data spirit;
  infile "U:\data\SpiritStLouis.csv" dsd firstobs=2;
  input gas weight headwind TO_distance; run;
title "Spirit of St. Louis Takeoff Distance";
  proc print data = spirit; run;
title "Scatter Plot with Regression Line";
  proc sgplot data=spirit;
    reg y = weight x=TO_distance; run;
title "Correlation";
  proc corr data = spirit;
    var weight TO_distance; run;
title "Simple Linear regression";
  proc reg data = spirit;
    model weight = TO_distance;
    plot rstudent.*predicted.; run;
### Spirit of St. Louis Takeoff Distance

<table>
<thead>
<tr>
<th>Obs</th>
<th>gas</th>
<th>weight</th>
<th>headwind</th>
<th>TO_distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
<td>2600</td>
<td>7</td>
<td>229</td>
</tr>
<tr>
<td>2</td>
<td>71</td>
<td>2800</td>
<td>9</td>
<td>287</td>
</tr>
<tr>
<td>3</td>
<td>111</td>
<td>3050</td>
<td>9</td>
<td>389</td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>3300</td>
<td>6</td>
<td>483</td>
</tr>
<tr>
<td>5</td>
<td>201</td>
<td>3600</td>
<td>4</td>
<td>615</td>
</tr>
<tr>
<td>6</td>
<td>251</td>
<td>3900</td>
<td>2</td>
<td>800</td>
</tr>
<tr>
<td>7</td>
<td>301</td>
<td>4200</td>
<td>0</td>
<td>1023</td>
</tr>
</tbody>
</table>
Correlation

The CORR Procedure

2 Variables: weight TO_distance

Simple Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>7</td>
<td>3350</td>
<td>583.80933</td>
<td>23450</td>
<td>2600</td>
<td>4200</td>
</tr>
<tr>
<td>TO_distance</td>
<td>7</td>
<td>546.57143</td>
<td>286.64488</td>
<td>3826</td>
<td>229.00000</td>
<td>1023</td>
</tr>
</tbody>
</table>

Pearson Correlation Coefficients, N = 7

Prob > |r| under H0: Rho=0

<table>
<thead>
<tr>
<th></th>
<th>weight</th>
<th>TO_distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>1.00000</td>
<td>0.98882</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>TO_distance</td>
<td>0.98882</td>
<td>1.00000</td>
</tr>
<tr>
<td></td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>
Simple Linear regression

The REG Procedure
Model: MODEL1
Dependent Variable: weight

Number of Observations Read 7
Number of Observations Used 7

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>1999529</td>
<td>1999529</td>
<td>219.87</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>45471</td>
<td>9094.24340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>6</td>
<td>2045000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE | 95.36374 | R-Square | 0.9778 |
Dependent Mean | 3350.00000 | Adj R-Sq | 0.9733 |
Coeff Var | 2.84668 |

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>TO_distance</td>
</tr>
</tbody>
</table>
Simple Linear regression

The REG Procedure
Model: MODEL1
Dependent Variable: weight

Fit Diagnostics for weight

Residual vs. Predicted Value

Residual vs. Quantile

Residual vs. Observation

Residual vs. Percent

Residual vs. Proportion Loss

Observations  7
Parameters   2
Error DF      5
MSE     5034.2
R-Square 0.9770
Adj R-Square 0.9733
data spirit;
infile "U:\data\SpiritStLouis.csv" dsd firstobs=2;
input gas weight headwind TO_distance;
TO_DistL10 = log10(TO_Distance);
weightL10 = log10(weight); run;
title "Spirit of St. Louis Takeoff Distance L10";
proc print data = spirit; run;
title "Scatter Plot with Regression Line L10";
proc sgplot data=spirit;
reg y = weightL10 x=TO_distL10; run;
title "Correlation L10";
proc corr data = spirit;
var weightL10 TO_distL10; run;
title "Simple Linear regression L10";
proc reg data = spirit;
model weightL10 = TO_distL10;
plot rstudent.*predicted.; run;
### Spirit of St. Louis Takeoff Distance L10

<table>
<thead>
<tr>
<th>Obs</th>
<th>gas</th>
<th>weight</th>
<th>headwind</th>
<th>TO_distance</th>
<th>TO_DistL10</th>
<th>weightL10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
<td>2600</td>
<td>7</td>
<td>229</td>
<td>2.35984</td>
<td>3.41497</td>
</tr>
<tr>
<td>2</td>
<td>71</td>
<td>2800</td>
<td>9</td>
<td>287</td>
<td>2.45788</td>
<td>3.44716</td>
</tr>
<tr>
<td>3</td>
<td>111</td>
<td>3050</td>
<td>9</td>
<td>389</td>
<td>2.58995</td>
<td>3.48430</td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>3300</td>
<td>6</td>
<td>483</td>
<td>2.68395</td>
<td>3.51851</td>
</tr>
<tr>
<td>5</td>
<td>201</td>
<td>3600</td>
<td>4</td>
<td>615</td>
<td>2.78888</td>
<td>3.55630</td>
</tr>
<tr>
<td>6</td>
<td>251</td>
<td>3900</td>
<td>2</td>
<td>800</td>
<td>2.90309</td>
<td>3.59106</td>
</tr>
<tr>
<td>7</td>
<td>301</td>
<td>4200</td>
<td>0</td>
<td>1023</td>
<td>3.00988</td>
<td>3.62325</td>
</tr>
</tbody>
</table>
## Correlation L10

The CORR Procedure

2 Variables: weightL10 TO_DistL10

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>weightL10</td>
<td>7</td>
<td>3.51937</td>
<td>0.07598</td>
<td>24.63556</td>
<td>3.41497</td>
<td>3.62325</td>
</tr>
<tr>
<td>TO_DistL10</td>
<td>7</td>
<td>2.68478</td>
<td>0.23461</td>
<td>18.79345</td>
<td>2.35984</td>
<td>3.00988</td>
</tr>
</tbody>
</table>

Pearson Correlation Coefficients, N = 7

Prob > |r| under H0: Rho=0

<table>
<thead>
<tr>
<th></th>
<th>weightL10</th>
<th>TO_DistL10</th>
</tr>
</thead>
<tbody>
<tr>
<td>weightL10</td>
<td>1.00000</td>
<td>0.99949 &lt;.0001</td>
</tr>
<tr>
<td>TO_DistL10</td>
<td>0.99949 &lt;.0001</td>
<td>1.00000</td>
</tr>
</tbody>
</table>
The REG Procedure 
Model: MODEL1 
Dependent Variable: weightL10 

<table>
<thead>
<tr>
<th>Number of Observations Read</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations Used</td>
<td>7</td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.03460</td>
<td>0.03460</td>
<td>4945.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>0.000003499</td>
<td>0.00000700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>6</td>
<td>0.03464</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.00265  R-Square 0.9990  
Dependent Mean 3.51937  Adj R-Sq 0.9988  
Coeff Var 0.07516  

Parameter Estimates

| Variable    | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|-------------|----|--------------------|----------------|---------|------|---|
| Intercept   | 1  | 2.65030            | 0.01240        | 213.76  | <.0001 |
| TO_DistL10  | 1  | 0.32370            | 0.00460        | 70.32   | <.0001 |
Simple Linear regression L10

The REG Procedure
Model: MODEL1
Dependent Variable: weightL10

- Residual vs Predicted Value
- Fit Diagnostics for weightL10
- Residual vs Quartile
- Residual vs Weights
- Cook's D vs Observation
- Fitted vs Residual
- Residual Spread

Observations: 7
Parameters: 2
Error DF: 5
MSE: 7E-6
R-Square: 0.989
Adj R-Square: 0.988
Simple Linear regression L10

weightL10 = 2.6603 + 0.3397 * DistL10

- N: 7
- Rsq: 0.9990
- Adj Rsq: 0.9988
- RMSE: 0.0026
The unfortunate aspect of this second analysis, based on log-transforms, is that the scatter plots show the log-value scales and not the properly scaled original units.

I was able to make a plot with original units, indicated on a distorted (log-transformed) scale.

I was not able to easily add any fitted line to it. In R this was easy via `abline`.

This link shows how to add a sloped line to a scatter plot, but it does not work for the log-scale version produced by the code on the next slide.

That it took some effort to add such a sloped line in SAS 9.2 is telling in itself. The link to it seems to have disappeared.

My solution to it seems klutzy, but it did the trick.
The Spirit of St. Louis (Using Log Transforms)

data spirit;
infile "U:\data\SpiritStLouis.csv" dsd firstobs=2;
input gas weight headwind TO_distance;
TO_DistL10 = log10(TO_Distance);
weightL10 = log10(weight);
run;

title "Scatter Plot with Log Scale";
proc sgplot data=spirit;
scatter y = weight x=TO_distance;
yaxis type=log logstyle=logexpand logbase=10
min =2000 max=6000;
xaxis type=log logstyle=logexpand logbase=10
min = 100 max=3000;
*lineparm x=500 y=3 slope=1.5 does not work;
run;
The Output from Code

Scatter Plot with Log Scale

weight

TO_distance
data spirit;
input gas weight headwind TO_distance x1;
TO_DistL10 = log10(TO_Distance);
weightL10 = log10(weight);
LS_line = 10**2.6503023 * x1**0.3237002;
datalines;
36 2600 7 229 .
71 2800 9 287 .
111 3050 9 389 .
151 3300 6 483 .
201 3600 4 615 .
251 3900 2 800 .
301 4200 0 1023 .
. . . . 100
. . . . 3000
run;
title "Spirit of St. Louis Takeoff Distance L10";
proc print data = spirit;
run;
title
   "Log10-Log10 Scatter Plot with Regression Line";
proc sgplot data=spirit ;
scatter y = weight x=TO_distance;
yaxis type=log logstyle=logexpand logbase=10
min  =2000 max=6000;
xaxis type=log logstyle=logexpand logbase=10
min  =100 max=3000;
   series x = x1 y=LS_line; * this connects points;
run;
### Spirit of St. Louis Takeoff Distance L10

<table>
<thead>
<tr>
<th>Obs</th>
<th>gas</th>
<th>weight</th>
<th>headwind</th>
<th>TO_distance</th>
<th>x1</th>
<th>TO_DistL10</th>
<th>weightL10</th>
<th>LS_line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
<td>2600</td>
<td>7</td>
<td>229</td>
<td>.</td>
<td>2.35984</td>
<td>3.41497</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>71</td>
<td>2800</td>
<td>9</td>
<td>287</td>
<td>.</td>
<td>2.45788</td>
<td>3.44716</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>111</td>
<td>3050</td>
<td>9</td>
<td>389</td>
<td>.</td>
<td>2.58995</td>
<td>3.48430</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>3300</td>
<td>6</td>
<td>483</td>
<td>.</td>
<td>2.68395</td>
<td>3.51851</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>201</td>
<td>3600</td>
<td>4</td>
<td>615</td>
<td>.</td>
<td>2.78888</td>
<td>3.55630</td>
<td>.</td>
</tr>
<tr>
<td>6</td>
<td>251</td>
<td>3900</td>
<td>2</td>
<td>800</td>
<td>.</td>
<td>2.90309</td>
<td>3.59106</td>
<td>.</td>
</tr>
<tr>
<td>7</td>
<td>301</td>
<td>4200</td>
<td>0</td>
<td>1023</td>
<td>.</td>
<td>3.00988</td>
<td>3.62325</td>
<td>.</td>
</tr>
<tr>
<td>8</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>100</td>
<td>.</td>
<td>.</td>
<td>1984.74</td>
</tr>
<tr>
<td>9</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>3000</td>
<td>.</td>
<td>.</td>
<td>5968.25</td>
</tr>
</tbody>
</table>
The Output from Code

Log10-Log10 Scatter Plot with Regression Line

weight

TO_distance

○ weight  —  LS_line
data multi;
do N = 10000 to 50000 by 10000;
   N2 = N*N;
   fit = .068 -.00000811*N + 1.228571E-9*N2;
   input time @; output;
end; datalines;
0.11  0.39  0.95  1.69  2.74;
run;
title "Concatenation Data: x <- c(x,i)" ;
proc print data=multi noobs;
var time N fit; run;
title "Quadratic Model Fit" ;
proc reg data=multi;
   model time = N N2; run; title "Quadratic Fit";
proc gplot data = multi; plot fit*N;
symbol value = dot interpol=sms line=1 width=2;
plot2 time*N; * this adds to a plot; run;
### Concatenation Data: \( x <- c(x,i) \)

<table>
<thead>
<tr>
<th>time</th>
<th>N</th>
<th>fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.11</td>
<td>10000</td>
<td>0.10976</td>
</tr>
<tr>
<td>0.39</td>
<td>20000</td>
<td>0.39723</td>
</tr>
<tr>
<td>0.95</td>
<td>30000</td>
<td>0.93041</td>
</tr>
<tr>
<td>1.69</td>
<td>40000</td>
<td>1.70931</td>
</tr>
<tr>
<td>2.74</td>
<td>50000</td>
<td>2.73393</td>
</tr>
</tbody>
</table>
**Quadratic Model Fit**

The REG Procedure
Model: MODEL1
Dependent Variable: time

Number of Observations Read 5
Number of Observations Used 5

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>4.51467</td>
<td>2.25734</td>
<td>5338.30</td>
<td>0.0002</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>0.00084571</td>
<td>0.00042286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>4</td>
<td>4.51552</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 0.02056
R-Square 0.9998
Dependent Mean 1.17600
Adj R-Sq 0.9996
Coeff Var 1.74860

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|--------------------|----------------|---------|------|---|
| Intercept| 1  | 0.06800            | 0.04410        | 1.54    | 0.2631|
| N        | 1  | -0.00000811        | 0.00000336     | -2.41   | 0.1371|
| N2       | 1  | 1.22857E-9         | 5.49582E-11    | 22.35   | 0.0020|
The Output from Code
The Output from Code

Quadratic Fit

fit
3

N
10000
20000
30000
40000
50000
time
3
2
1
0
-1
-2