Springbank Drive Compensation Analysis

Canning Consultants, Inc.

Total Word Count: 991

Contents

Page

1.Introduction ---------- 3

2.Data Description---------- 3

3.Model 1:Justification,model description and appropriateness----- 3

4. Model 1: Assessment of fit ----------4

5.Model 2:Justification,model description and appropriateness----- 5

6. Model 2: Assessment of fit ----------6

7. Conclusion and Application ----------6

8. Appendix----------7

9. Letter----------9

Introduction

We are from Canning Consultants Inc. and representing the City of London to determine the appropriate compensation for Springbank residents. We used multiple linear regression to construct our models under several assumptions, such as linear relationship between house price and its explanatory variables, multivariate normality, minimal collinearity, and homoscedasticity.

Data Description and Strategy

We were provided with 104 Springbank house sales data from 1998 to 2003 (Refer to appendix for full summary). Data showed that the average house price is $134,469. Area lot, lot frontage and area of frontage had skewed distributions, showing that their mean and median differ.

Our strategy was to obtain maximum adjusted-R-squared with minimal collinearity for our model.

Firstly, we analyzed the correlation of price with other explanatory variables, shown as below:

CORR TABLE FOR PRICE	
AGEYR	-0.13
LOTAREA	0.46
LOTFRT	0.33
LFA	0.4
DISTCURB	0.37
BSMTFINAREA	0.22
TRAFCOUNT	-0.4

Model 1

Justification

We selected our variables based on these conditions:

1) Explanatory variable having high correlation with the dependent variable.
2) Explanatory variable having low correlation with another explanatory variable.

Description and Appropriateness

Using this approach, we came up with our first MR model:

PRICE = 101,189.50 + 1.04(LOTAREA) + 33.58(LFA) – 0.80 (TRAFCOUNT) + 11,399.80(GOODINTCOND) + 17.91(BSMTFINAREA) – 14,581.51(BRICKEXFINFACTOR) + 10,868.66(HSETYPE3)

P-value = 8.77e-13 | R-squared = 0.51 | Adj. R-squared = 0.48

This model is comprised of three categorical and four numeric variables. Summary infers that before taking into account other variables, the house price in Springbank Area is roughly $101,189.50.

Further interpreting the coefficients, an increment of 1-square-feet of basement finished area would increase the price by $17.91 keeping other factors constant, 1-square-feet of lot area would increase house prices by $1.04 keeping other factors constant. Same interpretations for lot front area and traffic count.

Making the model more parsimonious, we manipulated the categorical variable to only select its most significant values. We made a dummy variable for brick by assigning the value 1 on the EXTFINFACTOR, and 0 otherwise. This means that houses made out of brick have $14.5K lesser value than non-brick houses keeping other factors constant. Applying the same concept, we added value 1 on HSETYPE only if the house has two-storey. Lastly, we assigned the value 1 on GOODINTCOND if the house interior condition was either 5 or 7. This means that house with superior condition is worth $11.4K more value than houses with fair interior condition, keeping other factors constant.

Our null hypothesis was *"The effect of these explanatory variables on price is zero"*.
Looking at the p-value of 8.77e^-13 of the F test, we reject the null hypothesis at 5% significance level.

		PRICE	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	101189.53	80143.10 – 122235.96	**<0.001**
LOTAREA	1.04	0.60 – 1.48	**<0.001**
LFA	33.58	16.79 – 50.37	**<0.001**
TRAFCOUNT	-0.80	-1.33 – -0.27	**0.003**
INTCOND == 5 | INTCOND == 7TRUE	11139.98	2687.06 – 19592.89	**0.010**
BSMTFINAREA	17.91	3.61 – 32.21	**0.015**
EXTFINFACTOR == 3TRUE	-14581.51	-22936.52 – -6226.50	**0.001**
HSETYPE == 3TRUE	10868.66	-1434.24 – 23171.56	0.083
Observations	104		
R^2 / R^2 adjusted	0.515 / 0.479		

Assessment of fit

From the R-squared value**,** we know that the variation of these variables explains **51.48%** of the variation in Springbank house price. It indicates a **moderate goodness of model fit** (Refer to appendix for the residual analysis).

Model 2

Justification

We improved the model by taking into account other explanatory variable: **Location**.

$$PRICE = 104755.70 + 1.08(LOTAREA) + 29.76(LFA) - 0.82(TRAFCOUNT) + 19.21(BSMTFINAREA) - 11444.51(EXTFINFACTOR3) + 7662.75(INTCOND57) + 101582.40(ADDRESS2 = OXFORD\ E)$$

P-value = 2.2e-16 | Multiple R-squared = 0.6177 | Adjusted R-squared = 0.5899

Description and Appropriateness

Similarly, this model is comprised of three categorical and four numeric variables. The interpretation of the coefficient follows the same logic as model 1. The primary difference in this model is *adding Address* and *removing House Type (HSETYPE)*.

Our method was separating the house number from the Address and recreate a dummy variable based on the Address street. Applying the same logic as the first model, we plugged in value 1 only if the house is located on Oxford E and 0 otherwise. Keeping other factors constant, the coefficient shows that the house value in Oxford E is $101,582.38 higher than other houses. In result, the adj-R-squared increased to 58.99%. This jump was because prime location significantly affects property value. Looking at the p-value of 2.2e^-16 from the F-test, we can reject the null hypothesis at 5% significance level.

		PRICE	
Predictors	*Estimates*	*CI*	*p*
(Intercept)	104755.68	86068.60 – 123442.75	**<0.001**
LOTAREA	1.09	0.70 – 1.48	**<0.001**
LFA	29.76	14.90 – 44.62	**<0.001**
TRAFCOUNT	-0.82	-1.28 – -0.37	**0.001**
BSMTFINAREA	19.21	6.76 – 31.67	**0.003**
EXTFINFACTOR == 3TRUE	-11444.51	-18808.39 – -4080.64	**0.003**
INTCOND == 5 | INTCOND == 7TRUE	7664.75	192.89 – 15136.61	**0.044**
Address2 == " Oxford E"TRUE	101582.38	64607.60 – 138557.16	**<0.001**
Observations	104		
R^2 / R^2 adjusted	0.618 / 0.590		

Assessment of fit

We formulated a model with seven variables for which R-squared increased from 51.48% to **61.77%**. The best indicator is by looking at the Adj-R-squared; this model has successfully raised the value from 47.95% to **58.99%**.

Conclusion and Application

By looking at the lower p-value, higher F-statistics, same count of explanatory-variables, and higher AdjRsquared, **we conclude that MODEL2 is a better fit** than the previous model.

We calculated our final compensation by adding the lump sum amount of $12,000 to the lost house value:

$$\text{LOST HOUSE VALUE} = \text{-LFA}(X_{\text{LFA}}) + \text{-TRAFCOUNT}(X_{\text{TRAFCOUNT}})$$

$$\text{COMPENSATION} = \$12{,}000 + \text{ABS (LOST HOUSE VALUE)}$$

Thereof, we applied both our models to five properties and compared the differences.

	Property 1	Property 2	Property 3	Property 4	Property 5	Total
PREVIOUS	$19,270.00	$18,112.00	$27,039.00	$21,728.00	$17,638.00	$103,787.00
MODEL 1	$29,232.00	$31,660.80	$33,174.40	$31,718.40	$38,889.60	$164,675.20
MODEL 2	$29,666.40	$32,156.16	$33,706.88	$32,215.68	$39,565.92	$167,311.04

APPENDIX

Variable Summary Table

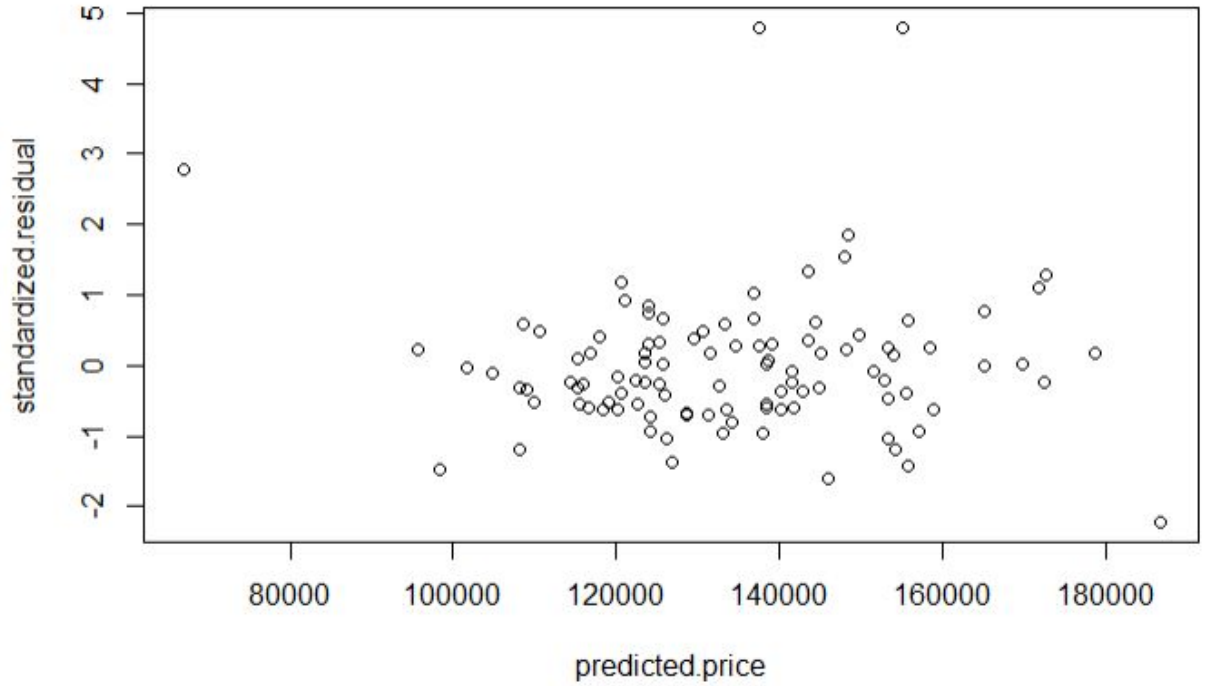
Property #	Address	Sales Date	HSETYPE	AGEYR	LOTAREA	LOTFRT	LFA
Min. : 1	Length:104	Min. :1998-01-19 00:00:00	Min. :0.00	Min. : 0	Min. : 760	Min. : 29	Min. : 0
1st Qu.: 27	Class :character	1st Qu.:1999-07-12 18:00:00	1st Qu.:0.00	1st Qu.: 41	1st Qu.: 6054	1st Qu.: 49	1st Qu.: 740
Median : 52	Mode :character	Median :2000-10-26 00:00:00	Median :0.00	Median : 48	Median : 8612	Median : 56	Median : 863
Mean : 52	NA	Mean :2000-10-15 17:32:18	Mean :0.63	Mean : 50	Mean :11960	Mean : 69	Mean : 903
3rd Qu.: 78	NA	3rd Qu.:2002-01-26 00:00:00	3rd Qu.:0.00	3rd Qu.: 55	3rd Qu.:13001	3rd Qu.: 76	3rd Qu.:1059
Max. :104	NA	Max. :2003-05-26 00:00:00	Max. :3.00	Max. :118	Max. :48787	Max. :617	Max. :1674

DISTCURB	EXTAMEN	EXTFINFACTOR	GAR	STSCAPE	CENAIR	POOL
Min. : 9	Min. :0.00	Min. :1.0	Min. :0.00	Min. :1	Min. :0.00	Min. :0.00
1st Qu.: 22	1st Qu.:1.00	1st Qu.:1.0	1st Qu.:1.00	1st Qu.:3	1st Qu.:0.00	1st Qu.:0.00
Median : 38	Median :2.00	Median :3.0	Median :1.00	Median :3	Median :1.00	Median :0.00
Mean : 42	Mean :1.47	Mean :2.2	Mean :0.92	Mean :3	Mean :0.52	Mean :0.13
3rd Qu.: 53	3rd Qu.:2.00	3rd Qu.:3.0	3rd Qu.:1.00	3rd Qu.:3	3rd Qu.:1.00	3rd Qu.:0.00
Max. :147	Max. :3.00	Max. :5.0	Max. :3.00	Max. :5	Max. :1.00	Max. :1.00

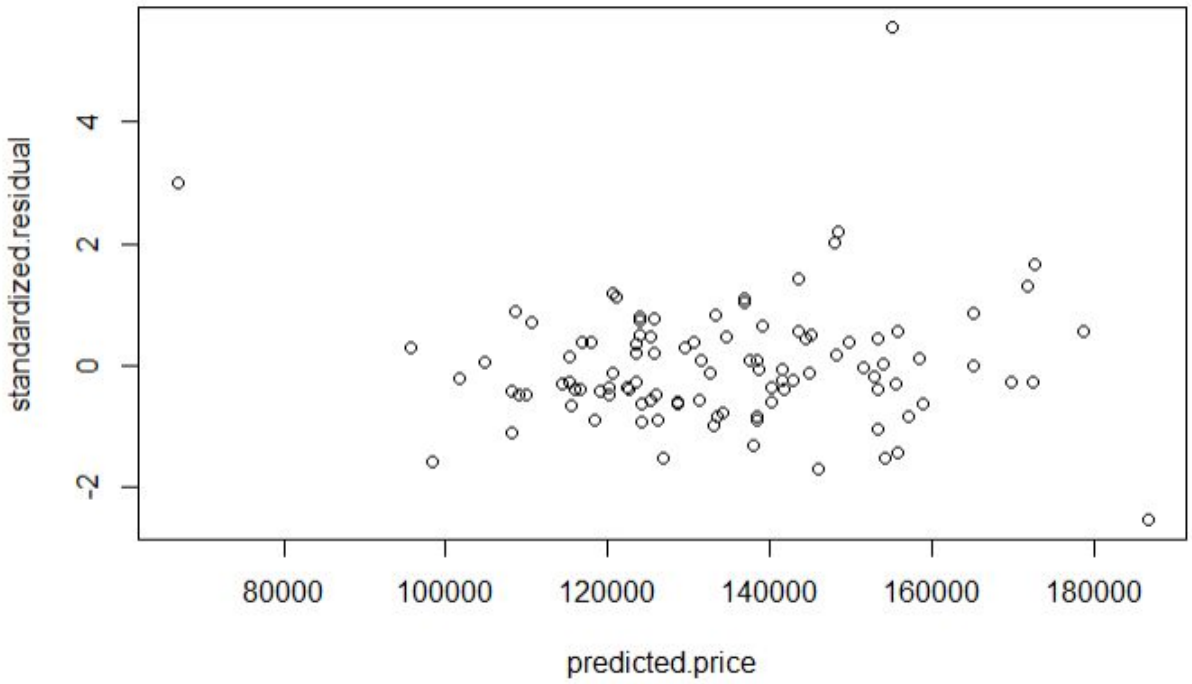
INTCOND	BASEMT	BSMTFINAREA	BI AMEN APPL	LANESRD	TRAFCOUNT	PRICE
Min. :1.0	Min. :1.00	Min. : 0	Min. :0.0	Min. :0.00	Min. : 0	Min. : 69900
1st Qu.:3.0	1st Qu.:1.00	1st Qu.: 75	1st Qu.:0.0	1st Qu.:0.00	1st Qu.:13750	1st Qu.:116550
Median :3.0	Median :1.00	Median : 471	Median :0.0	Median :0.00	Median :19000	Median :131400
Mean :3.9	Mean :1.01	Mean : 397	Mean :1.2	Mean :0.33	Mean :20221	Mean :134469
3rd Qu.:5.0	3rd Qu.:1.00	3rd Qu.: 645	3rd Qu.:2.2	3rd Qu.:1.00	3rd Qu.:25125	3rd Qu.:147250
Max. :7.0	Max. :2.00	Max. :1134	Max. :6.0	Max. :1.00	Max. :36000	Max. :249900

Constant variance and residual analysis

Model 1



Model 2



Word count = 698

Canning Consultants Inc.
9 Weybourne Crescent, London, ON N6H 4H2, Canada

9 December 2019

Officials of the City of London
300 Dufferin Ave, London, ON N6B 1Z2, Canada

Dear Sir or Madam,

I am writing to you about our new model that accounts for all the factors and for the total loss of the market value in Springbank houses to minimize any chance for the city being vulnerable to compensation claims. In contrast to the Automatic Valuation Model (AVM) which favored quantitative methods used before, our new method involved both quantitative and qualitative data to determine compensation prices.

We will refer to model 2 in our Technical Report for the intuitive explanation of the model. In short, houses located in Oxford East with brick and good interior, along with the quantitative factors (i.e. lot area, lost front area and basement finished area) best estimates the Springbank houses value.

Our new approach is better in the following ways:

- The price compensation carefully takes the concern of the properties in the respective location.
- The model takes into account the compensation to the interior condition and material of the houses.
- Traffic count in these specific location is taken into account which was not done before.

Summarizing, the five property owners' need to be compensated with the following amounts respectively: $29,666.40, $32,156.16, $33,706.88, $ 32,215.68 and $39,565.92. One important caveat is that there still might be the presence of omitted variable bias, where traffic count and number of lanes are correlated and both explain price. However, only traffic count is selected.

As a result of this model, the total compensation price would be $ 167,311.04. The explanatory variables in our model measures 58.99% of the variation in price, which is significantly higher than AVM. Therefore, I have great confidence that this model will eliminate vulnerability to any future compensation claims. Please contact us with any further questions.

Sincerely,

George Canning

Word count: 293