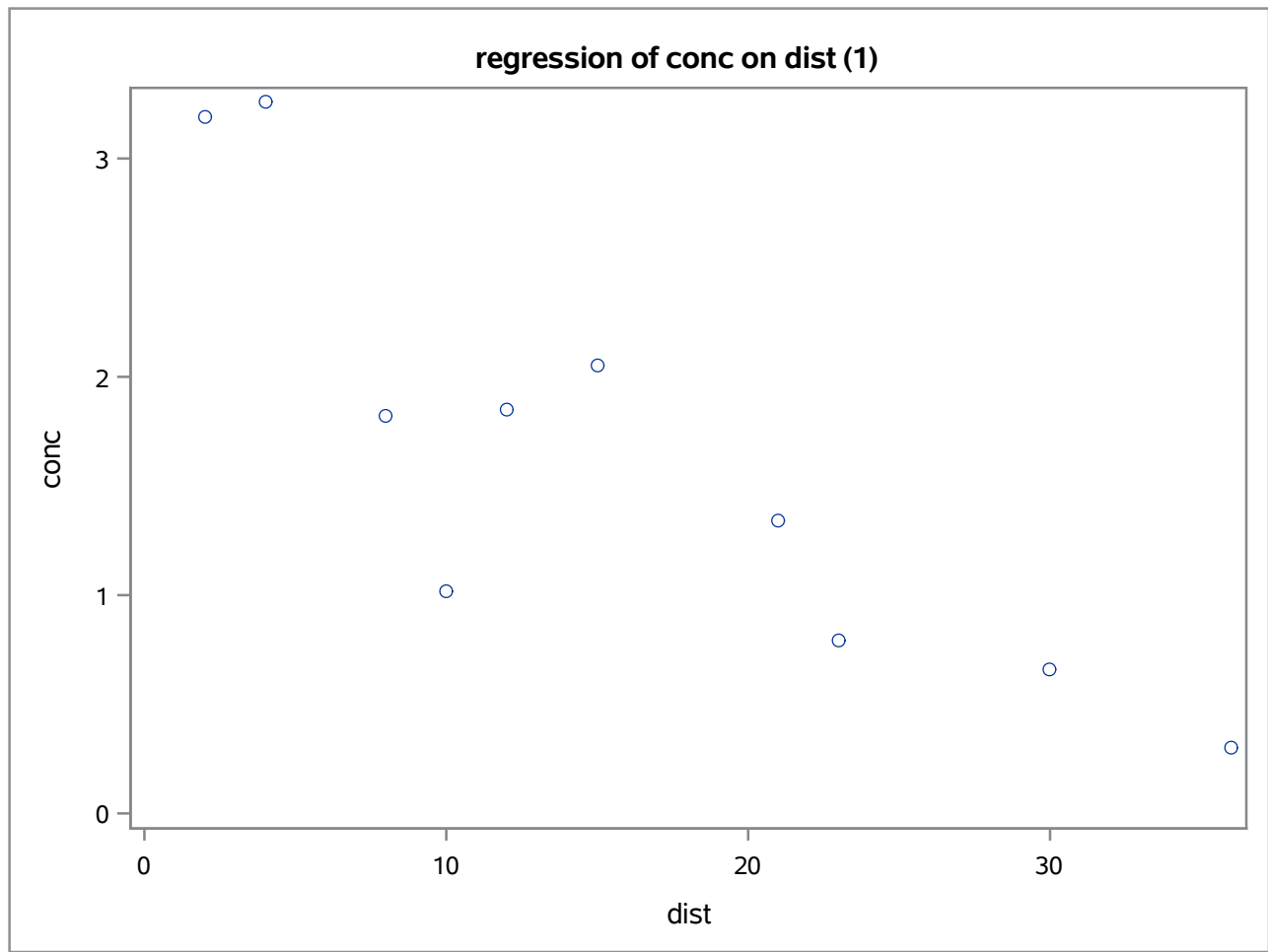


arsenic data

Obs	conc	dist
1	3.19	2
2	3.26	4
3	1.82	8
4	1.02	10
5	1.85	12
6	2.05	15
7	1.34	21
8	0.79	23
9	0.66	30
10	0.30	36

The response variable is conc which is the arsenic concentration in ppm

The explanatory variable, dist, is the the distance of the community from the power plant in km



This plot of arsenic concentration versus distance from the plant shows a reasonably strong decreasing linear trend.

The GLM Procedure

Number of Observations Read	10
Number of Observations Used	10

The GLM Procedure

Coefficients for Estimate meanresp20	
	Row 1
Intercept	1
dist	20

These coefficients which define the linear combination corresponding to an observation at a distance of 20 km will be used to estimate the mean response at this distance.

The GLM Procedure

Dependent Variable: conc

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.88257399	6.88257399	23.53	0.0013
Error	8	2.34038601	0.29254825		
Corrected Total	9	9.22296000			

This F test is for $H_0: \beta = 0$ (slope = 0). The small P-value .0013 shows strong evidence that the slope is not zero.

ANOVA for the regression of concentration on distance.

Model here indicates "slope"

R-Square	Coeff Var	Root MSE	conc Mean
0.746244	33.22342	0.540877	1.628000

$R^2 = .7462$

The linear trend relating arsenic concentration to distance from the plant accounts for 74.62% of the variability in arsenic concentration

Source	DF	Type I SS	Mean Square	F Value	Pr > F
dist	1	6.88257399	6.88257399	23.53	0.0013

Source	DF	Type II SS	Mean Square	F Value	Pr > F
dist	1	6.88257399	6.88257399	23.53	0.0013

Parameter	Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
intercept	2.88622593	0.31071996	9.29	<.0001	2.16970442	3.60274744
slope	-0.07815068	0.01611225	-4.85	0.0013	-0.11530558	-0.04099577
mean	1.62800000	0.17104042	9.52	<.0001	1.23358008	2.02241992
meanresp20	1.32321235	0.18221803	7.26	<.0001	0.90301683	1.74340787

Parameter	Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Intercept	2.886225930	0.31071996	-9.29	<.0001	2.169704415	3.602747444
dist	-0.078150679	0.01611225	-4.85	0.0013	-0.115305583	-0.040995775

value at $x=0$ is of no interest

$b_0 = 2.886$,
 $b_1 = -0.078$,
fitted line equation
 $\hat{y} = 2.886 - 0.078x$

test for $H_0: \beta_1 = 0$ (slope = 0) and 95% confidence interval for the slope

$\bar{y} = 1.628$
fitted line equation
 $\hat{y} = 1.628 - 0.078(x - 16.1)$
16.1 is the mean distance (\bar{x})

estimate of the mean response at $x=20$
 $\hat{y}(20) = 1.323$
95% confidence interval for $\mu(20)$
.903 $\leq \mu(20) \leq 1.743$

values of
 observed response: y
 predicted mean response: y-hat
 residual: y-y-hat

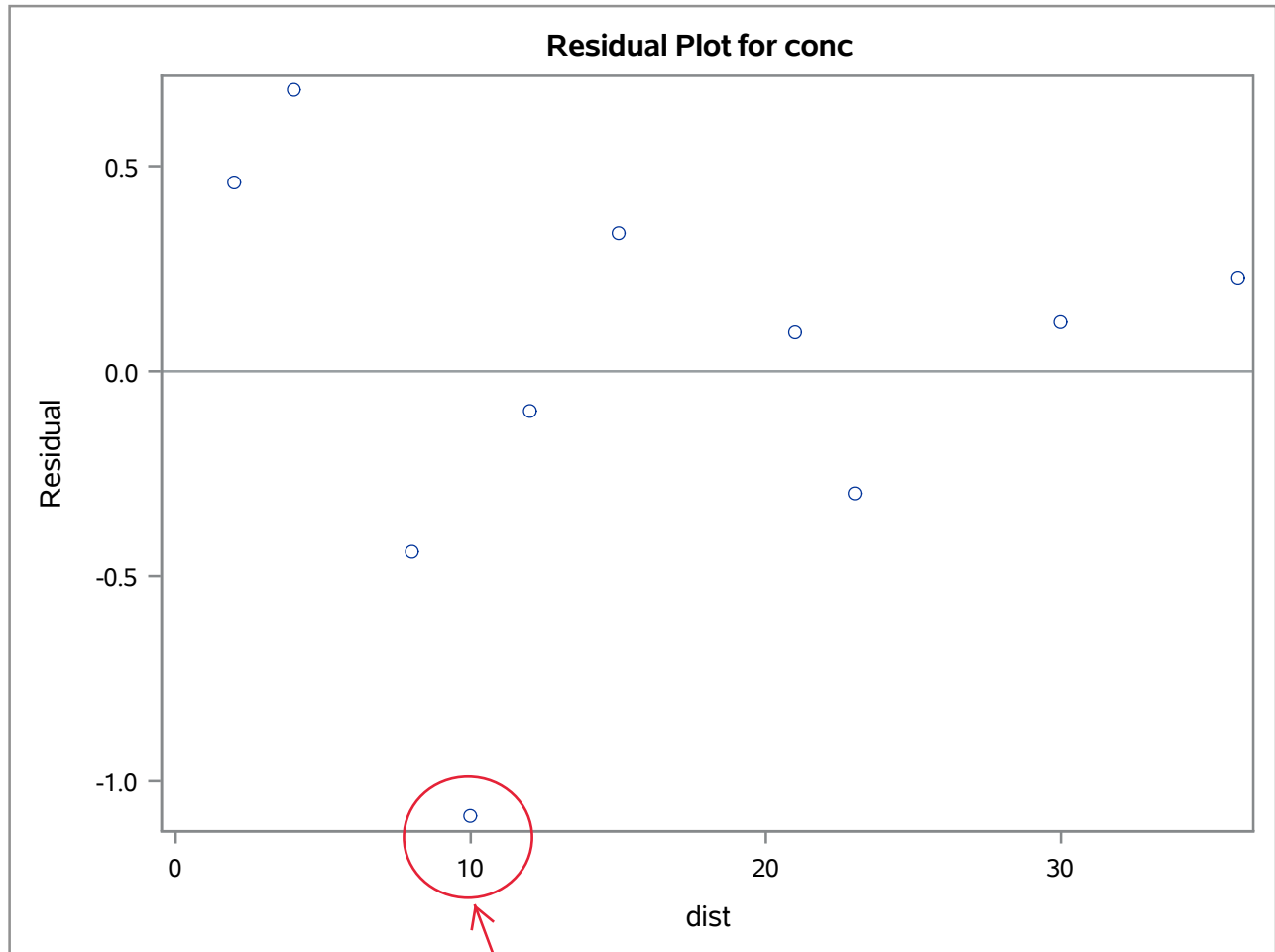
The GLM Procedure

Observation	Observed	Predicted	Residual	95% Confidence Limits for Mean Predicted Value	
1	3.19000000	2.72992457	0.46007543	2.07416420	3.38568495
2	3.26000000	2.57362321	0.68637679	1.97555620	3.17169023
3	1.82000000	2.26102050	-0.44102050	1.76489469	2.75714630
4	1.02000000	2.10471914	-1.08471914	1.64981807	2.55962022
5	1.85000000	1.94841778	-0.09841778	1.52560216	2.37123341
6	2.05000000	1.71396575	0.33603425	1.31743396	2.11049753
7	1.34000000	1.24506167	0.09493833	0.81065108	1.67947226
8	0.79000000	1.08876032	-0.29876032	0.61834343	1.55917721
9	0.66000000	0.54170556	0.11829444	-0.10813360	1.19154473
10	0.30000000	0.07280149	0.22719851	-0.76520429	0.91080727

Sum of Residuals	0.00000000
Sum of Squared Residuals	2.34038601
Sum of Squared Residuals - Error SS	0.00000000
PRESS Statistic	3.52155376
First Order Autocorrelation	0.23937056
Durbin-Watson D	1.40876096

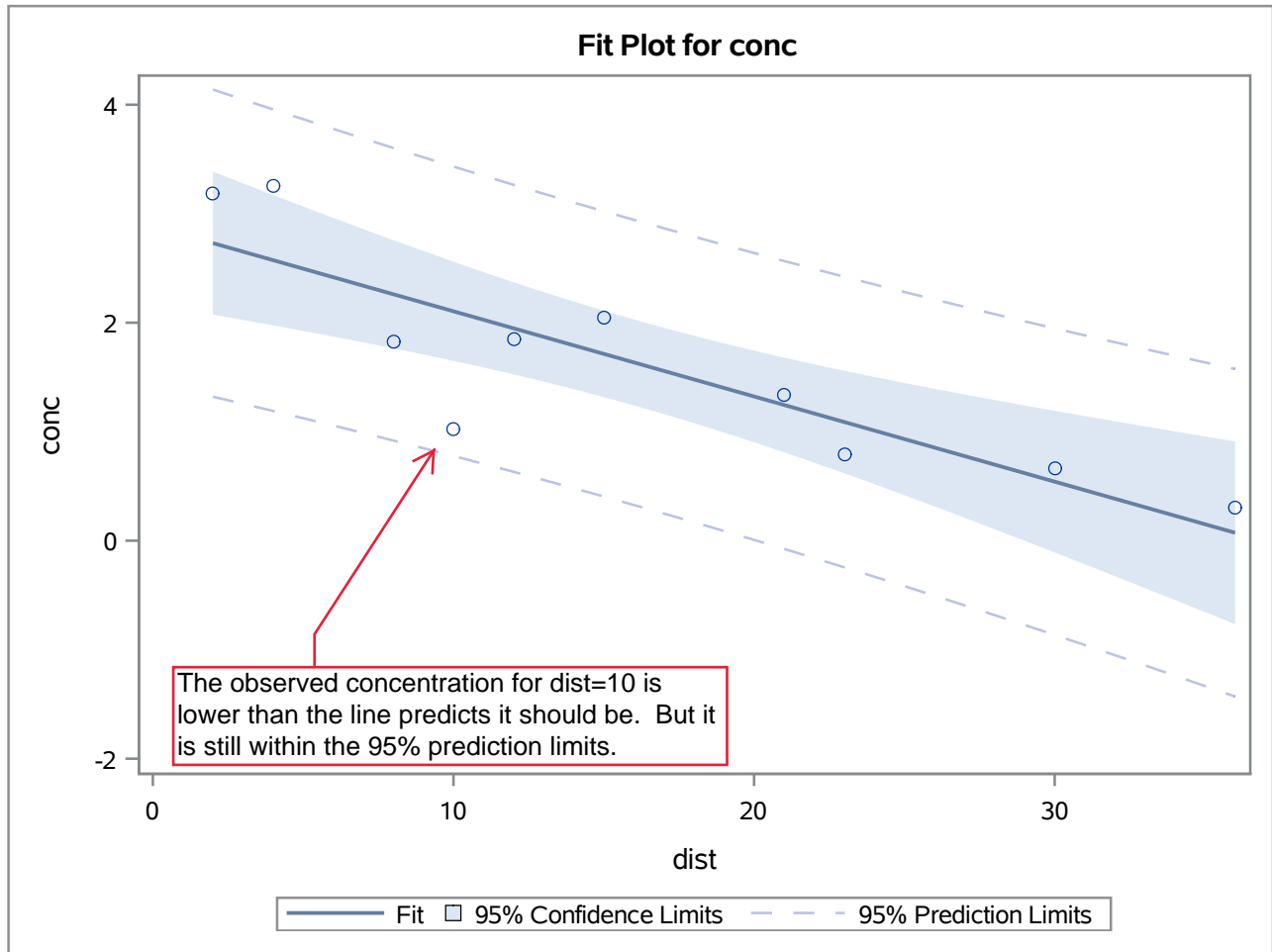
confidence intervals for the mean response at the observed values of x: i.e., confidence intervals for $\mu(x)$

The GLM Procedure



The residual for dist = 10 is far from zero. In this case, the observed concentration for the community at 10 km is lower than we would expect according to the linear trend in the data. See also the plot below with the fitted line.

The GLM Procedure



Obs	conc	dist	clml	clmu	clpl	clpu	yhat	resid
1	3.19	2	2.07416	3.38568	1.32078	4.13907	2.72992	0.46008
2	3.26	4	1.97556	3.17169	1.19038	3.95686	2.57362	0.68638
3	1.82	8	1.76489	2.75715	0.91870	3.60334	2.26102	-0.44102
4	1.02	10	1.64982	2.55962	0.77709	3.43235	2.10472	-1.08472
5	1.85	12	1.52560	2.37123	0.63143	3.26540	1.94842	-0.09842
6	2.05	15	1.31743	2.11050	0.40518	3.02275	1.71397	0.33603
7	1.34	21	0.81065	1.67947	-0.07569	2.56581	1.24506	0.09494
8	0.79	23	0.61834	1.55918	-0.24427	2.42179	1.08876	-0.29876
9	0.66	30	-0.10813	1.19154	-0.86469	1.94811	0.54171	0.11829
10	0.30	36	-0.76520	0.91081	-1.42984	1.57544	0.07280	0.22720

95% confidence intervals for the mean response $\mu(x)$

95% confidence prediction intervals for the actual response $\mu(x) + \text{error}$



The UNIVARIATE Procedure
Variable: resid

Basic Statistical Measures			
Location		Variability	
Mean	0.000000	Std Deviation	0.50994
Median	0.106616	Variance	0.26004
Mode	.	Range	1.77110
		Interquartile Range	0.63479

The normality assumption is reasonable.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.947998	Pr < W	0.6449
Kolmogorov-Smirnov	D	0.173846	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.037821	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.26181	Pr > A-Sq	>0.2500

Quantiles (Definition 5)	
Level	Quantile
100% Max	0.686377
99%	0.686377
95%	0.686377
90%	0.573226
75% Q3	0.336034
50% Median	0.106616
25% Q1	-0.298760
10%	-0.762870
5%	-1.084719
1%	-1.084719
0% Min	-1.084719

Extreme Values			
Lowest		Highest	
Order	Value	Order	Value
1	-1.0847191	6	0.118294
2	-0.4410205	7	0.227199
3	-0.2987603	8	0.336034
4	-0.0984178	9	0.460075
5	0.0949383	10	0.686377

regression of conc on dist (2) prediction interval for $x=20$ **The GLM Procedure**

Number of Observations Read	11
Number of Observations Used	10

an observation with $x=20$ and y missing has been added to generate information about the response at $x=20$

regression of conc on dist (2) prediction interval for x=20

same output as that on page 5

The GLM Procedure**Dependent Variable: conc**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.88257399	6.88257399	23.53	0.0013
Error	8	2.34038601	0.29254825		
Corrected Total	9	9.22296000			

R-Square	Coeff Var	Root MSE	conc Mean
0.746244	33.22342	0.540877	1.628000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
dist	1	6.88257399	6.88257399	23.53	0.0013

Source	DF	Type III SS	Mean Square	F Value	Pr > F
dist	1	6.88257399	6.88257399	23.53	0.0013

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	2.886225930	0.31071996	9.29	<.0001
dist	-0.078150679	0.01611225	-4.85	0.0013

regression of conc on dist (2) prediction interval for x=20

Obs	conc	dist	clml	clmu	clpl	clpu	yhat	resid
1	3.19	2	2.07416	3.38568	1.32078	4.13907	2.72992	0.46008
2	3.26	4	1.97556	3.17169	1.19038	3.95686	2.57362	0.68638
3	1.82	8	1.76489	2.75715	0.91870	3.60334	2.26102	-0.44102
4	1.02	10	1.64982	2.55962	0.77709	3.43235	2.10472	-1.08472
5	1.85	12	1.52560	2.37123	0.63143	3.26540	1.94842	-0.09842
6	2.05	15	1.31743	2.11050	0.40518	3.02275	1.71397	0.33603
7	1.34	21	0.81065	1.67947	-0.07569	2.56581	1.24506	0.09494
8	0.79	23	0.61834	1.55918	-0.24427	2.42179	1.08876	-0.29876
9	0.66	30	-0.10813	1.19154	-0.86469	1.94811	0.54171	0.11829
10	0.30	36	-0.76520	0.91081	-1.42984	1.57544	0.07280	0.22720
11	.	20	0.90302	1.74341	0.00707	2.63936	1.32321	.

x=20
y=. indicates missing value
i.e., no observation at x=20

95% confidence prediction interval for
the actual response at x=20

This estimates what we would
observe if we went to the community
at x=20 and measured y.

Compare to the output on page 5.

estimate of the mean response at x=20
y-hat(20)=1.323
95% confidence interval for mu(20)
.903 <= mu(20) <= 1.743

regression of conc on dist (3) ANOVA with uncorrected total SS**The GLM Procedure**

Number of Observations Read	10
Number of Observations Used	10

regression of conc on dist (3) ANOVA with uncorrected total SS

The GLM Procedure

Dependent Variable: conc

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	33.38641399	16.69320699	57.06	<.0001
Error	8	2.34038601	0.29254825		
Uncorrected Total	10	35.72680000			

modified output to show the ANOVA with the uncorrected total sum of squares

R-Square	Coeff Var	Root MSE	conc Mean
0.746244	33.22342	0.540877	1.628000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Intercept	1	26.50384000	26.50384000	90.60	<.0001
dist	1	6.88257399	6.88257399	23.53	0.0013

Source	DF	Type II SS	Mean Square	F Value	Pr > F
Intercept	1	25.24177252	25.24177252	86.28	<.0001
dist	1	6.88257399	6.88257399	23.53	0.0013

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	2.886225930	0.31071996	9.29	<.0001
dist	-0.078150679	0.01611225	-4.85	0.0013

regression of conc on dist (4) Standard output from proc glm**The GLM Procedure**

Number of Observations Read	10
Number of Observations Used	10

regression of conc on dist (4) Standard output from proc glm

The GLM Procedure

Dependent Variable: conc

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.88257399	6.88257399	23.53	0.0013
Error	8	2.34038601	0.29254825		
Corrected Total	9	9.22296000			

R-Square	Coeff Var	Root MSE	conc Mean
0.746244	33.22342	0.540877	1.628000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
dist	1	6.88257399	6.88257399	23.53	0.0013

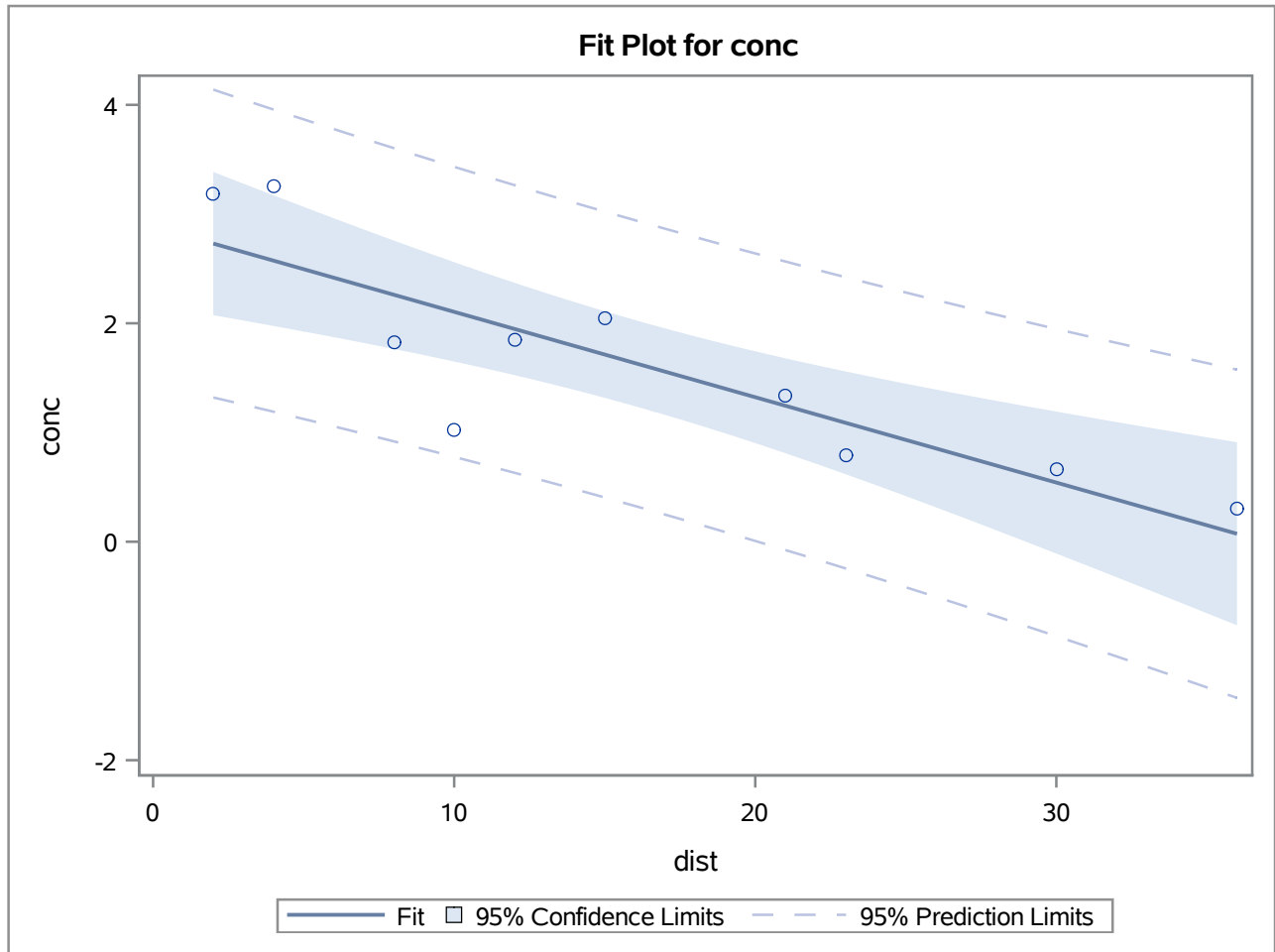
Source	DF	Type III SS	Mean Square	F Value	Pr > F
dist	1	6.88257399	6.88257399	23.53	0.0013

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	2.886225930	0.31071996	9.29	<.0001
dist	-0.078150679	0.01611225	-4.85	0.0013

regression of conc on dist (4) Standard output from proc glm

The GLM Procedure

Dependent Variable: conc



regression of conc on dist (5) Standard output from proc reg

The REG Procedure
 Model: MODEL1
 Dependent Variable: conc

Number of Observations Read	10
Number of Observations Used	10

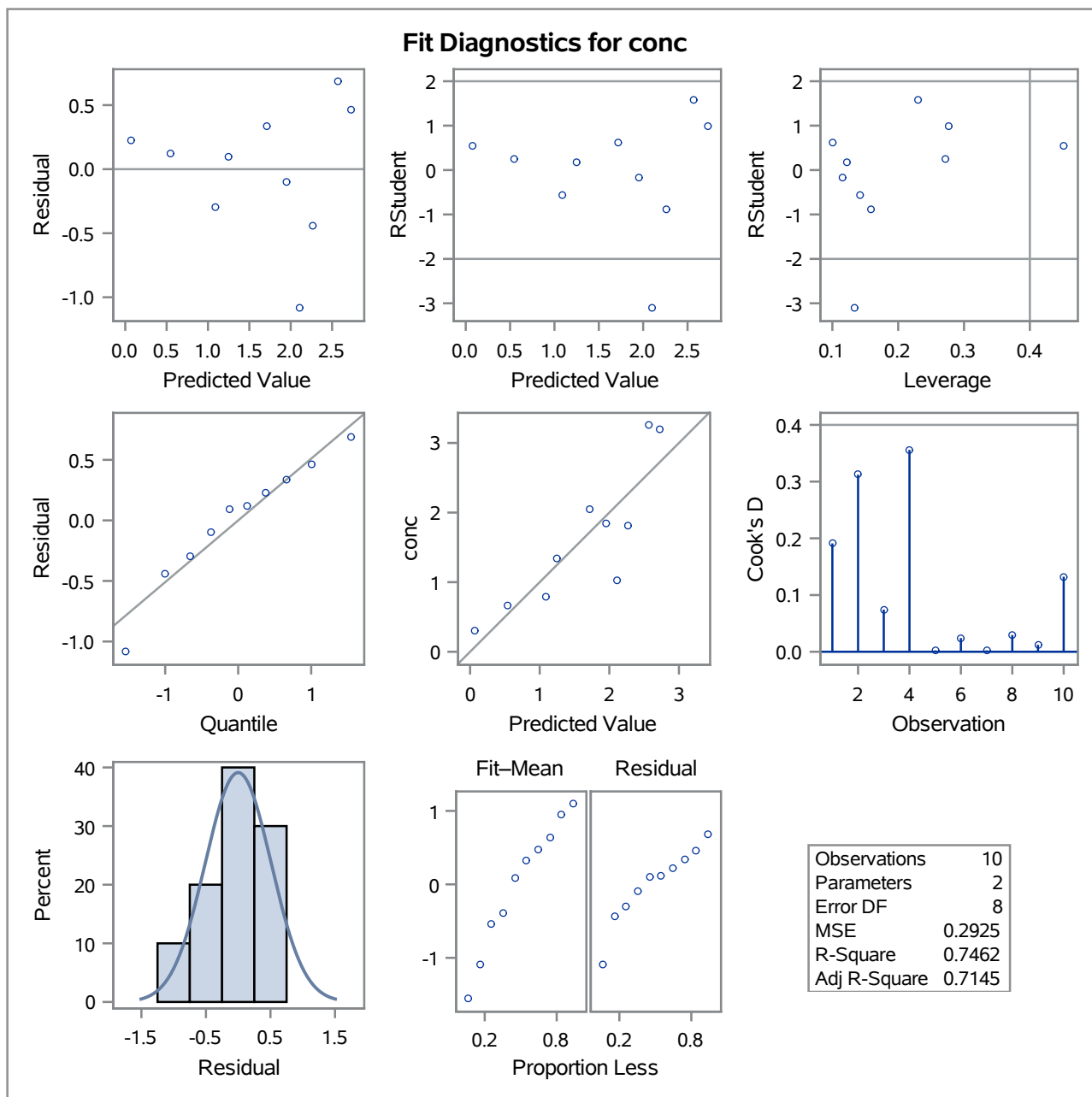
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.88257	6.88257	23.53	0.0013
Error	8	2.34039	0.29255		
Corrected Total	9	9.22296			

Root MSE	0.54088	R-Square	0.7462
Dependent Mean	1.62800	Adj R-Sq	0.7145
Coeff Var	33.22342		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.88623	0.31072	9.29	<.0001
dist	1	-0.07815	0.01611	-4.85	0.0013

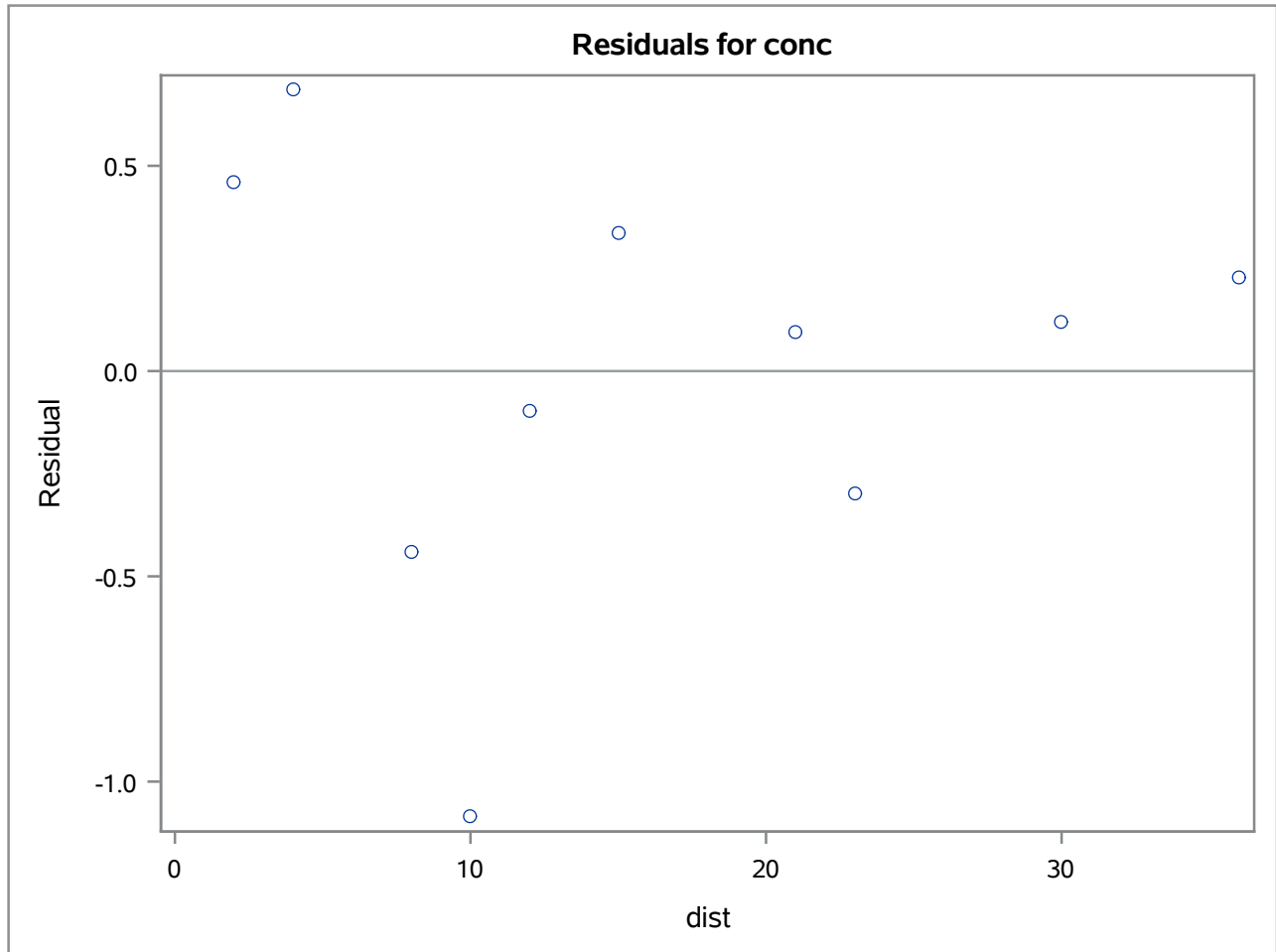
regression of conc on dist (5) Standard output from proc reg

The REG Procedure
 Model: MODEL1
 Dependent Variable: conc



regression of conc on dist (5) Standard output from proc reg

The REG Procedure
Model: MODEL1
Dependent Variable: conc



regression of conc on dist (5) Standard output from proc reg

The REG Procedure
Model: MODEL1
Dependent Variable: conc

