

Attached to this exam are some 9 pages of JMP-IN printout useful in the analysis of some data taken from *Statistics* by McClave and Sincich. The data concern the selling prices of $n = 25$ Minneapolis apartment buildings.

Begin by considering the prediction of $y =$ sale price (in \$1000) from only $x =$ building area (in ft²). And in particular, page 1 of the printout is from a simple linear regression of y on x .

a) What fraction of the raw variability in sale price is "accounted for" using area as a predictor variable?

b) What are the practical meanings of the SLR parameters β_1 and σ in this context?

β_1 :

σ :

c) Give 95% two-sided confidence limits for these two parameters (NO NEED TO SIMPLIFY, just plug correct numbers into correct formulas).

β_1 :

σ :

d) Give 95% two-sided confidence limits for the mean sales price of a 10,000 ft² apartment building in Minneapolis at the time of this study. (The sample mean of x is $\bar{x} = 11,423.4$ and the sample standard deviation of x is $s_x = 10,019.4$.) Show some work, but after plugging into a formula, NO NEED TO SIMPLIFY.

e) A Minneapolis landlord gets a property tax notice that places the valuation of an apartment building (that is supposed to be based on fair market value) at \$900,000. This building is of size 35,000 ft². The landlord wants to challenge the valuation. Based on the information on page 1 of the printout, does the landlord have a case? EXPLAIN. (More is required here than simply noting that \hat{y} is less than \$900,000. BUT YOU NEED NOT SHOW ANY CALCULATIONS.)

f) Temporarily suppose that the estimated parameter values from the SLR are exactly equal to the (actually unknown) parameter values. (This is too much to hope for, but humor me here.) Under this possibility, what fraction of sales of 20,000 ft² Minneapolis apartment buildings would be for more than \$500,000?

g) What about the plot on page 1 of the printout causes some concern that the SLR model relating y to x may not be a completely appropriate one? Page 2 of the printout shows a plot of \sqrt{y} versus \sqrt{x} and the associated least squares line. Why does or does not this 2nd plot look "better" in terms of appropriateness of the standard SLR model?

Comments about page 1:

Comments about page 2:

h) A SLR analysis of (\sqrt{x}, \sqrt{y}) pairs "transforms back" to the original scales (through squaring) to produce the report on page 3 of the printout. Consider again the issue raised in part e). Would your answer be different based on this 2nd analysis? Explain.

There was, in fact, more information available on the 25 buildings than simply their sizes and sales prices. The numbers of apartments in the buildings, the building ages (in years, according to McClave and Sincich), and the numbers of off-street parking spots were also available. The complete data set is on page 4 of the printout. Pages 5-9 of the printout are concerned with a regression of sales price on the other $k = 4$ variables. (No guarantee that this is a good model ... just the one we'll use here.)

i) The MLR estimate of " σ " is smaller than the SLR estimate of " σ ". Explain the difference in the two meanings of " σ " and why you are not surprised about the reduction in " s ".

j) What fraction of the raw variability in sales price is "accounted for" using all 4 predictor variables? If you wanted to attach a p -value this (or equivalently determine if these 4 predictors provide a "statistically significant" ability to predict sales price) where on the printout would you get such a p -value, and what is it?

fraction accounted for:

p -value and origin:

k) After accounting for building size, do the other 3 predictors "add significantly to one's ability to predict y "? Compute an appropriate F statistic, give the degrees of freedom and say what the F tables in the book reveal about the associated p -value.

$F =$ _____ d.f. = _____ p -value:

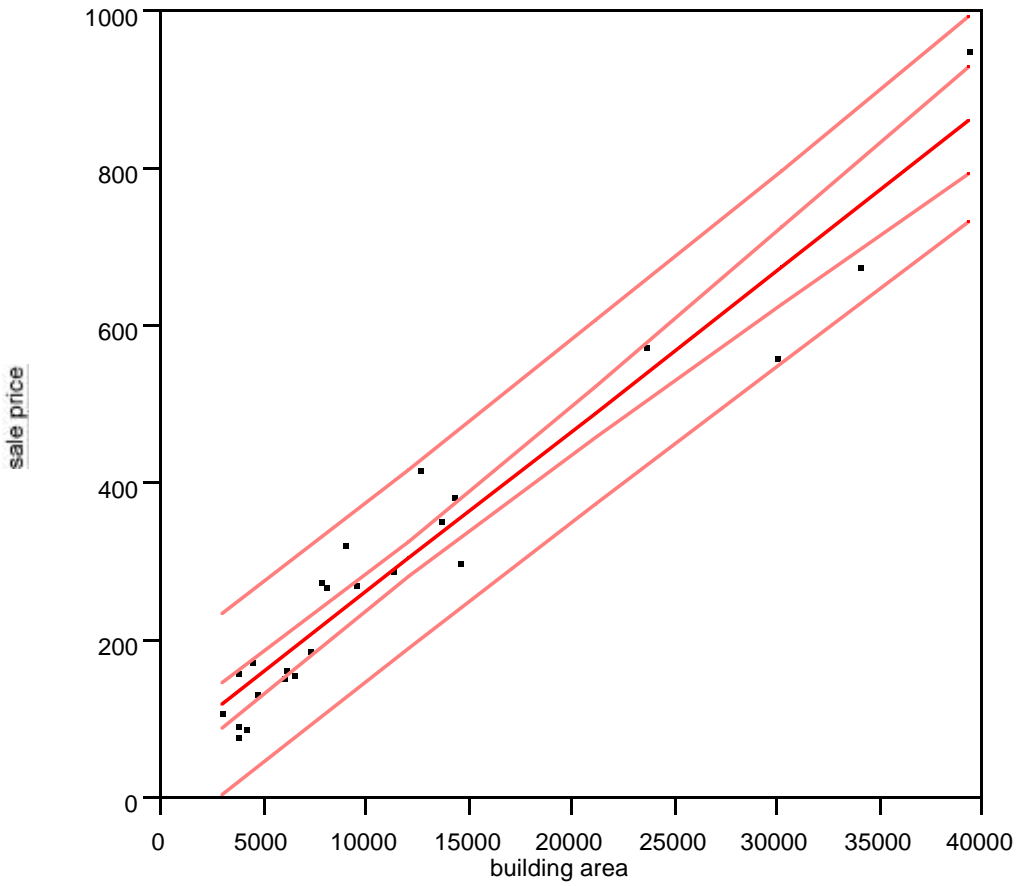
l) Suppose the current Minneapolis formula for making real estate valuations adds \$5,000 per off-street parking place to an apartment buildings' assessed valuation. What does this MLR analysis have to say about that part of the Minneapolis formula?

m) The largest (in absolute value) residual here is associated with building #25. Interestingly enough, the SLR formula on page 1 of the printout does a much better job of predicting its selling price. Looking at the predictor values for the case, what "explanation" can you offer for the large MLR residual?

n) What are 95% two-sided prediction limits for the selling price of a building with characteristics like that of building #25? (NO NEED TO SIMPLIFY.)

o) Look at the "prediction profile" part of the JMP report. Note that while these plots are set up around the values of predictors for building #12, they would just be shifted up or down (and not changed in fundamental character) by changing this point to some other values. At least across this data set, which predictor variable seems to have the largest impact on selling price? Explain.

sale price By building area



Linear Fit

Linear Fit

sale price = 57.0937 + 0.02044 building area

Summary of Fit

RSquare	0.93723
RSquare Adj	0.934501
Root Mean Square Error	54.13609
Mean of Response	290.5735
Observations (or Sum Wgts)	25

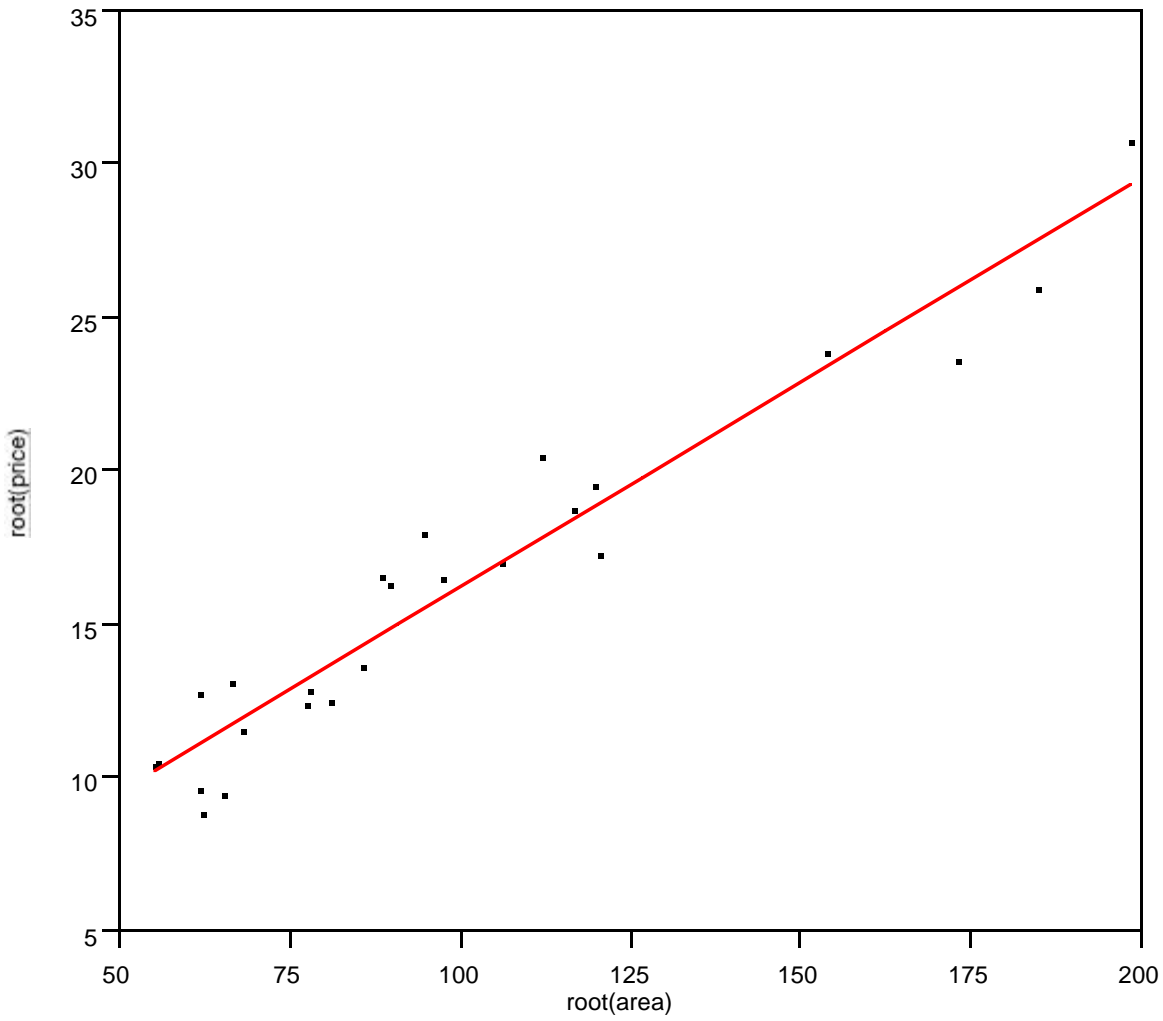
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1006463.5	1006463	343.4189
Error	23	67406.5	2931	Prob>F
C Total	24	1073869.9		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	57.093707	16.61217	3.44	0.0022
building area	0.0204387	0.001103	18.53	<.0001

root(price) By root(area)



— Linear Fit

Linear Fit

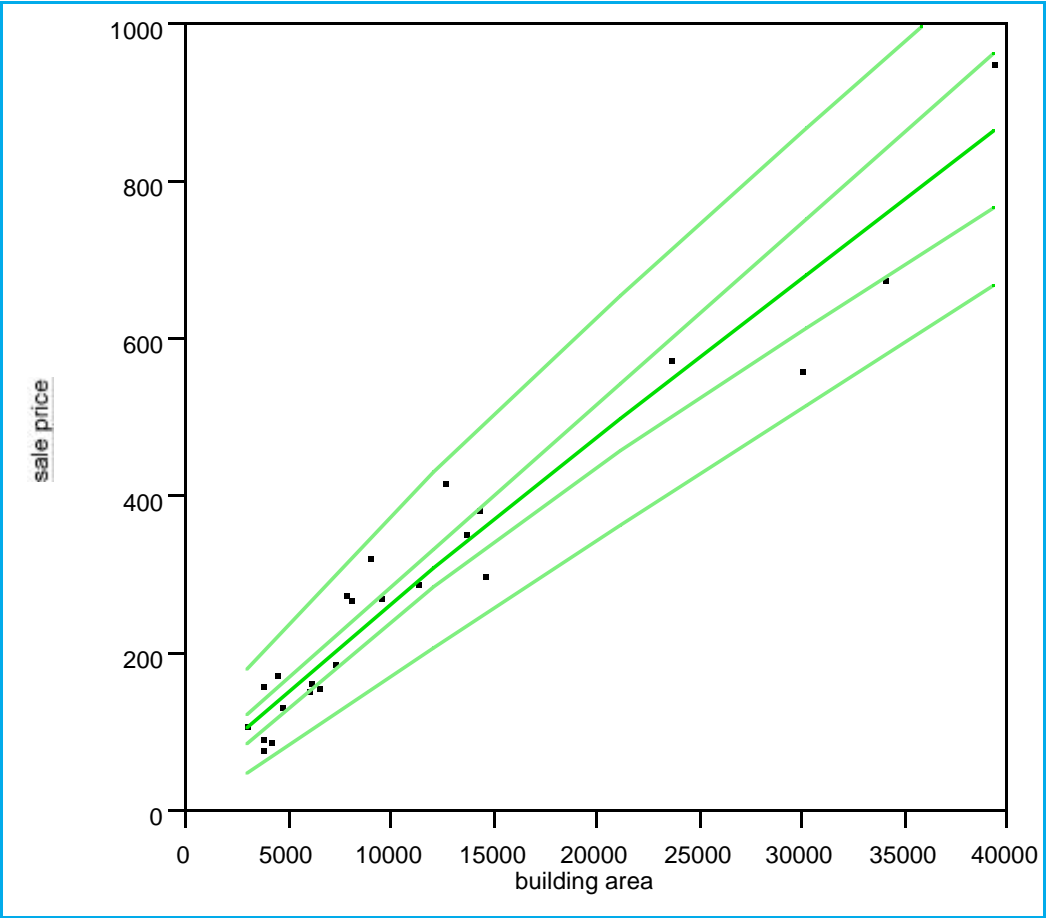
$$\text{root}(\text{price}) = 2.89891 + 0.13344 \text{ root}(\text{area})$$

Summary of Fit

Analysis of Variance

Parameter Estimates

sale price By building area



Transformed Fit Sqrt to Sqrt

$$\text{Sqrt}(\text{sale price}) = 2.89891 + 0.13344 \text{ Sqrt}(\text{building area})$$

Summary of Fit

Analysis of Variance

Parameter Estimates

Fit Measured on Original Scale

Sum of Squared Error	65352.659
Root Mean Square Error	53.304971
R-square	0.9391429
Sum of Residuals	51.720223

Mnsales

Rows	number of apartments	age	parking spaces	building area	sale price
1	4	82	0	4266	90.3
2	20	13	0	14391	384
3	5	66	0	6615	157.5
4	26	64	6	34144	676.2
5	5	55	0	6120	165
6	10	65	0	14552	300
7	4	82	0	3040	108.75
8	11	23	0	7881	276.538
9	20	18	20	12600	420
10	62	71	3	39448	950
11	26	74	0	30000	560
12	13	56	13	8088	268
13	9	76	0	11315	290
14	6	21	6	4461	173.2
15	11	24	8	9000	323.65
16	5	19	5	3828	162.5
17	20	62	2	13680	353.5
18	4	70	0	4680	134.4
19	8	19	0	7392	187
20	4	57	0	6030	155.7
21	4	82	0	3840	93.6
22	4	50	0	3092	110
23	14	10	0	23704	573.2
24	4	82	0	3876	79.3
25	5	82	0	9542	272

Response: sale price

Summary of Fit

RSquare	0.980361
RSquare Adj	0.976434
Root Mean Square Error	32.47245
Mean of Response	290.5735
Observations (or Sum Wgts)	25

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	99.966826	18.66129	5.36	<.0001
number of apart	4.4441508	1.156072	3.84	0.0010
age	-0.886493	0.275437	-3.22	0.0043
parking spaces	2.6059525	1.518891	1.72	0.1017
building area	0.0154868	0.00142	10.91	<.0001

Effect Test

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
number of apart	1	1	15582.53	14.7777	0.0010
age	1	1	10922.83	10.3587	0.0043
parking spaces	1	1	3103.92	2.9436	0.1017
building area	1	1	125506.63	119.0246	<.0001

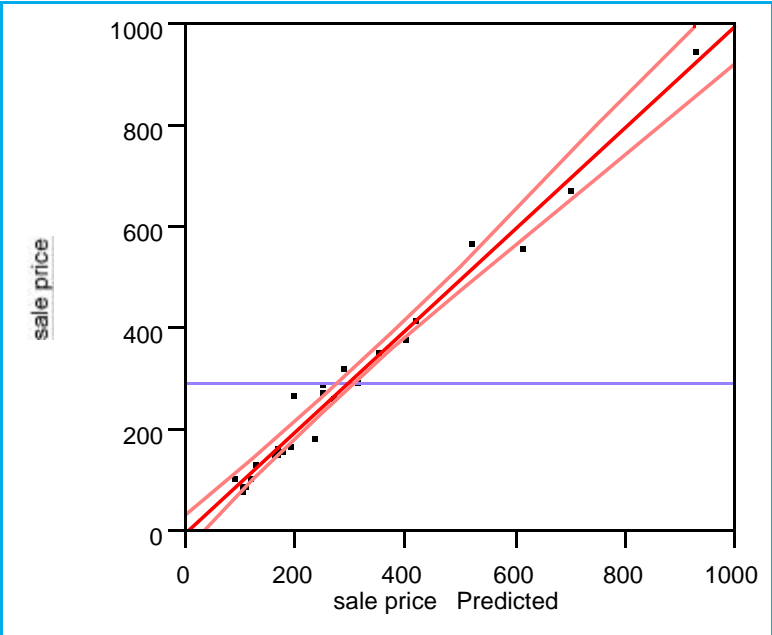
Whole-Model Test

number of apart

age

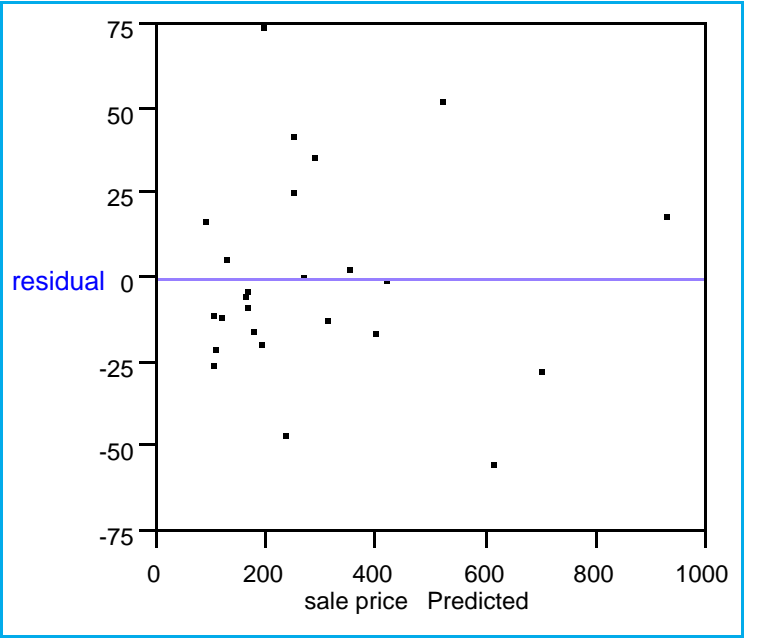
parking spaces

building area



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	1052780.7	263195	249.6019
Error	20	21089.2	1054	Prob>F
C Total	24	1073869.9		<.0001



Screening Fit

sale price

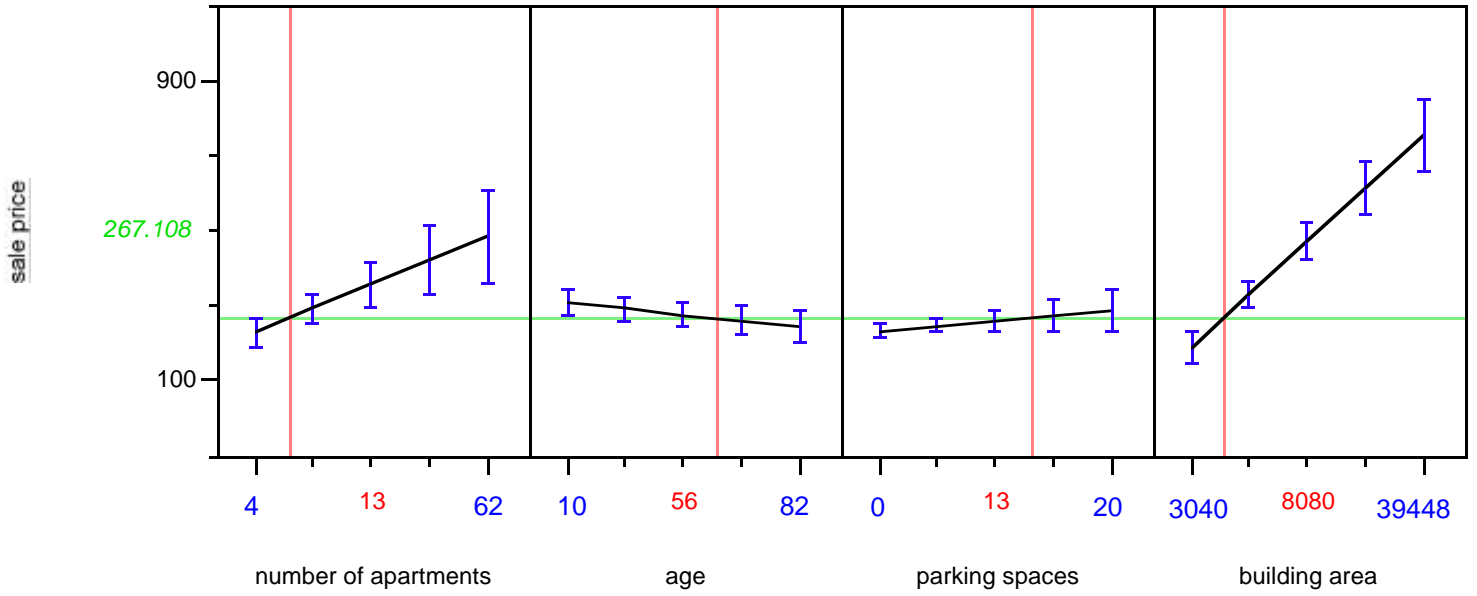
Summary of Fit

Analysis of Variance

Parameter Estimates

Effect Test

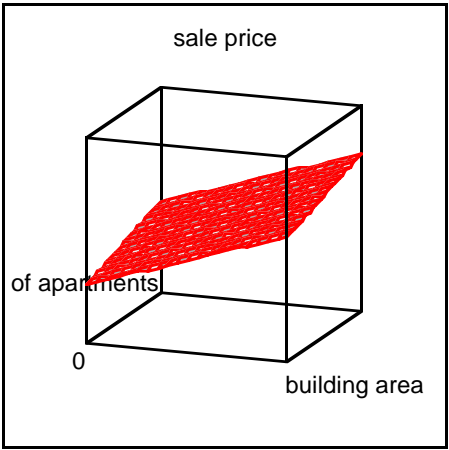
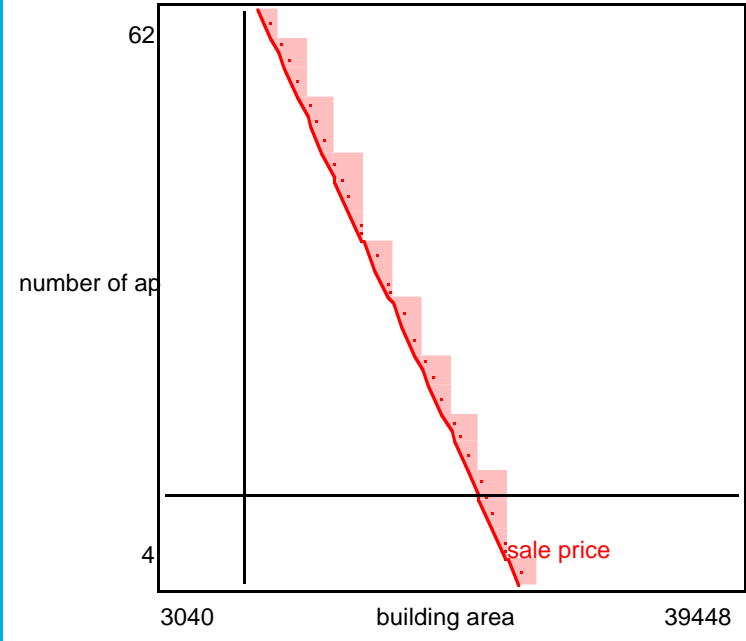
Prediction Profile



Contour Profiler

Horiz	Vert	Factor	Current X	Grid Density	Update Mode
<input type="checkbox"/>	<input checked="" type="checkbox"/>	number of apartments	13	20 x 20	Immediate
<input type="checkbox"/>	<input type="checkbox"/>	age	56		
<input type="checkbox"/>	<input type="checkbox"/>	parking spaces	13		
<input checked="" type="checkbox"/>	<input type="checkbox"/>	building area	8080		

Response	Contour	Current Y	Lo Limit	Hi Limit
sale price	500	267.108	?	500



Mnsales

Rows	Predicted sale price	StdErr Pred sale price
1	111.1178	11.05292
2	400.1961	16.16856
3	166.1243	8.34342
4	703.1965	22.26557
5	168.2098	8.034507
6	312.1504	9.58784
7	92.13093	11.59105
8	250.5147	12.93596
9	420.1458	24.40365
10	931.3047	29.84458
11	614.5186	16.35477
12	267.2319	17.19549
13	247.824	9.257175
14	192.7378	11.56224
15	287.8056	11.23431
16	177.6575	11.88533
17	350.9587	9.502139
18	128.1672	9.186342
19	233.1552	13.13862
20	160.5988	8.176873
21	104.5204	11.21224
22	121.304	9.43203
23	520.4193	21.18746
24	105.0779	11.1976
25	197.2703	11.21527