STA302 Final Project

Khizer Asad, Shirley Ching, Dionysius Indraatmadja, Daleep Singh

1 Introduction

The question we seek to address is: for each section of a house, in which spaces of a house is an increase in area most directly related to the increase in sale price? The model is developed to enable us to examine which areas of the house people consider to be most significant in contributing to a house's market value. We will use a hypothesis test with a multivariate linear model to test whether there is a linear relation between the house's sale price with the square footage of various sections in a house. Since the data-set contains many predictor variables, there are some we deem to be not relevant to our question and thus will be omitted from our model by analysis. The model will show in which room is an increase in square footage of area is associated with the greatest increase in sale price. [Replace last sentence with?: The models presented will determine whether certain parts of the house in area or category would be effective in predicting the final sale price.

2 Exploratory Data Analysis

The data set we will use in this regression analysis comes from `https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data`. We group certain predictor variables into 4 categories which correspond to different sections or aspects of a house. The response variable is SalePrice which is the property's sale price in dollars.

2.1 Premium

This grouping includes the nice-to-have or *premium* features of a house: masonry veneer area (MasVnrArea) in square feet, total basement finished area (BsmtFinSF) in square feet, remodel date (YearRemodAdd) in years, total rooms above grade (TotRmsAbvGrd) not including bathrooms. Note that BsmtFinSF is the sum of BsmtFinSF1 and BsmtFinSF2 from the data set.

In Figure 1 we qualitatively observe a linear relationship between SalePrice and each of the predictor variables. We do not see any non-random relationships between pairs of predictor variables. In Table 1 we do not observe any high correlation coefficients between any pair of predictor variables. This suggests no severe multi-collinearity between the predictor variables. We observe that the values of 0 in MasVnrArea, BsmtFinSF, and YearRemodAdd have higher variance and have higher density than at other values.

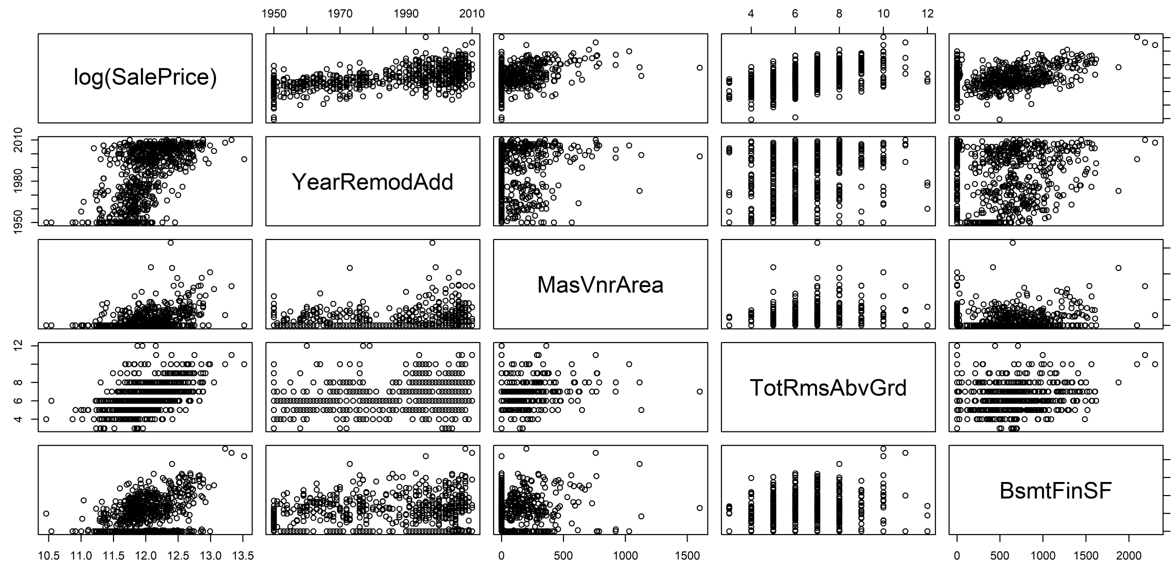

Figure 1: Pairs plot of premium predictor variables and sales price

	SalePrice	YearRemodAdd	MasVnrArea	TotRmsAbvGrd	BsmtFinSF
SalePrice	1.0000000	0.53108864	0.4520757	0.52273728	0.39959027
YearRemodAdd	0.5310886	1.00000000	0.1872555	0.20309413	0.06731494
MasVnrArea	0.4520757	0.18725550	1.0000000	0.23369016	0.21545434
TotRmsAbvGrd	0.5227373	0.20309413	0.2336902	1.00000000	0.01740391
BsmtFinSF	0.3995903	0.06731494	0.2154543	0.01740391	1.00000000

Table 1: Correlation matrix of premium predictor variables and sales price

We select the columns of data that correspond to the response and predictor variables. Since our goal is to predict SalePrice given the values for the predictor variables, we sample the data randomly into two halves: one for model building, and one for model validation.

2.2 Interior

For the interior, we are interested in garage square footage(GarageArea), first floor square footage(X1stFlrSF), second floor square footage(X2ndFlrSF), and total basement square footage (TotalBsmtSF) as the predictor variables. The response being sales price of a home. Looking at the following correlation matrix we can definitely see that there exists some multi-collinearity: The only two variables where this is majorly significant is X1stFlrSF vs TotalBsmtSF.

Looking at the pairs plot, the strength is visually evident:

2.3 Exterior

This grouping will include predictor variables regarding the area of porches for houses sold in square ft: wood deck (WoodDeckSF), open porch (OpenPorchSF), enclosed porch

	SalePrice	GarageArea	TotalBsmtSF	X1stFlrSF	X2ndFlrSF
SalePrice	1.0000000	0.6234314	0.6135806	0.6058522	0.3193338
GarageArea	0.6234314	1.0000000	0.4866655	0.4897817	0.1383470
TotalBsmtSF	0.6135806	0.4866655	1.0000000	0.8195300	−0.1745120
X1stFlrSF	0.6058522	0.4897817	0.8195300	1.0000000	−0.2026462
X2ndFlrSF	0.3193338	0.1383470	−0.1745120	−0.2026462	1.0000000

Table 2: Correlation matrix for interior predictor variables

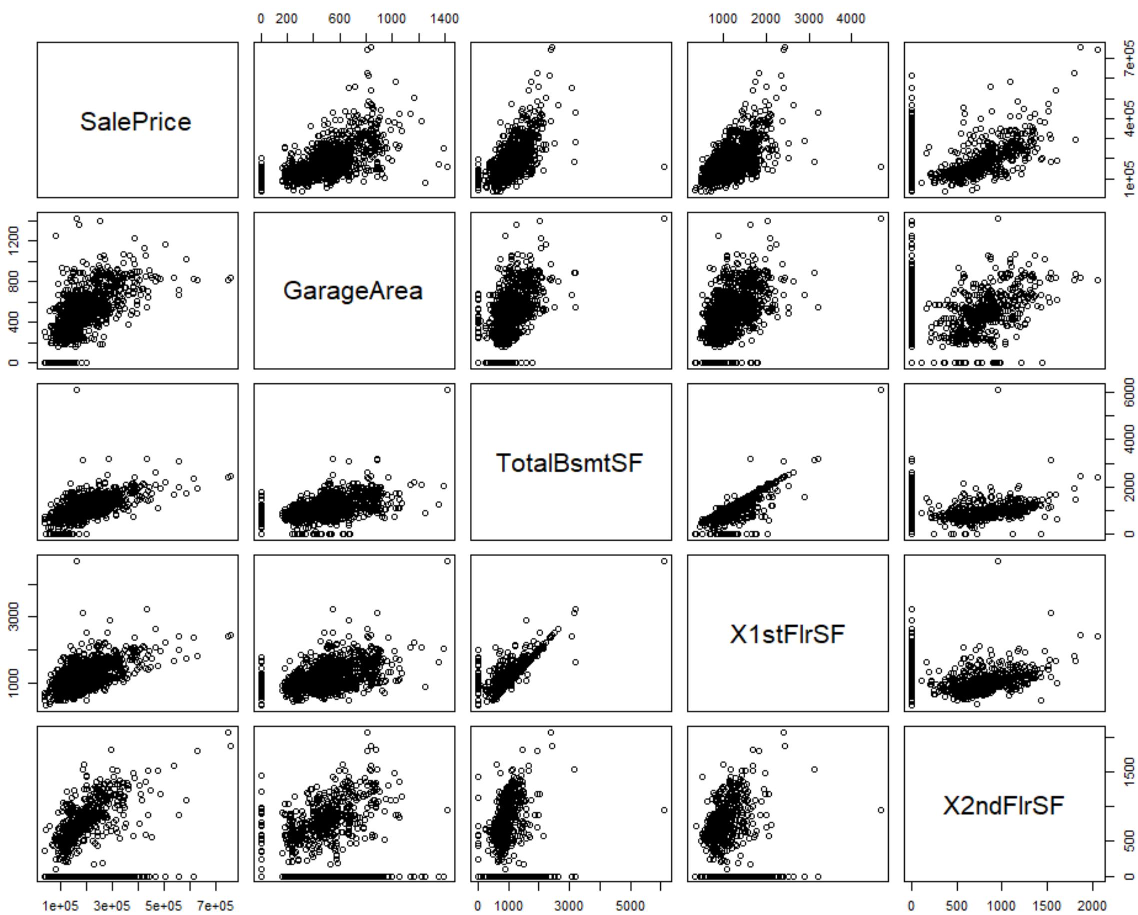

Figure 2: Pairs plot of Interior predictor variables and sales price

(EnclosedPorch), three season porch (X3SsnPorch), and screen porch (ScreenPorch).

From the correlation table (Figure 4), we can see that there is no multi-collinearity present. However, the three season porch area (X3SsnPorch) has a near close 0 correlation with the house sale price (SalePrice). This implies we cannot use this predictor variable in our model as the modelling assumption would not be valid that there is a link between the explanatory (X3SsnPorch) and response (SalePrice) variable by Gauss-Markov Theorem.

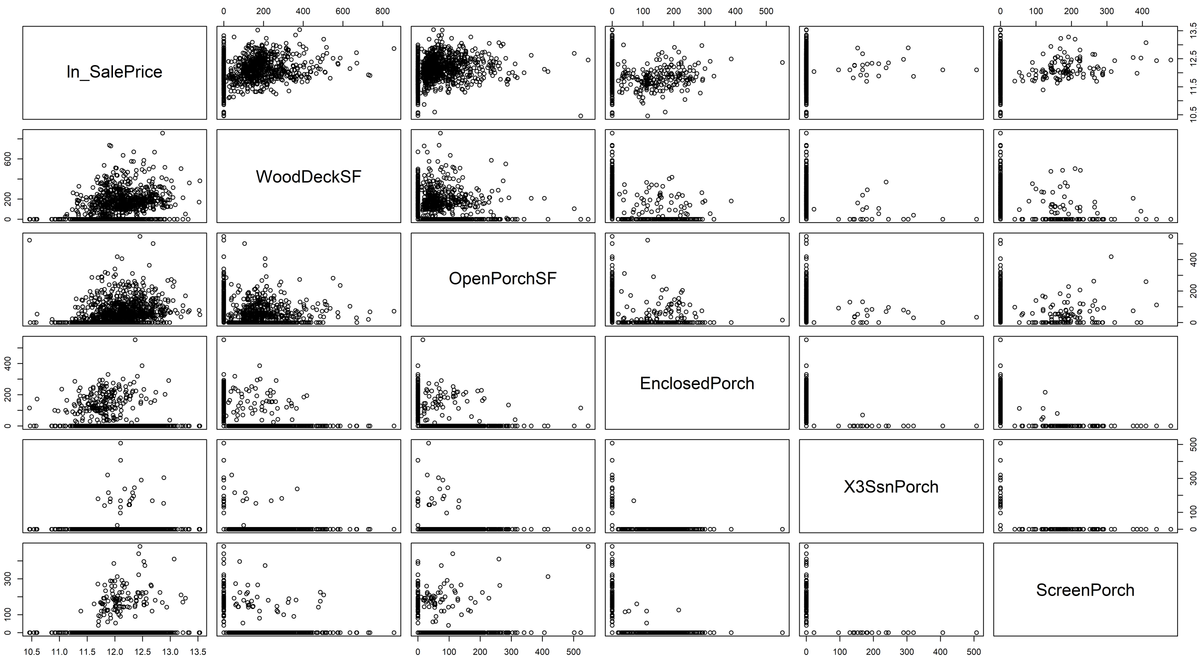

Figure 3: Pairs plot of exterior predictor variables and sales price.

	ln_SalePrice	WoodDeckSF	OpenPorchSF	EnclosedPorch	X3SsnPorch	ScreenPorch
ln_SalePrice	1.00000000	0.33413507	0.321052972	-0.14905028	0.054900226	0.12120760
WoodDeckSF	0.33413507	1.00000000	0.058660609	-0.12598889	-0.032770634	-0.07418135
OpenPorchSF	0.32105297	0.05866061	1.000000000	-0.09307932	-0.005842499	0.07430394
EnclosedPorch	-0.14905028	-0.12598889	-0.093079318	1.00000000	-0.037305283	-0.08286424
X3SsnPorch	0.05490023	-0.03277063	-0.005842499	-0.03730528	1.000000000	-0.03143585
ScreenPorch	0.12120760	-0.07418135	0.074303944	-0.08286424	-0.031435847	1.00000000

Figure 4: Correlation matrix for exterior predictor variables.

3 Model Development

We want to see if there is a linear relation between the house's sale price with the square footage of various sections of a house. We assume that the distribution follows a multivariate linear model:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1}$$

Alternatively, certain relationships may be better characterized as linear with a log-transformed response variable. In such cases we assume the model is:

$$\log(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

with

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

where $\boldsymbol{Y}$ is the response variable vector for n data points, $\boldsymbol{\beta}$ is the parameter vector, $\boldsymbol{X}$ is the predictor variable matrix for p predictor variables, and $\boldsymbol{\epsilon}$ is the error vector. We assume that the errors $\boldsymbol{\epsilon}$ are pairwise independent, have constant variance, and are normally distributed. We will check that these four assumptions are justified for our data set as follows:

- Plot the response variable against each explanatory variable to qualitatively check for a linear relationship.
- Check that the plot of the residuals of the multivariate linear model against each of the explanatory variables have no trend. This indicates linearity between the true mean of Y and X.
- Plot the residuals against the fitted values. A lack of a pattern indicates independence of errors.
- In the previous plot, check that the variance of the residuals is constant throughout the data. A pattern in the variability of the residuals indicates heteroscedasticity.
- Use a normal Q-Q plot to see if the residuals follow a normal distribution.

Hypothesis Test

We construct a two-sided F-test to see if there is a relationship between the chosen predictor variables with the response variable. We use the p-value and the standard error to examine the significance of individual predictor variables in the multiple linear regression model. High R^2 and R^2_{adj} values indicate how well the regression surface fits the data.

- **Two-sided test:** $\begin{cases} H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0 \\ H_1 : \text{ at least one of the } \beta_j, k = 1, \ldots, p \text{ is non null} \end{cases}$
- **Test Statistic:** $F^* = \text{MSReg}/\text{MSRes}$.
- **Decision rule:** Reject H_0 if $F^* > F_{1-\alpha;p'-1;n-p'}$, do not reject H_0 if $F^* \leq F_{1-\alpha;p'-1;n-p'}$.

Model Selection, Validation, Diagnostics

3.1 Premium

We use the log-transformed model (Equation 2) and seek to establish a linear relationship between the response variable and the multiple predictor variables. We choose a final model using stepwise backward elimination: after building the linear model with all relevant predictor variables, we eliminate the predictor variable with the lowest t-value if its corresponding p-value exceeds our specified significance level $\alpha = .05$.

3.1.1 Significance of Estimates

In the summary in Table 3 we observe that all predictor variables are significant at the $\alpha = 0.5$ level and we have that $R^2 = 0.6593$ and $R^2_{\text{adj}} = 0.6574$ which indicate that the regression surface fits the data well. We notice that the standard error of each estimate is an order of magnitude smaller than the estimate itself, which indicates that the estimates are relatively precise.

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	-5.743e+00	8.327e-01	-6.897	1.16e-11
YearRemodAdd	8.541e-03	4.228e-04	20.200	<2e-16
MasVnrArea	3.983e-04	5.105e-05	7.800	2.17e-14
TotRmsAbvGrd	9.912e-02	5.749e-03	17.241	<2e-16
BsmtFinSF	2.586e-04	1.951e-05	13.255	< 2e-16

Table 3: Summary of premium model estimates for the model building data set.

3.1.2 Model Selection

We find the F-statistic to be $F^* = 348.9$ and the critical value $F_{.95;4;721} = 2.384284$. Since $F^* > F_{.95;4;721}$, we reject the null hypothesis H_0 and accept the alternative hypothesis H_1. Using `stepAIC()` from the MASS library, we confirm that the final model which minimizes the Akaike Information Criterion ($AIC_p = -2137.005$) is the one that includes all 4 predictor variables. We include all variables in the model since we have that individual p-values are smaller than the significance level, standard errors are small relative to the estimates, F-test led us to accept the alternative hypothesis, and the AIC_p is minimized.

3.1.3 Model Validation

We use the mean squared prediction error (MSPE) as an indicator for the predictive ability of the model. Note that for our model, we have MSRes = 0.05231897. We find the predicted values using the coefficients from our model and the predictor variables from the validation data set. We find that:

$$\text{MSPE} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n^*} = 0.06830661,$$

where Y_i are the (log-transformed) response variables from the validation data set, and $\hat{Y}_i$ are the predicted values. Since MSPE is reasonably close to MSRes, we expect the model to have good predictive ability.

3.1.4 Diagnostics

Improper functional form

We check for an improper functional form by plotting the residuals against each predictor variable in Figure 5. We do not see non-random patterns in the residual plots, which indicates that the functional form of the model is appropriate.

Outliers and Influential Points

We find 2 outlying observations (large studentized deleted residuals) with $DFFITS$ values of 0.297 and 0.415. Since the values are less than 1, the observations are not considered influential to the fitted value. We observe later, in the Diagnostics Plots section, that no observations have a high Cook's distance. Therefore, there are no influential outlying observations in the model.

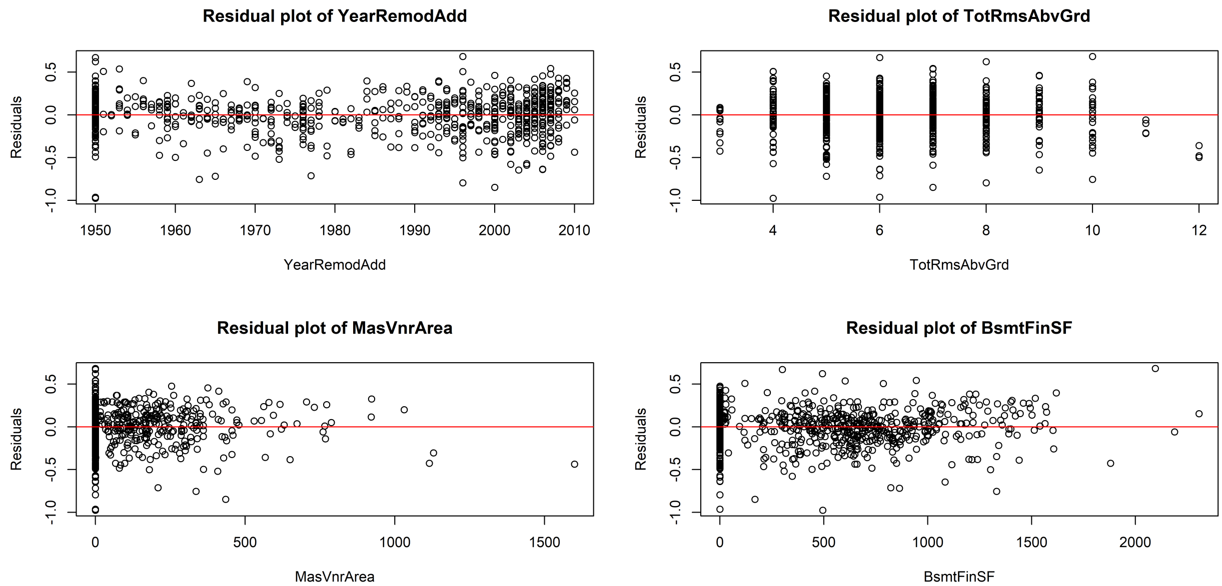

Figure 5: Residuals vs. predictor variables of the premium model.

Multi-collinearity

We compute the variance inflation factor (VIF) in Table 4 and observe that each individual VIF is smaller than 10. In addition to the fact that the mean VIF of 1.08 is not considerably larger than 1, we conclude that there are no indications of severe multi-collinearity.

	YearRemodAdd	MasVnrArea	TotRmsAbvGrd	BsmtFinSF
VIF	1.067269	1.131248	1.089786	1.051189

Table 4: VIF of the predictor variables.

Diagnostics Plots

We analyze the diagnostics plots displayed in Figure 6 as follows:

- **Residuals vs Fitted**: we observe equally spread residuals around a horizontal line at $y = 0$ without distinct patterns, so there is no indication that the linearity assumption is violated.
- **Scale-Location**: we observe equally spread points around a horizontal line, so there is no indication of heteroscedasticity.
- **Normal Q-Q**: the residuals follow a straight line except for the left tails, where it diverges significantly.
- **Residuals vs Leverage**: all observations have small Cook's distance, so we do not have high leverage of any outlying observation and no observations are influential.

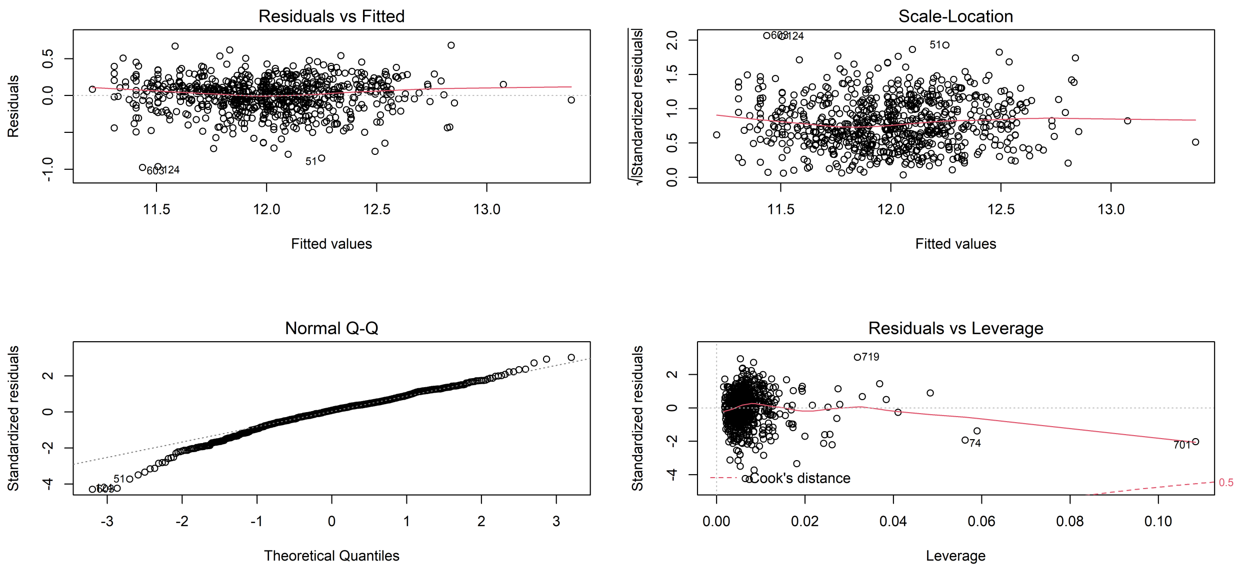

Figure 6: Diagnostics of the premium model.

3.2 Interior

3.3 Exterior

3.3.1 Model Selection

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.8593877  0.0190435 622.752   <2e-16 ***
WoodDeckSF    0.0008885  0.0001010   8.796   <2e-16 ***
OpenPorchSF   0.0016920  0.0001917   8.828   <2e-16 ***
EnclosedPorch -0.0004059 0.0002100  -1.933   0.0536 .
ScreenPorch   0.0005880  0.0002288   2.570   0.0104 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3412 on 725 degrees of freedom
Multiple R-squared:  0.2053,    Adjusted R-squared:  0.2009
F-statistic: 46.82 on 4 and 725 DF,  p-value: < 2.2e-16
```

Figure 7: Model-building data set model summary.

Using backwards elimination and AIC_p (Figure 8), the final model consists of the predictor variables: WoodDeckSF, OpenPorchSF, EnclosedPorch and ScreenPorch.

From the model-building dataset multiple regression model summary and F-test results (Figure 7), the model shows that it is significant, such that should reject H_0 and accept H_1. However, unlike the validation set, EnclosedPorch shows only marginal significance in the $\alpha = 0.1$, while the ScreenPorch exhibits significance at $\alpha = 0.05$, while WoodDeckSF, OpenPorchSF are significant at $\alpha = 0.001$. While predictability of a model is important in our investigation, it is not as important as the model estimate themselves being unbiased and satisfying the Gauss-Markov assumptions. However, the dataset used in our model is heavily zero-inflated, and to deal with such an ordeal is not covered in the scope of this

```
Start:  AIC=-1472.54
valid.ln_SalePrice ~ WoodDeckSF + OpenPorchSF + EnclosedPorch +
    ScreenPorch

                Df Sum of Sq     RSS     AIC
<none>                        95.791 -1472.5
- EnclosedPorch  1    0.7106  96.501 -1469.2
- ScreenPorch    1    2.7022  98.493 -1454.2
- OpenPorchSF    1    9.9734 105.764 -1402.2
- WoodDeckSF     1   13.8993 109.690 -1375.6
> step$anova # display results
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
valid.ln_SalePrice ~ WoodDeckSF + OpenPorchSF + EnclosedPorch +
    ScreenPorch

Final Model:
valid.ln_SalePrice ~ WoodDeckSF + OpenPorchSF + EnclosedPorch +
    ScreenPorch


  Step Df Deviance Resid. Df Resid. Dev       AIC
1                        725   95.79057 -1472.543
```

Figure 8: Model-building data set model AIC results.

course. The rest of this section will shed light to other possible reasons as to why our model might not be the best despite being significant.

3.3.2 Model Validation

We will be comparing the MSPE (mean squared prediction error) and MSres(mean squared residuals) to test the predictive ability of the model.

$$\text{MSPE} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n^*} = 0.13122,$$
$$\text{MSRes} = 0.1164042$$

We can see that the two values are fairly close to each other, meaning the predictive ability of our model is fairly good.

3.3.3 Diagnostics

- **Improper Functional Form**

  Diagnostic residual plots (Figure 9) between the model residuals and predictor variables are used to check for improper functional form. From the residual graphs, we can see that there is no non-random pattern, which is a sign of adequate functional form.

- **Outliers and Influential Points** Testing for outliers in the y-direction, we used studentized deleted residual values. In the y-direction, the outliers were observations: 899, 917 and 1183. Testing for outliers in the x-direction, we used each observation's Pii (projection matrix) value.

  $\text{P}_{ii} > 2\frac{\sum_{i=1}^{n}(P_{ii})}{n} = 0.01369863$. In the x-direction, the outliers were observations: 736, 748, 764, 765, 770, 776, 785, 786, 796, 800, 801, 804, 808, 814, 829, 831, 837, 841, 845,

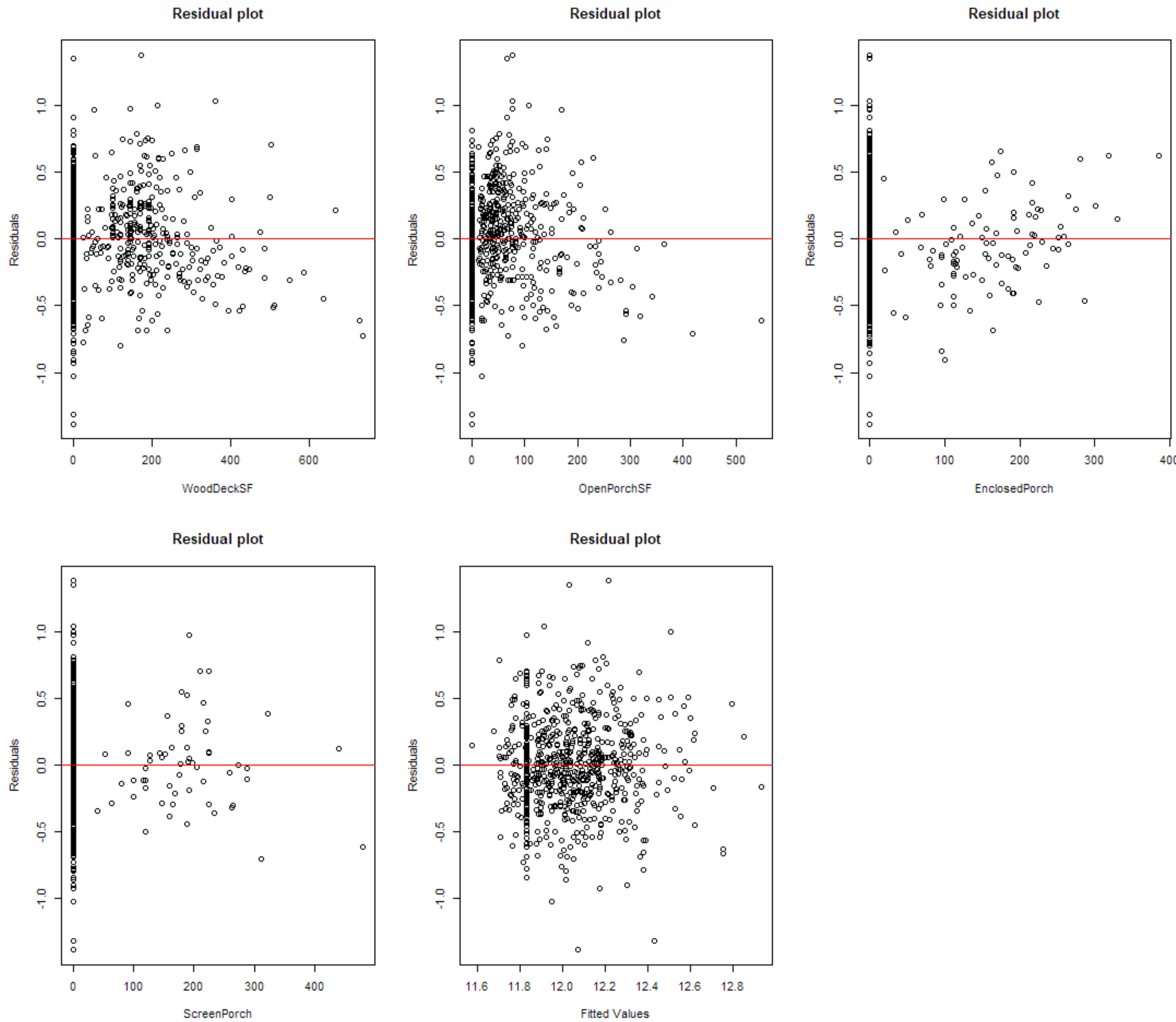

Figure 9: Diagnostic residual plots used to check improper functional form.

847, 849, 855, 860, 861, 876, 888, 889, 894, 908, 915, 919, 920, 940, 945, 946, 948, 962, 975, 997, 1014, 1031, 1038, 1045, 1056, 1068, 1069, 1071, 1082, 1095, 1107, 1120, 1140, 1151, 1153, 1155, 1156, 1172, 1185, 1186, 1188, 1194, 1198, 1203, 1211, 1228, 1229, 1249, 1267, 1283, 1293, 1294, 1299, 1302, 1311, 1313, 1314, 1318, 1321, 1327, 1329, 1361, 1370, 1372, 1383, 1387, 1394, 1415, 1420, 1424, 1439, 1440, 1446 and 1460.

Influential points: Cooks distance will be used to decide whether an observation is influential to our model. From R, F(5,725)=2.226458. When looped to find whether any of the observations' cooks distance are larger, none of them are flagged. This comes to no surprise as the model-buiding data set has 730 observations, it is a sample big enough that it would be unusual for data points to be highly influential. Since none of the points are influential, they have not been removed from the model despite being outliers.

- **Multi-collinearity**

	WoodDeckSF	OpenPorchSF	EnclosedPorch	ScreenPorch
VIF	1.015562	1.034292	1.027436	1.028064

Table 5: VIF of exterior predictor variables.

The VIF (variance inflation factor) is used to find any multi-collinearity within our model. If the VIF of a predictor variable is larger than 10, there is serious multi-collinearity. Also, the mean VIF value is 1.026338, which is close to 1. A good sign from the two VIF values that there is no multi-collinearity.

- **Diagnostic Plots**

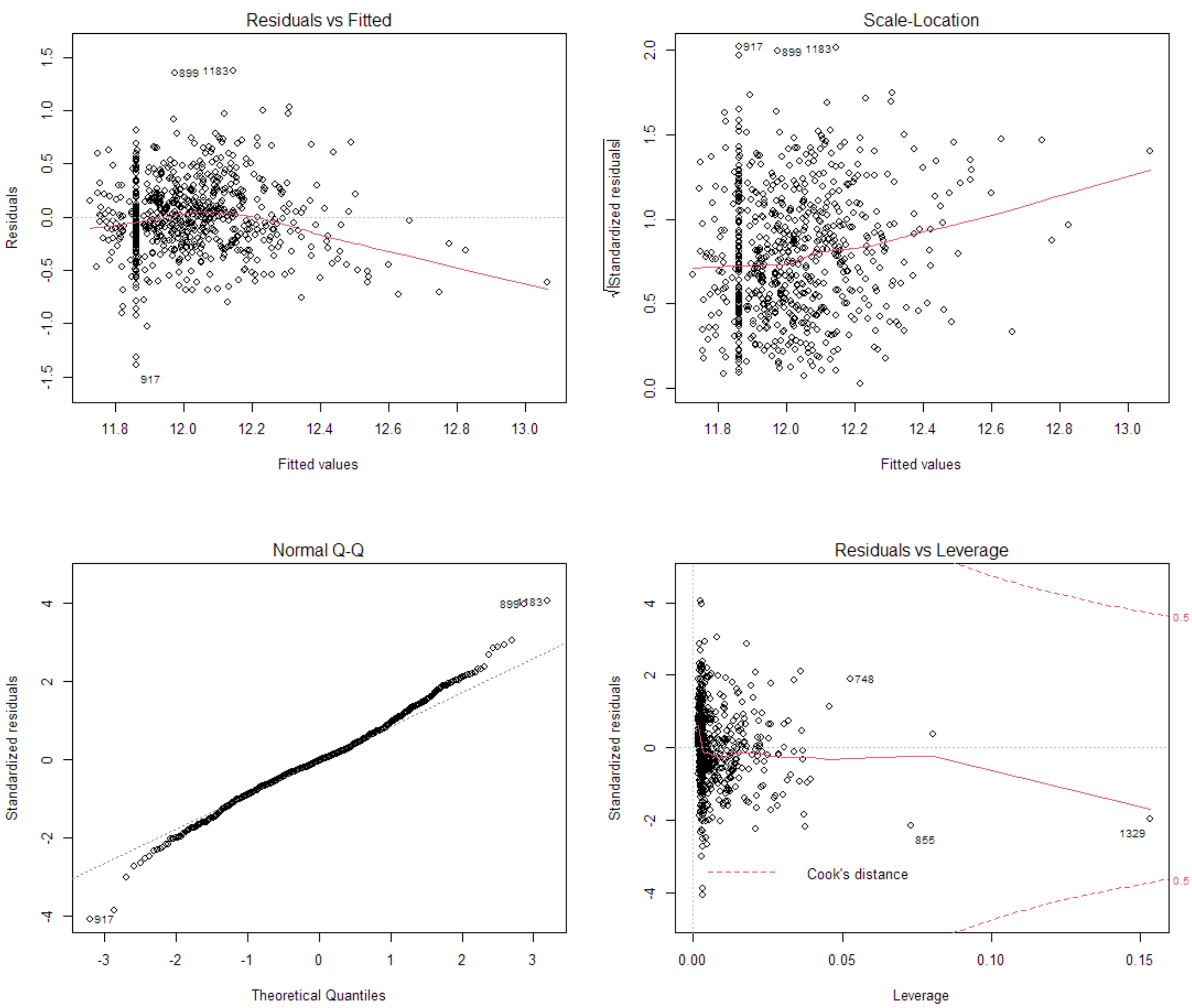

Figure 10: Diagnostic plots of our exterior predictors model used to Gauss-Markov assumptions.

From the diagnostic plots of our model (Figure 10), we will be checking for any violation of the Gauss-Markov Assumptions.

Checking Linearity: In the predictor-residual plots (Figure 9), we can see that the points are generally scattered along the horizontal line. This tells there is linearity between SalePrice and each predictor variable. It is safe to say a linear model is appropriate in this case.

Checking Independence of Errors: Back to the residual-fitted plot (Figure 9), the points appear random though it slightly goes below the horizontal line on the right. This could be the case if we do not have many observations for higher SalePrice observations. However, the points still seem non-patterned. implying there is an independence of errors to the response.

Checking Constant Variance of Errors: Using the Scale-Location, we can see that there is no pattern. And so homoscedasticity is satisfied for the model.

Checking Normality of Errors: Using the QQplot, we check for normality. Non-normal errors is present if they go off line. Most of the points are on top of the line, which is a sign of normal errors.

Therefore, our model satisfies the Gauss-Markov Assumptions.

3.4 Qualitative

4 Conclusion

We observe that a change in the log of the sale price of properties is directly associated with a change in the masonry veneer area, total basement finished area, remodel date, and the total rooms above grade. This model may be used to estimate the sale price of a house, given values for the predictor variables. We observed that an increase in masonry veneer area tends to drive the sale price up more than an increase in finished basement area.

The only significant technical issue with the model is the non-Gaussian errors, as seen in the normal Q-Q plot of Figure 6. The estimators are still unbiased and minimum variance, but since our data set is small-medium (<1000 for each of the the training set and validation set), the non-normal errors of our model indicate that the tests of significance and the construction of confidence and prediction intervals may not be valid. However, with the p-values of the predictor variables being much smaller than the significance level $\alpha = 0.05$, the F-statistic being much greater than the critical value, and the lack of issues with other diagnostic tests, the model is still be a justified representation of the data set.

From the hypothesis testing for the exterior predictors model, we can say that the exterior predictors do play a part in the prediction of the house sale price. However, we have to note that the predictor values were heavily zero-inflated. Another issue is the difference in significance between the two models for the predictor variables itself. However, that can most likely be solved with a larger sample size. A sample with more data on highly priced houses will also improve our model.

Group Member Contributions

Khizer Asad:
Shirley Ching:
Dionysius Indraatmadja:
Daleep Singh:

Appendix: R code for Premium Model

Dionysius Indraatmadja

The dataset: https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data. Use `setwd` to set the working directory to the extracted folder `house-prices-advanced-regression techniques`, and set `plotwd` to a folder where the images will be saved.

```r
library(car)
library(MASS)
set.seed(1003024416)

setwd("X:/Dropbox/sta302/final project/house-prices-advanced-regression-techniques")
plotwd<-"X:/Dropbox/sta302/final project/writeup/plots/"
table<-read.csv("train.csv", header=TRUE)
data<-table[, c("SalePrice", "YearRemodAdd", "MasVnrArea", "TotRmsAbvGrd")]
data[, "BsmtFinSF"]=table$BsmtFinSF1+table$BsmtFinSF2
data<-data[complete.cases(data), ]  # exclude incomplete data points

# Split data into model building set and validation set
sample_size<-floor(.5*nrow(data))
picked<-sample(seq_len(nrow(data)), size=sample_size)
train<-data[picked,]
attach(train)
test<-data[-picked,]

# perform the linear fit
multi.fit<-lm(log(SalePrice)~YearRemodAdd + MasVnrArea + TotRmsAbvGrd + BsmtFinSF)
predictors<-c("YearRemodAdd", "MasVnrArea", "TotRmsAbvGrd", "BsmtFinSF")
multi.res<-resid(multi.fit)
fitted<-fitted(multi.fit)
summary(multi.fit)

# Model selection using AIC
step<-stepAIC(multi.fit, direction="backward")
step$anova

# F critical value
crit<-qf(.95, length(multi.fit$coefficients)-1, multi.fit$df)
fstatistic<-summary(multi.fit)$fstatistic[[1]]  #dendf is denominator dof
print(crit)
print(fstatistic)

# Anova table
anova_table<-anova(multi.fit)
MSRes<-anova_table[nrow(anova_table), 3]
s2<-MSRes  # estimator of variance
```

```
# Model Validation
beta<-as.numeric(data.matrix(coef(multi.fit)))  # beta coefficients
X<-data.matrix(cbind("1"=vector("numeric", length=dim(test)[1])+1,
                     test[predictors]))  # test data matrix
predictions<-as.numeric(X %*% beta)
test_response<-as.numeric(data.matrix(test[1]))
residuals<-log(test_response) - predictions
MSPE<-sum((residuals)**2)/(dim(test)[1])
MSPE
MSRes

# Outliers
semistudentized<-residuals/var(residuals)
Pmatrix<-X %*% solve(t(X) %*% X) %*% t(X)  # projection matrix
del_res<-residuals/(1-diag(Pmatrix))
n<-dim(train)[1]
p_prime<-length(coef(multi.fit))
mean_leverage<-2*p_prime/n

t<-as.numeric(rstudent(multi.fit))
t_crit<-qt(1-(.05/(2*n)), n-p_prime-1)
outliers_i<-which(abs(t) > t_crit)
pii<-c(diag(Pmatrix)[[outliers_i[1]]], diag(Pmatrix)[[outliers_i[2]]])
t[outliers_i]  # outliers
mean_leverage
which(as.numeric(diag(Pmatrix))>mean_leverage)

# Measure of influence of outlying observations
DFFITS<-abs(t[outliers_i]*sqrt(pii/(1-pii)))
DFFITS

# VIF
vif(multi.fit)
mean(vif(multi.fit))

# Diagnostics plots
w=3*4
h=3*2
png(paste(plotwd, "diagnostics.png", sep=""), width=w, height=h, units="in", res=600)
layout(matrix(c(1,2,3,4),2,2)) # yields 4 graphs/page
plot(multi.fit)
dev.off()

# Pairs plot
png("X:/Dropbox/sta302/final project/writeup/plots/pairs-plot.png",
    width = w, height = h, units = 'in', res = 600)
pairs(log(SalePrice) ~ YearRemodAdd + MasVnrArea + TotRmsAbvGrd + BsmtFinSF, data = train)
dev.off()
cor(train)

# Residuals Plots
png(paste(plotwd, "a1.png", sep=""), width=w, height=h, units="in", res=600)
layout(matrix(c(1,2,3,4),2,2)) # yields 4 graphs/page
```

```
plot(x=YearRemodAdd, y=multi.res,
     ylab="Residuals", xlab="YearRemodAdd",
     main="Residual plot of YearRemodAdd")
abline(0, 0, col='red')
plot(x=MasVnrArea, y=multi.res,
     ylab="Residuals", xlab="MasVnrArea",
     main="Residual plot of MasVnrArea")
abline(0, 0, col='red')
plot(x=TotRmsAbvGrd, y=multi.res,
     ylab="Residuals", xlab="TotRmsAbvGrd",
     main="Residual plot of TotRmsAbvGrd")
abline(0, 0, col='red')
plot(x=BsmtFinSF, y=multi.res,
     ylab="Residuals", xlab="BsmtFinSF",
     main="Residual plot of BsmtFinSF")
abline(0, 0, col='red')
dev.off()
```