0.1241163792	-0.016091361	-0.112336367	0.0197832292	0.73677386
y	9.9107930382	0.306459682	-5.24022871	0.73677386	6.7990811384

d. (3 points)Implement a Durbin-Watson test for first-order autocorrelation test (H_0: no autocorrelation, H_a: positive autocorrelation). Specify the value of the test statistic and P-value on your output and give the conclusion.

Durbin-Watson D	1.702
Pr < DW	0.2927
Pr > DW	0.7073
Number of Observations	10
1st Order Autocorrelation	-0.075

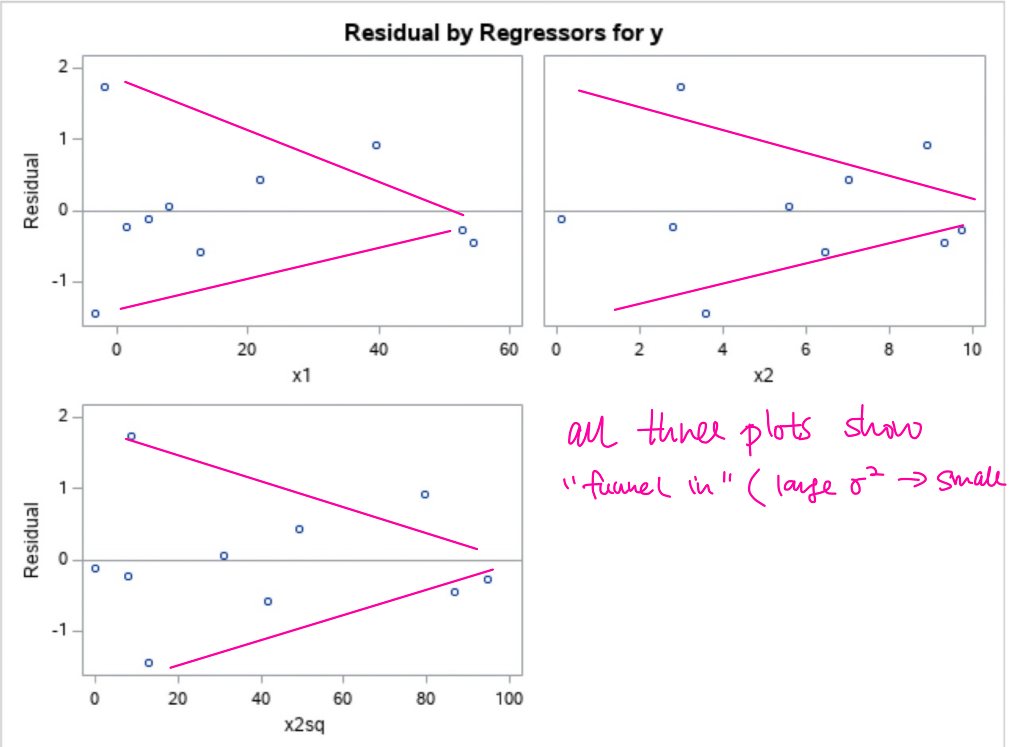
T.S.: d

P-value +A.C.

P-value −A.C.

P-values > $\alpha = 0.01$ ⇒ do not reject H_0 (conclusion)

e. (2 points)Use the "Residual by Regressors for y" panel of the regression model output to validate the constant variance assumption.



all three plots show "funnel in" (large $\sigma^2 \rightarrow$ small σ^2)

f. (2 points)Calculate the prediction of y_0 for $x_{01} = 10, x_{02} = 5$.

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	99% CL Mean		99% CL Predict		Residual
1	1.23	1.4639	0.6383	-0.9027	3.8305	-3.1379	6.0657	-0.2339
2	6.12	6.0617	0.5012	4.2034	7.9201	1.6995	10.4239	0.0583
3	-1.90	-0.4484	0.5709	-2.5649	1.6681	-4.9267	4.0299	-1.4516
4	44.61	44.8784	0.7555	42.0775	47.6793	40.0389	49.7179	-0.2684
5	41.28	41.7381	0.9124	38.3554	45.1209	36.5402	46.9361	-0.4581
6	10.56	10.6824	0.9792	7.0520	14.3128	5.3200	16.0448	-0.1224
7	34.78	33.8718	0.6465	31.4750	36.2685	29.2544	38.4891	0.9082
8	16.57	16.1495	0.5038	14.2818	18.0172	11.7833	20.5157	0.4205
9	10.09	10.6732	0.5371	8.6820	12.6644	6.2527	15.0937	-0.5832
10	2.01	0.2794	0.4751	-1.4822	2.0410	-4.0425	4.6013	1.7306
11	.	5.1936	0.8748	1.9502	8.4370	0.0852	10.3020	.

y_0 given $\tilde{x}_0' = [1 \quad 10 \quad 5 \quad 25]$

$= 5.1936$

2 (3 points) Fit the data to linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + \epsilon$, compare this with the above model (in sub-question 1.) using multiple coefficients of determination R^2.

Root MSE	1.06451	**R-Square**	0.9975
Dependent Mean	16.53500	**Adj R-Sq**	0.9962
Coeff Var	6.43792		

→ R^2 (1st model)

Root MSE	2.38424	**R-Square**	0.9874
Dependent Mean	16.53500	**Adj R-Sq**	0.9811
Coeff Var	14.41932		

→ R^2 (2nd model)

Both R^2 are close to 1, so they both indicate a good fit model. The R^2 of the first model in sub–question 1 is greater than the R^2 of the second model (0.9975 > 0.9874), so the first model is better.

```
data A2_Q3; /* Assigns name A2_Q3 to file*/
input i x1 x2 y; /* assigns variable names i x1 x2 y*/
x2sq = x2*x2; /*interaction term data*/
t = _N_;
cards; /* start to input data */
1 1.43 2.79 1.23
2 7.90 5.59 6.12
3 -3.40 3.58 -1.90
4 52.87 9.73 44.61
5 54.39 9.32 41.28
6 4.70 0.13 10.56
7 39.68 8.91 34.78
8 21.75 7.03 16.57
9 12.62 6.46 10.09
10 -1.85 2.97 2.01

; /* inputs data, .= missing value*/
proc reg data = A2_Q3; /* specifies regression procedure */
model y = x1 x2 x2sq /alpha=0.01 p clm cli clb xpx i dwprob;
mytest: test x2=0, x2sq=0; /* partial F-test */
run;
```

The REG Procedure
Model: MODEL1

Model Crossproducts X'X X'Y Y'Y					
Variable	**Intercept**	**x1**	**x2**	**x2sq**	**y**
Intercept	10	190.09	56.51	412.7623	165.35
x1	190.09	8061.8657	1640.4116	14679.655592	6574.034
x2	56.51	1640.4116	412.7623	3323.565419	1348.5262
x2sq	412.7623	14679.655592	3323.565419	28273.561383	12004.502344
y	165.35	6574.034	1348.5262	12004.502344	5438.2429

The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read	10
Number of Observations Used	10

X'X Inverse, Parameter Estimates, and SSE					
Variable	**Intercept**	**x1**	**x2**	**x2sq**	**y**
Intercept	1.4489080779	-0.085966558	-0.856100865	0.1241163792	9.9107930382
x1	-0.085966558	0.0140719141	0.0854121885	-0.016091361	0.306459682
x2	-0.856100865	0.0854121885	0.684714617	-0.112336367	-5.24022871
x2sq	0.1241163792	-0.016091361	-0.112336367	0.0197832292	0.73677386
y	9.9107930382	0.306459682	-5.24022871	0.73677386	6.7990811384

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	2697.38157	899.12719	793.45	<.0001
Error	6	6.79908	1.13318		
Corrected Total	9	2704.18065			

Root MSE	1.06451	**R-Square**	0.9975
Dependent Mean	16.53500	**Adj R-Sq**	0.9962
Coeff Var	6.43792		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**99% Confidence Limits**	
Intercept	1	9.91079	1.28136	7.73	0.0002	5.16026	14.66133
x1	1	0.30646	0.12628	2.43	0.0514	-0.16171	0.77462
x2	1	-5.24023	0.88085	-5.95	0.0010	-8.50593	-1.97452
x2sq	1	0.73677	0.14973	4.92	0.0027	0.18167	1.29187

The REG Procedure
Model: MODEL1
Dependent Variable: y

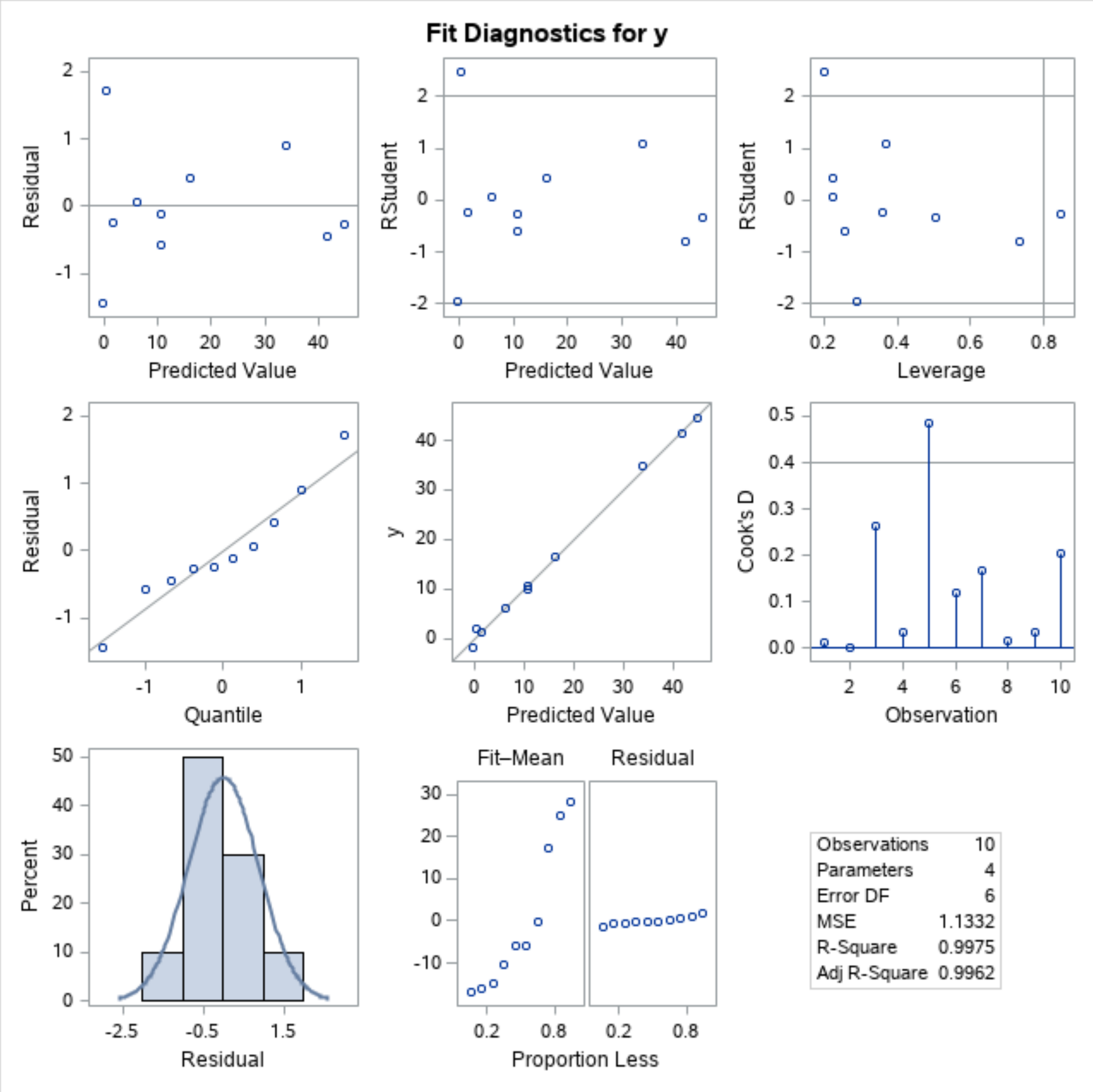
Durbin-Watson D	1.702
Pr < DW	0.2927
Pr > DW	0.7073
Number of Observations	10
1st Order Autocorrelation	-0.075

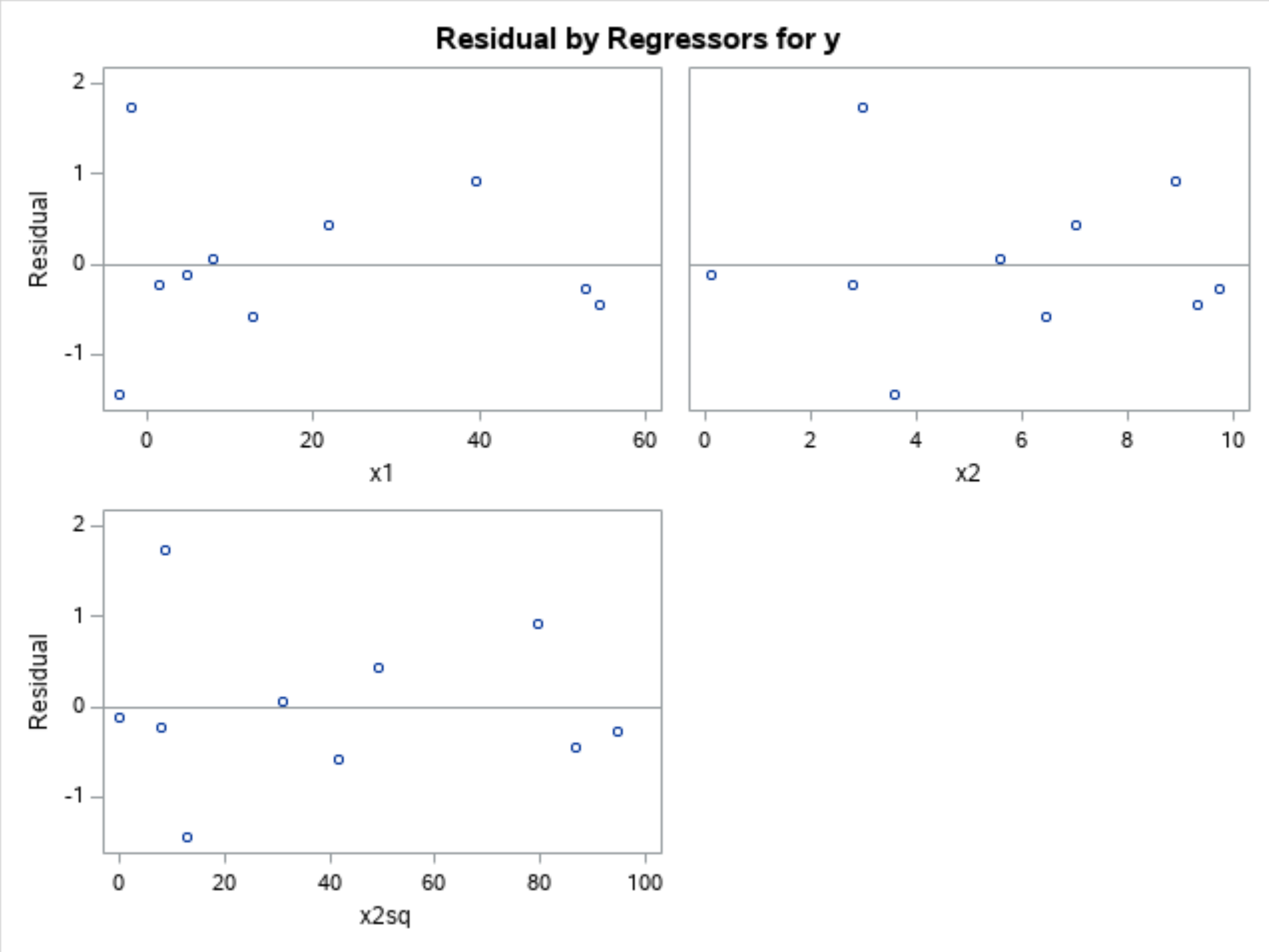
Note: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	99% CL Mean		99% CL Predict		Residual
1	1.23	1.4639	0.6383	-0.9027	3.8305	-3.1379	6.0657	-0.2339
2	6.12	6.0617	0.5012	4.2034	7.9201	1.6995	10.4239	0.0583
3	-1.90	-0.4484	0.5709	-2.5649	1.6681	-4.9267	4.0299	-1.4516
4	44.61	44.8784	0.7555	42.0775	47.6793	40.0389	49.7179	-0.2684
5	41.28	41.7381	0.9124	38.3554	45.1209	36.5402	46.9361	-0.4581
6	10.56	10.6824	0.9792	7.0520	14.3128	5.3200	16.0448	-0.1224
7	34.78	33.8718	0.6465	31.4750	36.2685	29.2544	38.4891	0.9082
8	16.57	16.1495	0.5038	14.2818	18.0172	11.7833	20.5157	0.4205
9	10.09	10.6732	0.5371	8.6820	12.6644	6.2527	15.0937	-0.5832
10	2.01	0.2794	0.4751	-1.4822	2.0410	-4.0425	4.6013	1.7306

Sum of Residuals	0
Sum of Squared Residuals	6.79908
Predicted Residual SS (PRESS)	15.84524


Fit Diagnostics for y
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
y
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 10
Parameters 4
Error DF 6
MSE 1.1332
R-Square 0.9975
Adj R-Square 0.9962



The REG Procedure
Model: MODEL1

Test mytest Results for Dependent Variable y				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	2	25.63895	22.63	0.0016
Denominator	6	1.13318		