

Homework 0 Key (not to be handed in) due? Jan. 10

The results of running diamond.sas is listed below: Note: I did slightly reduce the size of some of the graphs so that they would fit on the page.

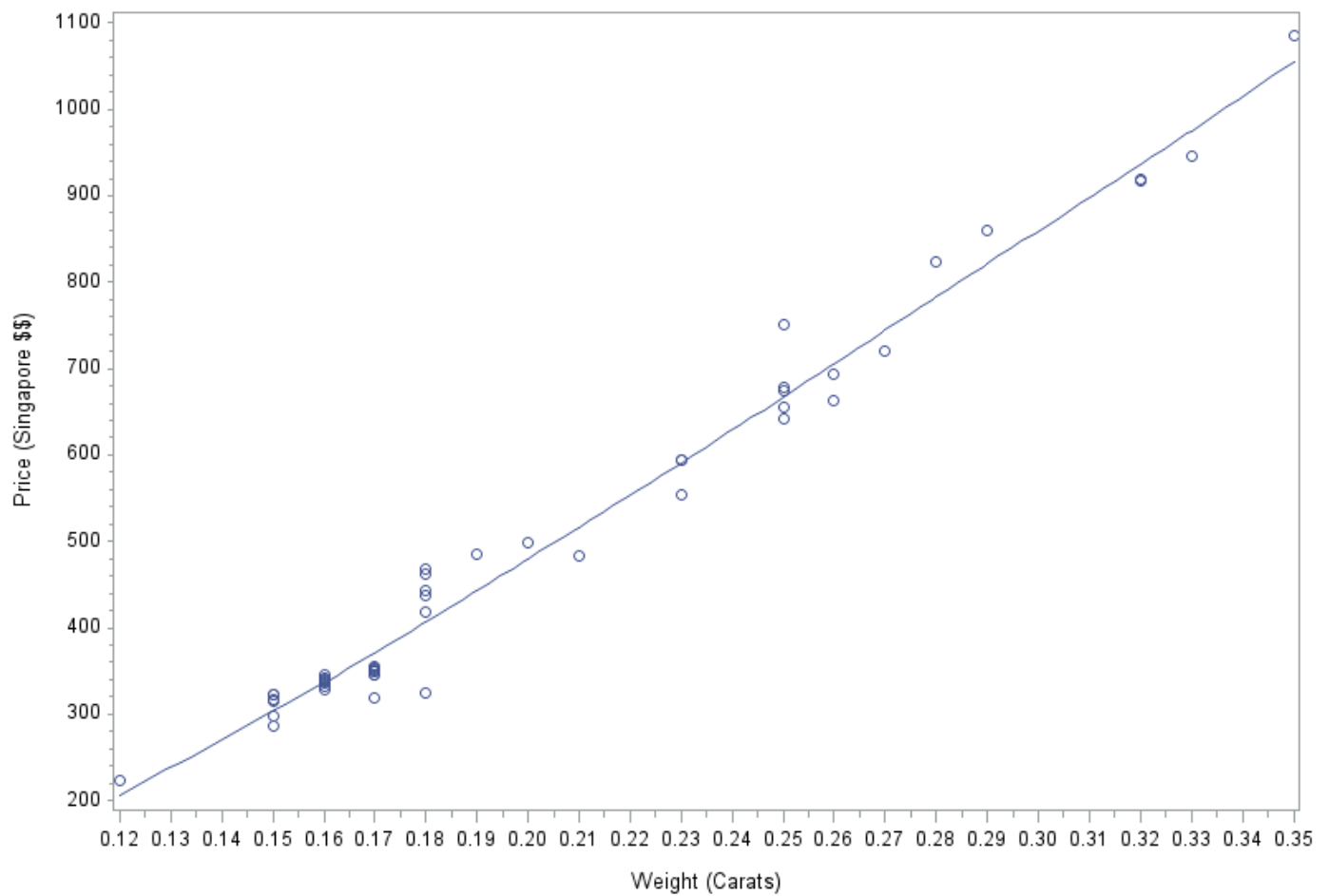
The SAS System

Obs	weight	price
1	0.17	355
2	0.16	328
3	0.17	350
4	0.18	325
5	0.25	642
6	0.16	342
7	0.15	322
8	0.19	485
9	0.21	483
10	0.15	323
11	0.18	462
12	0.28	823
13	0.16	336
14	0.20	498
15	0.23	595
16	0.29	860
17	0.12	223
18	0.26	663
19	0.25	750
20	0.27	720
21	0.18	468
22	0.16	345
23	0.17	352
24	0.16	332
25	0.17	353
26	0.18	438

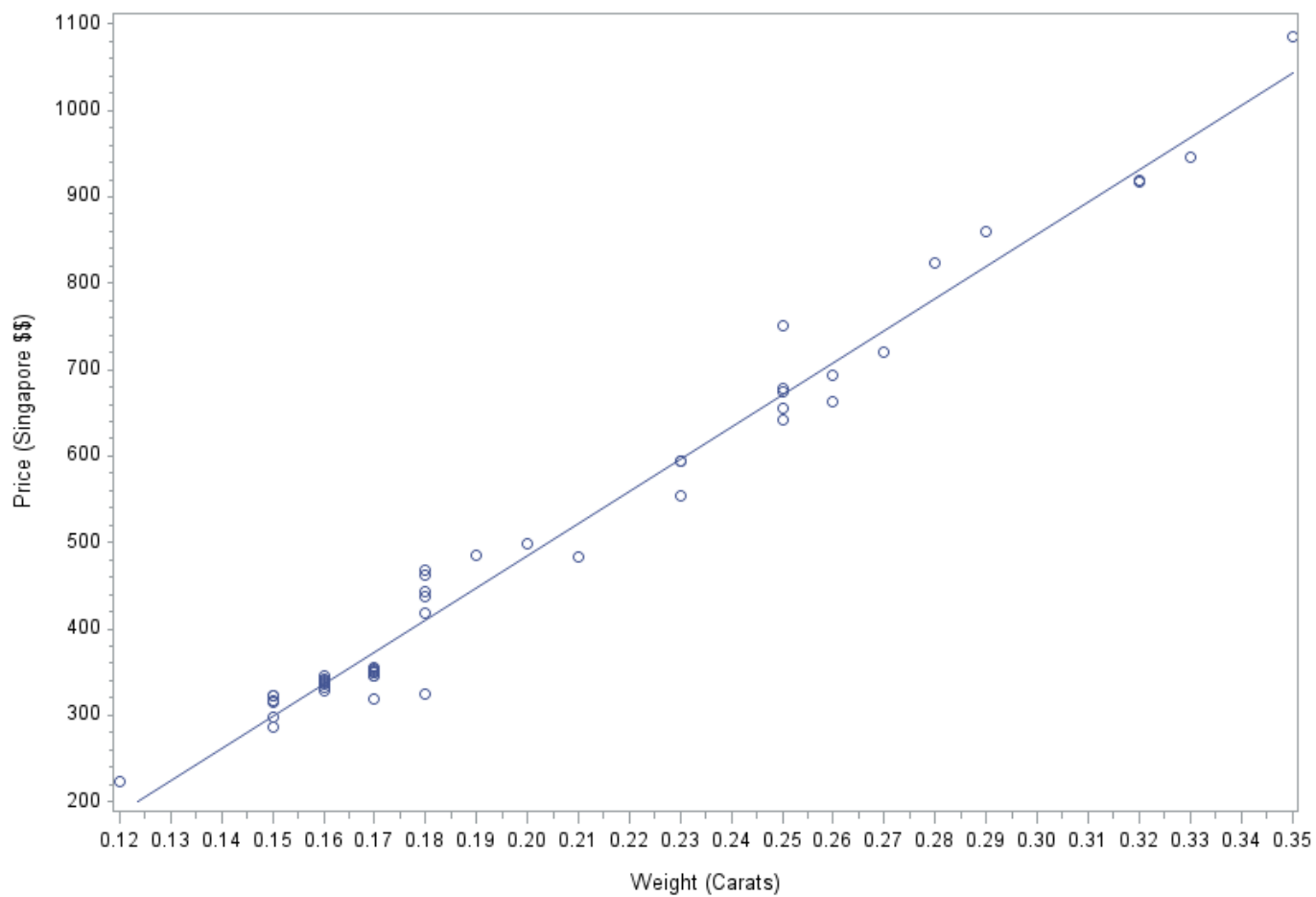
Obs	weight	price
27	0.17	318
28	0.18	419
29	0.17	346
30	0.15	315
31	0.17	350
32	0.32	918
33	0.32	919
34	0.15	298
35	0.16	339
36	0.16	338
37	0.23	595
38	0.23	553
39	0.17	345
40	0.33	945
41	0.25	655
42	0.35	1086
43	0.18	443
44	0.25	678
45	0.25	675
46	0.15	287
47	0.26	693
48	0.15	316
49	0.43	.

Diamond Ring Price Study

Scatter plot of Price vs. Weight with Smoothing Curve



Diamond Ring Price Study
Scatter plot of Price vs. Weight with Regression Line



Diamond Ring Price Study
Scatter plot of Price vs. Weight with Regression Line

The REG Procedure

Model: MODEL1

Dependent Variable: price

Number of Observations Read	49
Number of Observations Used	48
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2098596	2098596	2069.99	<.0001
Error	46	46636	1013.81886		
Corrected Total	47	2145232			

Root MSE	31.84052	R-Square	0.9783
Dependent Mean	500.08333	Adj R-Sq	0.9778
Coeff Var	6.36704		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Intercept	1	-259.62591	17.31886	-14.99	<.0001	-294.48696	-224.76486
weight	1	3721.02485	81.78588	45.50	<.0001	3556.39841	3885.65129

Diamond Ring Price Study

Scatter plot of Price vs. Weight with Regression Line

The REG Procedure

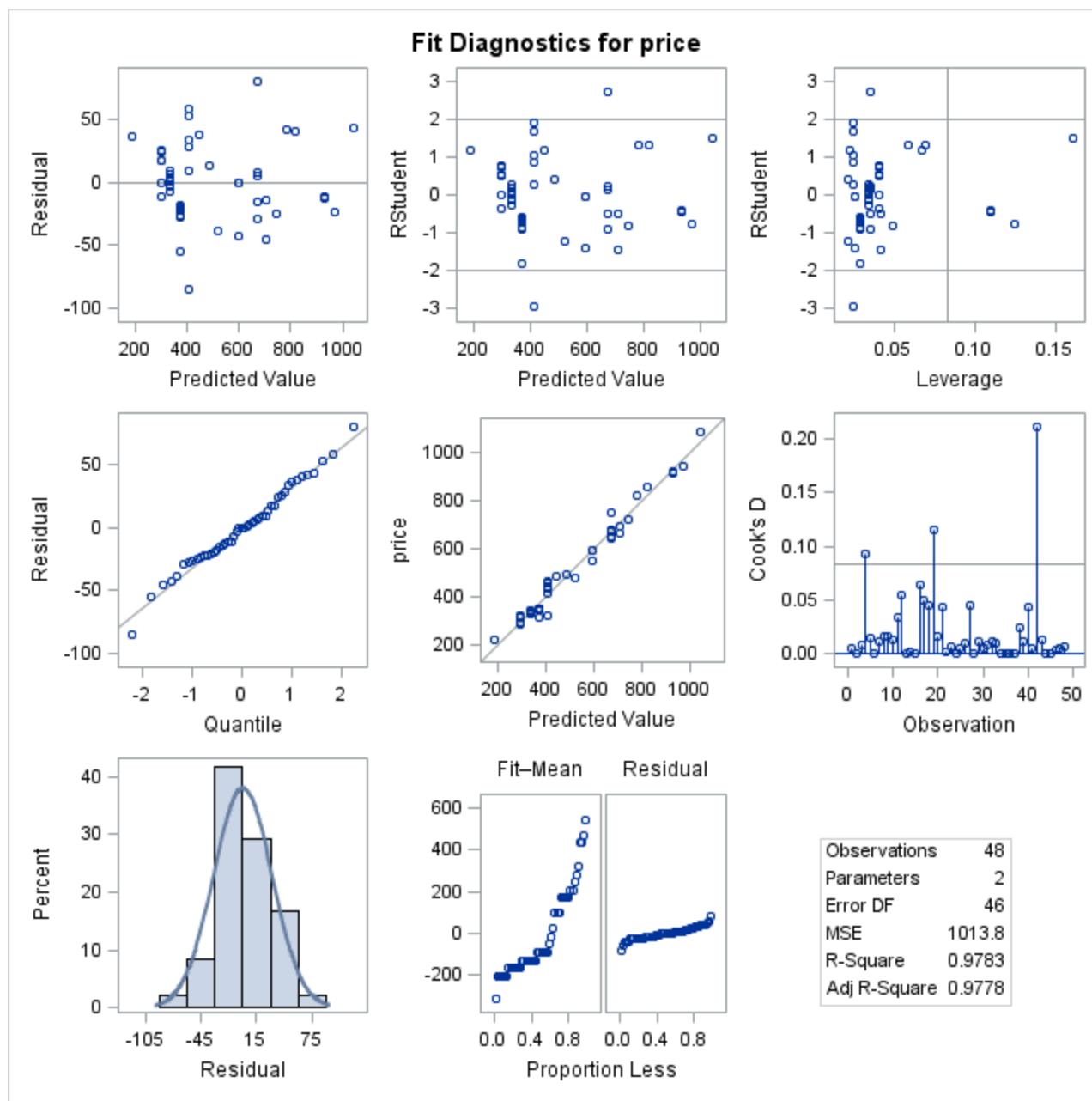
Model: MODEL1

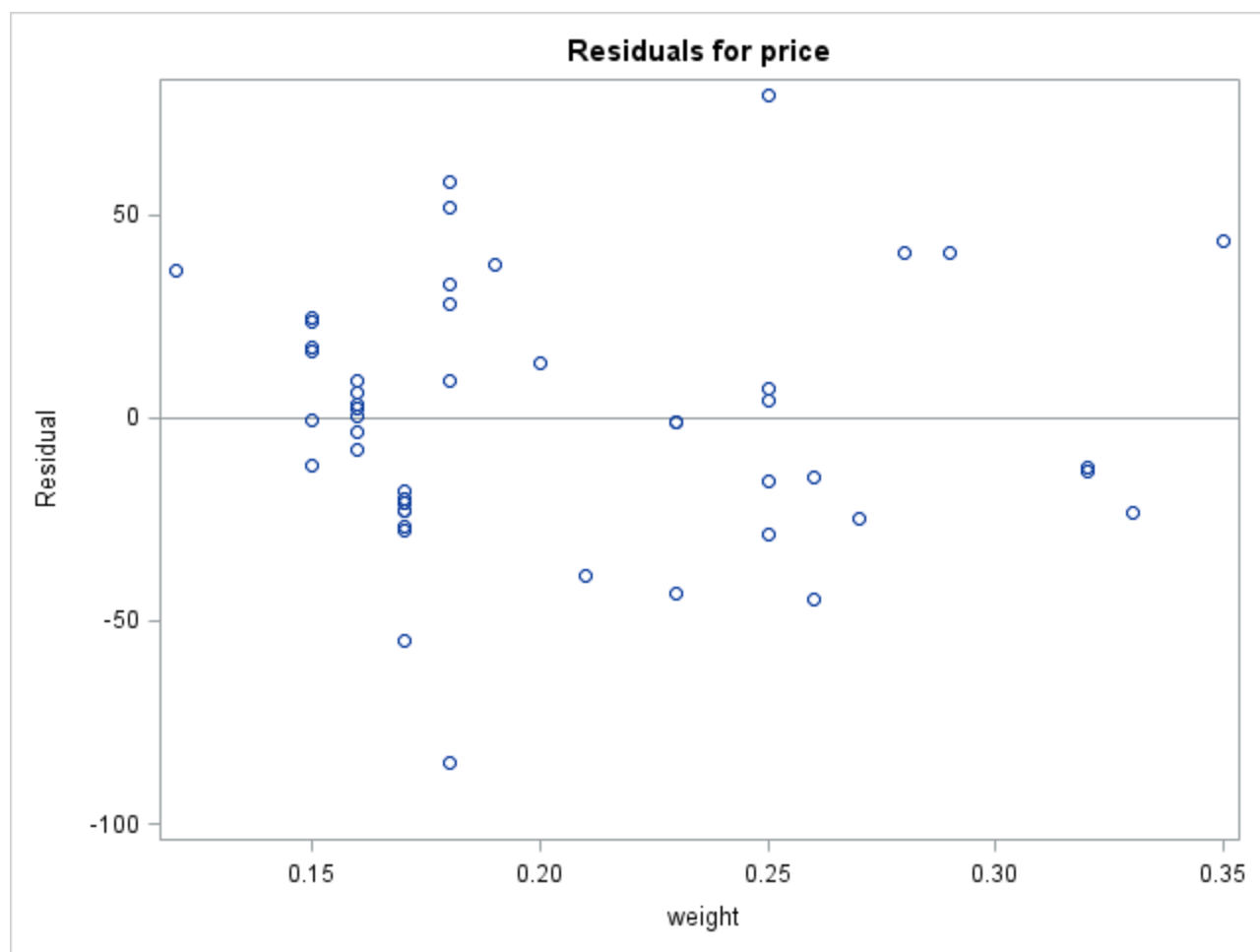
Dependent Variable: price

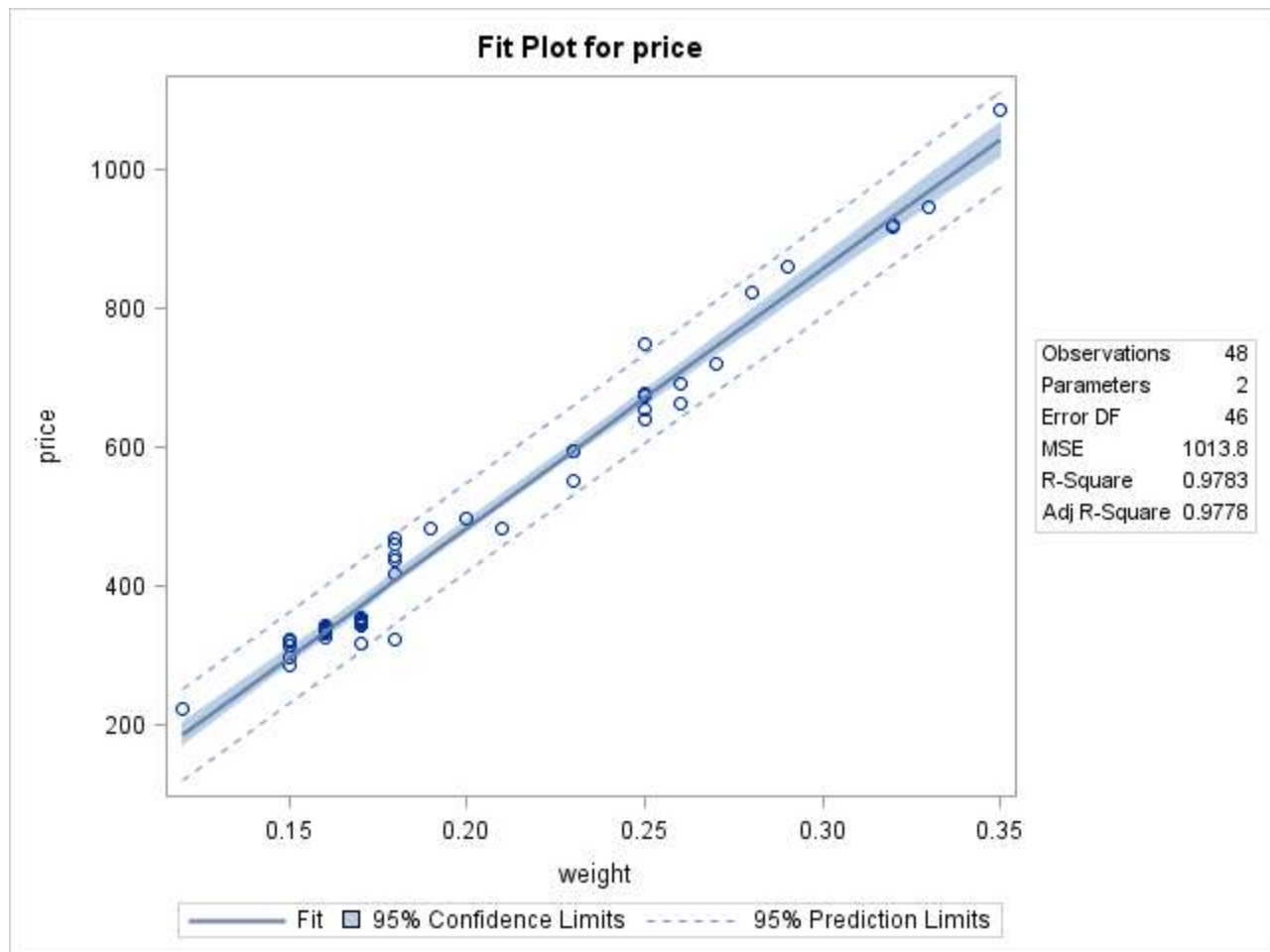
Output Statistics													
Obs	weight	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2	-1	0	1	2	Cook's D
1	0.17	355.0000	372.9483	5.3786	-17.9483	31.383	-0.572		*				0.005
2	0.16	328.0000	335.7381	5.8454	-7.7381	31.299	-0.247						0.001
3	0.17	350.0000	372.9483	5.3786	-22.9483	31.383	-0.731		*				0.008
4	0.18	325.0000	410.1586	5.0028	-85.1586	31.445	-2.708		*****				0.093
5	0.25	642.0000	670.6303	5.9307	-28.6303	31.283	-0.915		*				0.015
6	0.16	342.0000	335.7381	5.8454	6.2619	31.299	0.200						0.001
7	0.15	322.0000	298.5278	6.3833	23.4722	31.194	0.752			*			0.012
8	0.19	485.0000	447.3688	4.7396	37.6312	31.486	1.195			**			0.016
9	0.21	483.0000	521.7893	4.6205	-38.7893	31.503	-1.231		**				0.016
10	0.15	323.0000	298.5278	6.3833	24.4722	31.194	0.785			*			0.013
11	0.18	462.0000	410.1586	5.0028	51.8414	31.445	1.649			***			0.034
12	0.28	823.0000	782.2611	7.7193	40.7389	30.891	1.319			**			0.054
13	0.16	336.0000	335.7381	5.8454	0.2619	31.299	0.00837						0.000
14	0.20	498.0000	484.5791	4.6084	13.4209	31.505	0.426						0.002
15	0.23	595.0000	596.2098	5.0582	-1.2098	31.436	-0.0385						0.000
16	0.29	860.0000	819.4713	8.3905	40.5287	30.715	1.320			**			0.065
17	0.12	223.0000	186.8971	8.2768	36.1029	30.746	1.174			**			0.050
18	0.26	663.0000	707.8406	6.4787	-44.8406	31.174	-1.438		**				0.045
19	0.25	750.0000	670.6303	5.9307	79.3697	31.283	2.537			*****			0.116
20	0.27	720.0000	745.0508	7.0789	-25.0508	31.044	-0.807		*				0.017
21	0.18	468.0000	410.1586	5.0028	57.8414	31.445	1.839			***			0.043
22	0.16	345.0000	335.7381	5.8454	9.2619	31.299	0.296						0.002
23	0.17	352.0000	372.9483	5.3786	-20.9483	31.383	-0.668		*				0.007
24	0.16	332.0000	335.7381	5.8454	-3.7381	31.299	-0.119						0.000
25	0.17	353.0000	372.9483	5.3786	-19.9483	31.383	-0.636		*				0.006

Output Statistics													
Obs	weight	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2	-1	0	1	2	Cook's D
26	0.18	438.0000	410.1586	5.0028	27.8414	31.445	0.885		*				0.010
27	0.17	318.0000	372.9483	5.3786	-54.9483	31.383	-1.751		***				0.045
28	0.18	419.0000	410.1586	5.0028	8.8414	31.445	0.281						0.001
29	0.17	346.0000	372.9483	5.3786	-26.9483	31.383	-0.859		*				0.011
30	0.15	315.0000	298.5278	6.3833	16.4722	31.194	0.528		*				0.006
31	0.17	350.0000	372.9483	5.3786	-22.9483	31.383	-0.731		*				0.008
32	0.32	918.0000	931.1020	10.5294	-13.1020	30.049	-0.436						0.012
33	0.32	919.0000	931.1020	10.5294	-12.1020	30.049	-0.403						0.010
34	0.15	298.0000	298.5278	6.3833	-0.5278	31.194	-0.0169						0.000
35	0.16	339.0000	335.7381	5.8454	3.2619	31.299	0.104						0.000
36	0.16	338.0000	335.7381	5.8454	2.2619	31.299	0.0723						0.000
37	0.23	595.0000	596.2098	5.0582	-1.2098	31.436	-0.0385						0.000
38	0.23	553.0000	596.2098	5.0582	-43.2098	31.436	-1.375		**				0.024
39	0.17	345.0000	372.9483	5.3786	-27.9483	31.383	-0.891		*				0.012
40	0.33	945.0000	968.3123	11.2709	-23.3123	29.779	-0.783		*				0.044
41	0.25	655.0000	670.6303	5.9307	-15.6303	31.283	-0.500						0.004
42	0.35	1086	1043	12.7819	43.2672	29.162	1.484		**				0.211
43	0.18	443.0000	410.1586	5.0028	32.8414	31.445	1.044		**				0.014
44	0.25	678.0000	670.6303	5.9307	7.3697	31.283	0.236						0.001
45	0.25	675.0000	670.6303	5.9307	4.3697	31.283	0.140						0.000
46	0.15	287.0000	298.5278	6.3833	-11.5278	31.194	-0.370						0.003
47	0.26	693.0000	707.8406	6.4787	-14.8406	31.174	-0.476						0.005
48	0.15	316.0000	298.5278	6.3833	17.4722	31.194	0.560		*				0.007
49	0.43	.	1340	19.0332	.	.	.						.

Sum of Residuals	0
Sum of Squared Residuals	46636
Predicted Residual SS (PRESS)	50738







Diamond Ring Price Study

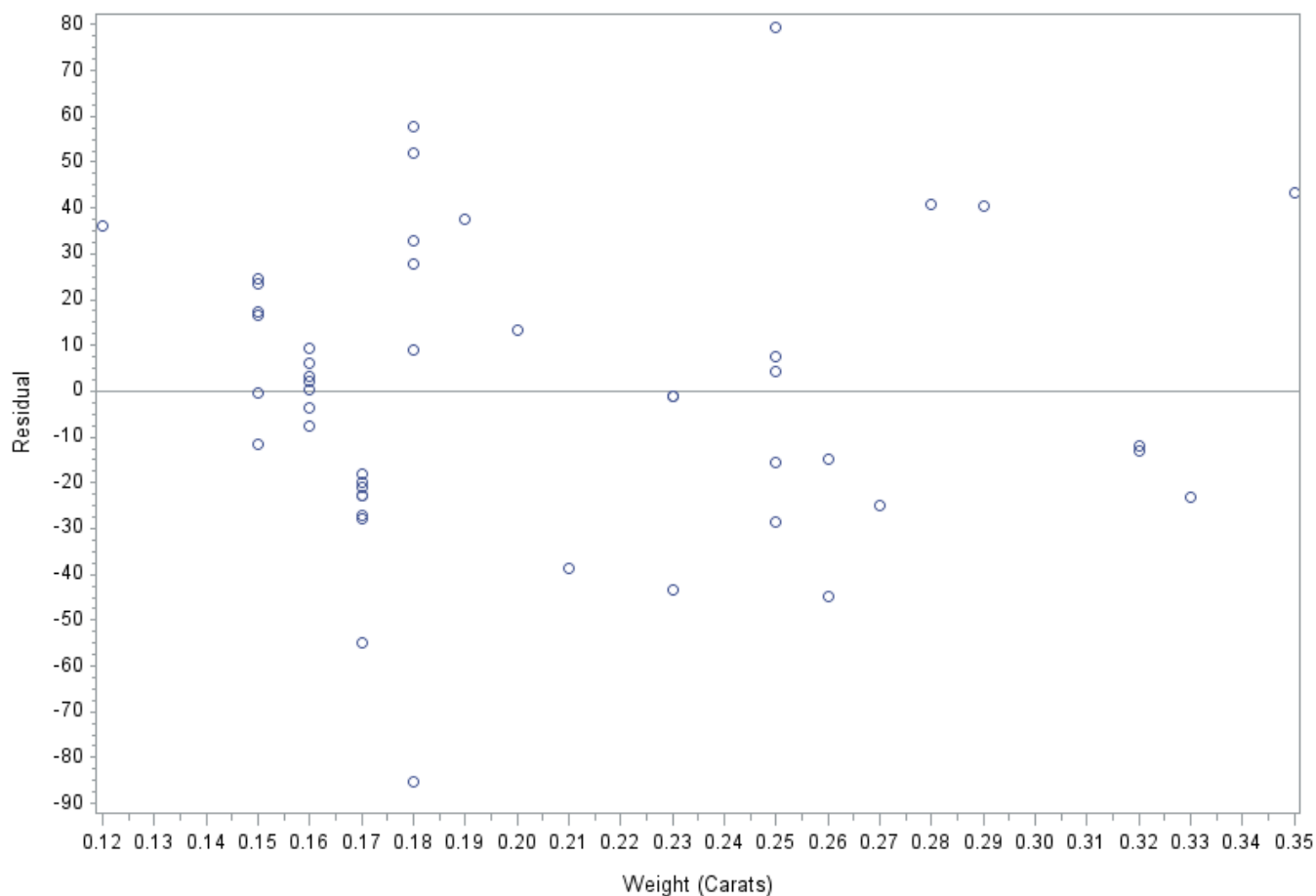
Scatter plot of Price vs. Weight with Regression Line

Obs	weight	price	pred	resid
1	0.17	355	372.95	-17.9483
2	0.16	328	335.74	-7.7381
3	0.17	350	372.95	-22.9483
4	0.18	325	410.16	-85.1586
5	0.25	642	670.63	-28.6303
6	0.16	342	335.74	6.2619
7	0.15	322	298.53	23.4722
8	0.19	485	447.37	37.6312
9	0.21	483	521.79	-38.7893
10	0.15	323	298.53	24.4722
11	0.18	462	410.16	51.8414
12	0.28	823	782.26	40.7389
13	0.16	336	335.74	0.2619
14	0.20	498	484.58	13.4209
15	0.23	595	596.21	-1.2098
16	0.29	860	819.47	40.5287
17	0.12	223	186.90	36.1029
18	0.26	663	707.84	-44.8406
19	0.25	750	670.63	79.3697
20	0.27	720	745.05	-25.0508
21	0.18	468	410.16	57.8414
22	0.16	345	335.74	9.2619
23	0.17	352	372.95	-20.9483
24	0.16	332	335.74	-3.7381
25	0.17	353	372.95	-19.9483
26	0.18	438	410.16	27.8414
27	0.17	318	372.95	-54.9483
28	0.18	419	410.16	8.8414

Obs	weight	price	pred	resid
29	0.17	346	372.95	-26.9483
30	0.15	315	298.53	16.4722
31	0.17	350	372.95	-22.9483
32	0.32	918	931.10	-13.1020
33	0.32	919	931.10	-12.1020
34	0.15	298	298.53	-0.5278
35	0.16	339	335.74	3.2619
36	0.16	338	335.74	2.2619
37	0.23	595	596.21	-1.2098
38	0.23	553	596.21	-43.2098
39	0.17	345	372.95	-27.9483
40	0.33	945	968.31	-23.3123
41	0.25	655	670.63	-15.6303
42	0.35	1086	1042.73	43.2672
43	0.18	443	410.16	32.8414
44	0.25	678	670.63	7.3697
45	0.25	675	670.63	4.3697
46	0.15	287	298.53	-11.5278
47	0.26	693	707.84	-14.8406
48	0.15	316	298.53	17.4722
49	0.43	.	1340.41	.

Diamond Ring Price Study

Residual Plot



log window

NOTE: Copyright (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software 9.3 (TS1M1)

Licensed to PURDUE UNIVERSITY - T&R, Site 70085364.

NOTE: This session is executing on the X64_7PRO platform.

NOTE: Updated analytical products:

SAS/STAT 9.3_M1, SAS/ETS 9.3_M1, SAS/OR 9.3_M1

NOTE: SAS initialization used:

real time 10.74 seconds

cpu time 1.45 seconds

```
1 *If you are running version 9.3 locally (either on your personal computer or on
2   an ITAP computer), the following will reset the output. The lines will NOT work
3   if you are using goremote.;
```

```
4 ods html close;
```

```
5 ods html;
```

NOTE: Writing HTML Body file: sashtml.htm

```
6
```

```
7  *The following linesize (ls) and pagesize (ps) options MAY work
8  well if you have your print setup (click file, print setup)
9  with 0.5 in margins and portrait selected on page setup and
10 and SAS Monospace, Roman, size 8 selected on font. The
11 print setup display will tell you the ls and ps for the
12 selections you have chosen. Some printers may be a little
13 different and you may need to play with these settings. ;
14
15 options ls=105 ps=60 nocenter;
16
17 *Read in the data using the cards (datalines) statement. The @@ allows more
18 than one case per line. The lone . represents a missing value
19 and we can use this for prediction of price at that weight;
20 data diamonds; input weight price @@;
21 cards;
```

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.

NOTE: The data set WORK.DIAMONDS has 49 observations and 2 variables.

NOTE: DATA statement used (Total process time):

real time	0.77 seconds
cpu time	0.03 seconds

```
29 ;
30
31 *Create new data set that does not include the last case (we do
32 this for plotting purposes since we don't want 0.43 included
33 on the x-axis in our plots);
34 data diamonds1; set diamonds; if price ne .;
35
36 *Print the data set diamonds1;
```

NOTE: There were 49 observations read from the data set WORK.DIAMONDS.

NOTE: The data set WORK.DIAMONDS1 has 48 observations and 2 variables.

NOTE: DATA statement used (Total process time):

real time	0.02 seconds
cpu time	0.01 seconds

```
37 proc print data=diamonds; run;
```

NOTE: There were 49 observations read from the data set WORK.DIAMONDS.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.29 seconds
cpu time	0.03 seconds

```
38
39 *Sort the data according to weight (if we don't, the smoothing
40 curve on our plot will not work correctly);
41 proc sort data=diamonds1; by weight;
42
43 *Generate a scatterplot with smooth curve fitted to
44 the data. Note that there are several preceding statements
45 that can be used to title the plot and axes.;
46 symbol1 v=circle i=sm70;
47 title1 'Diamond Ring Price Study';
48 title2 'Scatter plot of Price vs. Weight with Smoothing Curve';
49 axis1 label=('Weight (Carats)');
50 axis2 label=(angle=90 'Price (Singapore $$)');
```

NOTE: There were 48 observations read from the data set WORK.DIAMONDS1.

NOTE: The data set WORK.DIAMONDS1 has 48 observations and 2 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.39 seconds
cpu time	0.01 seconds

```
51 proc gplot data=diamonds1;
52     plot price*weight / haxis=axis1 vaxis=axis2;
53 run;
```

NOTE: Input data contained multiple vertical values for individual horizontal values. Parametric fit is required to force curve through individual observations.

NOTE: 4 records written to D:\Users\lfindsen\gplot.png.

```
54
55 *To copy plots from SAS to WORD: (1) In SAS, select the plot,
56     right click and choose COPY. (2) In WORD, put the cursor in the
57     desired location, PASTE SPECIAL and select
58     "HTML Format".
59
60 *We can also make a plot with a regression line;
61 symbol1 v=circle i=r1;
62 title2 'Scatter plot of Price vs. Weight with Regression Line';
```

NOTE: There were 48 observations read from the data set WORK.DIAMONDS1.

NOTE: PROCEDURE GPLOT used (Total process time):

real time	2.03 seconds
cpu time	0.32 seconds

```
63 proc gplot data=diamonds1;
64     plot price*weight / haxis=axis1 vaxis=axis2;
65 run;
```

NOTE: Regression equation : $\text{price} = -259.6259 + 3721.025 \times \text{weight}$.

NOTE: 5 records written to D:\Users\lfindsen\gplot1.png.

```
66
67 *Perform regression analysis using data set 'diamonds'. The clb option
68     generates confidence interval for the slope and intercept.
69     The p option generates fitted values and standard errors.
70     The r option does some residual analysis (i.e., check
71     assumptions). The output statement generates a new data set
72     that contains the residuals and predicted/fitted values. The
73     id statement adds the variable specified to the fitted values output;
```

NOTE: There were 48 observations read from the data set WORK.DIAMONDS1.

NOTE: PROCEDURE GPLOT used (Total process time):

real time	0.50 seconds
cpu time	0.34 seconds

```
74 proc reg data=diamonds; model price=weight/clb p r;
75     output out=diag p=pred r=resid;
76     id weight; run;
```

```
77
```

NOTE: The data set WORK.DIAG has 49 observations and 4 variables.

NOTE: PROCEDURE REG used (Total process time):

real time	5.65 seconds
cpu time	1.15 seconds

```
78  proc print data=diag; run;
```

NOTE: There were 49 observations read from the data set WORK.DIAG.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.08 seconds
cpu time	0.01 seconds

```
79  *generates a residual plot to assess model assumptions;
80  *the following code is not necessary if ODS graphics is on;
81  symbol1 v=circle i=NONE;
82  title2 color=blue 'Residual Plot';
83  axis2 label=(angle=90 'Residual');
84  proc gplot data=diag; plot resid*weight / haxis=axis1 vaxis=axis2 vref=0;
85     where price ne .;
86  run;
```

NOTE: 4 records written to D:\Users\lfindsen\gplot2.png.

```
87  quit;
```

NOTE: There were 48 observations read from the data set WORK.DIAG.

WHERE price not = .;

NOTE: PROCEDURE GPLOT used (Total process time):

real time	0.45 seconds
cpu time	0.24 seconds