A Multiple Regression Model for Predicting the Property Price around Kowloon Station in Hong Kong

YU Bangning, LIU Xiaohan, ZHU Yizhu

Abstract

This model is for predicting the price of properties around Kowloon Station in Hong Kong, with the command of both practical and theoretical purposes, through finding the relationship among the existing property transaction prices, property characteristics and the infrastructure level in the neighborhood.

Introduction

Due to the distance between domicile and employment, shopping and other necessary sites, transportation has a strong positive influence on property price (Du& Mulley, 2007；McMillen& McDonald, 2004). Accessibility has a relatively large influence on the property price, and this will count a huge percentage when residents purchasing properties.

Besides infrastructure, other factors including floor and property age cannot be neglected as well.

The Pearson Correlation of the independent variable floor level (F) and the property price is 0.608. The matrix scatter-plot shows that there is a linear trend between the observations. The general trend is 'the higher the floor, the higher the price'. According to research done by Apple Daily in 2013, one of the most important reasons is that higher floors enjoy more beautiful views. In addition, due to the wet weather in Hong Kong, it is better for ventilation.

The tendency of the Pearson Correlation shows that the property price tends to go up then fall down as the property age growth. We further deduce that the second-hand properties with complete decoration but not too old are the most welcome-choices among Hong Kong residents. This deduction could also be confirmed by the chart below, showing that the second-hand property transaction takes up 70.49% of the entire property transaction.

土地註冊處各類物業買賣註冊統計 – 近12個月

			2018年 4月	5月	6月	7月	8月	9月	10月	11月	12月	2019年 1月	2月	3月 (截至6日)	2019年3月 各類型物業 佔整體之百分率	2019年 (截至3月6日) 總計 宗數	金額(百萬元)
整體物業註冊宗數			8,741	7,788	9,252	8,466	6,688	4,799	5,374	3,953	3,038	5,589	4,999	688		11,276	119,379.2
整體物業註冊金額 (百萬元)			77,838	63,489	80,777	99,367	60,944	45,487	45,536	43,325	43,044	71,981	41,665	5,733			
一手住宅 註冊宗數	私人 住宅	香港	336	229	295	519	175	33	103	25	25	28	30	9		67	3,774.0
		九龍	117	194	270	489	96	622	182	83	82	1,531	340	13		1,884	26,770.1
		新界	1,674	352	1,512	752	1,226	608	2,072	621	290	638	1,290	93		2,021	17,179.7
		總數	2,127	775	2,077	1,760	1,497	1,263	2,357	729	397	2,197	1,660	115	16.72%	3,972	47,723.9
	公營 房屋	香港	0	0	0	0	0	0	0	0	0	0	0	0		0	0.0
		九龍	0	0	0	0	0	2	6	6	7	15	15	2		32	101.3
		新界	69	72	34	54	42	3	8	3	0	1	1	0		2	4.7
		總數	69	72	34	54	42	5	14	9	7	16	16	2	0.29%	34	106.0
二手住宅 註冊宗數		香港	1,009	1,129	1,054	1,001	808	524	441	523	388	470	506	102		1,078	11,087.3
		九龍	1,333	1,352	1,328	1,258	928	677	574	526	478	811	680	139		1,630	11,160.1
		新界	2,315	2,495	2,441	2,227	1,701	1,115	942	1,009	1,043	1,335	1,447	244		3,026	18,259.2
		總數	4,657	4,976	4,823	4,486	3,437	2,316	1,957	2,058	1,909	2,616	2,633	485	70.49%	5,734	40,506.6

The estimation is based on the record of the transaction in the past two years of Kowloon Station District, which was randomly picked from online lucky draw. After scrapped the data, we deleted the data which has a null value.

The factors we supposed at beginning including :
(i) the trading price(P);
(ii) floor level(F),
(iii) building age(A);
(iii) distance to the nearest railway station(M),
(iv) distance to the nearest hospital(H);
(vi) distance to the nearest supermarket(SU);
(vii) distance to the nearest school (primary or secondary school)(SC);
(viii) distance to the nearest Starbucks(ST).

Data Selection, Preprocessing, Analysis

The data we used are collected from (i) Centaline Property Agency web page; and web page of some main supermarket in Hong Kong, consisting of (ii) ParknShop, (iii) Manning, (iv) Wellcome; (v) Census and Statistics Department, and (vi) Google Maps with the help of following tools.

Tools:

(i) Pycharm -- Python:

To access longitude and latitude of the properties/ schools/ hospitals/ Starbucks

(ii) Eclipse -- Java:

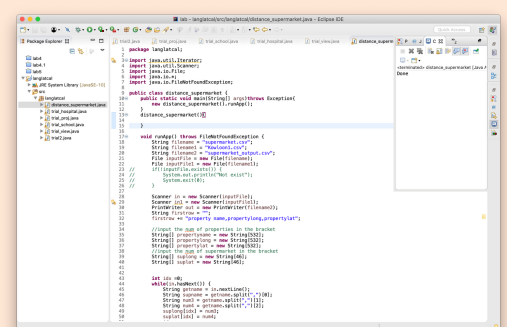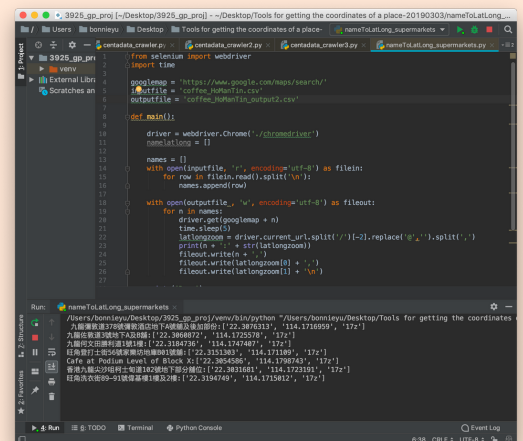To match and calculate the distance of properties with other places

(iii) Web scraper:

Web scraping plugin to get simple data

(iv) OpenRefine:

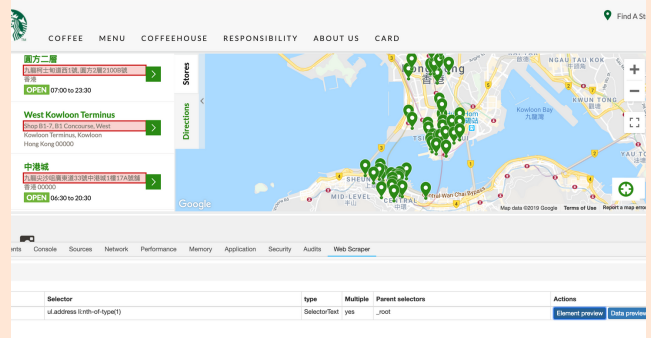To clean syntax(eg. $, /) and characters in the dataset. insights.

(v)SPSS

For data analysis
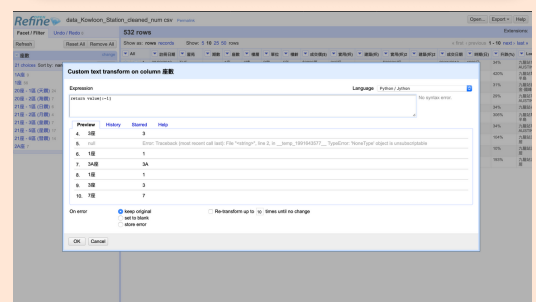
We scraped the transaction records over the past 2 years from the Centadata website which includes floor, property age already, and matched them with MTR station location to find the nearest MTR station. And we further searched the information about supermarkets, schools, hospitals and Starbucks branches through the official and verified websites.

To measure the accessibility, we use the distance between certain property and the referred place. The calculation is shown in the picture below.

```java
public double distance(String num1, String num2, String num3, String num4) {
    Double N1 = Double.valueOf(num1);
    Double N2 = Double.valueOf(num2);
    Double N3 = Double.valueOf(num3);
    Double N4 = Double.valueOf(num4);
    double slat = Math.toRadians(N3);
    double slon = Math.toRadians(N4);
    double elat = Math.toRadians(N1);
    double elon = Math.toRadians(N2);
    double distance = 6371.01*Math.acos(Math.sin(slat)*Math.sin(elat) +
            Math.cos(slat)*Math.cos(elat)*Math.cos(slon-elon));
    return distance;
}
```

The dataset(N = 532) contains 21 null values so the size of analyzed data is N = 512.

N		
	實用呎價格	512
	樓層	512
	樓齡	512
	最近地鐵站相距	512
	最近超市距離	512
	最近学校距離	512
	最近医院距離	512
	最近starbucks距離	512

Material & Methods

Descriptive Statistics

	Mean	Std. Deviation	N
實用呎價格	31447.38	11634.177	512
樓層	35.51	21.590	512
樓齡	11.95	4.359	512
最近地鐵站相距	.25705103479	.139545835893	512
最近超市距離	.50304688979	.194358553614	512
最近学校距離	.55872943013	.209564480715	512
最近医院距離	.84775860595	.199804755035	512
最近starbucks距離	.57398031290	.213128116681	512

The descriptive data is shown as above.

Model Summary[e]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.608[a]	.370	.369	9243.668	
2	.673[b]	.454	.451	8616.735	
3	.732[c]	.536	.534	7945.595	
4	.757[d]	.573	.569	7634.055	1.654

a. Predictors: (Constant), 樓層

b. Predictors: (Constant), 樓層, 最近医院距離

c. Predictors: (Constant), 樓層, 最近医院距離, 樓齡

d. Predictors: (Constant), 樓層, 最近医院距離, 樓齡, 最近starbucks距離

e. Dependent Variable: 實用呎價格

SPSS provided information about four models including R, the multiple correlation coefficient, R^2, standard error of the estimate and Durbin-Waston value.

The R of the fourth model = 0.757 > 0.7, which indicates most of the data located around the regression line. With larger R, the fourth value taking (i) floor, (ii) nearest hospital, (iii) property age, (iv) distance to nearest Starbucks as independent variables(IV) is proved to be the best model.

Also, we adopted the Durbin-Waston statistics to test the autocorrelation in the residuals. Since the value = 1.654, which is in between 1.5 and 2.5 and relatively closer to 2, it proves the autocorrelation in the model is acceptable.

Since three variables in the original dataset were excluded, we looked at the scatter-plot matrix to find the reason.

The excluded variables:
(i) distance to the nearest MTR stations (M);
(ii) distance to nearest school (SC);
(iii) distance to nearest supermarket (SU);

Here is the scatter-plot matrix and correlation table of the variables:
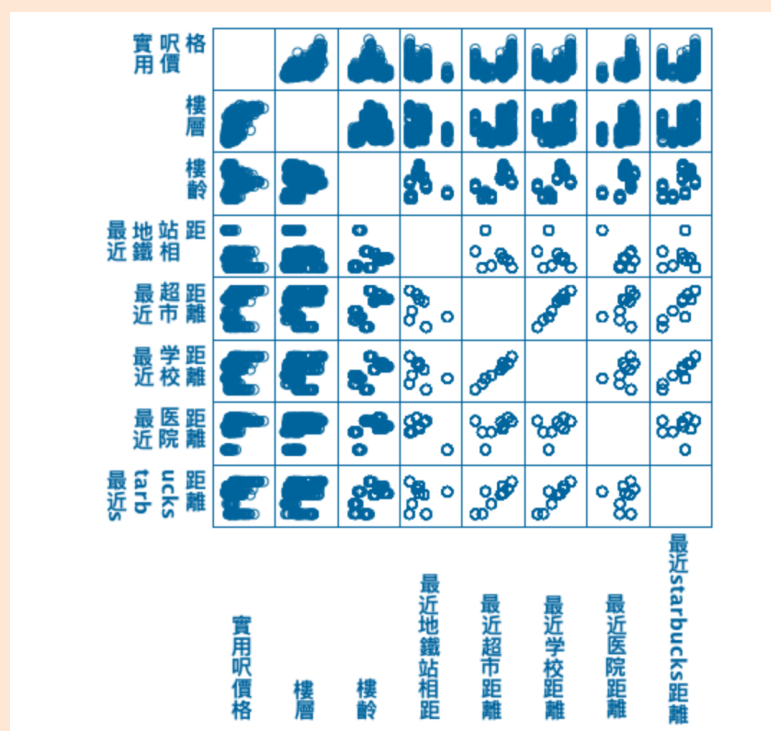
Correlations

		實用呎價格	樓層	樓齡	最近地鐵站相距	最近超市距離	最近學校距離	最近医院距離	最近starbucks距離
Pearson Correlation	實用呎價格	1.000	.608	-.050	-.353	.255	.236	.477	.141
	樓層	.608	1.000	.110	-.208	.210	.237	.337	.220
	樓齡	-.050	.110	1.000	-.274	.631	.586	.478	.450
	最近地鐵站相距	-.353	-.208	-.274	1.000	-.560	-.523	-.733	-.130
	最近超市距離	.255	.210	.631	-.560	1.000	.979	.500	.827
	最近學校距離	.236	.237	.586	-.523	.979	1.000	.390	.899
	最近医院距離	.477	.337	.478	-.733	.500	.390	1.000	.047
	最近starbucks距離	.141	.220	.450	-.130	.827	.899	.047	1.000

According to the Pearson correlation result, there is strong correlation existing between M and H, up to -73.3%. And the correlation is even strong regarding SC, SU, and ST, all around 90%. We assume the three elements are excluded the high correlation among variables though they have a relatively stronger relationship with regression linear separately, according to the first column.
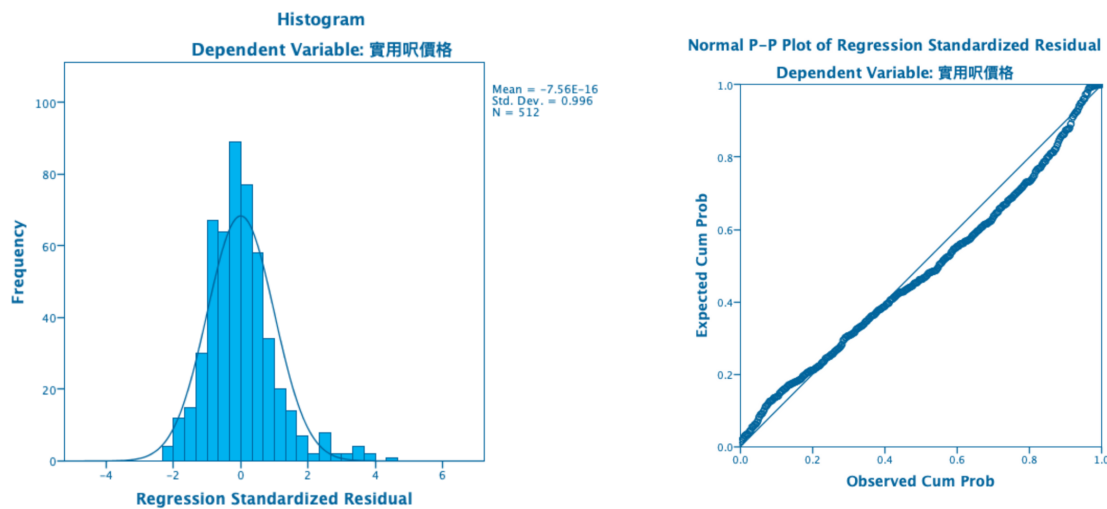
Coefficients^a

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	Upper Bound	Correlations Zero-order	Partial	Part	Collinearity Statistics Tolerance	VIF
1	(Constant)	19808.114	786.926		25.172	.000	18262.099	21354.129					
	樓層	327.758	18.940	.608	17.305	.000	290.549	364.968	.608	.608	.608	1.000	1.000
2	(Constant)	6626.519	1663.798		3.983	.000	3357.762	9895.275					
	樓層	271.900	18.755	.505	14.497	.000	235.054	308.747	.608	.541	.475	.886	1.128
	最近医院距離	17888.592	2026.621	.307	8.827	.000	13907.021	21870.164	.477	.364	.289	.886	1.128
3	(Constant)	9390.093	1561.433		6.014	.000	6322.431	12457.754					
	樓層	261.602	17.328	.485	15.097	.000	227.559	295.646	.608	.557	.456	.883	1.133
	最近医院距離	27398.277	2119.025	.471	12.930	.000	23235.146	31561.409	.477	.498	.391	.689	1.451
	樓齡	-875.639	91.985	-.328	-9.519	.000	-1056.356	-694.922	-.050	-.389	-.288	.768	1.301
4	(Constant)	3993.151	1709.732		2.336	.020	634.120	7352.183					
	樓層	229.179	17.362	.425	13.200	.000	195.068	263.290	.608	.506	.383	.812	1.232
	最近医院距離	31590.007	2133.251	.543	14.808	.000	27398.906	35781.108	.477	.549	.430	.628	1.593
	樓齡	-1223.830	103.005	-.459	-11.881	.000	-1426.199	-1021.460	-.050	-.467	-.345	.566	1.768
	最近starbucks距離	12463.851	1893.943	.228	6.581	.000	8742.909	16184.794	.141	.281	.191	.700	1.429

a. Dependent Variable: 實用呎價格

Since the significances(Sig.) of all variables are all smaller than 0.05, it proves the model is generally correct. And here we use the unstandardized coefficients to get the formula as following:

Predicted property price = 299.179*F + 31590.007*H - 1233.830*A +12463.851*ST + 3993.151

Relevant Test

the residual of the predicted (P) trend fitted normal distribution, and the point pattern on the P-P plot is linear through the origin and has the unit slope.

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	19808.114	786.926		25.172	.000	18262.099	21354.129					
	樓層	327.758	18.940	.608	17.305	.000	290.549	364.968	.608	.608	.608	1.000	1.000
2	(Constant)	6626.519	1663.798		3.983	.000	3357.762	9895.275					
	樓層	271.900	18.755	.505	14.497	.000	235.054	308.747	.608	.541	.475	.886	1.128
	最近医院距離	17888.592	2026.621	.307	8.827	.000	13907.021	21870.164	.477	.364	.289	.886	1.128
3	(Constant)	9390.093	1561.433		6.014	.000	6322.431	12457.754					
	樓層	261.602	17.328	.485	15.097	.000	227.559	295.646	.608	.557	.456	.883	1.133
	最近医院距離	27398.277	2119.025	.471	12.930	.000	23235.146	31561.409	.477	.498	.391	.689	1.451
	樓齡	-875.639	91.985	-.328	-9.519	.000	-1056.356	-694.922	-.050	-.389	-.288	.768	1.301
4	(Constant)	3993.151	1709.732		2.336	.020	634.120	7352.183					
	樓層	229.179	17.362	.425	13.200	.000	195.068	263.290	.608	.506	.383	.812	1.232
	最近医院距離	31590.007	2133.251	.543	14.808	.000	27398.906	35781.108	.477	.549	.430	.628	1.593
	樓齡	-1223.830	103.005	-.459	-11.881	.000	-1426.199	-1021.460	-.050	-.467	-.345	.566	1.768
	最近starbucks距離	12463.851	1893.943	.228	6.581	.000	8742.909	16184.794	.141	.281	.191	.700	1.429

a. Dependent Variable: 實用呎價格

Here we use the tolerance and VIF to measure its multicollinearity. According to Hair(2010), if the VIF exceeds 4.0 or the tolerance is less than 0.2, there is problem with the multicollinearity in the dataset. Since here all VIF are less than 2.0 and tolerance all exceeding 0.2, there should be no collinearity.

References

- 丰雷, 刘迎梅, & 包晓辉. (2011). 公共交通易达性对香港房价的影响分析——Hedonic 模型的应用 (Doctoral dissertation).
- 教育局分區學校名冊. (n.d.). Retrieved from https://www.edb.gov.hk/tc/student-parents/sch-info/sch-search/schlist-by-district/index.html
- 分店位置. (n.d.). Retrieved from https://www.mannings.com.hk/zh-hk/store-locator
- 店舖一覽 | 百佳網上超級市場. (n.d.). Retrieved from https://www.parknshop.com/zh-hk/storeLocator
- 惠康為您送. (n.d.). Retrieved from https://www.wellcome.com.hk/wd2shop/zh/html/customer-services/store-locator.html
- 教育局分區學校名冊. (n.d.). Retrieved from https://www.edb.gov.hk/tc/student-parents/sch-info/sch-search/schlist-by-district/index.html
- GovHK 香港政府一站通：醫院及診所. (2018, 11). Retrieved from https://www.gov.hk/tc/residents/health/hosp/index.htm
- 中原地產2017-2019九龍站附近房屋交易記錄 http://hk.centadata.com/ptest.aspx?type=22&code=202&ref=CD2_Detail_Nav
- 港樓價，高一層貴兩萬？. (2013, July 9). Retrieved from https://hk.news.appledaily.com/local/daily/article/20130709/18328081
- 美聯物業土地註冊處各類物業買賣註冊統計 - 近12個月. (n.d.). Retrieved from https://www.midland.com.hk/land-registry-record/12months.html

Work Distribution:

Data collection and analysis:

YU Bangning

LIU Xiaohan

Text:

ZHU Yizhu

LIU Xiaohan

YU Bangning