Math 3330 Project Report

Yiming Shao (215159932), Shuyu Fu (214184279), Xuanchen Guo (216427114)

Recap:

We chose the energy efficiency data set to analyze. This data set was published on November 30th, 2012. It was originally created by Angeliki Xifara /Structural engineer), The sample size is 768. This data set measures the heating load and the cooling load of different types of buildings, it contains 8 attributes and 2 response variables.

x1: Relative Compactness x2: Surface Area

x3: Wall Area x4: Roof Area

x5: Overall Height x6: Orientation

x7: Glazing Area x8: Glazing Area Distribution

y1: Heating Load

y2: Cooling Load

We decided to use the heating load as our response variable, with all eight attributes for our linear regression model.

In order to achieve a better model, we have done the following data analytic methods, removing outliers, dealing with multicollinearity, dropping insignificant variables, model comparison, model improvements, assumption checks and the evaluation of the final model.

But we decided to drop two attributes and add one interaction term. So the final model is: $y = \beta0 + x1\beta1 + x2\beta2 + x3\beta3 + x5\beta4 + x7\beta5 + x8\beta6 + x1x2\beta7$.

Build Model (Phase 1):

After looking Model comparison statistics for all reasonable regression models all possible models with full rank are compared by their R squared index and adjusted R square index, and there are three models have the equal best results, they are y = β0 + x1β1 + x3β2 + x4β3 + x5β4 + x7β5 + x8β6, y = β0 + x1β1 + x2β2 + x4β3 + x5β4 + x7β5 + x8β6 and y = β0 + x1β1 + x2β2 + x3β3 + x5β4 + x7β5 + x8β6, even though the best model analyzed by stepwise procedure is y =β0 + x1β1 + x3β2 + x4β3 + x5β4 + x7β5 + x8β6, however, we need to consider more factors to choose the best model among those three.

By observing the parameter estimates, we found that

x4 =	2.97E-8 * Intercept - 1.63E-8 * x1 + 0.5 * x2 - 0.5 * x3 + 342E-12 * x5

x4 is a linear combination of x1 x2 x3 x5 other parameters, so x4 should be removed here, the best model so far is the one showing below y = β0 + x1β1 + x2β2 + x3β3 + x5β4 + x7β5 + x8β6, x6 (Orientation) and x4 (Roof Area) are removed.

Analysis and Improvement (Phase 2):

Overall F-test:

Number of Observations Read	768
Number of Observations Used	768

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	71546	10221	1187.06	<.0001
Error	760	6543.76619	8.61022		
Corrected Total	767	78090			

so we applied f test for the model we have so far.

The overall F-test has a p-value that is smaller than 0.0001, which indicates that at least one of the β's is not zero.

there is no leverage value is smaller than 0.02344.

Outlier with respect to x: We did not find any outlier with respect to x values.

Outlier with respect to y:

t_0.025 (768-9-1) = 1.9631

Some obs` absolute value of studentized residual is greater than 1.9631. Indicate that there exists some outliers W.R.T. y value.

Covariance ratio references are 1+3k/n = 1.0352 1-3k/n = 0.9648

Eliminating observations that have covariance ratio < 0.9648 will enhance the model, and eliminating observations with covariance ratio > 1.0352 will damage the model, so we should remove obs that has covariance ratio > 1.0352.

In the end, 51 outliers and influential points were removed from the original data set, the new data set will be used to improve the model.

Correlation matrix:

Pearson Correlation Coefficients, N = 768
Prob > |r| under H0: Rho=0

	x1	x2	x3	x4	x5	x6	x7	x8	y1
x1	1.00000	-0.99190 <.0001	-0.20378 <.0001	-0.86882 <.0001	0.82775 <.0001	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.62227 <.0001
x2	-0.99190 <.0001	1.00000	0.19550 <.0001	0.88072 <.0001	-0.85815 <.0001	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	-0.65812 <.0001
x3	-0.20378 <.0001	0.19550 <.0001	1.00000	-0.29232 <.0001	0.28098 <.0001	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.45567 <.0001
x4	-0.86882 <.0001	0.88072 <.0001	-0.29232 <.0001	1.00000	-0.97251 <.0001	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	-0.86183 <.0001
x5	0.82775 <.0001	-0.85815 <.0001	0.28098 <.0001	-0.97251 <.0001	1.00000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.88943 <.0001
x6	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	1.00000	0.00000 1.0000	0.00000 1.0000	-0.00259 0.9429
x7	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	1.00000	0.21296 <.0001	0.26984 <.0001
x8	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.00000 1.0000	0.21296 <.0001	1.00000	0.08737 0.0154
y1	0.62227 <.0001	-0.65812 <.0001	0.45567 <.0001	-0.86183 <.0001	0.88943 <.0001	-0.00259 0.9429	0.26984 <.0001	0.08737 0.0154	1.00000

By looking at the matrix, x6 and x4 are removed now, x1 and x2 are clearly correlated, this information helps us to improve our model, a new possible interaction term x1*x2 can be considered, an improved model y = β0 + x1β1 + x2β2 + x3β3 +

x5β4 + x7β5 + x8β6 + x1x2β7 will be compared to the previous model in the next step.

Next part is the comparison between original model and improved model.

We have five criterions needed to be analyzed here.

R^2: The original model has a R-square of 0.9369 and an adjusted R-square of 0.9363. The improved model has a R-square of 0.9404 and an adjusted R-square of 0.9398. The larger values indicate a better fit for the improved model.

Standard error: The mean square error for the original model is 6.25209, so the standard error is √6.25209 = 2.5004. The mean square error for the improved model is 5.90884, which gives a lower standard error of √5.90884 = 2.4308.

VIF: variance inflation factor. The VIFs of the improved data are relatively smaller than the original ones, it indicates smaller Rj^2 and a less negative impact model.

C statistic: (small and <=k)

The C statistic for the improved model is 7.2093, and k = 7. A small C statistic that is close to k indicates a good model.

Press statistic: (small)

Before improvement, using all eight x variables.

Sum of Residuals	0
Sum of Squared Residuals	4432.72919
Predicted Residual SS (PRESS)	4533.24962

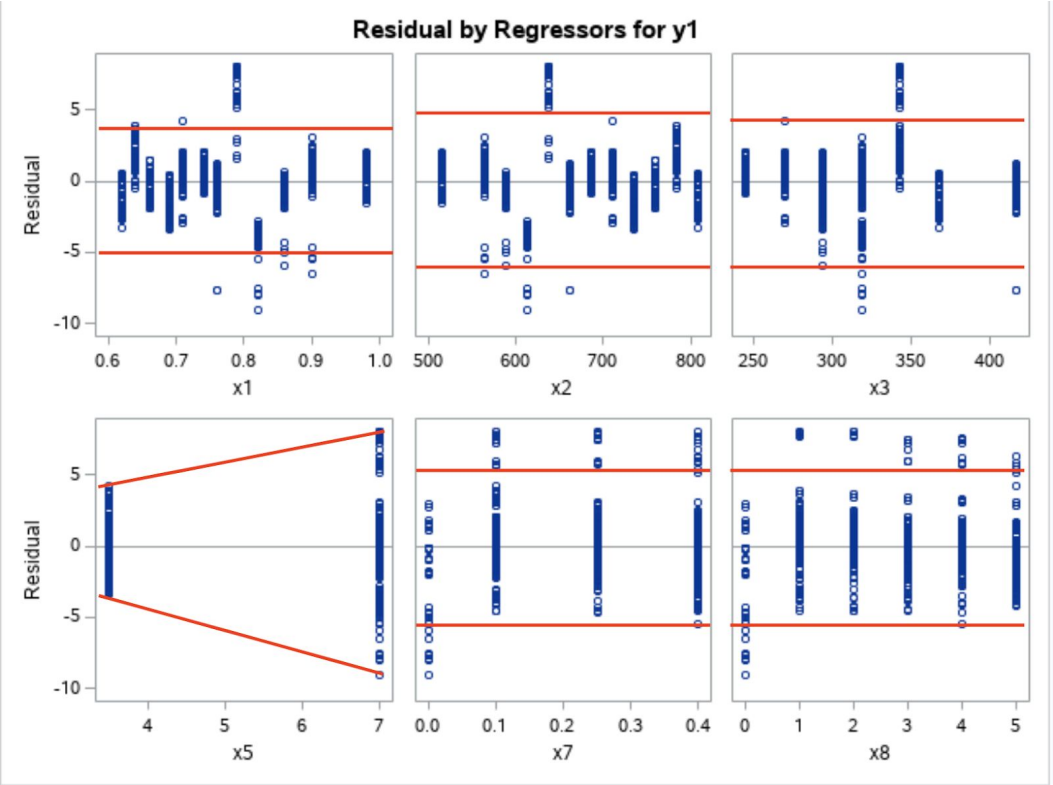
After improvement, using six x variables and one interaction term.

Sum of Residuals	4.26788E-10
Sum of Squared Residuals	4189.36499
Predicted Residual SS (PRESS)	4283.26356

The press statistic improved by 249.98606.

Assumptions behind regression:

1. **A1: Constant Variance**



We can see that x5 clearly shows a fan out pattern, so constant variance assumption for x5 is violated, as criterion increase , residual increase as well.

As for the other plots, they roughly show constant variance. Which is good

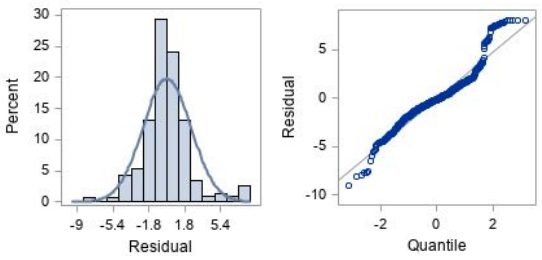
2. **A2: Independence**

Durbin-Watson D	0.907
Pr < DW	<.0001
Pr > DW	1.0000

H0 is the error terms are not autocorrelated.

Durbin-watson test rejects non hypothesis, indicates there is significant positive autocorrelation.

3. A3: Normal Population



Histogram plot (in bell shape) and residual quantile (Q-Q plot) are unfortunately not perfectly fit but roughly support the normal assumption.

Conclusion (Interpretation):

Our final model with parameter estimate is: -147.95753 - 43.28520(x1) - 0.05617(x2) + 0.06718(x3) + 4.34667(x5) + 19.44072(x7) + 0.19581(x8) + 0.38034(x1x2)

It indicates that larger x1 values will make heating load significantly smaller, which makes sense because x1 is the compactness of the building. The model also suggests that a larger x5 value will make heating load significantly larger, and x5 is the overall height of the building, so it means that a taller building will consume more

energy for heating. Also, the model indicates that larger x7 will make heating load significantly larger. x7 is the glazing area. Glazing is a technique that is used to alter the appearance of traditional paint, so according to the data and our model, it will significantly damage the heating efficiency.

Appendix:

SAS code

```
title 'Orginal data set';
data orginal_data; /* https://archive.ics.uci.edu/ml/datasets/Energy+efficiency */
    infile '/folders/myfolders/data/a1.csv' dlm="," firstobs=2;
    input x1 x2 x3 x4 x5 x6 x7 x8 y1;
run;

/* calculate correlation matrix */
title 'Correlation matrix';
proc corr data=orginal_data;
RUN;


/* Influential data check */
title 'Test for outliers and influencial points';
proc reg data=orginal_data;
    model y1 = x1 x2 x3 x5 x7 x8 /vif influence ;
run;

/*  Rstudent t_0.025 (768-9-1)=1.9631
    F_0.5(9,759)=0.93
    2*k\n=0.0234
    Cov Ratio: 1+3k/n=1.0352 1-3k/n=0.9648
    DFFITS: >2

    There are only outleirs under Cov Ratio and Rstudent found
    Keep in mind, observations that have Cov Ratio < 0.9648 will enhance the model
    We removed 51 observations.
*/

/*  Model Build: phase 1
*/
title 'Comparison between all possible models';
proc rsquare data=orginal_data cp mse sse adjrsq;
    model y1 = x1 x2 x3 x4 x5 x6 x7 x8;
run;

title 'Stepwise Procedure';
proc stepwise data=orginal_data;
    model y1 = x1 x2 x3 x4 x5 x6 x7 x8 / stepwise ;
run;

title 'Forward Procedure';
proc stepwise data=orginal_data;
    model y1 = x1 x2 x3 x4 x5 x6 x7 x8 / forward ;
run;
```

```
title 'Backward Procedure';
proc stepwise data=orginal_data;
    model y1 = x1 x2 x3 x4 x5 x6 x7 x8 / backward ;
run;

title 'Maxr Procedure';
proc stepwise data=orginal_data;
    model y1 = x1 x2 x3 x4 x5 x6 x7 x8 / maxr;
run;
/*
    y = x1 x2 x3 x5 x7 x8 is the best model so far

*/

/*
    Model Build: phase 2 (improvement)
*/

title 'Improved data set';
data improved_data; /* data set after outliers removed */
    infile '/folders/myfolders/data/a2.csv' dlm="," firstobs=2;
    input x1 x2 x3 x4 x5 x6 x7 x8 y1 x1x2;
    x1x2=x1*x2;
run;

/*
    Final model is y = x1 x2 x3 x5 x7 x8 x1x2
*/
title 'Improved model';
proc reg data=improved_data;
    model y1 = x1 x2 x3 x5 x7 x8 x1x2;
run;

title 'Orginal model';
proc reg data=orginal_data;
    model y1 = x1 x2 x3 x5 x7 x8;
run;

/*
    dw test
*/
title 'Dw test';
proc reg data = improved_data;
model y1 = x1 x2 x3 x5 x7 x8 x1x2 /dw dwprob;
run;
```

Partial output:

Some of outliers and influential points

The REG Procedure
Model: MODEL1
Dependent Variable: y1

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS	DFBETAS							
												Intercept	x1	x2	x3	x5	x6	x7	x8
1	15.55	22.6472	0.4265	-7.0972	2.903	-2.445	0.016	-2.4527	0.0211	0.9692	-0.3604	0.0666	-0.1080	-0.0444	-0.0530	0.0589	0.1200	0.1258	0.1318
2	15.55	22.6239	0.4050	-7.0739	2.906	-2.434	0.014	-2.4420	0.0190	0.9677	-0.3403	0.0676	-0.1074	-0.0441	-0.0527	0.0586	0.0398	0.1251	0.1311
3	15.55	22.6005	0.4050	-7.0505	2.906	-2.426	0.014	-2.4338	0.0190	0.9681	-0.3391	0.0687	-0.1071	-0.0440	-0.0526	0.0584	-0.0397	0.1247	0.1307
4	15.55	22.5772	0.4265	-7.0272	2.903	-2.421	0.016	-2.4283	0.0211	0.9704	-0.3568	0.0700	-0.1069	-0.0439	-0.0525	0.0583	-0.1188	0.1245	0.1305
5	20.84	25.0418	0.3376	-4.2018	2.915	-1.442	0.003	-1.4426	0.0132	1.0020	-0.1671	-0.0062	-0.0054	0.0103	-0.0292	0.0254	0.0703	0.0737	0.0772
6	21.46	25.0185	0.3099	-3.5585	2.918	-1.220	0.002	-1.2199	0.0112	1.0061	-0.1296	-0.0045	-0.0046	0.0087	-0.0247	0.0215	0.0198	0.0623	0.0652
7	20.71	24.9952	0.3099	-4.2852	2.918	-1.469	0.003	-1.4697	0.0112	0.9990	-0.1561	-0.0046	-0.0055	0.0105	-0.0297	0.0259	-0.0239	0.0750	0.0786
8	19.68	24.9718	0.3376	-5.2918	2.915	-1.815	0.006	-1.8182	0.0132	0.9891	-0.2106	-0.0047	-0.0068	0.0130	-0.0368	0.0320	-0.0886	0.0929	0.0973
9	19.50	24.0042	0.3707	-4.5042	2.911	-1.547	0.005	-1.5488	0.0160	1.0014	-0.1973	0.0561	-0.0489	-0.0682	0.0996	-0.0981	0.0756	0.0792	0.0830
10	19.95	23.9809	0.3457	-4.0309	2.914	-1.383	0.003	-1.3842	0.0139	1.0043	-0.1642	0.0509	-0.0436	-0.0609	0.0889	-0.0876	0.0225	0.0707	0.0741
11	19.34	23.9576	0.3457	-4.6176	2.914	-1.585	0.004	-1.5863	0.0139	0.9980	-0.1882	0.0592	-0.0500	-0.0698	0.1019	-0.1004	-0.0258	0.0811	0.0849
12	18.31	23.9343	0.3707	-5.6243	2.911	-1.932	0.008	-1.9357	0.0160	0.9873	-0.2465	0.0734	-0.0611	-0.0852	0.1245	-0.1226	-0.0945	0.0990	0.1038
13	17.05	25.9465	0.3506	-8.8965	2.913	-3.054	0.017	-3.0707	0.0143	0.9288	-0.3696	0.0056	0.0137	-0.0312	0.1135	-0.1245	0.1497	0.1569	0.1644
14	17.41	25.9232	0.3240	-8.5132	2.916	-2.919	0.013	-2.9337	0.0122	0.9348	-0.3260	0.0070	0.0131	-0.0298	0.1083	-0.1188	0.0476	0.1498	0.1569
15	16.95	25.8999	0.3240	-8.9499	2.916	-3.069	0.015	-3.0860	0.0122	0.9259	-0.3429	0.0091	0.0137	-0.0313	0.1139	-0.1249	-0.0501	0.1576	0.1651
16	15.98	25.8765	0.3506	-9.8965	2.913	-3.397	0.021	-3.4209	0.0143	0.9071	-0.4117	0.0121	0.0153	-0.0347	0.1264	-0.1387	-0.1668	0.1748	0.1832
17	28.52	27.2411	0.3529	1.2789	2.913	0.439	0.000	0.4388	0.0145	1.0233	0.0532	-0.0010	-0.0023	0.0044	-0.0131	0.0177	-0.0214	-0.0224	-0.0235
18	29.90	27.2177	0.3265	2.6823	2.916	0.920	0.001	0.9197	0.0124	1.0142	0.1030	-0.0026	-0.0047	0.0092	-0.0274	0.0371	-0.0149	-0.0470	-0.0492
19	29.63	27.1944	0.3265	2.4356	2.916	0.835	0.001	0.8351	0.0124	1.0158	0.0935	-0.0028	-0.0043	0.0084	-0.0248	0.0337	0.0136	-0.0426	-0.0447
20	28.75	27.1711	0.3529	1.5789	2.913	0.542	0.001	0.5418	0.0145	1.0223	0.0656	-0.0021	-0.0028	0.0054	-0.0161	0.0219	0.0264	-0.0277	-0.0290
21	24.77	31.5154	0.4439	-6.7454	2.901	-2.326	0.016	-2.3324	0.0229	0.9768	-0.3569	-0.1689	0.1630	0.1759	-0.2089	0.1489	0.1142	0.1197	0.1255
22	23.93	31.4921	0.4232	-7.5621	2.904	-2.604	0.018	-2.6143	0.0208	0.9606	-0.3810	-0.1877	0.1825	0.1969	-0.2340	0.1667	0.0426	0.1341	0.1405
23	24.77	31.4688	0.4232	-6.6988	2.904	-2.307	0.014	-2.3136	0.0208	0.9756	-0.3372	-0.1648	0.1615	0.1743	-0.2070	0.1476	-0.0377	0.1186	0.1243
24	23.93	31.4455	0.4439	-7.5155	2.901	-2.591	0.020	-2.6009	0.0229	0.9634	-0.3980	-0.1839	0.1818	0.1961	-0.2330	0.1661	-0.1274	0.1335	0.1399
25	6.07	5.6480	0.3730	0.4220	2.911	0.145	0.000	0.1449	0.0162	1.0270	0.0186	0.0064	-0.0057	-0.0057	0.0008	-0.0047	-0.0071	-0.0074	-0.0078
26	6.05	5.6247	0.3481	0.4253	2.914	0.146	0.000	0.1459	0.0141	1.0248	0.0174	0.0064	-0.0058	-0.0057	0.0008	-0.0047	-0.0024	-0.0075	-0.0078
27	6.01	5.6014	0.3481	0.4086	2.914	0.140	0.000	0.1402	0.0141	1.0248	0.0167	0.0061	-0.0055	-0.0055	0.0008	-0.0045	0.0023	-0.0072	-0.0075

Output Statistics																			
												DFBETAS							
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS	Intercept	x1	x2	x3	x5	x6	x7	x8
424	12.84	14.4084	0.2101	-1.5684	2.927	-0.536	0.000	-0.5356	0.0051	1.0127	-0.0385	-0.0029	0.0033	0.0030	-0.0037	0.0055	-0.0260	-0.0018	-0.0019
425	16.83	15.1252	0.2417	1.7048	2.924	0.583	0.000	0.5827	0.0068	1.0139	0.0482	-0.0163	0.0165	0.0162	-0.0035	0.0070	-0.0283	0.0020	0.0021
426	16.93	15.1019	0.2012	1.8281	2.927	0.624	0.000	0.6242	0.0047	1.0112	0.0429	-0.0178	0.0176	0.0174	-0.0038	0.0075	-0.0101	0.0021	0.0022
427	16.66	15.0785	0.2012	1.5815	2.927	0.540	0.000	0.5400	0.0047	1.0122	0.0371	-0.0157	0.0153	0.0150	-0.0032	0.0065	0.0087	0.0018	0.0019
428	16.86	15.0552	0.2417	1.8048	2.924	0.617	0.000	0.6169	0.0068	1.0134	0.0510	-0.0183	0.0175	0.0172	-0.0037	0.0074	0.0300	0.0021	0.0022
429	13.91	15.7720	0.3175	-1.8620	2.917	-0.638	0.001	-0.6381	0.0117	1.0182	-0.0694	0.0403	-0.0402	-0.0392	0.0121	-0.0219	0.0311	-0.0022	-0.0023
430	14.34	15.7486	0.2879	-1.4086	2.920	-0.482	0.000	-0.4821	0.0096	1.0179	-0.0475	0.0307	-0.0303	-0.0296	0.0092	-0.0166	0.0078	-0.0016	-0.0017
431	13.95	15.7253	0.2879	-1.7753	2.920	-0.608	0.000	-0.6077	0.0096	1.0164	-0.0599	0.0390	-0.0382	-0.0373	0.0115	-0.0209	-0.0099	-0.0021	-0.0022
432	13.99	15.7020	0.3175	-1.7120	2.917	-0.587	0.001	-0.5866	0.0117	1.0189	-0.0638	0.0381	-0.0369	-0.0361	0.0111	-0.0202	-0.0286	-0.0020	-0.0021
433	28.70	28.4455	0.3598	0.2545	2.912	0.087	0.000	0.0873	0.0150	1.0259	0.0108	-0.0025	0.0038	0.0016	0.0019	-0.0021	-0.0043	-0.0001	0.0024
434	28.55	28.4221	0.3339	0.1279	2.915	0.044	0.000	0.0438	0.0130	1.0238	0.0050	-0.0013	0.0019	0.0008	0.0009	-0.0010	-0.0007	-0.0001	0.0012
435	28.15	28.3988	0.3339	-0.2488	2.915	-0.085	0.000	-0.0853	0.0130	1.0238	-0.0098	0.0025	-0.0037	-0.0015	-0.0018	0.0020	-0.0014	0.0001	-0.0023
436	28.62	28.3755	0.3598	0.2445	2.912	0.084	0.000	0.0839	0.0150	1.0259	0.0104	-0.0025	0.0037	0.0015	0.0018	-0.0020	0.0041	-0.0001	0.0023
437	32.67	30.8401	0.2480	1.8299	2.924	0.626	0.000	0.6256	0.0071	1.0137	0.0531	0.0018	0.0023	-0.0045	0.0126	-0.0110	-0.0304	-0.0011	0.0172
438	31.69	30.8168	0.2087	0.8732	2.927	0.298	0.000	0.2982	0.0051	1.0148	0.0213	0.0007	0.0011	-0.0021	0.0060	-0.0052	-0.0048	-0.0005	0.0082
439	32.07	30.7934	0.2087	1.2766	2.927	0.436	0.000	0.4359	0.0051	1.0137	0.0311	0.0008	0.0016	-0.0031	0.0088	-0.0077	0.0071	-0.0007	0.0120
440	33.28	30.7701	0.2480	2.5099	2.924	0.858	0.001	0.8583	0.0071	1.0100	0.0728	0.0011	0.0032	-0.0061	0.0173	-0.0151	0.0417	-0.0015	0.0236
441	29.47	29.8025	0.2915	-0.3325	2.920	-0.114	0.000	-0.1138	0.0099	1.0205	-0.0114	0.0043	-0.0036	-0.0050	0.0073	-0.0072	0.0055	0.0002	-0.0031
442	28.42	29.7792	0.2589	-1.3592	2.923	-0.465	0.000	-0.4648	0.0078	1.0162	-0.0412	0.0176	-0.0146	-0.0204	0.0298	-0.0293	0.0075	0.0008	-0.0128
443	29.08	29.7558	0.2589	-0.6758	2.923	-0.231	0.000	-0.2311	0.0078	1.0179	-0.0205	0.0089	-0.0073	-0.0101	0.0148	-0.0146	-0.0037	0.0004	-0.0063
444	29.88	29.7325	0.2915	0.1475	2.920	0.051	0.000	0.0505	0.0099	1.0206	0.0050	-0.0020	0.0016	0.0022	-0.0032	0.0032	0.0025	-0.0001	0.0014
445	25.66	31.7448	0.2654	-6.0848	2.922	-2.082	0.004	-2.0868	0.0082	0.9733	-0.1896	0.0066	0.0093	-0.0211	0.0769	-0.0843	0.1014	0.0035	-0.0574
446	24.96	31.7214	0.2292	-6.7614	2.925	-2.311	0.004	-2.3180	0.0061	0.9610	-0.1816	0.0086	0.0103	-0.0234	0.0853	-0.0936	0.0375	0.0039	-0.0636
447	25.43	31.6981	0.2292	-6.2681	2.925	-2.143	0.004	-2.1478	0.0061	0.9687	-0.1683	0.0092	0.0095	-0.0217	0.0790	-0.0867	-0.0348	0.0036	-0.0590
448	26.00	31.6748	0.2654	-5.6748	2.922	-1.942	0.004	-1.9455	0.0082	0.9792	-0.1767	0.0094	0.0086	-0.0197	0.0717	-0.0786	-0.0946	0.0033	-0.0535
449	40.00	33.0393	0.2684	6.9607	2.922	2.382	0.006	2.3895	0.0084	0.9598	0.2195	-0.0086	-0.0122	0.0239	-0.0710	0.0961	-0.1162	-0.0041	0.0657
450	38.84	33.0160	0.2326	5.8240	2.925	1.991	0.003	1.9950	0.0063	0.9753	0.1587	-0.0083	-0.0102	0.0200	-0.0592	0.0802	-0.0323	-0.0034	0.0548
451	38.33	32.9927	0.2326	5.3373	2.925	1.825	0.003	1.8275	0.0063	0.9819	0.1453	-0.0086	-0.0094	0.0183	-0.0542	0.0734	0.0296	-0.0031	0.0502
452	40.12	32.9693	0.2684	7.1507	2.922	2.447	0.006	2.4553	0.0084	0.9566	0.2255	-0.0130	-0.0126	0.0246	-0.0729	0.0987	0.1194	-0.0042	0.0675
453	36.95	37.3137	0.3802	-0.3637	2.910	-0.125	0.000	-0.1249	0.0168	1.0277	-0.0163	-0.0089	0.0087	0.0094	-0.0112	0.0080	0.0061	0.0002	-0.0034
454	36.45	37.2904	0.3558	-0.8404	2.913	-0.289	0.000	-0.2883	0.0147	1.0248	-0.0352	-0.0202	0.0201	0.0217	-0.0257	0.0183	0.0047	0.0005	-0.0080
455	36.81	37.2670	0.3558	-0.4570	2.913	-0.157	0.000	-0.1568	0.0147	1.0254	-0.0192	-0.0109	0.0109	0.0118	-0.0140	0.0100	-0.0025	0.0003	-0.0043

Best models among all possible models generated by SAS.

Number in Model	R-Square	Adjusted R-Square	C(p)	MSE	SSE	Variables in Model
6	0.9162	0.9155	6.0607	8.59959	6544.28862	x1 x3 x4 x5 x7 x8
6	0.9162	0.9155	6.0607	8.59959	6544.28862	x1 x2 x4 x5 x7 x8
6	0.9162	0.9155	6.0607	8.59959	6544.28862	x1 x2 x3 x5 x7 x8

Model before improvement.

Orginal model

The REG Procedure
Model: MODEL1
Dependent Variable: y1

Number of Observations Read	768
Number of Observations Used	768

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	71546	11924	1386.61	<.0001
Error	761	6544.28862	8.59959		
Corrected Total	767	78090			

Root MSE	2.93251	R-Square	0.9162
Dependent Mean	22.30720	Adj R-Sq	0.9155
Coeff Var	13.14601		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	83.93287	19.01897	4.41	<.0001
x1	1	-64.77399	10.28309	-6.30	<.0001
x2	1	-0.08729	0.01706	-5.12	<.0001
x3	1	0.06081	0.00664	9.15	<.0001
x5	1	4.16994	0.33778	12.35	<.0001
x7	1	19.93268	0.81348	24.50	<.0001
x8	1	0.20377	0.06987	2.92	0.0036

Improved Model

Improved model

The REG Procedure
Model: MODEL1
Dependent Variable: y1

Number of Observations Read	717
Number of Observations Used	717

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	66109	9444.15930	1598.31	<.0001
Error	709	4189.36499	5.90884		
Corrected Total	716	70298			

Root MSE	2.43081	R-Square	0.9404
Dependent Mean	21.66234	Adj R-Sq	0.9398
Coeff Var	11.22136		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-147.95753	39.67020	-3.73	0.0002
x1	1	-43.28520	9.09487	-4.76	<.0001
x2	1	-0.05617	0.01535	-3.66	0.0003
x3	1	0.06718	0.00573	11.72	<.0001
x5	1	4.34667	0.30741	14.14	<.0001
x7	1	19.44072	0.69608	27.93	<.0001
x8	1	0.19581	0.06045	3.24	0.0013
x1x2	1	0.38034	0.05912	6.43	<.0001

DW test

Dw test

The REG Procedure
Model: MODEL1
Dependent Variable: y1

Durbin-Watson D	0.907
Pr < DW	<.0001
Pr > DW	1.0000
Number of Observations	717
1st Order Autocorrelation	0.542

VIF

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-147.95753	39.67020	-3.73	0.0002	0
x1	1	-43.28520	9.09487	-4.76	<.0001	116.28680
x2	1	-0.05617	0.01535	-3.66	0.0003	228.93804
x3	1	0.06718	0.00573	11.72	<.0001	7.91182
x5	1	4.34667	0.30741	14.14	<.0001	34.94115
x7	1	19.44072	0.69608	27.93	<.0001	1.04545
x8	1	0.19581	0.06045	3.24	0.0013	1.04723
x1x2	1	0.38034	0.05912	6.43	<.0001	2.26560