MSDS 6371 Project

Amy Adyanthaya & Nicholas Mueller

Introduction

Our group has been hired on to examine 79 explanatory variables describing every aspect of residential homes in Ames, Iowa. Our client would first like us to examine 3 neighborhoods of interest (Northwest Ames, Brookside, and Edwards) and compare ground living area and its impacts on each house's sales price in each of its respected neighborhood. Once providing our client with the information on sales price in the 3 desired neighborhoods, they would like us to build a predictive model for sales price for homes in all of Ames, Iowa. In order to successfully provided the most accurate data/model, our group will extensively examine all 79 explanatory variables and use a cross validation technique to provide our clients with important information.

Data Description

Our data was retrieved from the Ames Housing data set on Kaggle for a competition. The dataset contains 2930 observations divided into a testing and training set. The data set(s) consist of 79 explanatory variables that describe several aspects of a house which could have an influence on its sales price. For this dataset we cannot find more observations, however for general housing market data we can use several resources throughout different real-estate companies. Each variable used in the analysis was examined thoroughly with statistical evidence of collinearity. Our group used VIF scores, plots, excel spread sheets, p-values, and several other factors to decided which explanatory variable to keep or removed. For more information please refer to our appendix.

<u>Analysis Question 1</u>

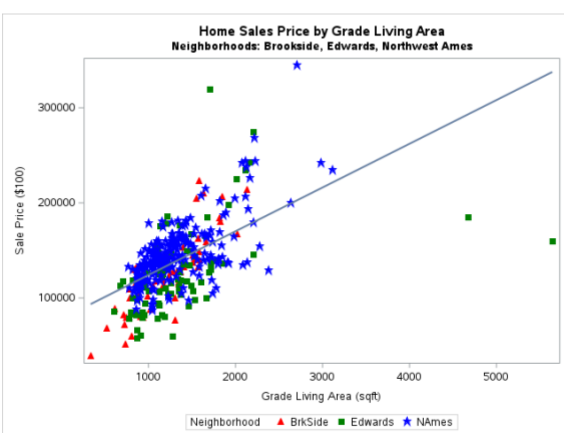
State the Problem:

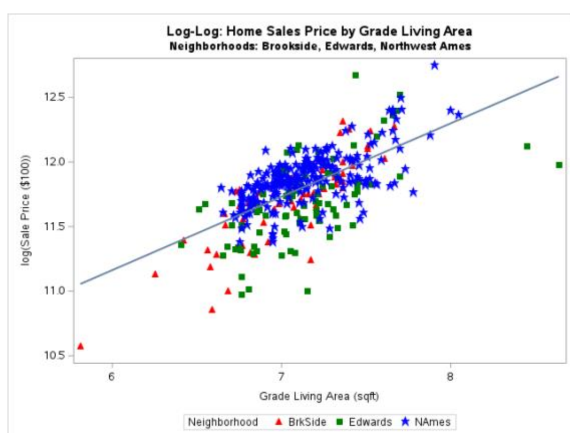
Century 21 Ames would like to obtain an estimate of how the sales price of a house was related to its square footage of the living area of the house; depending on its respective neighborhood of the Ames Iowa area.

Plot and Review the Data:

The scatter plot of the original data shows a limited linear relationship between the grade living area and sale price, with much of the data clustered between 1000 and 2000 sq ft. Both the Grade Living Area and Sale Price contain a couple influential datapoints possibly impacting the data (Appendix Figure 1). These datapoints were reviewed and determined that they were appropriate for the analysis. The neighborhood effect was associated with a 2^{B1} multiplicative increase in the median of sale price. A log-log transformation was used to control for the influential datapoints (Appendix Figure 4).

Fit





$\hat{\mu}$\{log(SalePrice) | Neighborhood , log(GrLivArea)\} = β_0 + β_1*Neighborhood = Brookside + β_2*Neighborhood = Edwards + β_3*log(GrLivArea)

Pred {log(SalePrice)|Neighborhood, GrLivArea} = 7.9021 + 0.5558*(Brookside) - 0.1328*(Edwards) - 0.1532* log(GrLivArea)

Pred {log(SalePrice)|Neighborhood=Brookside, GrLivArea} = 8.4579 - 0.1532* log(GrLivArea)

Pred {log(SalePrice)|Neighborhood=Edwards, GrLivArea} = 7.7693 - 0.1532* log(GrLivArea)

	Parameter Estimates						Cross Validation Estimates				
	Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	1	2	3	4	5
	Intercept	1	7.902150	0.231340	34.16	<.0001	7.836	7.709	7.977	8.113	7.878
*	log(Grade Living Are	1	0.555788	0.032369	17.17	<.0001	0.565	0.583	0.546	0.526	0.560
*	Neighborhood BrkSide	1	-0.132789	0.029061	-4.57	<.0001	-0.104	-0.125	-0.153	-0.142	-0.139
*	Neighborhood Edwards	1	-0.153226	0.023571	-6.50	<.0001	-0.159	-0.141	-0.174	-0.148	-0.142
*	Neighborhood NAmes	0	0	.	.	.	0.000	0.000	0.000	0.000	0.000

* Forced into the model by the INCLUDE= option

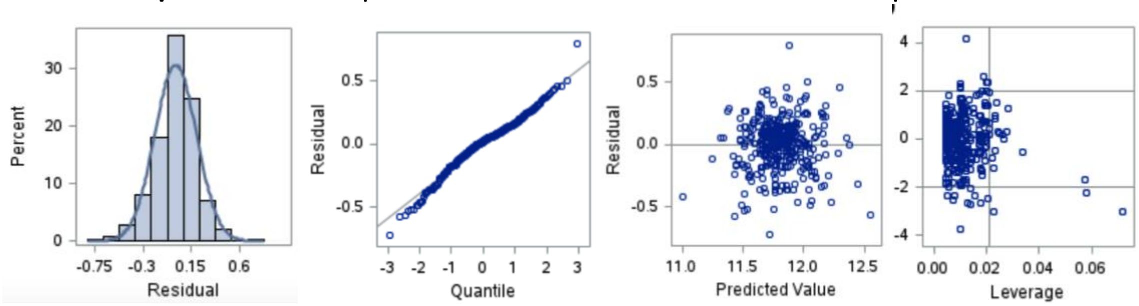
Cross Validation Details

Index	Observations: Fitted	Observations: Left Out	CV PRESS
1	312	71	2.5897
2	315	68	2.3306
3	305	78	3.0757
4	292	91	3.8508
5	308	75	3.1010
Total			14.9478

Root MSE	0.19612
Dependent Mean	11.79887
R-Square	0.4897
Adj R-Sq	0.4857
AIC	-858.86316
AICC	-858.70401
SBC	-1228.07102
CV PRESS	14.94781

Assumptions (log-log transformed data):

- **Normality:** The q-q plot and the histogram show slight right skewness however not strong enough evidence against normality. There were a couple of influential observations with both high leverage and residual as well as high Cook's D.
- **Linearity:** It was tough to check linearity in multiple dimensions however the scatterplot does show linearity with influential datapoints.
- **Constant Variance**: There was no visual evidence of differing standard deviation throughout the residual plot.
- **Independence**: Independence cannot be assumed so we will proceed with caution.



Interpretation:

There was overwhelming evidence to suggest the impact of a neighborhood has a 2^{B1} multiplicative increase impact on home sales price. Holding grade living area constant, it was estimated that the Brookside neighborhood sales prices was \$87.57 ($e^{-0.13279}$) more per 100 sq ft than the Northwest Ames neighborhood. A 95% confidence interval for this estimate was ($e^{-0.18975}$, $e^{-0.07583}$) = (\$82.72, \$ 92.70). Holding grade living area constant, it was estimated that the Edwards neighborhood sales prices was \$85.79 ($e^{-0.15323}$) more per 100 sq ft than the Northwest Ames neighborhood. A 95% confidence interval for this estimate was ($e^{-0.19943}$, $e^{-0.10703}$) = (\$81.92, \$ 89.85). Holding the neighborhood constant, it was estimated that the grand living area was

\$174.33 ($e^{0.55579}$) per 100 sq ft with a 95% confidence interval for the estimate of ($e^{0.49234}$, $e^{0.61923}$) = (\$163.61, \$ 185.75).

Conclusion:
The real estate data for the Ames Iowa overwhelming showed that the sales price of a house was related to its square footage of the living area of the house and the neighborhood of the home (p-value < 0.0001). The Brookside neighborhood average sale price was the highest, followed by the Edwards neighborhood and then the Northwest Adams neighborhood.

R-Shiny App
Scatter plot of the sale price of neighborhood sale price vs the home square footage can be seen here:
https://nickmueller2.shinyapps.io/Neighborhood3/?_ga=2.68375953.127761547.1680987412-408874730.1680987412

MSDS 6371 Project
Amy Adyanthaya & Nicholas Mueller

<u>Analysis Question 2</u>

State the Problem:

We would like to build the most predictive model for sale prices of homes in all of Ames Iowa.

Exploratory Data Analysis:

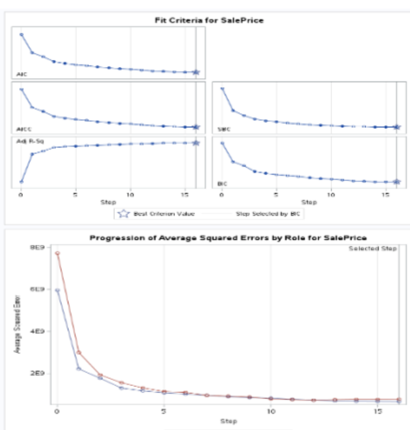
When exploring our data, we first started by examining the excel spreadsheet noticing obvious variables that could have a negative impact on our model. For example, we removed the variables SaleType, BsmtHalfBath, Functional, LowQualFinSF, BsmtFinSF2, ScreenPorch, EnclosedPorch, PoolArea, MiscVal, Utilities, and street since 85% to 99% of each column consisted of 1 main value. After examining the excel spreadsheet we inputted the data into SAS so that we could explore other competing factors against linearity. Factors considered included the p-value, VIF, sets with heavy outliers, unique identifiers (ex: ID), evidence against normality, and more. When addressing NA values, we decided that these were not "unknown" values, but values that were 0 (ex: The alley column NA= House didn't have an alley way). Therefore, all NA/missing values we converted to 0 and other string values were converted to 1, 2, 3, and so on. As a team we also used the SAS Proc Corr to visually show numeric identifying variables which needed to be removed.

Model Selection

Type of Selection: Stepwise

For our stepwise model we were able to get a Kaggle score of 0.1637 which was also our best competing model in the Kaggle competition. Therefore, we used it for our custom model as well. The adjusted R square value was a .88 pr 88% with a cv press of 1.546512E12 and BIC score of 24309(figure 7 screenshots and code provided in appendix).

STEPWISE.csv
Complete · 2d ago
0.1637



Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	78	6.28494E12	80576156554	111.93
Error	1109	7.983513E11	719883908	
Corrected Total	1187	7.083291E12		

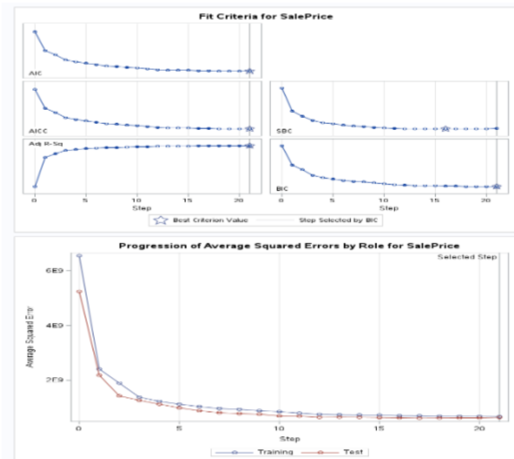
Root MSE	26831
Dependent Mean	181626
R-Square	0.8873
Adj R-Sq	0.8794
AIC	25495
AICC	25507
BIC	24309
C(p)	143.59423
SBC	24706
ASE (Train)	672012840
ASE (Test)	779410399

As we can see from the stepwise plot above our testing set seems to show a very close and similar path to the training set. There does not seem to be a large amount of variation, therefore we proceeded to submit out model to the completion for a Kaggle score.

Type of Selection: Forward

For our forward predicting model, we were able to get a Kaggle score of 0.16756. This Kaggle score was very close to our stepwise model, however, was not quick enough to pass it. The adjusted R square value was a .89 or 89% with a cv press of 1.607329E12 and BIC score of 23842 (figure 5 and 6 screenshots and code provided in appendix).

FORWARD3.csv
Complete · 31m ago
0.16756



Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	86	6.818803E12	79288404486	109.14
Error	1074	7.802169E11	726458977	
Corrected Total	1160	7.59902E12		

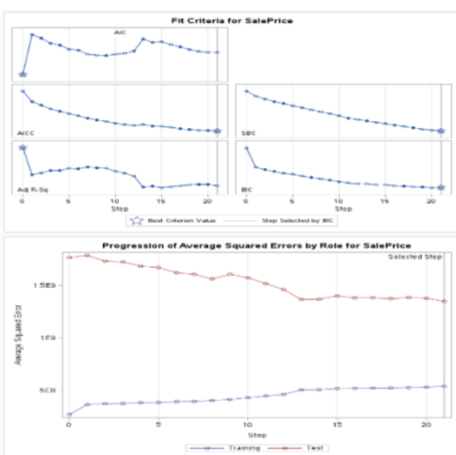
Root MSE	26953
Dependent Mean	181961
R-Square	0.8973
Adj R-Sq	0.8891
AIC	24935
AICC	24950
BIC	23784
C(p)	110.22173
SBC	24212
ASE (Train)	672021482
ASE (Test)	642762792

As we can see from the forward model plot above our testing set seems to show a very close and similar path to the training set. There does not seem to be a large amount of variation, therefore we will proceed with submission to the competition for a Kaggle score.

Type of Selection: Backward

For our backward model, it significantly underperformed compared to the other models. For unknown reasons the best Kaggle score we achieved with the backward model was a 0.38305. The adjusted R square values was .89 or 89% with a cv press of 2.650568E11 and a BIC score of 6197.8082123842 (figure 8, 9, and 10 screenshots and code provided in appendix).

BACKWARD2.csv
Complete · 2h ago
0.38305



Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	53	1.714696E12	32352746601	49.39
Error	249	1.631218E11	655107543	
Corrected Total	302	1.877817E12		

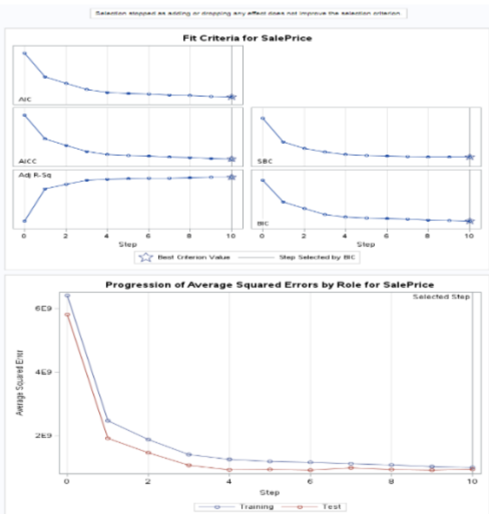
Root MSE	25595
Dependent Mean	177126
R-Square	0.9131
Adj R-Sq	0.8946
AIC	6504.52111
AICC	6529.46038
BIC	6197.80821
C(p)	118.78050
SBC	6400.06268
ASE (Train)	538355704
ASE (Test)	1348647959

Unlike the forward and stepwise models, the backward model starts with every variable. As we can see from the plot above our model seems to trend together, however never intersects, showing us that the model is not quite perfect. That being said, with the time allotted this was our best seed attempt. Therefore, we will proceed with submission to the competition for a Kaggle score.

Type of Selection: Custom

For our custom model we used domain knowledge in the housing market to determine which variable would have the highest influence on sales price. Therefore, with the collected effort of cook's D and other identifiers we were able to run a model that we personally thought would receive the best Kaggle score. However, our custom model only received a Kaggle score of 0.17581 which was worse than our original stepwise and forward model. The custom model also received an adjusted R square of 0.84 or 84% and a BIC of 23833(figure 11, 12 and 13 screenshots and code provided in appendix).

CUSTOM.csv
Complete · now
0.17581



Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	48	6.181642E12	1.287842E11	122.08
Error	1096	1.156183E12	1054911386	
Corrected Total	1144	7.337826E12		

Root MSE	32479
Dependent Mean	181320
R-Square	0.8424
Adj R-Sq	0.8355
AIC	24984
AICC	24989
BIC	23833
C(p)	174.03534
SBC	24084
ASE (Train)	1009766707
ASE (Test)	955089793

As we can see from the custom(stepwise) model plot above our testing set seems to show a very close and similar path to the training set. There does not seem to be a large amount of variation, therefore we will proceed with submission to the competition for a Kaggle score.

Checking Assumptions:

Once removing the values with visual evidence against normality we were able to obtain the residual plot, high leveraging point, and evidence against linearity. We were able to get plots that showed normality and collinearity.

- **Residual Plots:** We were able to see visual evidence of equal standard deviations with no evidence of curvature in the residual plot.
- **Influential point analysis (Cook's D and Leverage)**: We were able to remove high Cook's D and leverage variable. With our final model Cook's D we had the highest Cook's D being around .2 to .3, however our team decided that they did not have enough leverage to be removed. We wanted to have as many variables as possible, so we proceeded with caution and took note of those higher Cook's D 23842 (screenshots and code provided in appendix).
- **Address each Assumption:** Our group also reviewed a histogram that visually shows a normal bell shape curve. We also used a QQ plot to show linearity and possible outlier, for our final model we

saw no evidence of significant outliers and majority variables seemed to follow a linear trend 23842 (screenshots and code provided in appendix).

Comparing Competing Models:

Predictive Models	Adjusted R2	CV PRESS	Kaggle Score	BIC
Forward	.89	1.607329E12	0.16756	23842
Backward	.89	2.650568E11	0.3836	6197.80821
Stepwise	.88	1.546512E12	0.1637	24309
CUSTOM	.84	1.469603E12	0.17581	23833

Conclusion:

In conclusion, we observed 4 models which included a forward, backward, stepwise, and custom(stepwise) model. Our overall goal was to check the assumptions of the data, identify high leveraged values, and other factors that have a negative impact on our model. We were able to remove these high influential values. After removing the high influential values/variables we were able to run 4 successful models. In the end the best preforming model was the stepwise model with a Kaggle score of 0.1637 ranking our group number 2700.

Appendix
MSDS 6371 Project
Amy Adyanthaya & Nicholas Mueller

Analysis Question 1

Figure 1: Proc SGSCATTER
Figure 2: Proc GLMSELECT Stepwise Model

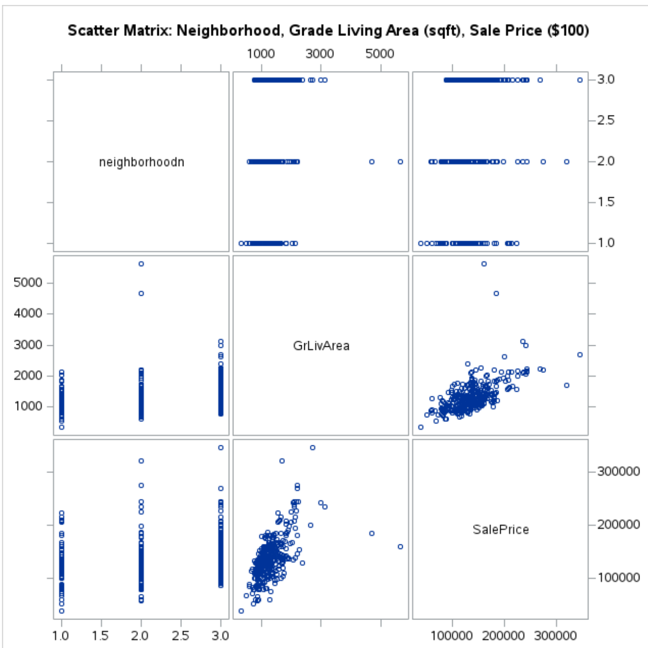


Figure 1

Stepwise Model for Reporting

The GLMSELECT Procedure
Selected Model

The selected model, based on Cross Validation, is the model at Step 0.

Effects:	Intercept log(Grade Living Are Neighborhood

Note: The p-values for parameters and effects are not adjusted for the fact that the terms in the model have been selected and so are generally liberal.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	13.98907	4.66302	121.24	<.0001
Error	379	14.57726	0.03846		
Corrected Total	382	28.56633			

Root MSE	0.19612
Dependent Mean	11.79887
R-Square	0.4897
Adj R-Sq	0.4857
AIC	-858.86316
AICC	-858.70401
SBC	-1228.07102
CV PRESS	14.94781

Cross Validation Details			
	Observations		
Index	Fitted	Left Out	CV PRESS
1	312	71	2.5897
2	315	68	2.3306
3	305	78	3.0757
4	292	91	3.8508
5	308	75	3.1010
Total			14.9478

Parameter Estimates											
							Cross Validation Estimates				
	Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	1	2	3	4	5
	Intercept	1	7.902150	0.231340	34.16	<.0001	7.836	7.709	7.977	8.113	7.878
*	log(Grade Living Are	1	0.555788	0.032369	17.17	<.0001	0.565	0.583	0.546	0.526	0.560
*	Neighborhood BrkSide	1	-0.132789	0.029061	-4.57	<.0001	-0.104	-0.125	-0.153	-0.142	-0.139
*	Neighborhood Edwards	1	-0.153226	0.023571	-6.50	<.0001	-0.159	-0.141	-0.174	-0.148	-0.142
*	Neighborhood NAmes	0	0	.	.	.	0.000	0.000	0.000	0.000	0.000

* Forced into the model by the INCLUDE= option

Figure 2

Appendix
MSDS 6371 Project
Amy Adyanthaya & Nicholas Mueller

Figure 3: SGPLOT of Untransformed Data
Figure 4: SGPLOT of Transformed Data

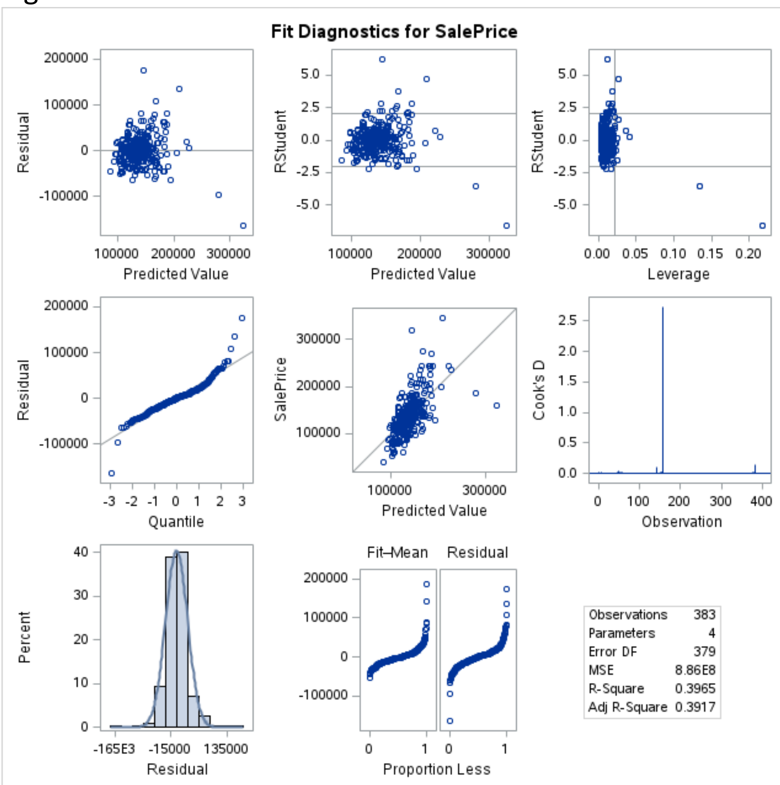


Figure 3

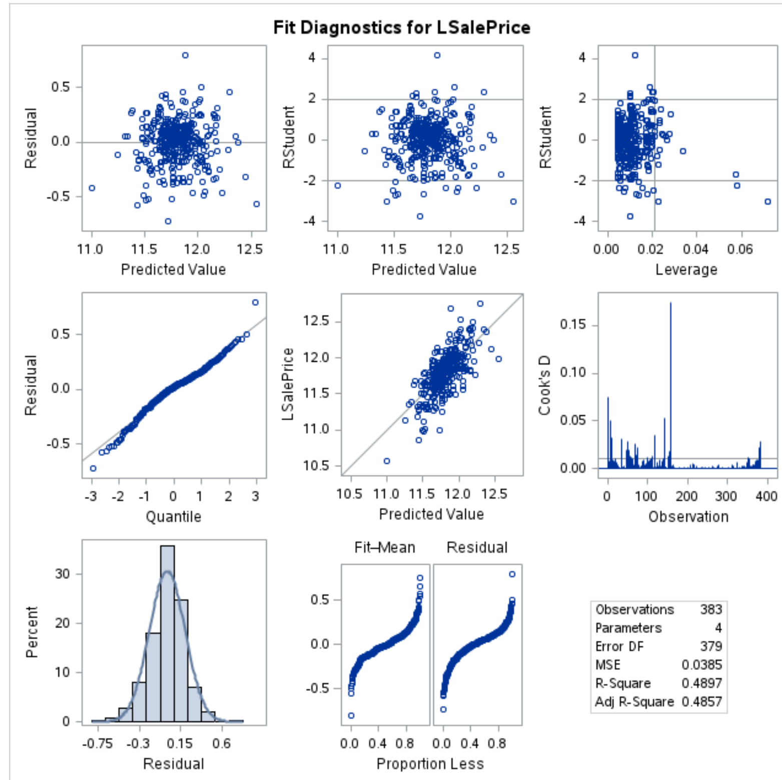


Figure 4

Appendix

MSDS 6371 Project

Amy Adyanthaya & Nicholas Mueller

Figure 5: Forward Model Selection Summary

Figure 6: Forward Model Procedure

Forward Selection Summary

Step	Effect Entered	Number Effects In	Number Parms In	BIC	SBC	ASE	Test ASE
0	Intercept	1	1	26241.5542	26247.9830	6545236629	5240146401
1	OverallQual	2	2	25081.9947	25093.9205	2407610075	2183734340
2	GrLivArea	3	3	24805.8483	24823.5501	1895872246	1429062155
3	Neighborhood	4	26	24463.1069	24618.2931	1381378934	1267681724
4	BsmtQual	5	30	24336.2152	24511.5898	1229813156	1125859284
5	RoofMatl	6	37	24246.7142	24459.3969	1126773478	992553971
6	BsmtExposure	7	41	24161.2624	24392.6392	1038258063	887798190
7	BldgType	8	45	24090.5129	24340.1432	968520658	814819681
8	GarageArea	9	46	24055.7053	24308.6166	936862726	783103258
9	BsmtFinType1	10	51	24005.1044	24281.3913	887754085	761308325
10	ExterQual	11	54	23968.8200	24258.0695	854376460	696263651
11	Condition2	12	60	23906.4544	24220.9513	797857915	689549176
12	HouseStyle	13	67	23861.0360	24205.5865	754569578	646923249
13	PoolQC	14	70	23843.9898	24201.1057	738080491	652576202
14	CentralAir	15	71	23837.2600	24198.3431	731864152	647648726
15	MasVnrArea	16	72	23832.1043	24197.3239	726790799	638244853
16	SaleCondition	17	77	23811.0731	24197.1159*	704908084	630389106
17	Fireplaces	18	78	23807.4864	24197.6873	700981312	616828008
18	LotConfig	19	82	23794.7532	24201.5393	686417051	622107216
19	FullBath	20	83	23792.9572	24204.0552	683737406	628163858
20	GarageCars	21	84	23792.7385	24208.4062	682145631	628792633
21	GarageFinish	22	87	23784.3622*	24212.2170	672021482	642762792

* Optimal Value of Criterion

Selection stopped at a local minimum of the BIC criterion.

Stop Details

Candidate For	Effect	Candidate BIC		Compare BIC
Entry	OpenPorchSF	23784.9416	>	23784.3622

Figure 5

The GLMSELECT Procedure

Data Set	WORK.HOUSETRAIN2
Dependent Variable	SalePrice
Selection Method	Forward
Select Criterion	SBC
Stop Criterion	BIC
Choose Criterion	BIC
Effect Hierarchy Enforced	None
Random Number Seed	733974034

Number of Observations Read	1460
Number of Observations Used	1452
Number of Observations Used for Training	1161
Number of Observations Used for Testing	291

Figure 6

Appendix
MSDS 6371 Project
Amy Adyanthaya & Nicholas Mueller

Figure 7: Stepwise Model Selection Summary

Stepwise Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	BIC	SBC	ASE	Test ASE
0	Intercept		1	1	26741.0337	26747.4552	5962366553	7725243940
1	OverallQual		2	2	25569.6518	25581.5495	2221329111	3001251169
2	GrLivArea		3	3	25306.8265	25324.5344	1778558972	1937356458
3	Neighborhood		4	27	24966.6182	25128.1246	1306621164	1563422223
4	BsmtQual		5	31	24852.2460	25034.3108	1178966455	1314602007
5	RoofMatl		6	38	24764.6421	24984.2739	1084156517	1147780081
6	GarageArea		7	39	24703.6409	24926.2926	1026378772	1103531761
7	BsmtExposure		8	43	24636.5386	24878.3751	962581769	960194599
8	BldgType		9	47	24570.7164	24830.7858	902999718	922552535
9	BsmtFinType1		10	52	24520.1367	24803.8128	856812917	882823574
10	ExterQual		11	55	24490.3672	24787.7114	830300219	793366216
11	Condition2		12	62	24431.6642	24760.2611	778184564	761813595
12	HouseStyle		13	69	24380.1392	24738.5813	732891290	731737614
13	SaleCondition		14	74	24350.1533	24729.4601	705933866	760244367
14	PoolQC		15	77	24326.4197	24716.8994	686131803	774524960
15	Fireplaces		16	78	24316.7218	24710.5388	678412850	771366889
16	CentralAir		17	79	24308.9753*	24706.3582*	672012840	779410399

* Optimal Value of Criterion

Figure 7

Appendix

MSDS 6371 Project

Amy Adyanthaya & Nicholas Mueller

Figure 8: Backward Model Selection Summary
Figure 9: Backward Model Procedure
Figure 10: Backward Model performance statistics

The GLMSELECT Procedure

Backward Selection Summary							
Step	Effect Removed	Number Effects In	Number Parms In	BIC	SBC	ASE	Test ASE
0		45	146	6445.5212	6715.9144	269366149	1766365074
1	Neighborhood	44	123	6325.7061	6677.0563	365601043	1787598053
2	BsmtFinType2	43	118	6310.9716	6654.2470	372616846	1733462327
3	HeatingQC	42	114	6299.7344	6633.3240	375000226	1725984059
4	Foundation	41	110	6288.7084	6616.6553	382735146	1684816852
5	LandContour	40	107	6280.6063	6600.4910	383971130	1671257935
6	BldgType	39	103	6270.4554	6584.7108	393041878	1622570572
7	Heating	38	100	6262.6436	6568.4969	394246612	1606835959
8	Fence	37	96	6253.4122	6552.9746	403903862	1562522686
9	RoofStyle	36	92	6244.7879	6537.3226	413620458	1607501090
10	GarageType	35	87	6235.6346	6520.3064	429695056	1573897345
11	FireplaceQu	34	82	6227.6178	6503.0687	446068090	1519424694
12	LotConfig	33	78	6222.6138	6490.6868	461755759	1461424610
13	Condition1	32	71	6220.6396	6478.4372	506036547	1366416013
14	ExterCond	31	69	6216.7895	6467.4962	506849552	1370332806
15	BsmtExposure	30	66	6214.2617	6456.8521	517835092	1401301340
16	PavedDrive	29	64	6210.6997	6446.2114	519181433	1382915258
17	Alley	28	62	6207.2977	6435.8048	520933570	1383361967
18	MiscFeature	27	60	6203.9451	6425.4332	522752060	1374530163
19	GarageFinish	26	58	6201.1987	6415.9905	526187565	1385478194
20	Electrical	25	56	6198.9376	6407.2207	530823186	1378954274
21	BsmtCond	24	54	6197.8082*	6400.0627*	538355704	1348647959

* Optimal Value of Criterion

Selection stopped at a local minimum of the BIC criterion.

Stop Details				
Candidate For	Effect	Candidate BIC		Compare BIC
Removal	MasVnrType	6199.1413	>	6197.8082

Figure 8

The GLMSELECT Procedure

Data Set	WORK.HOUSETRAIN2
Dependent Variable	SalePrice
Selection Method	Backward
Select Criterion	SBC
Stop Criterion	BIC
Choose Criterion	BIC
Effect Hierarchy Enforced	None
Random Number Seed	70560727

Number of Observations Read	1460
Number of Observations Used	1452
Number of Observations Used for Training	303
Number of Observations Used for Testing	1149

Figure 9

Root MSE	23299
Dependent Mean	174646
R-Square	0.9066
Adj R-Sq	0.8753
AIC	6144.17401
AICC	6196.28668
SBC	6121.57012
ASE (Train)	405263744
ASE (Test)	1632045137
CV PRESS	2.650568E11

Figure 10

Appendix
MSDS 6371 Project

Amy Adyanthaya & Nicholas Mueller

Figure 11: Custom Stepwise Model Procedure

Figure 12: Custom Stepwise Model Selection Summary

Figure 13: Custom stepwise Model performance Statistics

The GLMSELECT Procedure

Data Set	WORK.HOUSETRAIN2
Dependent Variable	SalePrice
Selection Method	Stepwise
Select Criterion	SBC
Stop Criterion	BIC
Choose Criterion	BIC
Effect Hierarchy Enforced	None
Random Number Seed	835413291

Number of Observations Read	1460
Number of Observations Used	1452
Number of Observations Used for Training	1145
Number of Observations Used for Testing	307

Figure 11

The GLMSELECT Procedure

Stepwise Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	BIC	SBC	ASE	Test ASE
0	Intercept		1	1	25855.9776	25862.1779	6408580915	5811622598
1	OverallQual		2	2	24770.0649	24781.3866	2478278295	1926992906
2	GrLivArea		3	3	24460.1504	24476.7671	1887719753	1473114964
3	Neighborhood		4	27	24164.1033	24316.3713	1415745779	1080659551
4	BsmtQual		5	31	24045.5319	24215.5856	1264945488	937837851
5	GarageCars		6	32	23989.9292	24162.9276	1200680311	950070851
6	Fireplaces		7	33	23965.9828	24143.2008	1172934110	928740515
7	PoolQC		8	36	23930.1438	24121.4053	1129775986	999394767
8	ExterQual		9	39	23896.9781	24101.9648	1090446288	947364117
9	HouseStyle		10	46	23856.5910	24094.6545	1037842308	928186230
10	LandContour		11	49	23833.1577*	24084.3828*	1009766707	955089793

* Optimal Value of Criterion

Selection stopped as adding or dropping any effect does not improve the selection criterion.

Figure 12

Root MSE	34302
Dependent Mean	179879
R-Square	0.8074
Adj R-Sq	0.8023
AIC	25705
AICC	25707
SBC	24687
ASE (Train)	1145550298
ASE (Test)	1206106267
CV PRESS	1.469603E12

Figure 13