**MSDS 6371 Group Project - D&B All-Stars**

Vanessa Torres

Michael Landrum

Travis Deason

Brian Coari

# Introduction

Introduction to the paper

# Data Description

(Where did the data come from? How big is it? How many observations? Where can we find out more? What are the specific variables that we need to know with respect to your analysis?)

# Analysis Question 1

Mr. Mason,

Thank you for contacting DABAS LLC; we feel confident we can answer your questions about the Ames, Iowa housing market. We were able to track down the most recent 2,930 home sales in Ames as well as 79 characteristics of each house. Since you are only interested in three specific neighborhoods, North Ames (NA), Eastwood (EW) and Brookside (BS), and how square footage (SF) relates to their sale price (SP, in dollars), we will restrict our model to only focus on those variables.

By examining the scatterplot (see Appendix Plot 1A), we concluded we would need to take the log of the sales price and square footage to create a linear relationship. Also, the slope of the regression lines of different neighborhoods would be different, so we included interaction terms. This means our model would look like this:

ln(SP) = B0 + B1 \* ln(SF) + B2 \* BS + B3 \* EW + B4 \* NA + B5 \* ln(SF) \* BS + B6 \* ln(SF) \* EW + B7 \* ln(SF) \* NA

To test our model, we used a null hypothesis that all Bi are zero, or that there is no correlation between sales price and any of the other variables. And our alternative hypothesis is that at least one Bi is not zero, or that there is a correlation between at least one of the variables and sales price. We received a p-value well below 0.05 (it actually was below 0.001, see Appendix Table 1B column 6), so we rejected the null hypothesis and concluded there was a relationship between the sales price and the other variables.

We then needed to test whether each individual variable was needed in our model. For each variable, our null hypothesis was that given the other variables contribution to the model, that particular variable did not help our model (or that the coefficient in front of that variable was 0). And our alternative hypothesis was that the variable contributed to the model (or that the coefficient in front of the variable was not 0). For each variable aside from BS, including the interaction terms, we received p-value below 0.05 (see Appendix Table 1C column 6), which means for each variable we rejected the null hypothesis and concluded that they were all beneficial to our model. Since the interaction term was significant that included BS, we kept BS in the model. The full regression equation is:

**ln(SP) = 5.16 + 0.95 ln(SF) + 1.04 BS + 2.40 EW + 3.33 NA – 0.17 BS ln(SF) - 0.36 EW ln(SF) - 0.48 NA ln(SF)**

Note: A zero for all categorical variables (BS, EW, NA) would represent the houses outside those three neighborhoods, but still in Ames. We decided to use them in an effort to reduce the overall variance to our model.

This model has an r-squared of 0.5649 and an adjusted r-squared of 0.5368. The r-squared means 56.49% of the variation in sales price, can be explained the variation in neighborhood and square footage. As always, these coefficients are just estimates. For instance, we are 95% confident the true coefficient for ln(Square Footage) is in the interval (0.914,0.984) (see Appendix 1C, row 3 columns 7 and 8. The other confidence intervals for the rest of the coefficients can be found in the same table.).

For us to use this model, we had to assume independence even though that likely isn't completely true. Also, this model gave us relatively normal residuals (slightly curved and skewed left, but with 2000+ observations, the Central Limit Theorem should kick in), but it did show some outliers and leverage points (see Appendix Plot 1D). We concluded the outliers were all accurate measurements so we had to include them. The model created after removing the leverage point gave very similar results to our model, so we decided to keep our model and include the leverage points **(**all confidence intervals for the coefficients contain the coefficients from our full model. See Appendix Table 1E).

We can simplify this model a little by separating it by neighborhood. This leaves us with three different equations you can look at depending on which neighborhood you are working with.

**{ ln(SP) | Brookside } = 6.20 + 0.78 ln(SF)**

**{ ln(SP) | Edwards } = 7.56 + 0.59 ln(SF)**

**{ ln(SP) | North Ames } = 8.49 + 0.47 ln(SF)**

In general, when you double the size of a house in Brookside, the sale price increases by a factor of 2.16. When you double the size of a house in Edwards, the sale price increases by a factor of 1.79. And when you double the size of a house in North Ames, the sale price increases by a factor of 1.56.

Note: The model created by removing the other neighborhoods produced the same regression equations for the individual neighborhoods (See Appendix Table 1F and Reference 1G).

Due to the nature of this study (it was an observational study by nature), we cannot infer causation and say the square footage or the neighborhood causes a change in the sales price of the house; there are obviously many confounding variables that play a part. But, we can say there is a correlation between the sales price and the square footage and neighborhood.

# Analysis Question 2

**Restatement of Problem**

**Overall Data Observations:**

* Number of Observations Read: 1460
* Number of Observations with Missing Values: 339
* Number of Observations Used: 1121

**Checking Assumptions:**

First we will look at the distribution of our response variable, the Sale Price using all of our identified continuous and indicator variables through proc reg.

***(****All plots for the SalePrice are available in the Appendix. Please see Appendix Table 2A).*

In the SalePrice plot group we see a few worrying things. The residual plot seems clustered tightly together with a few outliers.

1. **Normality**: Judging from the qq plot, there seems to be a curvature at the tail ends of the data, which violates normality.
2. **Constant variance:** Judging from the scatter plots we do not see a large concern with variance. There are many outliers that are concerning, but with so many observations that itself wouldn’t be out of the question… still, given the violation in normality we should try a transform and hope this gets better.
3. **Linear trend:** From the predicted value line and an adjusted R-Square of .91 our linear trend looks good here, so we will compare this to our transformed data.
4. We will assume all observations are **independent**.

***(****All plots for the scatterplots for the individual variables are available in the Appendix. Please see Appendix Table 2B).*

Viewing the scatterplots for the individual variables we see many variables with uneven distributions and distant outliers. These data points might also benefit from a log transformation

After taking the logs we see a few improvements in the model:

***(****All plots for the scatterplots for the SalePrice and the individual variables after the log transformations are available in the Appendix. Please see Appendix Tables 2C and 2D).*

After the transformations we see many improvements in the shape of our data. In general the data has become less affected by outliers, and closer to a random cloud.

1. **Normality**: Judging from the qq plot, the curvature at both ends is still present, but far less pronounced. The tails are mostly going to be formed by our large number of categorical variables that have a small number of discrete values. In the middle we see a straight line indicating normality, and the histogram shows a much more normal distribution.
2. **Constant variance:** Across the board the scatter plots are showing data that are more cloud-like and normal in their variance. We still have a large number of variables with values of zero which negatively affects our constant variance, but we do not see a way around this, we will accept it and proceed.
3. **Linear trend:** From the predicted value line we see a slight improvement here, which is also reflected by an increase in our adjusted R-Square to .93. Our linear trend looks even better than before the transformation.
4. We will assume all observations are **independent**.

**Comparing Competing Models**

**Forward**

***(****The Table output of the Forward Selection process is available in the Appendix. Please see Appendix Table 2E).*

**Backward**

***(****The Table output of the Backward Selection process is available in the Appendix. Please see Appendix Table 2F).*

**Stepwise**

***(****The Table output of the Stepwise Selection process is available in the Appendix. Please see Appendix Table 2G).*

**CUSTOM**

***(****The Table output of the Stepwise Selection process is available in the Appendix. Please see Appendix Table 2H).*

**Conclusion**:

|  |  |  |  |
| --- | --- | --- | --- |
| **Predictive Models** | Adjusted R2 | CV PRESS | Kaggle Score |
| Forward | 0.8488 | 27.26 |  |
| Backward | 0.9342 | 54.25 |  |
| Stepwise | 0.8488 | 27.26 |  |
| CUSTOM |  |  |  |

A short summary of the analysis.

# Appendix

Plot 1A:

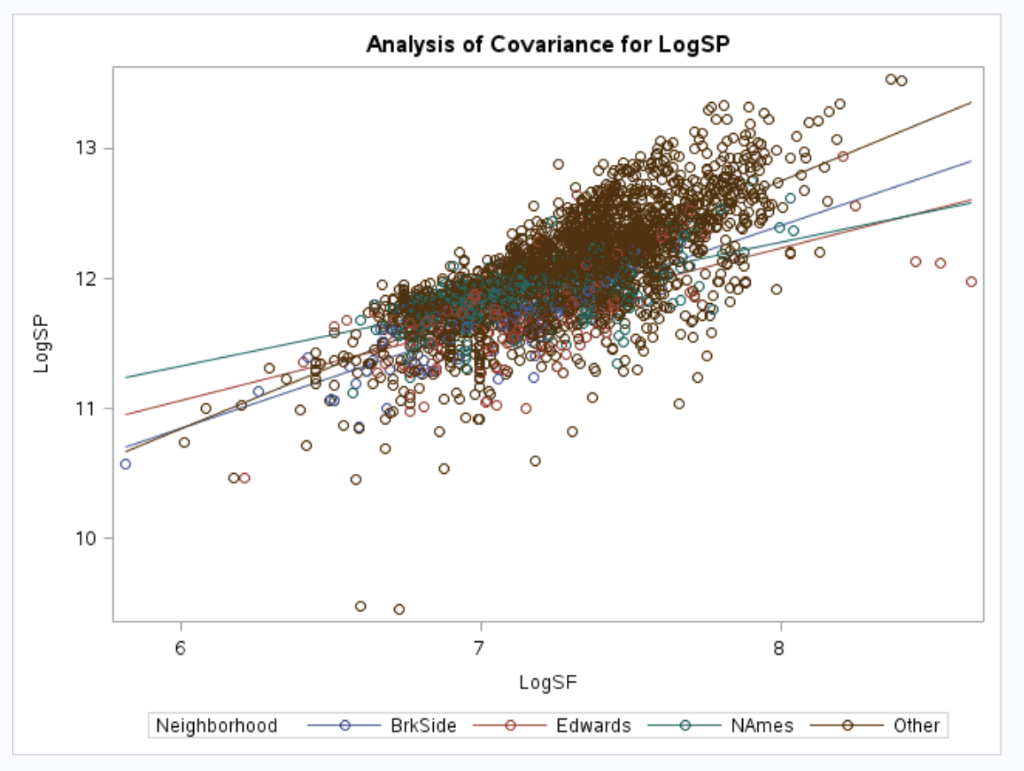


Table 1B:

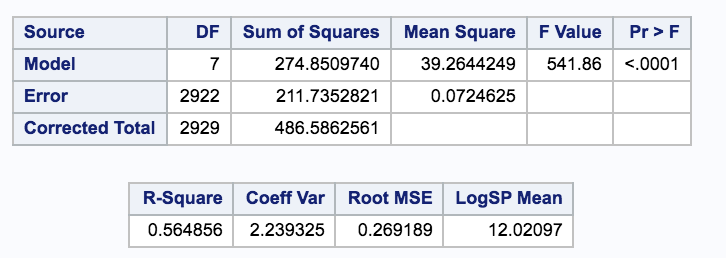
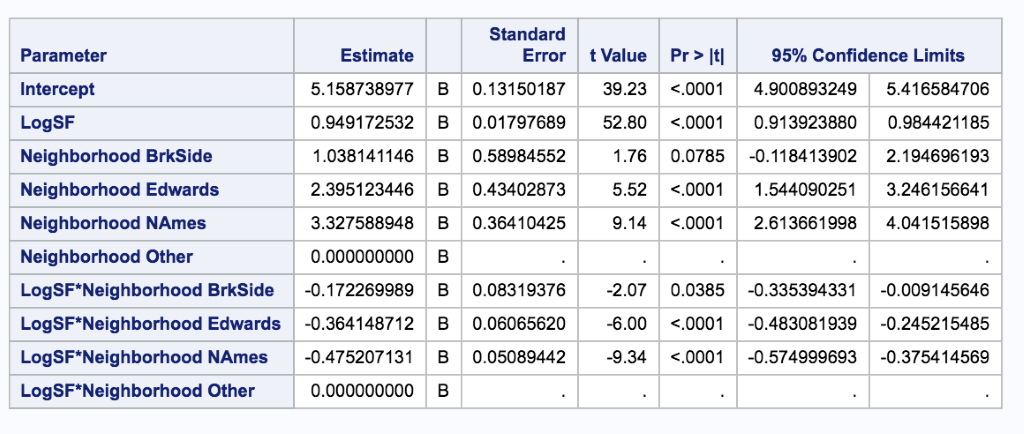


Table 1C:



Plot 1D:

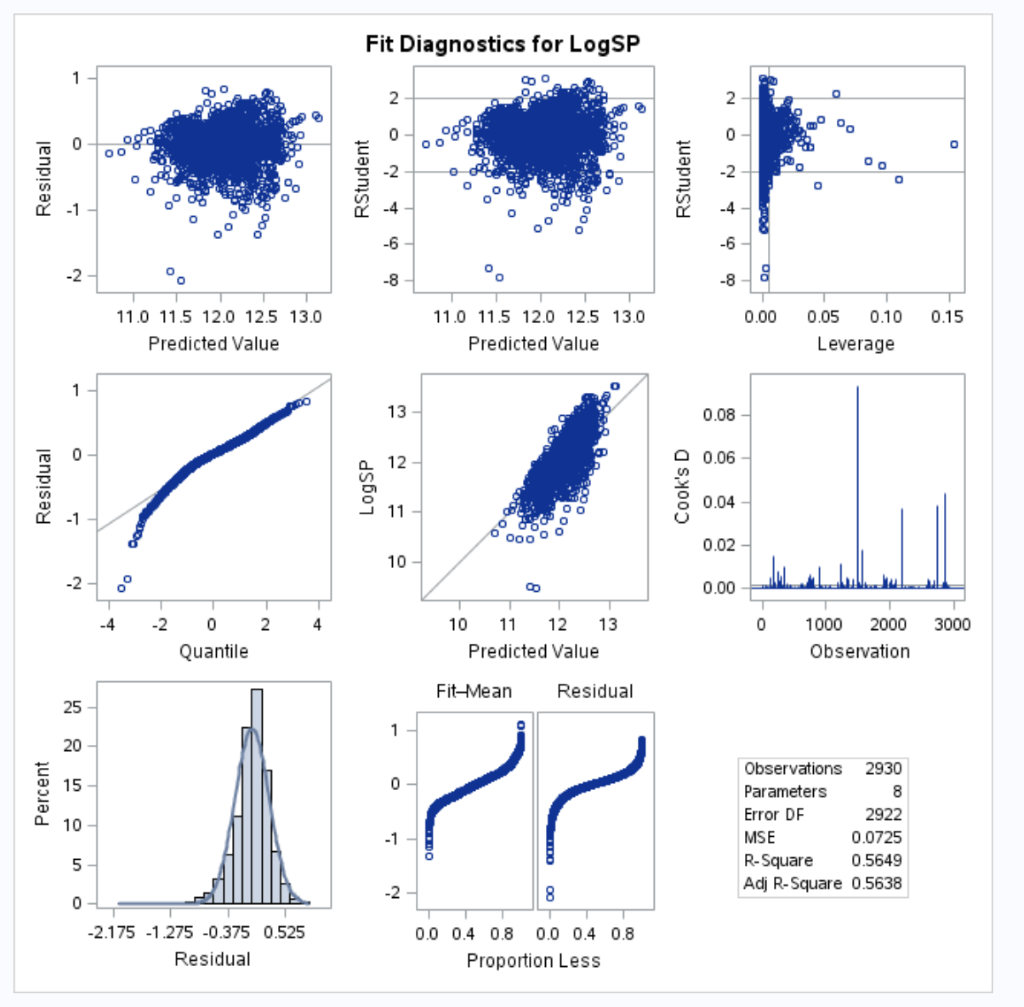


Table 1E:

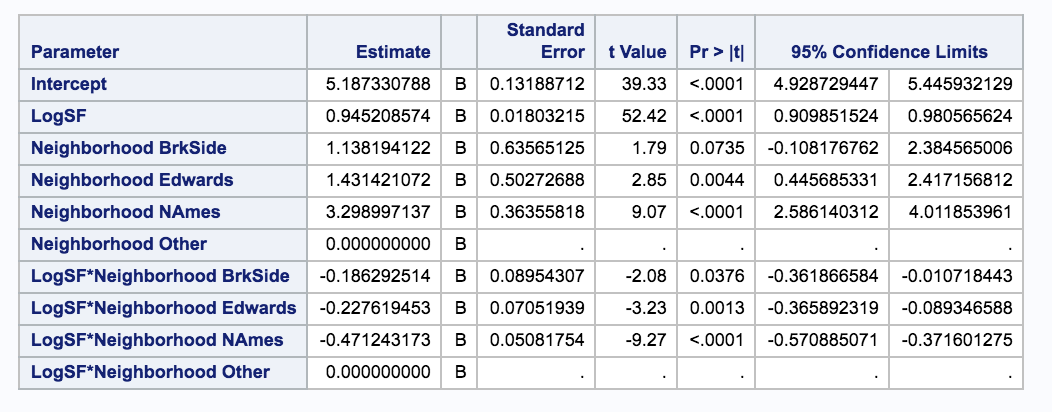
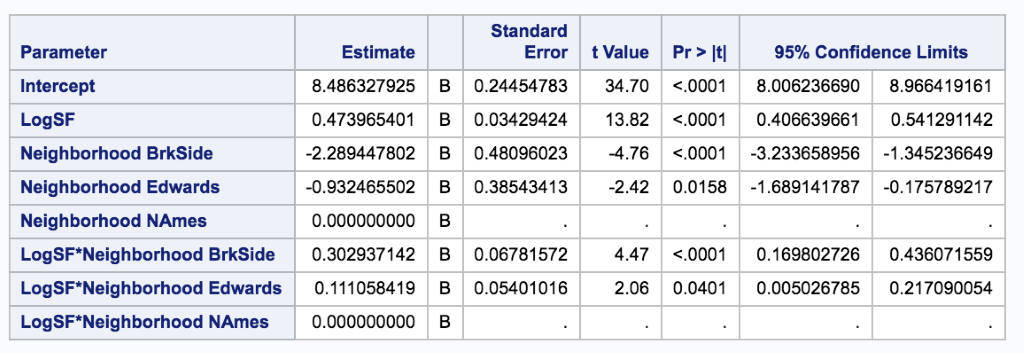


Table 1F:



Reference 1G:

Separated regression equations:

{ ln(SP) | Brookside } = 6.19 + 0.77 ln(SF)

{ ln(SP) | Edwards } = 7.55 + 0.58 ln(SF)

{ ln(SP) | North Ames } = 8.48 + 0.47 ln(SF)

**Question 1 Code:**

R code to turn all other neighborhoods into "other", removed unwanted variables to help with speed of SAS and created log variables:

ah <- read.csv("~/Documents/AmesHousing.csv")

q1 <- ah[,c("Neighborhood","Gr.Liv.Area","SalePrice")] ## removes unwanted variables

colnames(q1) <- c("Neighborhood","SquareFeet","SalePrice") ## Renamed Gr.Liv.Area to Sqaure Feet

q1$LogSF <- log( q1$SquareFeet )

q1$LogSP <- log( q1$SalePrice ) ## create log variables for square feet and sale price

q1$Neighborhood[ Q1$Neighborhood != "BrkSide" & Q1$Neighborhood != "Edwards" & q1$Neighborhood = "NAmes" ] <- "Other" ## turns all the other neighborhoods into "Other"

write.csv(q1,"q1all.csv")

R code to remove leverage points:

sorted <- q1[ order(q1$LogSF,decreasing=TRUE ), ]

row.names(sorted) <- 1:nrow(sorted) ## sorted and changed row names to easily find highest/lowest

sorted2 <- sorted[-c(1,2,3,4,5,2930),] ## 5 highest were above 4000 sq ft. Lowest was also a leverage pt

write.csv(sorted2,"nl.csv") ## that's an L, for "no leverage"

R code to remove all other neighborhoods:

q1some <- q1[ q1$Neighborhood=="NAmes" | q1$Neighborhood=="Edwards" | q1$Neighborhood == "BrkSide" , ]

write.csv(q1some,"q1some.csv")

All SAS code was the same, but with different DATANAME

proc glm data=DATANAME plots=all;

class Neighborhood;

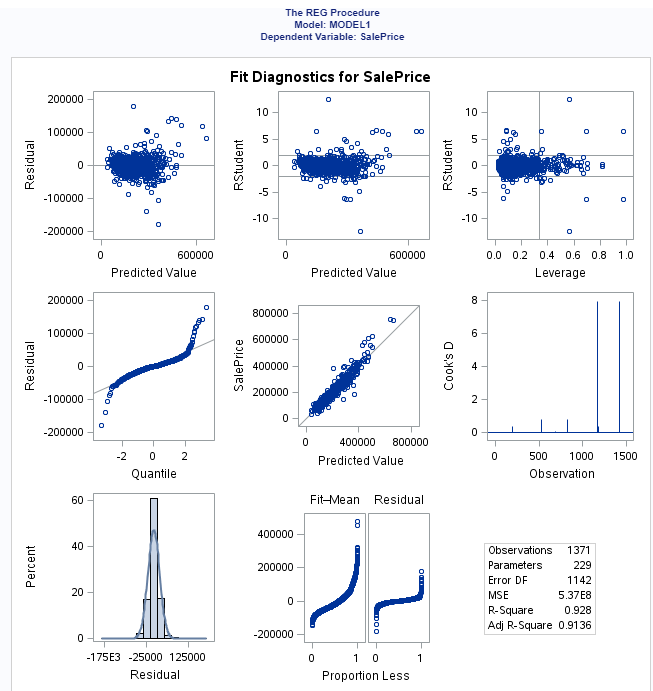
model LogSP = LogSF | Neighborhood / solution clparm;

run;

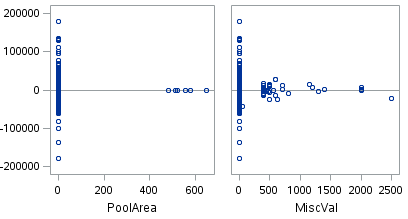
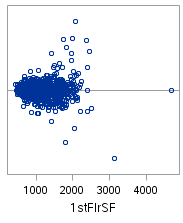
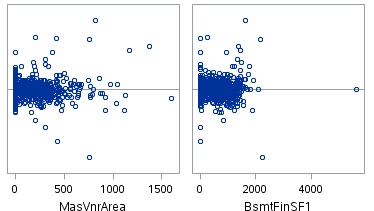
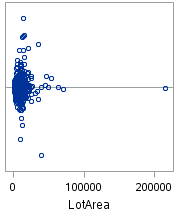
Plot 1A, Table 1B, Table 1C, and plot for 1D were generated using "q1all". Table 1E used "nl". Table 1F used "q1some".

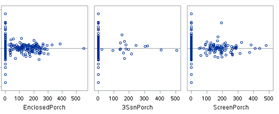
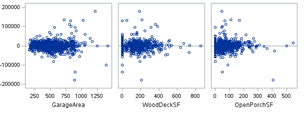
**Question 2 Prog Reg Output Before Log Transformations:**

Plot Group 2A:



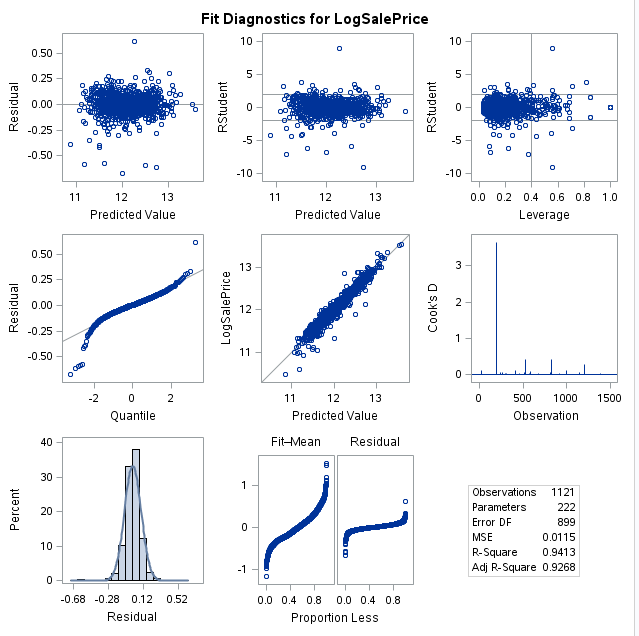
Plot Group 2B:



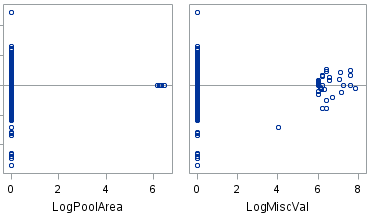
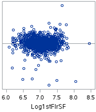
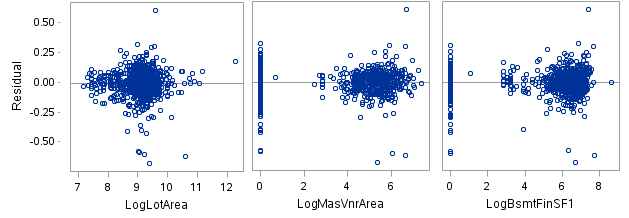


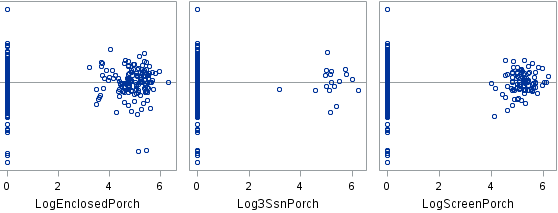
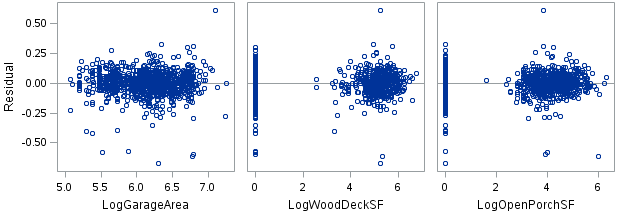
**Question 2 Prog Reg Output After Log Transformations:**

Plot Group 2C:



Plot Group 2D:





**Question 2 Forward Selection Output:**

Table 2E:

**The GLMSELECT Procedure**

|  |  |
| --- | --- |
| **Data Set** | WORK.TRAIN\_DUMMY |
| **Dependent Variable** | LogSalePrice |
| **Selection Method** | Forward |
| **Select Criterion** | SBC |
| **Stop Criterion** | Cross Validation |
| **Cross Validation Method** | Random |
| **Cross Validation Fold** | 5 |
| **Effect Hierarchy Enforced** | None |
| **Random Number Seed** | 46344627 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 1460 |
| **Number of Observations Used** | 1121 |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **MSZoning** | 5 | C FV RH RL RM |
| **Street** | 2 | Grvl Pave |
| **Alley** | 3 | Gr NA Pa |
| **LotShape** | 4 | IR1 IR2 IR3 Reg |
| **LandContour** | 4 | Bnk HLS Low Lvl |
| **Utilities** | 1 | AllPub |
| **LotConfig** | 5 | Corner CulDSac FR2 FR3 Inside |
| **LandSlope** | 3 | Gtl Mod Sev |
| **Neighborhood** | 25 | Blmngtn Blueste BrDale BrkSide ClearCr CollgCr Crawfor Edwards Gilbert IDOTRR MeadowV Mitchel NAmes NPkVill NWAmes NoRidge NridgHt OldTown SWISU Sawyer SawyerW Somerst StoneBr Timber Veenker |
| **Condition1** | 9 | Artery Feedr Norm PosA PosN RRAe RRAn RRNe RRNn |
| **Condition2** | 6 | Artery Feedr Norm PosA PosN RRNn |
| **BldgType** | 5 | 1Fam 2fmCon Duplex Twnhs TwnhsE |
| **HouseStyle** | 8 | 1.5Fin 1.5Unf 1Story 2.5Fin 2.5Unf 2Story SFoyer SLvl |
| **RoofStyle** | 5 | Flat Gable Gambr Hip Mansa |
| **RoofMatl** | 7 | ClyTile CompShg Membran Roll Tar&Grv WdShake WdShngl |
| **Exterior1st** | 14 | AsbShng BrkComm BrkFace CBlock CemntBd HdBoard ImStucc MetalSd Plywood Stone Stucco VinylSd Wd Sdng WdShing |
| **Exterior2nd** | 16 | AsbShng AsphShn Brk Cmn BrkFace CBlock CmentBd HdBoard ImStucc MetalSd Other Plywood Stone Stucco VinylSd Wd Sdng Wd Shng |
| **MasVnrType** | 4 | BrkCmn BrkFace None Stone |
| **ExterQual** | 4 | Ex Fa Gd TA |
| **ExterCond** | 4 | Ex Fa Gd TA |
| **Foundation** | 6 | BrkTil CBlock PConc Slab Stone Wood |
| **BsmtQual** | 5 | Ex Fa Gd NA TA |
| **BsmtCond** | 5 | Fa Gd NA Po TA |
| **BsmtExposure** | 5 | Av Gd Mn NA No |
| **BsmtFinType1** | 7 | ALQ BLQ GLQ LwQ NA Rec Unf |
| **BsmtFinType2** | 7 | ALQ BLQ GLQ LwQ NA Rec Unf |
| **Heating** | 5 | GasA GasW Grav OthW Wall |
| **HeatingQC** | 5 | Ex Fa Gd Po TA |
| **CentralAir** | 2 | N Y |
| **Electrical** | 6 | FuseA FuseF FuseP Mix NA SBrkr |
| **KitchenQual** | 4 | Ex Fa Gd TA |
| **Functional** | 6 | Maj1 Maj2 Min1 Min2 Mod Typ |
| **FireplaceQu** | 6 | Ex Fa Gd NA Po TA |
| **GarageType** | 6 | 2Types Attchd Basment BuiltIn CarPort Detchd |
| **GarageFinish** | 3 | Fin RFn Unf |
| **GarageQual** | 5 | Ex Fa Gd Po TA |
| **GarageCond** | 5 | Ex Fa Gd Po TA |
| **PavedDrive** | 3 | N P Y |
| **PoolQC** | 4 | Ex Fa Gd NA |
| **Fence** | 5 | GdPrv GdWo MnPrv MnWw NA |
| **MiscFeature** | 4 | NA Othr Shed TenC |
| **SaleType** | 6 | COD CWD Con New Oth WD |
| **SaleCondition** | 6 | Abnorml AdjLand Alloca Family Normal Partial |

| **Dimensions** | |
| --- | --- |
| **Number of Effects** | 78 |
| **Number of Parameters** | 285 |

**The GLMSELECT Procedure**

| **Forward Selection Summary** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Effect Entered** | **Number Effects In** | **Number Parms In** | **Adjusted R-Square** | **SBC** | **CV PRESS** |
| **\* Optimal Value of Criterion** | | | | | | |
| **0** | **Intercept** | 1 | 1 | 0.0000 | -2066.4335 | 176.5671 |
| **1** | **OverallQual** | 2 | 2 | 0.6848 | -3354.5767 | 55.7194 |
| **2** | **Log1stFlrSF** | 3 | 3 | 0.7407 | -3567.6573 | 45.9320 |
| **3** | **2ndFlrSF** | 4 | 4 | 0.7812 | -3751.9351 | 38.8164 |
| **4** | **YearBuilt** | 5 | 5 | 0.8140 | -3928.0564 | 33.1151 |
| **5** | **OverallCond** | 6 | 6 | 0.8375 | -4073.1053 | 29.0716 |
| **6** | **LogLotArea** | 7 | 7 | 0.8488\* | -4147.7963\* | 27.1116\* |

|  |
| --- |
| Selection stopped at a local minimum of the cross validation PRESS. |

| **Stop Details** | | | | |
| --- | --- | --- | --- | --- |
| **Candidate For** | **Effect** | **Candidate CV PRESS** |  | **Compare CV PRESS** |
| **Entry** | RoofMatl | 27.5160 | > | 27.1116 |

**The GLMSELECT Procedure**

**Selected Model**

**The selected model is the model at the last step (Step 6).**

|  |  |
| --- | --- |
| **Effects:** | Intercept LogLotArea OverallQual OverallCond YearBuilt Log1stFlrSF 2ndFlrSF |

| **Analysis of Variance** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** |
| **Model** | 6 | 149.80112 | 24.96685 | 1048.61 |
| **Error** | 1114 | 26.52380 | 0.02381 |  |
| **Corrected Total** | 1120 | 176.32491 |  |  |

|  |  |
| --- | --- |
| **Root MSE** | 0.15430 |
| **Dependent Mean** | 12.04841 |
| **R-Square** | 0.8496 |
| **Adj R-Sq** | 0.8488 |
| **AIC** | -3059.95014 |
| **AICC** | -3059.82064 |
| **SBC** | -4147.79630 |
| **CV PRESS** | 27.11158 |

| **Parameter Estimates** | | | | |
| --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **t Value** |
| **Intercept** | 1 | -1.550928 | 0.446433 | -3.47 |
| **LogLotArea** | 1 | 0.101900 | 0.011103 | 9.18 |
| **OverallQual** | 1 | 0.110904 | 0.005312 | 20.88 |
| **OverallCond** | 1 | 0.063287 | 0.004850 | 13.05 |
| **YearBuilt** | 1 | 0.004269 | 0.000210 | 20.32 |
| **Log1stFlrSF** | 1 | 0.445317 | 0.021389 | 20.82 |
| **2ndFlrSF** | 1 | 0.000245 | 0.000012953 | 18.89 |

Table 2F:

**The GLMSELECT Procedure**

|  |  |
| --- | --- |
| **Data Set** | WORK.TRAIN\_DUMMY |
| **Dependent Variable** | LogSalePrice |
| **Selection Method** | Backward |
| **Select Criterion** | SBC |
| **Stop Criterion** | Cross Validation |
| **Cross Validation Method** | Random |
| **Cross Validation Fold** | 5 |
| **Effect Hierarchy Enforced** | None |
| **Random Number Seed** | 46497073 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 1460 |
| **Number of Observations Used** | 1121 |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **MSZoning** | 5 | C FV RH RL RM |
| **Street** | 2 | Grvl Pave |
| **Alley** | 3 | Gr NA Pa |
| **LotShape** | 4 | IR1 IR2 IR3 Reg |
| **LandContour** | 4 | Bnk HLS Low Lvl |
| **Utilities** | 1 | AllPub |
| **LotConfig** | 5 | Corner CulDSac FR2 FR3 Inside |
| **LandSlope** | 3 | Gtl Mod Sev |
| **Neighborhood** | 25 | Blmngtn Blueste BrDale BrkSide ClearCr CollgCr Crawfor Edwards Gilbert IDOTRR MeadowV Mitchel NAmes NPkVill NWAmes NoRidge NridgHt OldTown SWISU Sawyer SawyerW Somerst StoneBr Timber Veenker |
| **Condition1** | 9 | Artery Feedr Norm PosA PosN RRAe RRAn RRNe RRNn |
| **Condition2** | 6 | Artery Feedr Norm PosA PosN RRNn |
| **BldgType** | 5 | 1Fam 2fmCon Duplex Twnhs TwnhsE |
| **HouseStyle** | 8 | 1.5Fin 1.5Unf 1Story 2.5Fin 2.5Unf 2Story SFoyer SLvl |
| **RoofStyle** | 5 | Flat Gable Gambr Hip Mansa |
| **RoofMatl** | 7 | ClyTile CompShg Membran Roll Tar&Grv WdShake WdShngl |
| **Exterior1st** | 14 | AsbShng BrkComm BrkFace CBlock CemntBd HdBoard ImStucc MetalSd Plywood Stone Stucco VinylSd Wd Sdng WdShing |
| **Exterior2nd** | 16 | AsbShng AsphShn Brk Cmn BrkFace CBlock CmentBd HdBoard ImStucc MetalSd Other Plywood Stone Stucco VinylSd Wd Sdng Wd Shng |
| **MasVnrType** | 4 | BrkCmn BrkFace None Stone |
| **ExterQual** | 4 | Ex Fa Gd TA |
| **ExterCond** | 4 | Ex Fa Gd TA |
| **Foundation** | 6 | BrkTil CBlock PConc Slab Stone Wood |
| **BsmtQual** | 5 | Ex Fa Gd NA TA |
| **BsmtCond** | 5 | Fa Gd NA Po TA |
| **BsmtExposure** | 5 | Av Gd Mn NA No |
| **BsmtFinType1** | 7 | ALQ BLQ GLQ LwQ NA Rec Unf |
| **BsmtFinType2** | 7 | ALQ BLQ GLQ LwQ NA Rec Unf |
| **Heating** | 5 | GasA GasW Grav OthW Wall |
| **HeatingQC** | 5 | Ex Fa Gd Po TA |
| **CentralAir** | 2 | N Y |
| **Electrical** | 6 | FuseA FuseF FuseP Mix NA SBrkr |
| **KitchenQual** | 4 | Ex Fa Gd TA |
| **Functional** | 6 | Maj1 Maj2 Min1 Min2 Mod Typ |
| **FireplaceQu** | 6 | Ex Fa Gd NA Po TA |
| **GarageType** | 6 | 2Types Attchd Basment BuiltIn CarPort Detchd |
| **GarageFinish** | 3 | Fin RFn Unf |
| **GarageQual** | 5 | Ex Fa Gd Po TA |
| **GarageCond** | 5 | Ex Fa Gd Po TA |
| **PavedDrive** | 3 | N P Y |
| **PoolQC** | 4 | Ex Fa Gd NA |
| **Fence** | 5 | GdPrv GdWo MnPrv MnWw NA |
| **MiscFeature** | 4 | NA Othr Shed TenC |
| **SaleType** | 6 | COD CWD Con New Oth WD |
| **SaleCondition** | 6 | Abnorml AdjLand Alloca Family Normal Partial |

| **Dimensions** | |
| --- | --- |
| **Number of Effects** | 78 |
| **Number of Parameters** | 285 |

**The GLMSELECT Procedure**

| **Backward Selection Summary** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Effect Removed** | **Number Effects In** | **Number Parms In** | **Adjusted R-Square** | **SBC** | **CV PRESS** |
| **\* Optimal Value of Criterion** | | | | | | |
| **0** |  | 78 | 238 | 0.9355\* | -3741.3243 | 113.4878 |
|  | **Utilities** | 77 | 238 | 0.9355\* | -3741.3243 | 113.4878 |
| **1** | **Exterior2nd** | 76 | 224 | 0.9350 | -3813.1678 | 90.8970 |
| **2** | **Exterior1st** | 75 | 211 | 0.9341 | -3873.6103 | 62.2042 |
| **3** | **HouseStyle** | 74 | 204 | 0.9342 | -3915.2076\* | 54.2512\* |

**Note:**Effects dropped at step 0 are redundant.

|  |
| --- |
| Selection stopped at a local minimum of the cross validation PRESS. |

| **Stop Details** | | | | |
| --- | --- | --- | --- | --- |
| **Candidate For** | **Effect** | **Candidate CV PRESS** |  | **Compare CV PRESS** |
| **Removal** | BsmtFinType2 | 98.5113 | > | 54.2512 |

**The GLMSELECT Procedure**

**Selected Model**

**The selected model is the model at the last step (Step 3).**

|  |  |
| --- | --- |
| **Effects:** | Intercept IntLotFrontage MSSubClass MSZoning LogLotArea Street Alley LotShape LandContour LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl MasVnrType LogMasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 LogBsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF Heating HeatingQC CentralAir Electrical Log1stFlrSF 2ndFlrSF LowQualFinSF BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars LogGarageArea GarageQual GarageCond PavedDrive LogWoodDeckSF LogOpenPorchSF LogEnclosedPorch Log3SsnPorch LogScreenPorch LogPoolArea PoolQC Fence MiscFeature LogMiscVal MoSold YrSold SaleType SaleCondition |

| **Analysis of Variance** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** |
| **Model** | 203 | 166.82282 | 0.82179 | 79.31 |
| **Error** | 917 | 9.50209 | 0.01036 |  |
| **Corrected Total** | 1120 | 176.32491 |  |  |

|  |  |
| --- | --- |
| **Root MSE** | 0.10179 |
| **Dependent Mean** | 12.04841 |
| **R-Square** | 0.9461 |
| **Adj R-Sq** | 0.9342 |
| **AIC** | -3816.69074 |
| **AICC** | -3724.38473 |
| **SBC** | -3915.20755 |
| **CV PRESS** | 54.25118 |

| **Parameter Estimates** | | | | |
| --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **t Value** |
| **Intercept** | 1 | 3.638448 | 5.489206 | 0.66 |
| **IntLotFrontage** | 1 | -0.000231 | 0.000180 | -1.29 |
| **MSSubClass** | 1 | -0.000139 | 0.000271 | -0.51 |
| **MSZoning C** | 1 | -0.363794 | 0.049672 | -7.32 |
| **MSZoning FV** | 1 | 0.089339 | 0.036685 | 2.44 |
| **MSZoning RH** | 1 | 0.082802 | 0.041817 | 1.98 |
| **MSZoning RL** | 1 | 0.042371 | 0.019576 | 2.16 |
| **MSZoning RM** | 0 | 0 | . | . |
| **LogLotArea** | 1 | 0.073297 | 0.014811 | 4.95 |
| **Street Grvl** | 1 | -0.104974 | 0.081294 | -1.29 |
| **Street Pave** | 0 | 0 | . | . |
| **Alley Gr** | 1 | -0.030194 | 0.031107 | -0.97 |
| **Alley NA** | 1 | -0.023189 | 0.024085 | -0.96 |
| **Alley Pa** | 0 | 0 | . | . |
| **LotShape IR1** | 1 | -0.005405 | 0.008779 | -0.62 |
| **LotShape IR2** | 1 | -0.002317 | 0.025241 | -0.09 |
| **LotShape IR3** | 1 | 0.104030 | 0.051881 | 2.01 |
| **LotShape Reg** | 0 | 0 | . | . |
| **LandContour Bnk** | 1 | -0.016821 | 0.019728 | -0.85 |
| **LandContour HLS** | 1 | 0.002394 | 0.020205 | 0.12 |
| **LandContour Low** | 1 | -0.019266 | 0.035364 | -0.54 |
| **LandContour Lvl** | 0 | 0 | . | . |
| **LotConfig Corner** | 1 | 0.011597 | 0.009262 | 1.25 |
| **LotConfig CulDSac** | 1 | 0.038674 | 0.018839 | 2.05 |
| **LotConfig FR2** | 1 | -0.034966 | 0.020854 | -1.68 |
| **LotConfig FR3** | 1 | -0.058528 | 0.057155 | -1.02 |
| **LotConfig Inside** | 0 | 0 | . | . |
| **LandSlope Gtl** | 1 | 0.162705 | 0.062692 | 2.60 |
| **LandSlope Mod** | 1 | 0.194058 | 0.063807 | 3.04 |
| **LandSlope Sev** | 0 | 0 | . | . |
| **Neighborhood Blmngtn** | 1 | -0.038976 | 0.055737 | -0.70 |
| **Neighborhood Blueste** | 1 | 0.048870 | 0.094765 | 0.52 |
| **Neighborhood BrDale** | 1 | 0.015966 | 0.061695 | 0.26 |
| **Neighborhood BrkSide** | 1 | 0.017606 | 0.052995 | 0.33 |
| **Neighborhood ClearCr** | 1 | -0.041058 | 0.055257 | -0.74 |
| **Neighborhood CollgCr** | 1 | -0.068908 | 0.045604 | -1.51 |
| **Neighborhood Crawfor** | 1 | 0.082490 | 0.049230 | 1.68 |
| **Neighborhood Edwards** | 1 | -0.131454 | 0.047121 | -2.79 |
| **Neighborhood Gilbert** | 1 | -0.075630 | 0.047938 | -1.58 |
| **Neighborhood IDOTRR** | 1 | 0.016598 | 0.059375 | 0.28 |
| **Neighborhood MeadowV** | 1 | -0.111492 | 0.061584 | -1.81 |
| **Neighborhood Mitchel** | 1 | -0.091530 | 0.048559 | -1.88 |
| **Neighborhood NAmes** | 1 | -0.071000 | 0.045743 | -1.55 |
| **Neighborhood NPkVill** | 1 | -0.039215 | 0.063236 | -0.62 |
| **Neighborhood NWAmes** | 1 | -0.105459 | 0.047771 | -2.21 |
| **Neighborhood NoRidge** | 1 | 0.009716 | 0.048844 | 0.20 |
| **Neighborhood NridgHt** | 1 | 0.029834 | 0.047341 | 0.63 |
| **Neighborhood OldTown** | 1 | -0.042704 | 0.053329 | -0.80 |
| **Neighborhood SWISU** | 1 | -0.035793 | 0.055234 | -0.65 |
| **Neighborhood Sawyer** | 1 | -0.074587 | 0.047434 | -1.57 |
| **Neighborhood SawyerW** | 1 | -0.070324 | 0.046769 | -1.50 |
| **Neighborhood Somerst** | 1 | -0.028006 | 0.051304 | -0.55 |
| **Neighborhood StoneBr** | 1 | 0.085158 | 0.050105 | 1.70 |
| **Neighborhood Timber** | 1 | -0.052101 | 0.048589 | -1.07 |
| **Neighborhood Veenker** | 0 | 0 | . | . |
| **Condition1 Artery** | 1 | -0.120588 | 0.066915 | -1.80 |
| **Condition1 Feedr** | 1 | -0.083044 | 0.065658 | -1.26 |
| **Condition1 Norm** | 1 | -0.042370 | 0.063864 | -0.66 |
| **Condition1 PosA** | 1 | -0.031439 | 0.089463 | -0.35 |
| **Condition1 PosN** | 1 | -0.005419 | 0.077645 | -0.07 |
| **Condition1 RRAe** | 1 | -0.142894 | 0.076861 | -1.86 |
| **Condition1 RRAn** | 1 | -0.058972 | 0.065657 | -0.90 |
| **Condition1 RRNe** | 1 | -0.058083 | 0.122582 | -0.47 |
| **Condition1 RRNn** | 0 | 0 | . | . |
| **Condition2 Artery** | 1 | -0.012564 | 0.126690 | -0.10 |
| **Condition2 Feedr** | 1 | -0.018903 | 0.099750 | -0.19 |
| **Condition2 Norm** | 1 | 0.043119 | 0.082313 | 0.52 |
| **Condition2 PosA** | 1 | 0.228834 | 0.183816 | 1.24 |
| **Condition2 PosN** | 1 | -0.689906 | 0.121804 | -5.66 |
| **Condition2 RRNn** | 0 | 0 | . | . |
| **BldgType 1Fam** | 1 | 0.026622 | 0.033047 | 0.81 |
| **BldgType 2fmCon** | 1 | 0.042170 | 0.041585 | 1.01 |
| **BldgType Duplex** | 1 | -0.018980 | 0.043540 | -0.44 |
| **BldgType Twnhs** | 1 | -0.024079 | 0.025305 | -0.95 |
| **BldgType TwnhsE** | 0 | 0 | . | . |
| **OverallQual** | 1 | 0.044027 | 0.005294 | 8.32 |
| **OverallCond** | 1 | 0.040149 | 0.004891 | 8.21 |
| **YearBuilt** | 1 | 0.001907 | 0.000409 | 4.66 |
| **YearRemodAdd** | 1 | 0.000586 | 0.000295 | 1.98 |
| **RoofStyle Flat** | 1 | -0.186115 | 0.139418 | -1.33 |
| **RoofStyle Gable** | 1 | -0.085519 | 0.058796 | -1.45 |
| **RoofStyle Gambr** | 1 | -0.055062 | 0.070190 | -0.78 |
| **RoofStyle Hip** | 1 | -0.083502 | 0.059053 | -1.41 |
| **RoofStyle Mansa** | 0 | 0 | . | . |
| **RoofMatl ClyTile** | 1 | -1.749226 | 0.420396 | -4.16 |
| **RoofMatl CompShg** | 1 | -0.085917 | 0.053156 | -1.62 |
| **RoofMatl Membran** | 1 | 0.272881 | 0.187920 | 1.45 |
| **RoofMatl Roll** | 1 | -0.026816 | 0.124628 | -0.22 |
| **RoofMatl Tar&Grv** | 1 | -0.101021 | 0.122420 | -0.83 |
| **RoofMatl WdShake** | 1 | -0.261501 | 0.113863 | -2.30 |
| **RoofMatl WdShngl** | 0 | 0 | . | . |
| **MasVnrType BrkCmn** | 1 | -0.081327 | 0.040034 | -2.03 |
| **MasVnrType BrkFace** | 1 | -0.017529 | 0.013884 | -1.26 |
| **MasVnrType None** | 1 | -0.022256 | 0.033506 | -0.66 |
| **MasVnrType Stone** | 0 | 0 | . | . |
| **LogMasVnrArea** | 1 | -0.001304 | 0.006204 | -0.21 |
| **ExterQual Ex** | 1 | 0.022336 | 0.026661 | 0.84 |
| **ExterQual Fa** | 1 | -0.002942 | 0.057697 | -0.05 |
| **ExterQual Gd** | 1 | 0.003759 | 0.012884 | 0.29 |
| **ExterQual TA** | 0 | 0 | . | . |
| **ExterCond Ex** | 1 | -0.010044 | 0.114281 | -0.09 |
| **ExterCond Fa** | 1 | -0.040625 | 0.030393 | -1.34 |
| **ExterCond Gd** | 1 | -0.022517 | 0.012737 | -1.77 |
| **ExterCond TA** | 0 | 0 | . | . |
| **Foundation BrkTil** | 1 | 0.229864 | 0.082227 | 2.80 |
| **Foundation CBlock** | 1 | 0.261097 | 0.081086 | 3.22 |
| **Foundation PConc** | 1 | 0.260480 | 0.080440 | 3.24 |
| **Foundation Slab** | 1 | 0.307087 | 0.096005 | 3.20 |
| **Foundation Stone** | 1 | 0.303611 | 0.095519 | 3.18 |
| **Foundation Wood** | 0 | 0 | . | . |
| **BsmtQual Ex** | 1 | 0.054654 | 0.020534 | 2.66 |
| **BsmtQual Fa** | 1 | 0.006872 | 0.023613 | 0.29 |
| **BsmtQual Gd** | 1 | 0.020876 | 0.013776 | 1.52 |
| **BsmtQual NA** | 1 | 0.013270 | 0.161787 | 0.08 |
| **BsmtQual TA** | 0 | 0 | . | . |
| **BsmtCond Fa** | 1 | -0.024310 | 0.021311 | -1.14 |
| **BsmtCond Gd** | 1 | -0.005121 | 0.016417 | -0.31 |
| **BsmtCond NA** | 0 | 0 | . | . |
| **BsmtCond Po** | 1 | -0.140531 | 0.176863 | -0.79 |
| **BsmtCond TA** | 0 | 0 | . | . |
| **BsmtExposure Av** | 1 | 0.003930 | 0.011005 | 0.36 |
| **BsmtExposure Gd** | 1 | 0.056599 | 0.015312 | 3.70 |
| **BsmtExposure Mn** | 1 | -0.001748 | 0.012855 | -0.14 |
| **BsmtExposure NA** | 1 | -0.043053 | 0.103885 | -0.41 |
| **BsmtExposure No** | 0 | 0 | . | . |
| **BsmtFinType1 ALQ** | 1 | -0.107786 | 0.041240 | -2.61 |
| **BsmtFinType1 BLQ** | 1 | -0.109806 | 0.042014 | -2.61 |
| **BsmtFinType1 GLQ** | 1 | -0.090393 | 0.038578 | -2.34 |
| **BsmtFinType1 LwQ** | 1 | -0.128291 | 0.043649 | -2.94 |
| **BsmtFinType1 NA** | 0 | 0 | . | . |
| **BsmtFinType1 Rec** | 1 | -0.112851 | 0.042772 | -2.64 |
| **BsmtFinType1 Unf** | 0 | 0 | . | . |
| **LogBsmtFinSF1** | 1 | 0.029535 | 0.007817 | 3.78 |
| **BsmtFinType2 ALQ** | 1 | -0.034792 | 0.041170 | -0.85 |
| **BsmtFinType2 BLQ** | 1 | -0.050318 | 0.026057 | -1.93 |
| **BsmtFinType2 GLQ** | 1 | -0.017188 | 0.045000 | -0.38 |
| **BsmtFinType2 LwQ** | 1 | -0.039668 | 0.024592 | -1.61 |
| **BsmtFinType2 NA** | 1 | -0.021893 | 0.111296 | -0.20 |
| **BsmtFinType2 Rec** | 1 | -0.030538 | 0.026407 | -1.16 |
| **BsmtFinType2 Unf** | 0 | 0 | . | . |
| **BsmtFinSF2** | 1 | 0.000122 | 0.000045227 | 2.71 |
| **BsmtUnfSF** | 1 | 0.000031933 | 0.000020717 | 1.54 |
| **Heating GasA** | 1 | -0.050248 | 0.075515 | -0.67 |
| **Heating GasW** | 1 | 0.048846 | 0.078600 | 0.62 |
| **Heating Grav** | 1 | -0.143928 | 0.122248 | -1.18 |
| **Heating OthW** | 1 | -0.096249 | 0.133496 | -0.72 |
| **Heating Wall** | 0 | 0 | . | . |
| **HeatingQC Ex** | 1 | 0.035009 | 0.010804 | 3.24 |
| **HeatingQC Fa** | 1 | 0.037581 | 0.024420 | 1.54 |
| **HeatingQC Gd** | 1 | 0.011937 | 0.011326 | 1.05 |
| **HeatingQC Po** | 1 | -0.089259 | 0.120914 | -0.74 |
| **HeatingQC TA** | 0 | 0 | . | . |
| **CentralAir N** | 1 | -0.092112 | 0.020777 | -4.43 |
| **CentralAir Y** | 0 | 0 | . | . |
| **Electrical FuseA** | 1 | 0.001597 | 0.015285 | 0.10 |
| **Electrical FuseF** | 1 | 0.019199 | 0.029504 | 0.65 |
| **Electrical FuseP** | 1 | -0.119960 | 0.110790 | -1.08 |
| **Electrical Mix** | 0 | 0 | . | . |
| **Electrical NA** | 1 | 0.120597 | 0.107686 | 1.12 |
| **Electrical SBrkr** | 0 | 0 | . | . |
| **Log1stFlrSF** | 1 | 0.344678 | 0.028893 | 11.93 |
| **2ndFlrSF** | 1 | 0.000209 | 0.000021459 | 9.72 |
| **LowQualFinSF** | 1 | -0.000018252 | 0.000104 | -0.18 |
| **BsmtFullBath** | 1 | 0.022313 | 0.010116 | 2.21 |
| **BsmtHalfBath** | 1 | 0.004788 | 0.015744 | 0.30 |
| **FullBath** | 1 | 0.030695 | 0.011713 | 2.62 |
| **HalfBath** | 1 | 0.028019 | 0.010802 | 2.59 |
| **BedroomAbvGr** | 1 | -0.003198 | 0.007248 | -0.44 |
| **KitchenAbvGr** | 1 | -0.064513 | 0.033929 | -1.90 |
| **KitchenQual Ex** | 1 | 0.061194 | 0.019749 | 3.10 |
| **KitchenQual Fa** | 1 | 0.007549 | 0.026288 | 0.29 |
| **KitchenQual Gd** | 1 | -0.006787 | 0.011606 | -0.58 |
| **KitchenQual TA** | 0 | 0 | . | . |
| **TotRmsAbvGrd** | 1 | 0.005301 | 0.004912 | 1.08 |
| **Functional Maj1** | 1 | -0.061635 | 0.037687 | -1.64 |
| **Functional Maj2** | 1 | -0.203220 | 0.065626 | -3.10 |
| **Functional Min1** | 1 | -0.059856 | 0.024635 | -2.43 |
| **Functional Min2** | 1 | -0.054473 | 0.023437 | -2.32 |
| **Functional Mod** | 1 | -0.102736 | 0.036201 | -2.84 |
| **Functional Typ** | 0 | 0 | . | . |
| **Fireplaces** | 1 | 0.022887 | 0.013773 | 1.66 |
| **FireplaceQu Ex** | 1 | 0.026033 | 0.027367 | 0.95 |
| **FireplaceQu Fa** | 1 | 0.004146 | 0.024105 | 0.17 |
| **FireplaceQu Gd** | 1 | 0.020437 | 0.011893 | 1.72 |
| **FireplaceQu NA** | 1 | 0.019084 | 0.019197 | 0.99 |
| **FireplaceQu Po** | 1 | -0.023480 | 0.033817 | -0.69 |
| **FireplaceQu TA** | 0 | 0 | . | . |
| **GarageType 2Types** | 1 | -0.062054 | 0.056092 | -1.11 |
| **GarageType Attchd** | 1 | -0.010666 | 0.011695 | -0.91 |
| **GarageType Basment** | 1 | -0.019466 | 0.034412 | -0.57 |
| **GarageType BuiltIn** | 1 | -0.035576 | 0.019883 | -1.79 |
| **GarageType CarPort** | 1 | -0.053959 | 0.046977 | -1.15 |
| **GarageType Detchd** | 0 | 0 | . | . |
| **GarageYrBlt** | 1 | -0.000028731 | 0.000317 | -0.09 |
| **GarageFinish Fin** | 1 | 0.016274 | 0.012488 | 1.30 |
| **GarageFinish RFn** | 1 | 0.014855 | 0.011130 | 1.33 |
| **GarageFinish Unf** | 0 | 0 | . | . |
| **GarageCars** | 1 | 0.033126 | 0.011809 | 2.81 |
| **LogGarageArea** | 1 | 0.042170 | 0.020628 | 2.04 |
| **GarageQual Ex** | 1 | 0.351354 | 0.141652 | 2.48 |
| **GarageQual Fa** | 1 | -0.020751 | 0.022772 | -0.91 |
| **GarageQual Gd** | 1 | 0.024613 | 0.038207 | 0.64 |
| **GarageQual Po** | 1 | -0.021328 | 0.134131 | -0.16 |
| **GarageQual TA** | 0 | 0 | . | . |
| **GarageCond Ex** | 1 | -0.347378 | 0.162528 | -2.14 |
| **GarageCond Fa** | 1 | -0.033402 | 0.026039 | -1.28 |
| **GarageCond Gd** | 1 | 0.005137 | 0.053207 | 0.10 |
| **GarageCond Po** | 1 | 0.098331 | 0.071266 | 1.38 |
| **GarageCond TA** | 0 | 0 | . | . |
| **PavedDrive N** | 1 | -0.011619 | 0.019782 | -0.59 |
| **PavedDrive P** | 1 | -0.036717 | 0.026513 | -1.38 |
| **PavedDrive Y** | 0 | 0 | . | . |
| **LogWoodDeckSF** | 1 | 0.003761 | 0.001526 | 2.47 |
| **LogOpenPorchSF** | 1 | 0.003836 | 0.002012 | 1.91 |
| **LogEnclosedPorch** | 1 | 0.004597 | 0.002362 | 1.95 |
| **Log3SsnPorch** | 1 | 0.010776 | 0.005059 | 2.13 |
| **LogScreenPorch** | 1 | 0.007850 | 0.002453 | 3.20 |
| **LogPoolArea** | 1 | 1.003531 | 2.085809 | 0.48 |
| **PoolQC Ex** | 1 | -6.132658 | 13.112358 | -0.47 |
| **PoolQC Fa** | 1 | -6.428927 | 13.511355 | -0.48 |
| **PoolQC Gd** | 1 | -6.185390 | 13.261629 | -0.47 |
| **PoolQC NA** | 0 | 0 | . | . |
| **Fence GdPrv** | 1 | -0.017311 | 0.018717 | -0.92 |
| **Fence GdWo** | 1 | -0.011866 | 0.018801 | -0.63 |
| **Fence MnPrv** | 1 | -0.004832 | 0.012002 | -0.40 |
| **Fence MnWw** | 1 | -0.031468 | 0.036210 | -0.87 |
| **Fence NA** | 0 | 0 | . | . |
| **MiscFeature NA** | 1 | 0.396031 | 0.558417 | 0.71 |
| **MiscFeature Othr** | 1 | 0.400160 | 0.573758 | 0.70 |
| **MiscFeature Shed** | 1 | 0.010316 | 0.494370 | 0.02 |
| **MiscFeature TenC** | 0 | 0 | . | . |
| **LogMiscVal** | 1 | 0.059647 | 0.034786 | 1.71 |
| **MoSold** | 1 | -0.000850 | 0.001279 | -0.66 |
| **YrSold** | 1 | -0.000584 | 0.002677 | -0.22 |
| **SaleType COD** | 1 | 0.008781 | 0.021391 | 0.41 |
| **SaleType CWD** | 1 | 0.072266 | 0.055441 | 1.30 |
| **SaleType Con** | 1 | 0.084790 | 0.030512 | 2.78 |
| **SaleType New** | 1 | 0.065862 | 0.069659 | 0.95 |
| **SaleType Oth** | 1 | 0.566736 | 0.247585 | 2.29 |
| **SaleType WD** | 0 | 0 | . | . |
| **SaleCondition Abnorml** | 1 | -0.046818 | 0.069727 | -0.67 |
| **SaleCondition AdjLand** | 1 | 0.239192 | 0.132941 | 1.80 |
| **SaleCondition Alloca** | 1 | 0.014814 | 0.081666 | 0.18 |
| **SaleCondition Family** | 1 | -0.027156 | 0.073347 | -0.37 |
| **SaleCondition Normal** | 1 | 0.022716 | 0.068740 | 0.33 |
| **SaleCondition Partial** | 0 | 0 | . | . |

Table 2G:

**The GLMSELECT Procedure**

|  |  |
| --- | --- |
| **Data Set** | WORK.TRAIN\_DUMMY |
| **Dependent Variable** | LogSalePrice |
| **Selection Method** | Stepwise |
| **Select Criterion** | SBC |
| **Stop Criterion** | Cross Validation |
| **Cross Validation Method** | Random |
| **Cross Validation Fold** | 5 |
| **Effect Hierarchy Enforced** | None |
| **Random Number Seed** | 590572116 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 1460 |
| **Number of Observations Used** | 1121 |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **MSZoning** | 5 | C FV RH RL RM |
| **Street** | 2 | Grvl Pave |
| **Alley** | 3 | Gr NA Pa |
| **LotShape** | 4 | IR1 IR2 IR3 Reg |
| **LandContour** | 4 | Bnk HLS Low Lvl |
| **Utilities** | 1 | AllPub |
| **LotConfig** | 5 | Corner CulDSac FR2 FR3 Inside |
| **LandSlope** | 3 | Gtl Mod Sev |
| **Neighborhood** | 25 | Blmngtn Blueste BrDale BrkSide ClearCr CollgCr Crawfor Edwards Gilbert IDOTRR MeadowV Mitchel NAmes NPkVill NWAmes NoRidge NridgHt OldTown SWISU Sawyer SawyerW Somerst StoneBr Timber Veenker |
| **Condition1** | 9 | Artery Feedr Norm PosA PosN RRAe RRAn RRNe RRNn |
| **Condition2** | 6 | Artery Feedr Norm PosA PosN RRNn |
| **BldgType** | 5 | 1Fam 2fmCon Duplex Twnhs TwnhsE |
| **HouseStyle** | 8 | 1.5Fin 1.5Unf 1Story 2.5Fin 2.5Unf 2Story SFoyer SLvl |
| **RoofStyle** | 5 | Flat Gable Gambr Hip Mansa |
| **RoofMatl** | 7 | ClyTile CompShg Membran Roll Tar&Grv WdShake WdShngl |
| **Exterior1st** | 14 | AsbShng BrkComm BrkFace CBlock CemntBd HdBoard ImStucc MetalSd Plywood Stone Stucco VinylSd Wd Sdng WdShing |
| **Exterior2nd** | 16 | AsbShng AsphShn Brk Cmn BrkFace CBlock CmentBd HdBoard ImStucc MetalSd Other Plywood Stone Stucco VinylSd Wd Sdng Wd Shng |
| **MasVnrType** | 4 | BrkCmn BrkFace None Stone |
| **ExterQual** | 4 | Ex Fa Gd TA |
| **ExterCond** | 4 | Ex Fa Gd TA |
| **Foundation** | 6 | BrkTil CBlock PConc Slab Stone Wood |
| **BsmtQual** | 5 | Ex Fa Gd NA TA |
| **BsmtCond** | 5 | Fa Gd NA Po TA |
| **BsmtExposure** | 5 | Av Gd Mn NA No |
| **BsmtFinType1** | 7 | ALQ BLQ GLQ LwQ NA Rec Unf |
| **BsmtFinType2** | 7 | ALQ BLQ GLQ LwQ NA Rec Unf |
| **Heating** | 5 | GasA GasW Grav OthW Wall |
| **HeatingQC** | 5 | Ex Fa Gd Po TA |
| **CentralAir** | 2 | N Y |
| **Electrical** | 6 | FuseA FuseF FuseP Mix NA SBrkr |
| **KitchenQual** | 4 | Ex Fa Gd TA |
| **Functional** | 6 | Maj1 Maj2 Min1 Min2 Mod Typ |
| **FireplaceQu** | 6 | Ex Fa Gd NA Po TA |
| **GarageType** | 6 | 2Types Attchd Basment BuiltIn CarPort Detchd |
| **GarageFinish** | 3 | Fin RFn Unf |
| **GarageQual** | 5 | Ex Fa Gd Po TA |
| **GarageCond** | 5 | Ex Fa Gd Po TA |
| **PavedDrive** | 3 | N P Y |
| **PoolQC** | 4 | Ex Fa Gd NA |
| **Fence** | 5 | GdPrv GdWo MnPrv MnWw NA |
| **MiscFeature** | 4 | NA Othr Shed TenC |
| **SaleType** | 6 | COD CWD Con New Oth WD |
| **SaleCondition** | 6 | Abnorml AdjLand Alloca Family Normal Partial |

| **Dimensions** | |
| --- | --- |
| **Number of Effects** | 78 |
| **Number of Parameters** | 285 |

**The GLMSELECT Procedure**

| **Stepwise Selection Summary** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Effect Entered** | **Effect Removed** | **Number Effects In** | **Number Parms In** | **Adjusted R-Square** | **SBC** | **CV PRESS** |
| **\* Optimal Value of Criterion** | | | | | | | |
| **0** | **Intercept** |  | 1 | 1 | 0.0000 | -2066.4335 | 176.5015 |
| **1** | **OverallQual** |  | 2 | 2 | 0.6848 | -3354.5767 | 55.7969 |
| **2** | **Log1stFlrSF** |  | 3 | 3 | 0.7407 | -3567.6573 | 45.9899 |
| **3** | **2ndFlrSF** |  | 4 | 4 | 0.7812 | -3751.9351 | 38.9924 |
| **4** | **YearBuilt** |  | 5 | 5 | 0.8140 | -3928.0564 | 33.2951 |
| **5** | **OverallCond** |  | 6 | 6 | 0.8375 | -4073.1053 | 29.2610 |
| **6** | **LogLotArea** |  | 7 | 7 | 0.8488\* | -4147.7963\* | 27.2617\* |

|  |
| --- |
| Selection stopped at a local minimum of the cross validation PRESS. |

| **Stop Details** | | | | |
| --- | --- | --- | --- | --- |
| **Candidate For** | **Effect** | **Candidate CV PRESS** |  | **Compare CV PRESS** |
| **Entry** | RoofMatl | 27.5198 | > | 27.2617 |
| **Removal** | LogLotArea | 29.2610 | > | 27.2617 |

**The GLMSELECT Procedure**

**Selected Model**

**The selected model is the model at the last step (Step 6).**

|  |  |
| --- | --- |
| **Effects:** | Intercept LogLotArea OverallQual OverallCond YearBuilt Log1stFlrSF 2ndFlrSF |

| **Analysis of Variance** | | | | |
| --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** |
| **Model** | 6 | 149.80112 | 24.96685 | 1048.61 |
| **Error** | 1114 | 26.52380 | 0.02381 |  |
| **Corrected Total** | 1120 | 176.32491 |  |  |

|  |  |
| --- | --- |
| **Root MSE** | 0.15430 |
| **Dependent Mean** | 12.04841 |
| **R-Square** | 0.8496 |
| **Adj R-Sq** | 0.8488 |
| **AIC** | -3059.95014 |
| **AICC** | -3059.82064 |
| **SBC** | -4147.79630 |
| **CV PRESS** | 27.26170 |

| **Parameter Estimates** | | | | |
| --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **t Value** |
| **Intercept** | 1 | -1.550928 | 0.446433 | -3.47 |
| **LogLotArea** | 1 | 0.101900 | 0.011103 | 9.18 |
| **OverallQual** | 1 | 0.110904 | 0.005312 | 20.88 |
| **OverallCond** | 1 | 0.063287 | 0.004850 | 13.05 |
| **YearBuilt** | 1 | 0.004269 | 0.000210 | 20.32 |
| **Log1stFlrSF** | 1 | 0.445317 | 0.021389 | 20.82 |
| **2ndFlrSF** | 1 | 0.000245 | 0.000012953 | 18.89 |

**Question 2 SAS Code:**

\*import training data for SalePrice analysis;

proc import datafile="/home/bcoari0/sasuser.v94/train.csv"

out=train dbms=csv replace;

getnames=yes;

run;

\*Creating Dummy Variables and performing log transforms, adding 1 to the base before the log transform to account for 0 values if necessary;

Data train\_dummy;

set train;

LogSalePrice = Log(SalePrice);

IntLotFrontage = int(LotFrontage);

LogLotArea = Log(LotArea);

LogMasVnrArea = Log(MasVnrArea+1);

LogBsmtFinSF1 = Log(BsmtFinSF1+1);

Log1stFlrSF = Log('1stFlrSF'n);

LogPoolArea = Log(PoolArea+1);

LogMiscVal = Log(MiscVal+1);

LogGarageArea = Log(GarageArea);

LogWoodDeckSF = Log(WoodDeckSF+1);

LogOpenPorchSF = Log(OpenPorchSF+1);

LogEnclosedPorch = Log(EnclosedPorch+1);

Log3SsnPorch = Log('3SsnPorch'n+1);

LogScreenPorch = Log(ScreenPorch+1);

Alley\_dum0 = 0;

Alley\_dum1 = 0;

LotShape\_dum0 = 0;

LotShape\_dum1 = 0;

LotShape\_dum2 = 0;

LandContour\_dum0 = 0;

LandContour\_dum1 = 0;

LandContour\_dum2 = 0;

Utilities\_dum0 = 0;

LotConfig\_dum0 = 0;

LotConfig\_dum1 = 0;

LotConfig\_dum2 = 0;

LotConfig\_dum3 = 0;

LandSlope\_dum0 = 0;

LandSlope\_dum1 = 0;

Neighborhood\_dum0 = 0;

Neighborhood\_dum1 = 0;

Neighborhood\_dum2 = 0;

Neighborhood\_dum3 = 0;

Neighborhood\_dum4 = 0;

Neighborhood\_dum5 = 0;

Neighborhood\_dum6 = 0;

Neighborhood\_dum7 = 0;

Neighborhood\_dum8 = 0;

Neighborhood\_dum9 = 0;

Neighborhood\_dum10 = 0;

Neighborhood\_dum11 = 0;

Neighborhood\_dum12 = 0;

Neighborhood\_dum13 = 0;

Neighborhood\_dum14 = 0;

Neighborhood\_dum15 = 0;

Neighborhood\_dum16 = 0;

Neighborhood\_dum17 = 0;

Neighborhood\_dum18 = 0;

Neighborhood\_dum19 = 0;

Neighborhood\_dum20 = 0;

Neighborhood\_dum21 = 0;

Neighborhood\_dum22 = 0;

Neighborhood\_dum23 = 0;

Condition1\_dum0 = 0;

Condition1\_dum1 = 0;

Condition1\_dum2 = 0;

Condition1\_dum3 = 0;

Condition1\_dum4 = 0;

Condition1\_dum5 = 0;

Condition1\_dum6 = 0;

Condition1\_dum7 = 0;

Condition2\_dum0 = 0;

Condition2\_dum1 = 0;

Condition2\_dum2 = 0;

Condition2\_dum3 = 0;

Condition2\_dum4 = 0;

Condition2\_dum5 = 0;

Condition2\_dum6 = 0;

BldgType\_dum0 = 0;

BldgType\_dum1 = 0;

BldgType\_dum2 = 0;

BldgType\_dum3 = 0;

HouseStyle\_dum0 = 0;

HouseStyle\_dum1 = 0;

HouseStyle\_dum2 = 0;

HouseStyle\_dum3 = 0;

HouseStyle\_dum4 = 0;

HouseStyle\_dum5 = 0;

HouseStyle\_dum6 = 0;

RoofStyle\_dum0 = 0;

RoofStyle\_dum1 = 0;

RoofStyle\_dum2 = 0;

RoofStyle\_dum3 = 0;

RoofStyle\_dum4 = 0;

RoofMatl\_dum0 = 0;

RoofMatl\_dum1 = 0;

RoofMatl\_dum2 = 0;

RoofMatl\_dum3 = 0;

RoofMatl\_dum4 = 0;

RoofMatl\_dum5 = 0;

RoofMatl\_dum6 = 0;

Exterior1st\_dum0 = 0;

Exterior1st\_dum1 = 0;

Exterior1st\_dum2 = 0;

Exterior1st\_dum3 = 0;

Exterior1st\_dum4 = 0;

Exterior1st\_dum5 = 0;

Exterior1st\_dum6 = 0;

Exterior1st\_dum7 = 0;

Exterior1st\_dum8 = 0;

Exterior1st\_dum9 = 0;

Exterior1st\_dum10 = 0;

Exterior1st\_dum11 = 0;

Exterior1st\_dum12 = 0;

Exterior2nd\_dum0 = 0;

Exterior2nd\_dum1 = 0;

Exterior2nd\_dum2 = 0;

Exterior2nd\_dum3 = 0;

Exterior2nd\_dum4 = 0;

Exterior2nd\_dum5 = 0;

Exterior2nd\_dum6 = 0;

Exterior2nd\_dum7 = 0;

Exterior2nd\_dum8 = 0;

Exterior2nd\_dum9 = 0;

Exterior2nd\_dum10 = 0;

Exterior2nd\_dum11 = 0;

Exterior2nd\_dum12 = 0;

Exterior2nd\_dum13 = 0;

MasVnrType\_dum0 = 0;

MasVnrType\_dum1 = 0;

MasVnrType\_dum2 = 0;

ExterQual\_dum0 = 0;

ExterQual\_dum1 = 0;

ExterQual\_dum2 = 0;

ExterCond\_dum0 = 0;

ExterCond\_dum1 = 0;

ExterCond\_dum2 = 0;

Foundation\_dum0 = 0;

Foundation\_dum1 = 0;

Foundation\_dum2 = 0;

Foundation\_dum3 = 0;

Foundation\_dum4 = 0;

BsmtQual\_dum0 = 0;

BsmtQual\_dum1 = 0;

BsmtQual\_dum2 = 0;

BsmtQual\_dum3 = 0;

BsmtCond\_dum0 = 0;

BsmtCond\_dum1 = 0;

BsmtCond\_dum2 = 0;

BsmtExposure\_dum0 = 0;

BsmtExposure\_dum1 = 0;

BsmtExposure\_dum2 = 0;

BsmtExposure\_dum3 = 0;

BsmtFinType1\_dum0 = 0;

BsmtFinType1\_dum1 = 0;

BsmtFinType1\_dum2 = 0;

BsmtFinType1\_dum3 = 0;

BsmtFinType1\_dum4 = 0;

BsmtFinType1\_dum5 = 0;

BsmtFinType2\_dum0 = 0;

BsmtFinType2\_dum1 = 0;

BsmtFinType2\_dum2 = 0;

BsmtFinType2\_dum3 = 0;

BsmtFinType2\_dum4 = 0;

Heating\_dum0 = 0;

Heating\_dum1 = 0;

Heating\_dum2 = 0;

Heating\_dum3 = 0;

Heating\_dum4 = 0;

HeatingQC\_dum0 = 0;

HeatingQC\_dum1 = 0;

HeatingQC\_dum2 = 0;

HeatingQC\_dum3 = 0;

CentralAir\_dum0 = 0;

Electrical\_dum0 = 0;

Electrical\_dum1 = 0;

Electrical\_dum2 = 0;

Electrical\_dum3 = 0;

KitchenQual\_dum0 = 0;

KitchenQual\_dum1 = 0;

KitchenQual\_dum2 = 0;

Functional\_dum0 = 0;

Functional\_dum1 = 0;

Functional\_dum2 = 0;

Functional\_dum3 = 0;

Functional\_dum4 = 0;

Functional\_dum5 = 0;

FireplaceQu\_dum0 = 0;

FireplaceQu\_dum1 = 0;

FireplaceQu\_dum2 = 0;

FireplaceQu\_dum3 = 0;

FireplaceQu\_dum4 = 0;

GarageType\_dum0 = 0;

GarageType\_dum1 = 0;

GarageType\_dum2 = 0;

GarageType\_dum3 = 0;

GarageType\_dum4 = 0;

GarageFinish\_dum0 = 0;

GarageFinish\_dum1 = 0;

GarageQual\_dum0 = 0;

GarageQual\_dum1 = 0;

GarageQual\_dum2 = 0;

GarageQual\_dum3 = 0;

GarageCond\_dum0 = 0;

GarageCond\_dum1 = 0;

GarageCond\_dum2 = 0;

GarageCond\_dum3 = 0;

PavedDrive\_dum0 = 0;

PavedDrive\_dum1 = 0;

PoolQC\_dum0 = 0;

PoolQC\_dum1 = 0;

PoolQC\_dum2 = 0;

Fence\_dum0 = 0;

Fence\_dum1 = 0;

Fence\_dum2 = 0;

Fence\_dum3 = 0;

MiscFeature\_dum0 = 0;

MiscFeature\_dum1 = 0;

MiscFeature\_dum2 = 0;

MiscFeature\_dum3 = 0;

SaleType\_dum0 = 0;

SaleType\_dum1 = 0;

SaleType\_dum2 = 0;

SaleType\_dum3 = 0;

SaleType\_dum4 = 0;

if MSZoning = 'C' then MSZoning\_dum0 = 1;

if MSZoning = 'FV' then MSZoning\_dum1 = 1;

if MSZoning = 'RH' then MSZoning\_dum2 = 1;

if MSZoning = 'RL' then MSZoning\_dum3 = 1;

if Street = 'Grvl' then Street\_dum0 = 1;

if Alley = 'Gr' then Alley\_dum0 = 1;

if Alley = 'NA' then Alley\_dum1 = 1;

if LotShape = 'IR1' then LotShape\_dum0 = 1;

if LotShape = 'IR2' then LotShape\_dum1 = 1;

if LotShape = 'IR3' then LotShape\_dum2 = 1;

if LandContour = 'Bnk' then LandContour\_dum0 = 1;

if LandContour = 'HLS' then LandContour\_dum1 = 1;

if LandContour = 'Low' then LandContour\_dum2 = 1;

if Utilities = 'AllPub' then Utilities\_dum0 = 1;

if LotConfig = 'Corner' then LotConfig\_dum0 = 1;

if LotConfig = 'CulDSac' then LotConfig\_dum1 = 1;

if LotConfig = 'FR2' then LotConfig\_dum2 = 1;

if LotConfig = 'FR3' then LotConfig\_dum3 = 1;

if LandSlope = 'Gtl' then LandSlope\_dum0 = 1;

if LandSlope = 'Mod' then LandSlope\_dum1 = 1;

if Neighborhood = 'Blmngtn' then Neighborhood\_dum0 = 1;

if Neighborhood = 'Blueste' then Neighborhood\_dum1 = 1;

if Neighborhood = 'BrDale' then Neighborhood\_dum2 = 1;

if Neighborhood = 'BrkSide' then Neighborhood\_dum3 = 1;

if Neighborhood = 'ClearCr' then Neighborhood\_dum4 = 1;

if Neighborhood = 'CollgCr' then Neighborhood\_dum5 = 1;

if Neighborhood = 'Crawfor' then Neighborhood\_dum6 = 1;

if Neighborhood = 'Edwards' then Neighborhood\_dum7 = 1;

if Neighborhood = 'Gilbert' then Neighborhood\_dum8 = 1;

if Neighborhood = 'IDOTRR' then Neighborhood\_dum9 = 1;

if Neighborhood = 'MeadowV' then Neighborhood\_dum10 = 1;

if Neighborhood = 'Mitchel' then Neighborhood\_dum11 = 1;

if Neighborhood = 'NAmes' then Neighborhood\_dum12 = 1;

if Neighborhood = 'NPkVill' then Neighborhood\_dum13 = 1;

if Neighborhood = 'NWAmes' then Neighborhood\_dum14 = 1;

if Neighborhood = 'NoRidge' then Neighborhood\_dum15 = 1;

if Neighborhood = 'NridgHt' then Neighborhood\_dum16 = 1;

if Neighborhood = 'OldTown' then Neighborhood\_dum17 = 1;

if Neighborhood = 'SWISU' then Neighborhood\_dum18 = 1;

if Neighborhood = 'Sawyer' then Neighborhood\_dum19 = 1;

if Neighborhood = 'SawyerW' then Neighborhood\_dum20 = 1;

if Neighborhood = 'Somerst' then Neighborhood\_dum21 = 1;

if Neighborhood = 'StoneBr' then Neighborhood\_dum22 = 1;

if Neighborhood = 'Timber' then Neighborhood\_dum23 = 1;

if Condition1 = 'Artery' then Condition1\_dum0 = 1;

if Condition1 = 'Feedr' then Condition1\_dum1 = 1;

if Condition1 = 'Norm' then Condition1\_dum2 = 1;

if Condition1 = 'PosA' then Condition1\_dum3 = 1;

if Condition1 = 'PosN' then Condition1\_dum4 = 1;

if Condition1 = 'RRAe' then Condition1\_dum5 = 1;

if Condition1 = 'RRAn' then Condition1\_dum6 = 1;

if Condition1 = 'RRNe' then Condition1\_dum7 = 1;

if Condition2 = 'Artery' then Condition2\_dum0 = 1;

if Condition2 = 'Feedr' then Condition2\_dum1 = 1;

if Condition2 = 'Norm' then Condition2\_dum2 = 1;

if Condition2 = 'PosA' then Condition2\_dum3 = 1;

if Condition2 = 'PosN' then Condition2\_dum4 = 1;

if Condition2 = 'RRAe' then Condition2\_dum5 = 1;

if Condition2 = 'RRAn' then Condition2\_dum6 = 1;

if BldgType = '1Fam' then BldgType\_dum0 = 1;

if BldgType = '2fmCon' then BldgType\_dum1 = 1;

if BldgType = 'Duplex' then BldgType\_dum2 = 1;

if BldgType = 'Twnhs' then BldgType\_dum3 = 1;

if HouseStyle = '1.5Fin' then HouseStyle\_dum0 = 1;

if HouseStyle = '1.5Unf' then HouseStyle\_dum1 = 1;

if HouseStyle = '1Story' then HouseStyle\_dum2 = 1;

if HouseStyle = '2.5Fin' then HouseStyle\_dum3 = 1;

if HouseStyle = '2.5Unf' then HouseStyle\_dum4 = 1;

if HouseStyle = '2Story' then HouseStyle\_dum5 = 1;

if HouseStyle = 'SFoyer' then HouseStyle\_dum6 = 1;

if RoofStyle = 'Flat' then RoofStyle\_dum0 = 1;

if RoofStyle = 'Gable' then RoofStyle\_dum1 = 1;

if RoofStyle = 'Gambr' then RoofStyle\_dum2 = 1;

if RoofStyle = 'Hip' then RoofStyle\_dum3 = 1;

if RoofStyle = 'Mansard' then RoofStyle\_dum4 = 1;

if RoofMatl = 'ClyTile' then RoofMatl\_dum0 = 1;

if RoofMatl = 'CompShg' then RoofMatl\_dum1 = 1;

if RoofMatl = 'Membran' then RoofMatl\_dum2 = 1;

if RoofMatl = 'Metal' then RoofMatl\_dum3 = 1;

if RoofMatl = 'Roll' then RoofMatl\_dum4 = 1;

if RoofMatl = 'Tar&Grv' then RoofMatl\_dum5 = 1;

if RoofMatl = 'WdShake' then RoofMatl\_dum6 = 1;

if Exterior1st = 'AsbShng' then Exterior1st\_dum0 = 1;

if Exterior1st = 'BrkComm' then Exterior1st\_dum1 = 1;

if Exterior1st = 'BrkFace' then Exterior1st\_dum2 = 1;

if Exterior1st = 'CBlock' then Exterior1st\_dum3 = 1;

if Exterior1st = 'CemntBd' then Exterior1st\_dum4 = 1;

if Exterior1st = 'HdBoard' then Exterior1st\_dum5 = 1;

if Exterior1st = 'ImStucc' then Exterior1st\_dum6 = 1;

if Exterior1st = 'MetalSd' then Exterior1st\_dum7 = 1;

if Exterior1st = 'Plywood' then Exterior1st\_dum8 = 1;

if Exterior1st = 'Stone' then Exterior1st\_dum9 = 1;

if Exterior1st = 'Stucco' then Exterior1st\_dum10 = 1;

if Exterior1st = 'VinylSd' then Exterior1st\_dum11 = 1;

if Exterior1st = 'Wd Sdng' then Exterior1st\_dum12 = 1;

if Exterior2nd = 'AsbShng' then Exterior2nd\_dum0 = 1;

if Exterior2nd = 'AsphShn' then Exterior2nd\_dum1 = 1;

if Exterior2nd = 'Brk Cmn' then Exterior2nd\_dum2 = 1;

if Exterior2nd = 'BrkFace' then Exterior2nd\_dum3 = 1;

if Exterior2nd = 'CmentBd' then Exterior2nd\_dum5 = 1;

if Exterior2nd = 'HdBoard' then Exterior2nd\_dum6 = 1;

if Exterior2nd = 'ImStucc' then Exterior2nd\_dum7 = 1;

if Exterior2nd = 'MetalSd' then Exterior2nd\_dum8 = 1;

if Exterior2nd = 'Other' then Exterior2nd\_dum9 = 1;

if Exterior2nd = 'Plywood' then Exterior2nd\_dum10 = 1;

if Exterior2nd = 'Stone' then Exterior2nd\_dum11 = 1;

if Exterior2nd = 'Stucco' then Exterior2nd\_dum12 = 1;

if Exterior2nd = 'VinylSd' then Exterior2nd\_dum13 = 1;

if Exterior2nd = 'Wd Sdng' then Exterior2nd\_dum4 = 1;

if MasVnrType = 'BrkCmn' then MasVnrType\_dum0 = 1;

if MasVnrType = 'BrkFace' then MasVnrType\_dum1 = 1;

if MasVnrType = 'None' then MasVnrType\_dum2 = 1;

if ExterQual = 'Ex' then ExterQual\_dum0 = 1;

if ExterQual = 'Fa' then ExterQual\_dum1 = 1;

if ExterQual = 'Gd' then ExterQual\_dum2 = 1;

if ExterCond = 'Ex' then ExterCond\_dum0 = 1;

if ExterCond = 'Fa' then ExterCond\_dum1 = 1;

if ExterCond = 'Gd' then ExterCond\_dum2 = 1;

if Foundation = 'BrkTil' then Foundation\_dum0 = 1;

if Foundation = 'CBlock' then Foundation\_dum1 = 1;

if Foundation = 'PConc' then Foundation\_dum2 = 1;

if Foundation = 'Slab' then Foundation\_dum3 = 1;

if Foundation = 'Stone' then Foundation\_dum4 = 1;

if BsmtQual = 'Ex' then BsmtQual\_dum0 = 1;

if BsmtQual = 'Fa' then BsmtQual\_dum1 = 1;

if BsmtQual = 'Gd' then BsmtQual\_dum2 = 1;

if BsmtQual = 'NA' then BsmtQual\_dum3 = 1;

if BsmtCond = 'Fa' then BsmtCond\_dum0 = 1;

if BsmtCond = 'Gd' then BsmtCond\_dum1 = 1;

if BsmtCond = 'Po' then BsmtCond\_dum2 = 1;

if BsmtExposure = 'Av' then BsmtExposure\_dum0 = 1;

if BsmtExposure = 'Gd' then BsmtExposure\_dum1 = 1;

if BsmtExposure = 'Mn' then BsmtExposure\_dum2 = 1;

if BsmtExposure = 'NA' then BsmtExposure\_dum3 = 1;

if BsmtFinType1 = 'ALQ' then BsmtFinType1\_dum0 = 1;

if BsmtFinType1 = 'BLQ' then BsmtFinType1\_dum1 = 1;

if BsmtFinType1 = 'GLQ' then BsmtFinType1\_dum2 = 1;

if BsmtFinType1 = 'LwQ' then BsmtFinType1\_dum3 = 1;

if BsmtFinType1 = 'Rec' then BsmtFinType1\_dum4 = 1;

if BsmtFinType1 = 'NA' then BsmtFinType1\_dum5 = 1;

if BsmtFinType2 = 'ALQ' then BsmtFinType2\_dum0 = 1;

if BsmtFinType2 = 'BLQ' then BsmtFinType2\_dum1 = 1;

if BsmtFinType2 = 'GLQ' then BsmtFinType2\_dum2 = 1;

if BsmtFinType2 = 'LwQ' then BsmtFinType2\_dum3 = 1;

if BsmtFinType2 = 'Rec' then BsmtFinType2\_dum4 = 1;

if Heating = 'Floor' then Heating\_dum0 = 1;

if Heating = 'GasA' then Heating\_dum1 = 1;

if Heating = 'GasW' then Heating\_dum2 = 1;

if Heating = 'Grav' then Heating\_dum3 = 1;

if Heating = 'OthW' then Heating\_dum4 = 1;

if HeatingQC = 'Ex' then HeatingQC\_dum0 = 1;

if HeatingQC = 'Fa' then HeatingQC\_dum1 = 1;

if HeatingQC = 'Gd' then HeatingQC\_dum2 = 1;

if HeatingQC = 'Po' then HeatingQC\_dum3 = 1;

if CentralAir = 'N' then CentralAir\_dum0 = 1;

if Electrical = 'FuseA' then Electrical\_dum0 = 1;

if Electrical = 'FuseF' then Electrical\_dum1 = 1;

if Electrical = 'FuseP' then Electrical\_dum2 = 1;

if Electrical = 'NA' then Electrical\_dum3 = 1;

if KitchenQual = 'Ex' then KitchenQual\_dum0 = 1;

if KitchenQual = 'Fa' then KitchenQual\_dum1 = 1;

if KitchenQual = 'Gd' then KitchenQual\_dum2 = 1;

if Functional = 'Maj1' then Functional\_dum0 = 1;

if Functional = 'Maj2' then Functional\_dum1 = 1;

if Functional = 'Min1' then Functional\_dum2 = 1;

if Functional = 'Min2' then Functional\_dum3 = 1;

if Functional = 'Mod' then Functional\_dum4 = 1;

if Functional = 'Sev' then Functional\_dum5 = 1;

if FireplaceQu = 'Ex' then FireplaceQu\_dum0 = 1;

if FireplaceQu = 'Fa' then FireplaceQu\_dum1 = 1;

if FireplaceQu = 'Gd' then FireplaceQu\_dum2 = 1;

if FireplaceQu = 'NA' then FireplaceQu\_dum3 = 1;

if FireplaceQu = 'Po' then FireplaceQu\_dum4 = 1;

if GarageType = '2Types' then GarageType\_dum0 = 1;

if GarageType = 'Attchd' then GarageType\_dum1 = 1;

if GarageType = 'Basment' then GarageType\_dum2 = 1;

if GarageType = 'BuiltIn' then GarageType\_dum3 = 1;

if GarageType = 'CarPort' then GarageType\_dum4 = 1;

if GarageFinish = 'Fin' then GarageFinish\_dum0 = 1;

if GarageFinish = 'RFn' then GarageFinish\_dum1 = 1;

if GarageQual = 'Ex' then GarageQual\_dum0 = 1;

if GarageQual = 'Fa' then GarageQual\_dum1 = 1;

if GarageQual = 'Gd' then GarageQual\_dum2 = 1;

if GarageQual = 'Po' then GarageQual\_dum3 = 1;

if GarageCond = 'Ex' then GarageCond\_dum0 = 1;

if GarageCond = 'Fa' then GarageCond\_dum1 = 1;

if GarageCond = 'Gd' then GarageCond\_dum2 = 1;

if GarageCond = 'Po' then GarageCond\_dum3 = 1;

if PavedDrive = 'N' then PavedDrive\_dum0 = 1;

if PavedDrive = 'P' then PavedDrive\_dum1 = 1;

if PoolQC = 'Ex' then PoolQC\_dum0 = 1;

if PoolQC = 'Fa' then PoolQC\_dum1 = 1;

if PoolQC = 'Gd' then PoolQC\_dum2 = 1;

if Fence = 'GdPrv' then Fence\_dum0 = 1;

if Fence = 'GdWo' then Fence\_dum1 = 1;

if Fence = 'MnPrv' then Fence\_dum2 = 1;

if Fence = 'MnWw' then Fence\_dum3 = 1;

if MiscFeature = 'Gar2' then MiscFeature\_dum0 = 1;

if MiscFeature = 'NA' then MiscFeature\_dum1 = 1;

if MiscFeature = 'Othr' then MiscFeature\_dum2 = 1;

if MiscFeature = 'Shed' then MiscFeature\_dum3 = 1;

if SaleType = 'COD' then SaleType\_dum0 = 1;

if SaleType = 'CWD' then SaleType\_dum1 = 1;

if SaleType = 'Con' then SaleType\_dum2 = 1;

if SaleType = 'New' then SaleType\_dum3 = 1;

if SaleType = 'Oth' then SaleType\_dum4 = 1;

run;

\*Regression model to generate residual plots and Cook's D leverage values;

\*This is the plot after the log transforms were completed and erroneous variables were removed;

\*results and justifications for all adjustments are documented in the paper;

Proc reg data = train\_dummy;

Model LogSalePrice = IntLotFrontage

MSSUBCLASS

LogLotArea

OVERALLQUAL

OVERALLCOND

YEARBUILT

YEARREMODADD

LogMasVnrArea

LogBsmtFinSF1

BSMTFINSF2

BSMTUNFSF

Log1stFlrSF

'2ndFlrSF'n

LOWQUALFINSF

BSMTFULLBATH

BSMTHALFBATH

FULLBATH

HALFBATH

BEDROOMABVGR

KITCHENABVGR

TOTRMSABVGRD

FIREPLACES

GARAGEYRBLT

GARAGECARS

LogGarageArea

LogWoodDeckSF

LogOpenPorchSF

LogEnclosedPorch

Log3SsnPorch

LogScreenPorch

LogPoolArea

LogMiscVal

MOSOLD

YRSOLD

ALLEY\_DUM0

ALLEY\_DUM1

LOTSHAPE\_DUM0

LOTSHAPE\_DUM1

LOTSHAPE\_DUM2

LANDCONTOUR\_DUM0

LANDCONTOUR\_DUM1

LANDCONTOUR\_DUM2

LOTCONFIG\_DUM0

LOTCONFIG\_DUM1

LOTCONFIG\_DUM2

LOTCONFIG\_DUM3

LANDSLOPE\_DUM0

LANDSLOPE\_DUM1

NEIGHBORHOOD\_DUM0

NEIGHBORHOOD\_DUM1

NEIGHBORHOOD\_DUM2

NEIGHBORHOOD\_DUM3

NEIGHBORHOOD\_DUM4

NEIGHBORHOOD\_DUM5

NEIGHBORHOOD\_DUM6

NEIGHBORHOOD\_DUM7

NEIGHBORHOOD\_DUM8

NEIGHBORHOOD\_DUM9

NEIGHBORHOOD\_DUM10

NEIGHBORHOOD\_DUM11

NEIGHBORHOOD\_DUM12

NEIGHBORHOOD\_DUM13

NEIGHBORHOOD\_DUM14

NEIGHBORHOOD\_DUM15

NEIGHBORHOOD\_DUM16

NEIGHBORHOOD\_DUM17

NEIGHBORHOOD\_DUM18

NEIGHBORHOOD\_DUM19

NEIGHBORHOOD\_DUM20

NEIGHBORHOOD\_DUM21

NEIGHBORHOOD\_DUM22

NEIGHBORHOOD\_DUM23

CONDITION1\_DUM0

CONDITION1\_DUM1

CONDITION1\_DUM2

CONDITION1\_DUM3

CONDITION1\_DUM4

CONDITION1\_DUM5

CONDITION1\_DUM6

CONDITION1\_DUM7

CONDITION2\_DUM0

CONDITION2\_DUM1

CONDITION2\_DUM2

CONDITION2\_DUM3

CONDITION2\_DUM4

BLDGTYPE\_DUM0

BLDGTYPE\_DUM1

BLDGTYPE\_DUM2

BLDGTYPE\_DUM3

HOUSESTYLE\_DUM0

HOUSESTYLE\_DUM1

HOUSESTYLE\_DUM2

HOUSESTYLE\_DUM3

HOUSESTYLE\_DUM4

HOUSESTYLE\_DUM5

HOUSESTYLE\_DUM6

ROOFSTYLE\_DUM0

ROOFSTYLE\_DUM1

ROOFSTYLE\_DUM2

ROOFSTYLE\_DUM3

ROOFSTYLE\_DUM4

ROOFMATL\_DUM0

ROOFMATL\_DUM1

ROOFMATL\_DUM2

ROOFMATL\_DUM4

ROOFMATL\_DUM5

ROOFMATL\_DUM6

EXTERIOR1ST\_DUM0

EXTERIOR1ST\_DUM1

EXTERIOR1ST\_DUM2

EXTERIOR1ST\_DUM3

EXTERIOR1ST\_DUM4

EXTERIOR1ST\_DUM5

EXTERIOR1ST\_DUM6

EXTERIOR1ST\_DUM7

EXTERIOR1ST\_DUM8

EXTERIOR1ST\_DUM9

EXTERIOR1ST\_DUM10

EXTERIOR1ST\_DUM11

EXTERIOR1ST\_DUM12

EXTERIOR2ND\_DUM0

EXTERIOR2ND\_DUM1

EXTERIOR2ND\_DUM2

EXTERIOR2ND\_DUM3

EXTERIOR2ND\_DUM4

EXTERIOR2ND\_DUM5

EXTERIOR2ND\_DUM6

EXTERIOR2ND\_DUM7

EXTERIOR2ND\_DUM8

EXTERIOR2ND\_DUM9

EXTERIOR2ND\_DUM10

EXTERIOR2ND\_DUM11

EXTERIOR2ND\_DUM12

EXTERIOR2ND\_DUM13

MASVNRTYPE\_DUM0

MASVNRTYPE\_DUM1

MASVNRTYPE\_DUM2

EXTERQUAL\_DUM0

EXTERQUAL\_DUM1

EXTERQUAL\_DUM2

EXTERCOND\_DUM0

EXTERCOND\_DUM1

EXTERCOND\_DUM2

FOUNDATION\_DUM0

FOUNDATION\_DUM1

FOUNDATION\_DUM2

FOUNDATION\_DUM3

FOUNDATION\_DUM4

BSMTQUAL\_DUM0

BSMTQUAL\_DUM1

BSMTQUAL\_DUM2

BSMTQUAL\_DUM3

BSMTCOND\_DUM0

BSMTCOND\_DUM1

BSMTCOND\_DUM2

BSMTEXPOSURE\_DUM0

BSMTEXPOSURE\_DUM1

BSMTEXPOSURE\_DUM2

BSMTEXPOSURE\_DUM3

BSMTFINTYPE1\_DUM0

BSMTFINTYPE1\_DUM1

BSMTFINTYPE1\_DUM2

BSMTFINTYPE1\_DUM3

BSMTFINTYPE1\_DUM4

BSMTFINTYPE2\_DUM0

BSMTFINTYPE2\_DUM1

BSMTFINTYPE2\_DUM2

BSMTFINTYPE2\_DUM3

BSMTFINTYPE2\_DUM4

HEATING\_DUM1

HEATING\_DUM2

HEATING\_DUM3

HEATING\_DUM4

HEATINGQC\_DUM0

HEATINGQC\_DUM1

HEATINGQC\_DUM2

HEATINGQC\_DUM3

CENTRALAIR\_DUM0

ELECTRICAL\_DUM0

ELECTRICAL\_DUM1

ELECTRICAL\_DUM2

ELECTRICAL\_DUM3

KITCHENQUAL\_DUM0

KITCHENQUAL\_DUM1

KITCHENQUAL\_DUM2

FUNCTIONAL\_DUM0

FUNCTIONAL\_DUM1

FUNCTIONAL\_DUM2

FUNCTIONAL\_DUM3

FUNCTIONAL\_DUM4

FIREPLACEQU\_DUM0

FIREPLACEQU\_DUM1

FIREPLACEQU\_DUM2

FIREPLACEQU\_DUM3

FIREPLACEQU\_DUM4

GARAGETYPE\_DUM0

GARAGETYPE\_DUM1

GARAGETYPE\_DUM2

GARAGETYPE\_DUM3

GARAGETYPE\_DUM4

GARAGEFINISH\_DUM0

GARAGEFINISH\_DUM1

GARAGEQUAL\_DUM0

GARAGEQUAL\_DUM1

GARAGEQUAL\_DUM2

GARAGEQUAL\_DUM3

GARAGECOND\_DUM0

GARAGECOND\_DUM1

GARAGECOND\_DUM2

GARAGECOND\_DUM3

PAVEDDRIVE\_DUM0

PAVEDDRIVE\_DUM1

POOLQC\_DUM0

POOLQC\_DUM1

POOLQC\_DUM2

FENCE\_DUM0

FENCE\_DUM1

FENCE\_DUM2

FENCE\_DUM3

MISCFEATURE\_DUM1

MISCFEATURE\_DUM2

MISCFEATURE\_DUM3;

Run;

\*Forward Selection method using variables obtained through the proc reg cleanup;

\*(stop=CV) cvmethod=random(5) used to get CVPress Score per the Unit Learning presentations;

Proc glmselect data=train\_dummy;

class MSZoning

Street

Alley

LotShape

LandContour

Utilities

LotConfig

LandSlope

Neighborhood

Condition1

Condition2

BldgType

HouseStyle

RoofStyle

RoofMatl

Exterior1st

Exterior2nd

MasVnrType

ExterQual

ExterCond

Foundation

BsmtQual

BsmtCond

BsmtExposure

BsmtFinType1

BsmtFinType2

Heating

HeatingQC

CentralAir

Electrical

KitchenQual

Functional

FireplaceQu

GarageType

GarageFinish

GarageQual

GarageCond

PavedDrive

PoolQC

Fence

MiscFeature

SaleType

SaleCondition;

model LogSalePrice = IntLotFrontage MSSubClass MSZoning LogLotArea Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType LogMasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 LogBsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF Heating HeatingQC CentralAir Electrical Log1stFlrSF '2ndFlrSF'n LowQualFinSF BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars LogGarageArea GarageQual GarageCond PavedDrive LogWoodDeckSF LogOpenPorchSF LogEnclosedPorch Log3SsnPorch LogScreenPorch LogPoolArea PoolQC Fence MiscFeature LogMiscVal MoSold YrSold SaleType SaleCondition

/selection=Forward(stop=CV) cvmethod=random(5) stats =adjrsq;

Run;

\*Backward Selection method using variables obtained through the proc reg cleanup;

\*(stop=CV) cvmethod=random(5) used to get CVPress Score per the Unit Learning presentations;

Proc glmselect data=train\_dummy;

class MSZoning

Street

Alley

LotShape

LandContour

Utilities

LotConfig

LandSlope

Neighborhood

Condition1

Condition2

BldgType

HouseStyle

RoofStyle

RoofMatl

Exterior1st

Exterior2nd

MasVnrType

ExterQual

ExterCond

Foundation

BsmtQual

BsmtCond

BsmtExposure

BsmtFinType1

BsmtFinType2

Heating

HeatingQC

CentralAir

Electrical

KitchenQual

Functional

FireplaceQu

GarageType

GarageFinish

GarageQual

GarageCond

PavedDrive

PoolQC

Fence

MiscFeature

SaleType

SaleCondition;

model LogSalePrice = IntLotFrontage MSSubClass MSZoning LogLotArea Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType LogMasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 LogBsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF Heating HeatingQC CentralAir Electrical Log1stFlrSF '2ndFlrSF'n LowQualFinSF BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars LogGarageArea GarageQual GarageCond PavedDrive LogWoodDeckSF LogOpenPorchSF LogEnclosedPorch Log3SsnPorch LogScreenPorch LogPoolArea PoolQC Fence MiscFeature LogMiscVal MoSold YrSold SaleType SaleCondition

/selection=Backward(stop=CV) cvmethod=random(5) stats =adjrsq;

Run;

\*Stepwise Selection method using variables obtained through the proc reg cleanup;

\*(stop=CV) cvmethod=random(5) used to get CVPress Score per the Unit Learning presentations;

Proc glmselect data=train\_dummy;

class MSZoning

Street

Alley

LotShape

LandContour

Utilities

LotConfig

LandSlope

Neighborhood

Condition1

Condition2

BldgType

HouseStyle

RoofStyle

RoofMatl

Exterior1st

Exterior2nd

MasVnrType

ExterQual

ExterCond

Foundation

BsmtQual

BsmtCond

BsmtExposure

BsmtFinType1

BsmtFinType2

Heating

HeatingQC

CentralAir

Electrical

KitchenQual

Functional

FireplaceQu

GarageType

GarageFinish

GarageQual

GarageCond

PavedDrive

PoolQC

Fence

MiscFeature

SaleType

SaleCondition;

model LogSalePrice = IntLotFrontage MSSubClass MSZoning LogLotArea Street Alley LotShape LandContour Utilities LotConfig LandSlope Neighborhood Condition1 Condition2 BldgType HouseStyle OverallQual OverallCond YearBuilt YearRemodAdd RoofStyle RoofMatl Exterior1st Exterior2nd MasVnrType LogMasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure BsmtFinType1 LogBsmtFinSF1 BsmtFinType2 BsmtFinSF2 BsmtUnfSF Heating HeatingQC CentralAir Electrical Log1stFlrSF '2ndFlrSF'n LowQualFinSF BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces FireplaceQu GarageType GarageYrBlt GarageFinish GarageCars LogGarageArea GarageQual GarageCond PavedDrive LogWoodDeckSF LogOpenPorchSF LogEnclosedPorch Log3SsnPorch LogScreenPorch LogPoolArea PoolQC Fence MiscFeature LogMiscVal MoSold YrSold SaleType SaleCondition

/selection=Stepwise(stop=CV) cvmethod=random(5) stats =adjrsq;

Run;