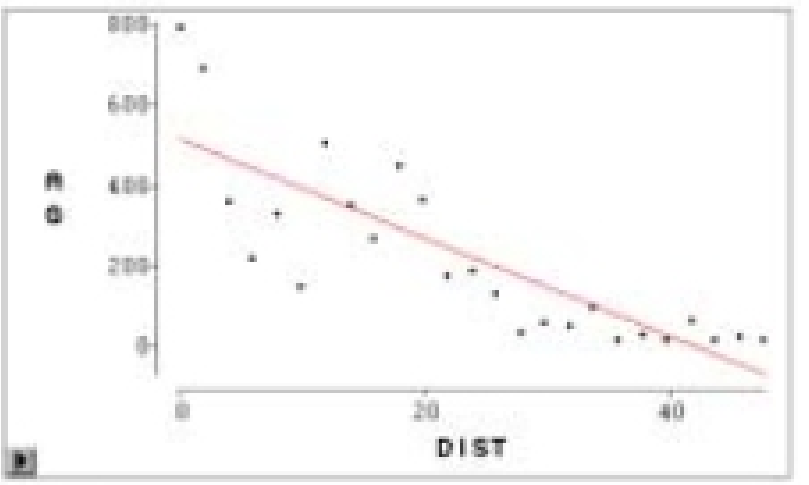
Polynomial Regression and Transformations Using SAS Insight

In INSIGHT, click on **Analyze** and then **Fit (Y X)**. Click on the response **AG** and then on **Y**. Next click on the predictor **DIST** and then on **X**. Clicking on output will produce the linear regression:

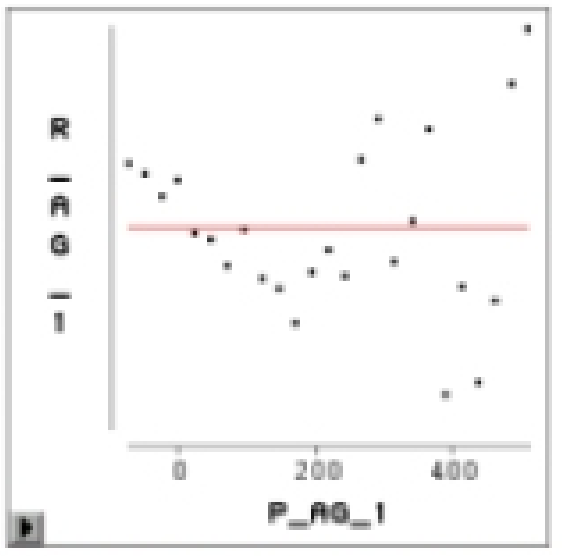
AG = DIST
Response Distribution: Normal
Link Function: Identity

Model Equation
AG = 515.425 - 12.2427 DIST

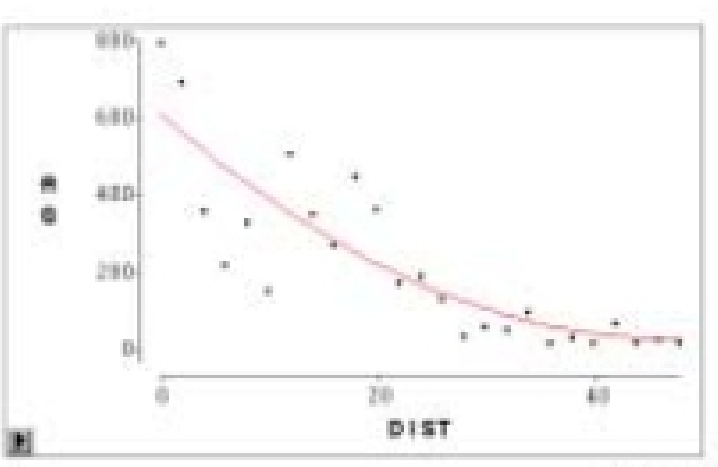


Parametric Regression Fit

		Model		Error				
Curve	Degree(Polynomial)	DF	Mean Square	DF	Mean Square	R-Square	F Stat	Prob > F
	1	1	779384.278	23	15987.3792	0.6783	48.7201	0.0001



The residual plot looks nonlinear, so you can move the slide-bar and fit a quadratic function to the data:



Parametric Regression Fit

		Model		Error				
Curve	Degree(Polynomial)	DF	Mean Square	DF	Mean Square	R-Square	F Stat	Prob > F
	2	2	419613.663	22	14068.3943	0.7384	29.8031	0.0001

Summary of Fit			
Mean of Response	221.6000	R-Square	0.6783
Root MSE	126.4507	Adj R-Sq	0.6554

Notice that the plot changed and the output near the slide-bar changed, but the fitted model did not. To get the regression equation for the quadratic model, you need to create x^2. On the worksheet with the data, click on the column heading, DIST. Go to the EDIT menu and click on **Variables** and then **Y*Y**. This creates a variable called B_DIST which is x^2. Next go back to the dialog box for Fit (**Y X**). Click on the response **AG** and then on **Y**. Next click on the predictor **DIST** and then on **X**. Also, click on **B_DIST** and then on **X**. Clicking on output will produce the quadratic regression output:

AG = DIST B_DIST
Response Distribution: Normal
Link Function: Identity

Model Equation

AG = 611.450 - 24.7678 DIST + 0.2609 B_DIST

Summary of Fit

Mean of Response	221.6000	R-Square	0.7304
Root MSE	118.5723	Adj R-Sq	0.7059

Analysis of Variance

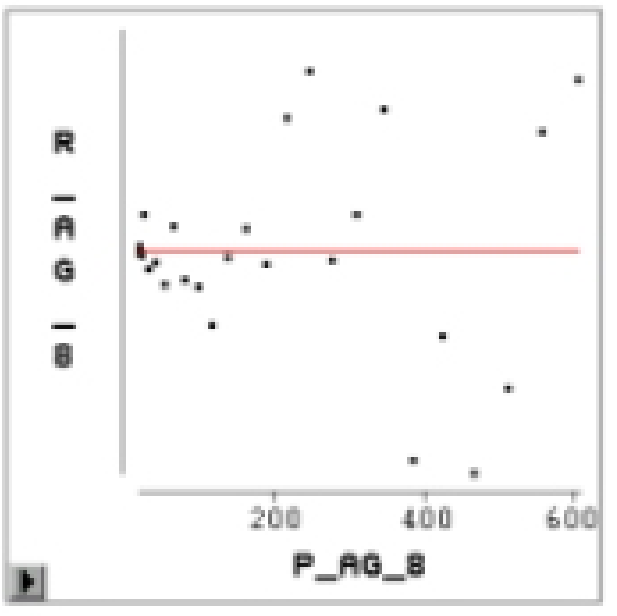
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	2	838027.325	419013.663	29.8031	0.0001
Error	22	309306.675	14059.3943		
C Total	24	1147334.00			

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
DIST	1	213900.145	213900.145	15.2140	0.0008
B_DIST	1	58633.0477	58633.0477	4.1704	0.0533

Parameter Estimates

Variable	DF	Estimate	Std Error	T Stat	Prob >|T|	Tolerance	Var Inflation
INTERCEPT	1	611.4503	65.8086	9.2913	0.0001	.	0
DIST	1	-24.7678	6.3499	-3.9005	0.0008	0.0671	14.9130
B_DIST	1	0.2609	0.1278	2.0422	0.0533	0.0671	14.9130

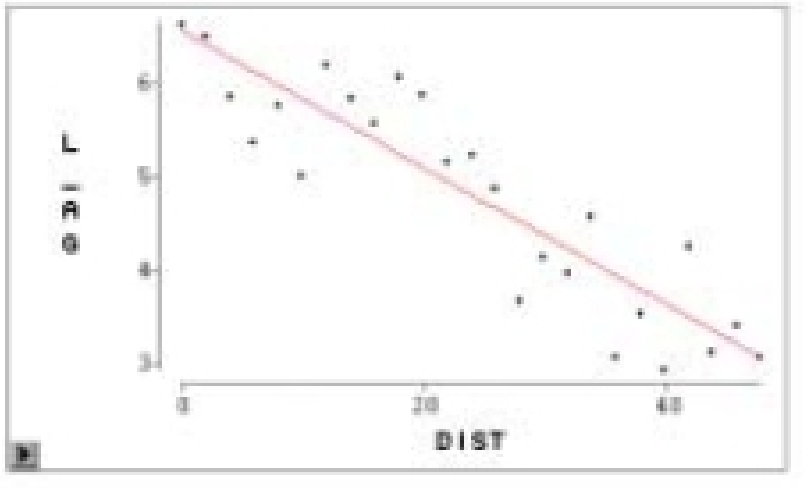


We next transform **AG** and fit the linear model relating **log(AG)** to **DIST**. On the worksheet with the data, click on the column heading, **AG**. Go to the **EDIT** menu and click on **Variables** and then **log(Y)**. This creates a variable called L_AG which is **log(AG)**. Carry out the linear regression as before using **DIST** as the predictor and L_AG as the response.

The output from the regression for the transformed model follows:

L_AG = DIST
Response Distribution: Normal
Link Function: Identity

Model Equation			
L_AG =	6.5837	-	0.0734 DIST



Parametric Regression Fit

		Model		Error				
Curve	Degree(Polynomial)	DF	Mean Square	DF	Mean Square	R-Square	F Stat	Prob > F
	1	1	28.0253	23	0.2842	0.8109	98.6148	0.0001

Summary of Fit

Mean of Response	4.8218	R-Square	0.8109
Root MSE	0.5331	Adj R-Sq	0.8027

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	1	28.0253	28.0253	98.6148	0.0001
Error	23	6.5384	0.2842		
C Total	24	34.5637			

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
DIST	1	28.0253	28.0253	98.6148	0.0001

Parameter Estimates

Variable	DF	Estimate	Std Error	T Stat	Prob >|T|	Tolerance	Var Inflation
INTERCEPT	1	6.5837	0.2079	31.6659	0.0001	.	0
DIST	1	-0.0734	0.0074	-9.9305	0.0001	1.0000	1.0000