Mengjie Shen

Jeffrey Simonoff

Regression and Multivariate Data Analysis

April 28, 2021

Modeling Hotel Prices

If we are traveling to a new city, we always hope to get the hotels with better living experience and more convenient location within the price range that we could accept. It would be interesting to know what factors can have an influence on the price of hotels. This will help us choose better hotels with lower costs.

I acquired data from one of the most famous Chinese hotels booking websites Ctrip. (https://hotels.ctrip.com/hotels/). I focus on the price of a king-bed room at each hotel on the night of May 10. It is worth noting that hotel prices may fluctuate on different dates and all data are collected on April 25th. I collected thirty-nine hotels in Shanghai from four different areas that tourists usually choose (Disney Resort, Hongqiao, Jing'an Temple, Lujiazui) in Shanghai. For each hotel, I also recorded the star rating (three-star, four-star, five-star). The following statistics summarize the price for different star ratings in different areas:

TESTLEVENE

Descriptive Statistics: Price

Results for Area = Disney Resort

Statistics

Variable	Stars	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	five	2	0	722	118	166	604	*	722	*	839
	four	2	0	481.0	78.0	110.3	403.0	*	481.0	*	559.0
	three	2	0	312	116	163	196	*	312	*	427

Results for Area = Hongqiao

Statistics

Variable	Stars	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	five	4	0	787.0	49.9	99.9	701.0	702.0	773.5	885.5	900.0
	four	3	0	435.3	34.4	59.6	381.0	381.0	426.0	499.0	499.0
	three	3	0	302.3	30.3	52.5	263.0	263.0	282.0	362.0	362.0

Results for Area = Jingan Temple

Statistics

Variable	Stars	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	five	4	0	1435	187	373	880	1048	1600	1657	1660
	four	3	0	489.7	40.7	70.5	418.0	418.0	492.0	559.0	559.0
	three	4	0	404.0	45.5	91.0	314.0	326.0	388.0	498.0	526.0

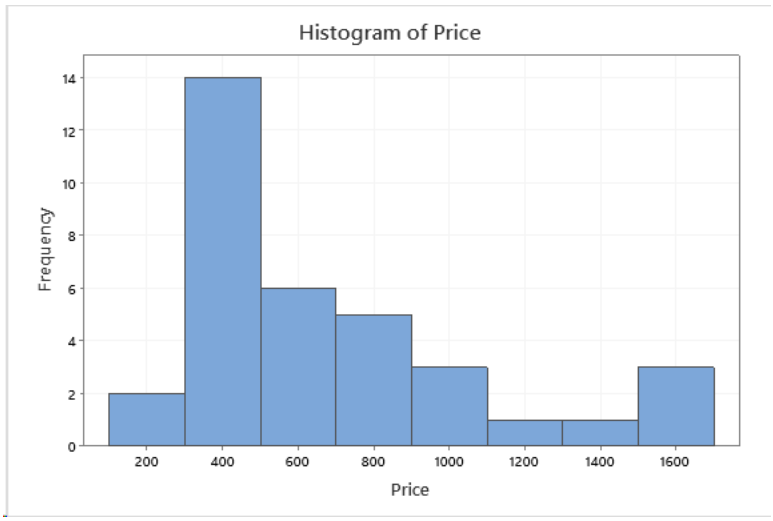
Results for Area = Lujiazui

Statistics

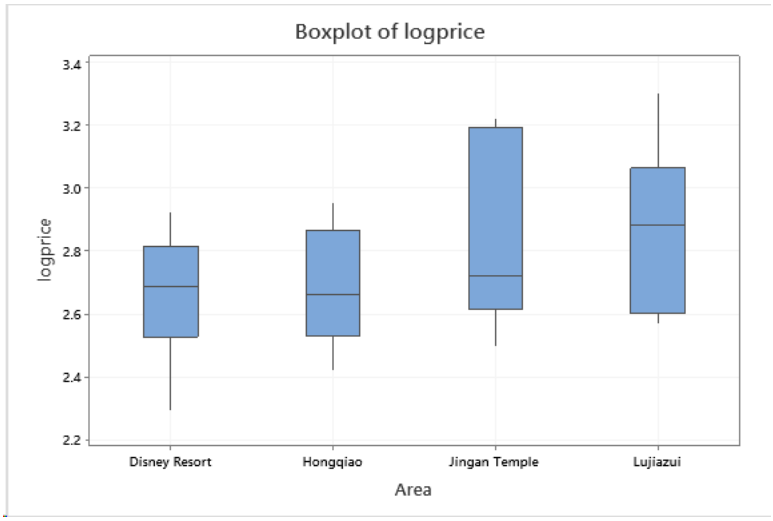
Variable	Stars	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	five	5	0	1322	177	396	1018	1043	1190	1668	1994
	four	4	0	638.0	81.6	163.2	462.0	486.0	627.0	801.0	836.0
	three	3	0	373.67	3.18	5.51	370.00	370.00	371.00	380.00	380.00

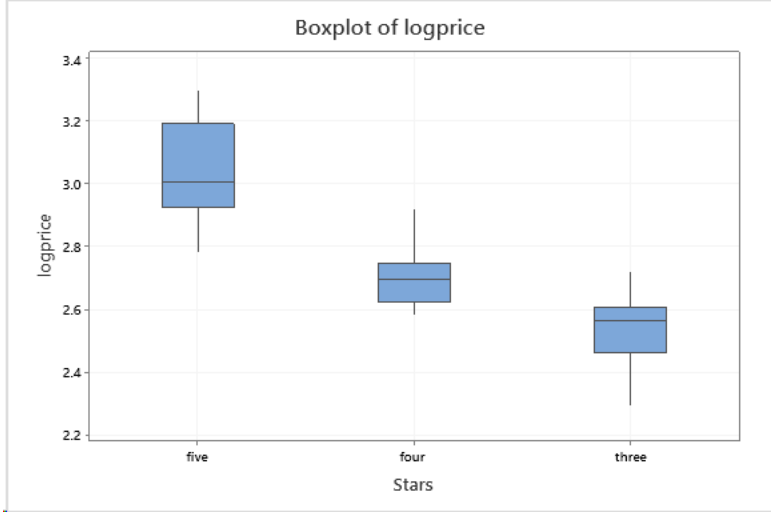
The pattern that the price of five-star hotels is the most expensive and the price of three-star hotels is the cheapest is quite consistent throughout each area. This is also in line with our common sense that five-star hotels are usually luxurious and three-star hotels are more economical. In each area, the price of a five-star hotel is a higher standard deviation and a three-star hotel has the lowest.

Here is the histogram of the price. We can witness the pattern of right tail. Besides, price is also a monetary variable. This suggests we take logs on our response variable.



Here are the side-by-side boxplots. There are area and star rating effects. In terms of star rating, five-star hotels have the highest price, four-star hotels have a relatively lower price, and three-star hotels have the lowest price. Hotels in Lujiazui are more expensive, and the prices of hotels in the area of Jing'an Temple are relatively lower. Hotels in the area of Hongqiao and Disney Resort are similar and the cheapest. We can see there are different amounts of variability in price, especially across different areas.





Here is a table that cross-classifies the hotels by star rating and areas. Since there are no holes in the table, we will continue to use the data to fit the two-way ANOVA with interaction effects.

Rows: Stars Columns: Area

	Disney Resort	Hongqiao	Jingan Temple	Lujiazui	All
five	2	4	4	5	15
four	2	3	3	4	12
three	2	3	4	3	12
All	6	10	11	12	39

Cell Contents
Count

Here is the two-way ANOVA for the 39 three-star, four-star, five-star hotels in areas of Disney Resort, Hongqiao, Jing'an Temple and Lujiazui.

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Area	Fixed	4	Disney Resort, Hongqiao, Jingan Temple, Lujiazui
Stars	Fixed	3	five, four, three

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	0.21156	0.07052	6.68	0.002
Stars	2	1.41745	0.70873	67.16	0.000
Area*Stars	6	0.06976	0.01163	1.10	0.387
Error	27	0.28492	0.01055		
Total	38	2.26238			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.102726	87.41%	82.28%	69.99%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.7415	0.0172	159.46	0.000	
Area					
Disney Resort	-0.0782	0.0343	-2.28	0.031	1.90
Hongqiao	-0.0729	0.0289	-2.53	0.018	1.73
Jingan Temple	0.0679	0.0280	2.43	0.022	1.71
Stars					
five	0.2576	0.0237	10.89	0.000	1.42
four	-0.0432	0.0246	-1.75	0.091	1.38
Area*Stars					
Disney Resort five	-0.0686	0.0481	-1.42	0.166	2.61
Disney Resort four	0.0562	0.0486	1.16	0.258	2.46
Hongqiao five	-0.0328	0.0392	-0.84	0.409	2.18
Hongqiao four	0.0107	0.0416	0.26	0.798	2.13
Jingan Temple five	0.0760	0.0386	1.97	0.059	2.24
Jingan Temple four	-0.0794	0.0410	-1.94	0.063	2.21

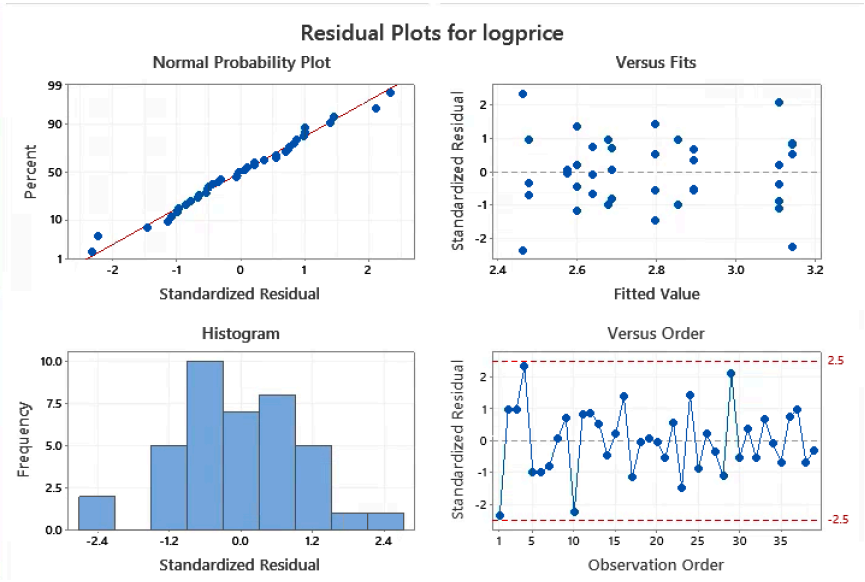
Regression Equation

logprice = 2.7415 - 0.0782 Area_Disney Resort - 0.0729 Area_Hongqiao + 0.0679 Area_Jingan Temple + 0.0832 Area_Lujiazui + 0.2576 Stars_five - 0.0432 Stars_four
- 0.2144 Stars_three - 0.0686 Area*Stars_Disney Resort five
+ 0.0562 Area*Stars_Disney Resort four + 0.0124 Area*Stars_Disney Resort three
- 0.0328 Area*Stars_Hongqiao five + 0.0107 Area*Stars_Hongqiao four
+ 0.0221 Area*Stars_Hongqiao three + 0.0760 Area*Stars_Jingan Temple five
- 0.0794 Area*Stars_Jingan Temple four + 0.0034 Area*Stars_Jingan Temple three
+ 0.0254 Area*Stars_Lujiazui five + 0.0125 Area*Stars_Lujiazui four
- 0.0379 Area*Stars_Lujiazui three

Means

Term	Fitted Mean	SE Mean
Area		
Disney Resort	2.6634	0.0419
Hongqiao	2.6686	0.0328
Jingan Temple	2.8095	0.0313
Lujiazui	2.8247	0.0303
Stars		
five	2.9991	0.0281
four	2.6983	0.0306
three	2.5271	0.0306
Area*Stars		
Disney Resort five	2.8524	0.0726
Disney Resort four	2.6764	0.0726
Disney Resort three	2.4613	0.0726
Hongqiao five	2.8934	0.0514
Hongqiao four	2.6361	0.0593
Hongqiao three	2.4763	0.0593
Jingan Temple five	3.1431	0.0514
Jingan Temple four	2.6869	0.0593
Jingan Temple three	2.5984	0.0514
Lujiazui five	3.1077	0.0459
Lujiazui four	2.7940	0.0514
Lujiazui three	2.5725	0.0593

We can see that the interaction effect is not statistically significant.



There are no outliers since the standardized residual for every observation is under 2.5, but residual plots show that there is a non-normality and non-constant pattern. Since the interaction effect is not significant here, we rerun the two-way ANOVA without the interaction effect.

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Area	Fixed	4	Disney Resort, Hongqiao, Jingan Temple, Lujiazui
Stars	Fixed	3	five, four, three

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	0.23889	0.07963	7.41	0.001
Stars	2	1.63323	0.81662	75.98	0.000
Error	33	0.35468	0.01075		
Lack-of-Fit	6	0.06976	0.01163	1.10	0.387
Pure Error	27	0.28492	0.01055		
Total	38	2.26238			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.103672	84.32%	81.95%	77.63%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.7432	0.0173	158.77	0.000	
Area					
Disney Resort	-0.0799	0.0346	-2.31	0.027	1.90
Hongqiao	-0.0788	0.0289	-2.73	0.010	1.70
Jingan Temple	0.0729	0.0281	2.60	0.014	1.69
Stars					
five	0.2663	0.0228	11.69	0.000	1.29
four	-0.0489	0.0241	-2.03	0.050	1.29

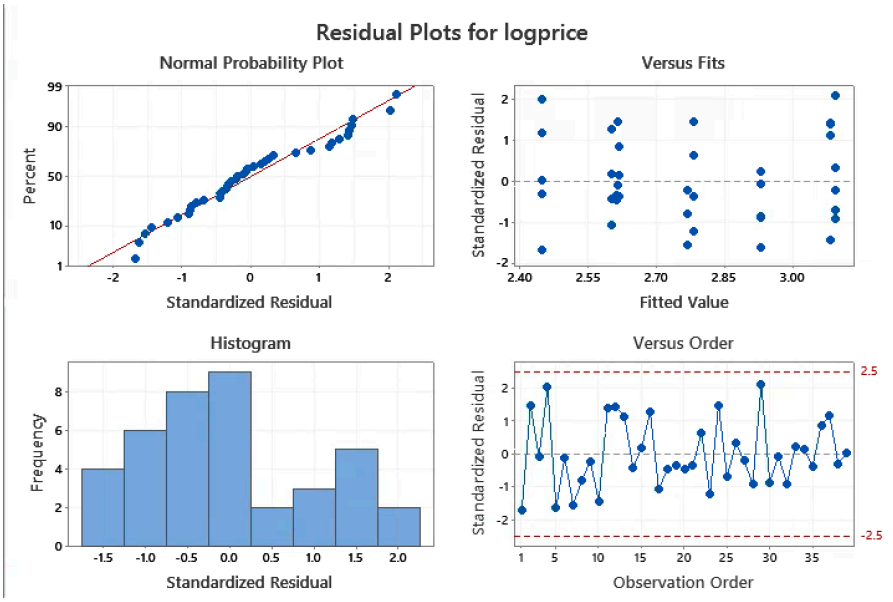
Regression Equation

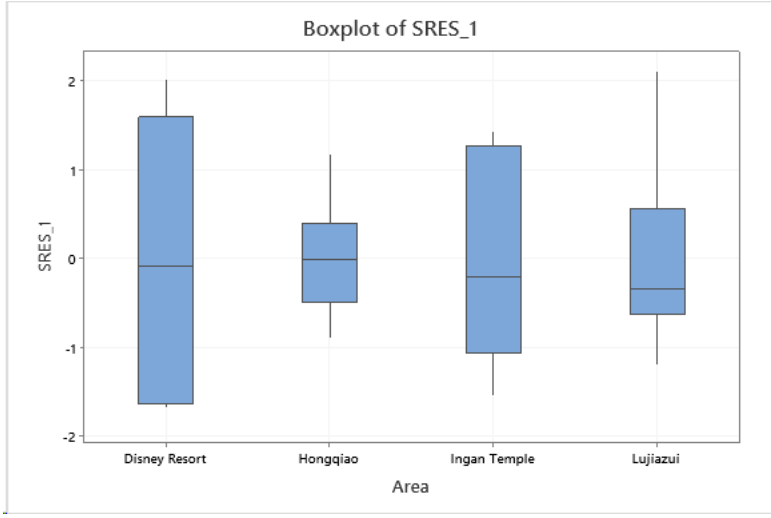
logprice = 2.7432 - 0.0799 Area_Disney Resort - 0.0788 Area_Hongqiao + 0.0729 Area_Jingan Temple + 0.0858 Area_Lujiazui + 0.2663 Stars_five - 0.0489 Stars_four - 0.2174 Stars_three

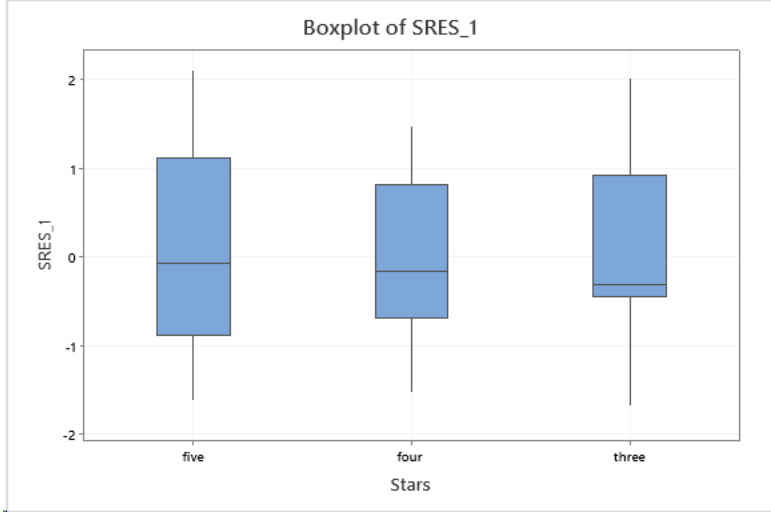
Means

Term	Fitted Mean	SE Mean
Area		
Disney Resort	2.6634	0.0423
Hongqiao	2.6644	0.0329
Jingan Temple	2.8162	0.0313
Lujiazui	2.8290	0.0301
Stars		
five	3.0096	0.0273
four	2.6943	0.0302
three	2.5258	0.0302

The two main effects are highly statistically significant. There are no outliers showed in residual plots. We can witness the pattern of non-normality and some evidence of nonconstant variance. The boxplot of standardized residual for areas shows the pattern of non-constant variance.







We can apply Levene's test to see if the nonconstant variance is indicated by it. We will first try the Levene's test including the interaction effect since the factors driving potential nonconstant variance do not necessarily have anything to do with the factors driving the relationship with the response variable.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	2.0968	0.69893	2.21	0.110
Stars	2	0.1325	0.06625	0.21	0.812
Area*Stars	6	2.5943	0.43238	1.37	0.263
Error	27	8.5292	0.31590		
Total	38	13.4070			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.562046	36.38%	10.46%	0.00%

We can see that the p-value for the interaction effect is insignificant. Therefore, we rerun the Levene's test with only the main effects.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	2.1518	0.71728	2.13	0.115
Stars	2	0.1862	0.09310	0.28	0.760
Error	33	11.1235	0.33708		
Lack-of-Fit	6	2.5943	0.43238	1.37	0.263
Pure Error	27	8.5292	0.31590		
Total	38	13.4070			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.580582	17.03%	4.46%	0.00%

Surprisingly, the p-value for both main effects is insignificant. It doesn't indicate the pattern of non-constant variance. This result doesn't agree with the pattern we observed from the box plots.

Even though the pattern of heteroscedasticity is not consistent, we still try to use weighted least squares. I will stick with logged price here. Since the box plot for standardized residual shows that only area matters, I calculate the standard deviations of the standardized residuals separated by area only and use the inverse of the squared entries given under StDev as the weight.

Statistics

Variable	Area	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
SRES	Disney Resort	6	0	0.001	0.623	1.525	-1.679	-1.634	-0.082	1.593
	Hongqiao	10	0	0.002	0.210	0.663	-0.884	-0.488	-0.011	0.400
	Jingan Temple	11	0	-0.003	0.347	1.152	-1.533	-1.059	-0.208	1.272
	Lujiazui	12	0	-0.001	0.282	0.978	-1.199	-0.632	-0.340	0.569

Variable	Area	Maximum
SRES	Disney Resort	2.016
	Hongqiao	1.175
	Jingan Temple	1.428
	Lujiazui	2.102

We first run the WLS including the interaction effect:

Method

Factor coding	(-1, 0, +1)
Weights	wt

Factor Information

Factor	Type	Levels	Values
Area	Fixed	4	Disney Resort, Hongqiao, Jingan Temple, Lujiazui
Stars	Fixed	3	five, four, three

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	0.25353	0.084511	8.87	0.000
Stars	2	0.98002	0.490009	51.45	0.000
Area*Stars	6	0.05990	0.009983	1.05	0.417
Error	27	0.25712	0.009523		
Total	38	2.44336			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0975864	89.48%	85.19%	77.87%

Coefficients

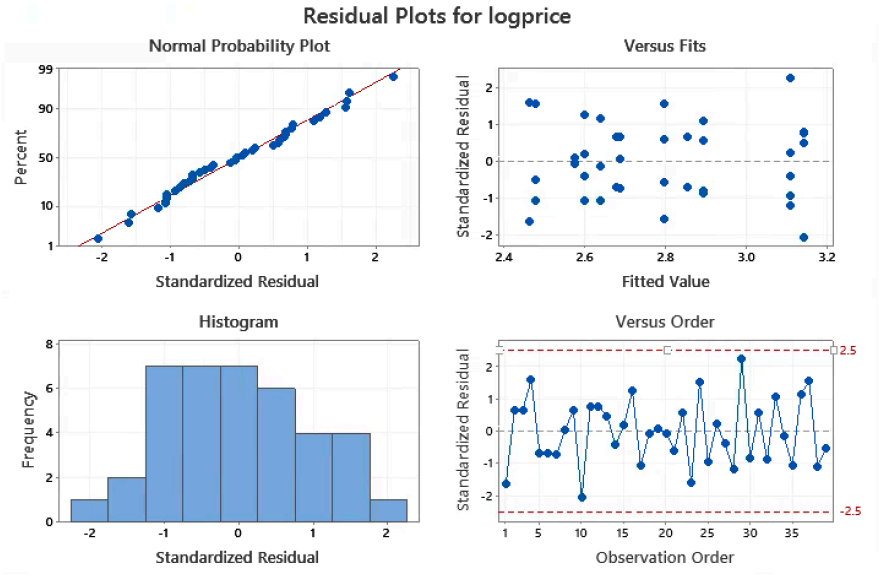
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.7415	0.0195	140.63	0.000	
Area					
Disney Resort	-0.0782	0.0472	-1.66	0.109	3.03
Hongqiao	-0.0729	0.0244	-2.99	0.006	2.06
Jingan Temple	0.0679	0.0311	2.19	0.038	2.07
Stars					
five	0.2576	0.0271	9.49	0.000	2.45
four	-0.0432	0.0278	-1.55	0.132	2.28
Area*Stars					
Disney Resort five	-0.0686	0.0665	-1.03	0.312	4.64
Disney Resort four	0.0562	0.0668	0.84	0.408	4.23
Hongqiao five	-0.0328	0.0335	-0.98	0.336	2.87
Hongqiao four	0.0107	0.0349	0.31	0.761	2.68
Jingan Temple five	0.0760	0.0430	1.77	0.088	2.77
Jingan Temple four	-0.0794	0.0454	-1.75	0.092	2.71

Regression Equation

logprice = 2.7415 - 0.0782 Area_Disney Resort - 0.0729 Area_Hongqiao + 0.0679 Area_Jingan Temple + 0.0832 Area_Lujiazui + 0.2576 Stars_five - 0.0432 Stars_four - 0.2144 Stars_three - 0.0686 Area*Stars_Disney Resort five + 0.0562 Area*Stars_Disney Resort four + 0.0124 Area*Stars_Disney Resort three - 0.0328 Area*Stars_Hongqiao five + 0.0107 Area*Stars_Hongqiao four + 0.0221 Area*Stars_Hongqiao three + 0.0760 Area*Stars_Jingan Temple five - 0.0794 Area*Stars_Jingan Temple four + 0.0034 Area*Stars_Jingan Temple three + 0.0254 Area*Stars_Lujiazui five + 0.0125 Area*Stars_Lujiazui four - 0.0379 Area*Stars_Lujiazui three

Means

Term	Fitted Mean	SE Mean
Area		
Disney Resort	2.6634	0.0608
Hongqiao	2.6686	0.0206
Jingan Temple	2.8095	0.0342
Lujiazui	2.8247	0.0282
Stars		
five	2.9991	0.0327
four	2.6983	0.0344
three	2.5271	0.0342
Area*Stars		
Disney Resort five	2.852	0.105
Disney Resort four	2.676	0.105
Disney Resort three	2.461	0.105
Hongqiao five	2.8934	0.0323
Hongqiao four	2.6361	0.0374
Hongqiao three	2.4763	0.0374
Jingan Temple five	3.1431	0.0562
Jingan Temple four	2.6869	0.0649
Jingan Temple three	2.5984	0.0562
Lujiazui five	3.1077	0.0427
Lujiazui four	2.7940	0.0477
Lujiazui three	2.5725	0.0551



The interaction effect is still not statistically significant, therefore, we won't check the assumptions right now and continue to rerun the WLS without the interaction effect:

Method

Factor coding (-1, 0, +1)
Weights wt

Factor Information

Factor	Type	Levels	Values
Area	Fixed	4	Disney Resort, Hongqiao, Jingan Temple, Lujiazui
Stars	Fixed	3	five, four, three

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	0.29213	0.097378	10.14	0.000
Stars	2	1.81384	0.906919	94.41	0.000
Error	33	0.31702	0.009607		
Lack-of-Fit	6	0.05990	0.009983	1.05	0.417
Pure Error	27	0.25712	0.009523		
Total	38	2.44336			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0980137	87.03%	85.06%	81.99%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.7439	0.0195	140.54	0.000	
Area					
Disney Resort	-0.0805	0.0474	-1.70	0.098	3.03
Hongqiao	-0.0785	0.0243	-3.23	0.003	2.03
Jingan Temple	0.0726	0.0310	2.34	0.025	2.05
Stars					
five	0.2567	0.0197	13.03	0.000	1.28
four	-0.0448	0.0210	-2.14	0.040	1.28

Regression Equation

logprice = 2.7439 - 0.0805 Area_Disney Resort - 0.0785 Area_Hongqiao + 0.0726 Area_Jingan Temple + 0.0864 Area_Lujiazui + 0.2567 Stars_five - 0.0448 Stars_four - 0.2118 Stars_three

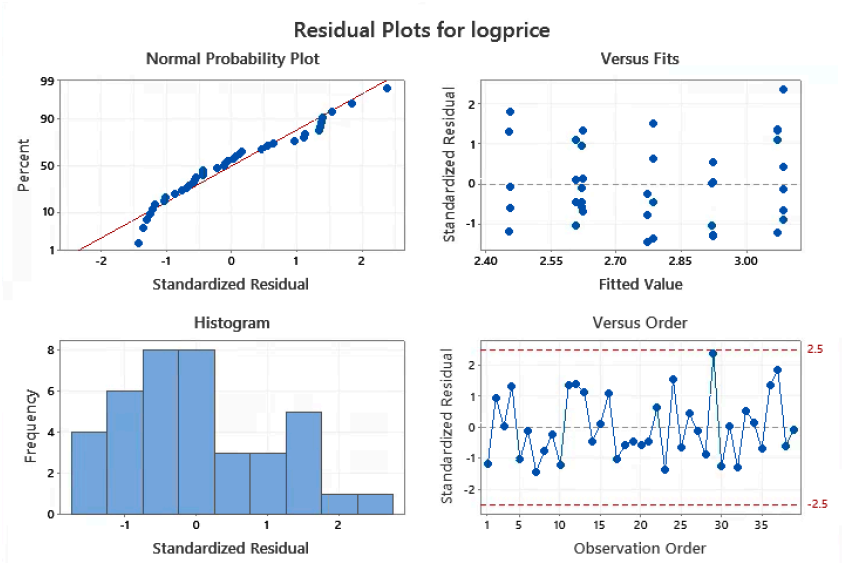
Means

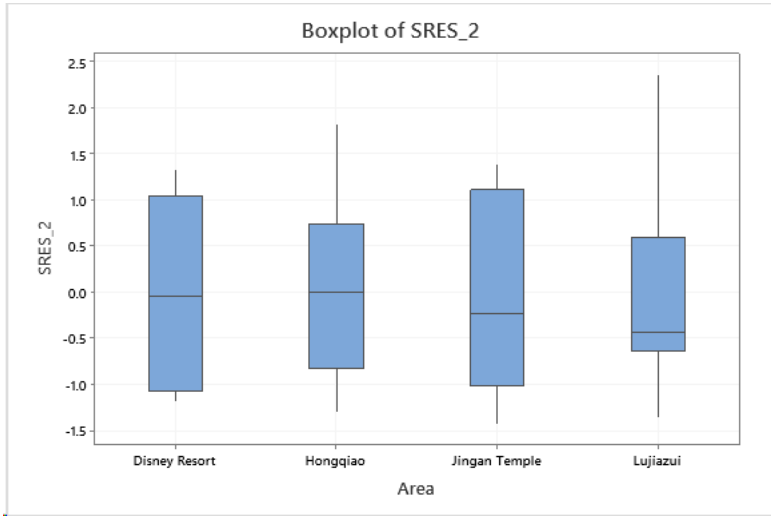
Term	Fitted Mean	SE Mean
Area		
Disney Resort	2.6634	0.0610
Hongqiao	2.6654	0.0206
Jingan Temple	2.8165	0.0341
Lujiazui	2.8303	0.0278
Stars		
five	3.0006	0.0268
four	2.6991	0.0291
three	2.5321	0.0292

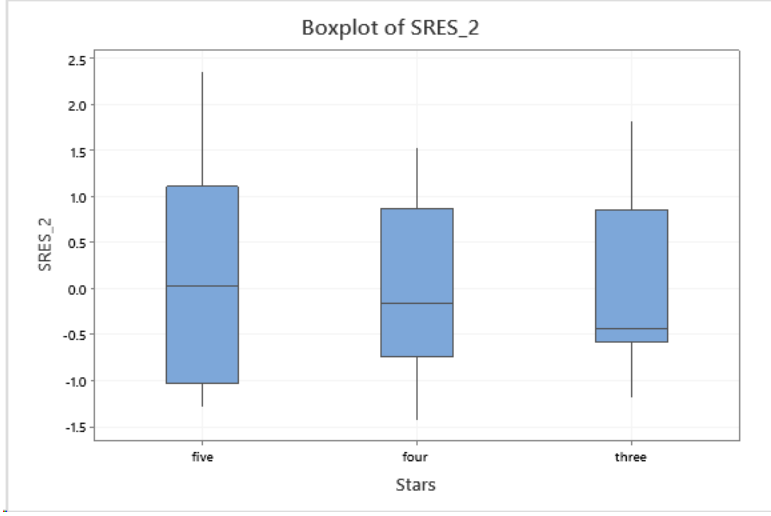
Here, the means are in the logged scale, so we will antilog them before we look into them. For example, the "typical" price for a hotel in the Disney Resort area is ¥460.68 ($10^{2.6634}$). The "typical" price for a three-star hotel is ¥340.49 ($10^{2.5321}$). The

differences between coefficients can be interpreted as estimated multiplicative differences. For example, a five-star hotel is estimated to cost 2.94 times of a three-star hotel ($10^{3.006-2.5321} = 2.94$). A hotel located in the Lujiazui area costs 1.46 times of a hotel located in Hongqiao ($10^{2.8303-2.6654} = 1.46$).

The pattern of non-normality persists, but pattern of non-constant variance seems to be alleviated in both residual plots and box plots.







Let's rerun the Levene's test based on the WLS standardized residuals to assess whether the weighting has addressed the nonconstant variance. We first run the Levene's with interaction effect.

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Area	Fixed	4	Disney Resort, Hongqiao, Jingan Temple, Lujiazui
Stars	Fixed	3	five, four, three

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	0.1299	0.04331	0.12	0.947
Stars	2	0.0687	0.03433	0.10	0.908
Area*Stars	6	1.7300	0.28834	0.81	0.571
Error	27	9.6111	0.35597		
Total	38	11.6565			

We can see that the p-value for the interaction effect is not significant. Therefore, we rerun the Levene's test with only the main effects.

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Area	Fixed	4	Disney Resort, Hongqiao, Jingan Temple, Lujiazui
Stars	Fixed	3	five, four, three

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Area	3	0.1487	0.04955	0.14	0.933
Stars	2	0.1739	0.08693	0.25	0.778
Error	33	11.3412	0.34367		
Lack-of-Fit	6	1.7300	0.28834	0.81	0.571
Pure Error	27	9.6111	0.35597		
Total	38	11.6565			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.586235	2.71%	0.00%	0.00%

The p-value for the two variables is larger than the two-way ANOVA model without weights (previously the p-value is 0.115 and 0.760). We can say that heteroscedasticity is not a problem here. Therefore, we will adopt the WLS model without the interaction effect as our final model.

Lastly, let's look at comparisons to see how the areas are different from each other and how the star ratings are different from each other. In terms of areas, there are two groups. Taking the star rating into account, the two categories Lujiazui and Jing'an Temple and Disney Resort are not statistically different from each other. Similarly, taking the star rating into account, Hongqiao and Disney Resort is not statistically different from each other. It is worth noticing that the category Disney Resort is not statistically different from either group. In terms of star rating, taking the area into account, the categories of the five-star, four-star and three-star are all statistically different from each other.

Tukey Pairwise Comparisons: Area

Grouping Information Using the Tukey Method and 95% Confidence

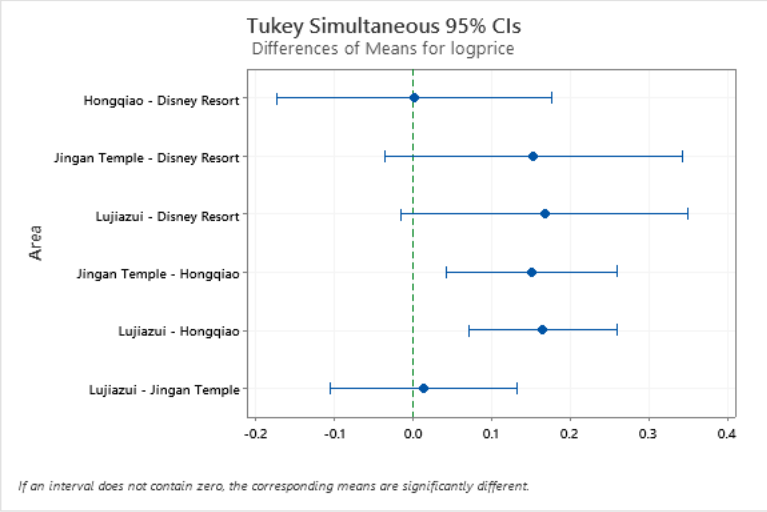
Area	N	Mean	Grouping	
Lujiazui	12	2.83030	A	
Jingan Temple	11	2.81653	A	
Hongqiao	10	2.66542		B
Disney Resort	6	2.66337	A	B

Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

Difference of Area Levels	Difference of Means	SE of Difference	Simultaneous 95% CI	T-Value	Adjusted P-Value
Hongqiao - Disney Resort	0.0020	0.0644	(-0.1724, 0.1765)	0.03	1.000
Jingan Temple - Disney Resort	0.1532	0.0699	(-0.0361, 0.3425)	2.19	0.147
Lujiazui - Disney Resort	0.1669	0.0671	(-0.0147, 0.3486)	2.49	0.080
Jingan Temple - Hongqiao	0.1511	0.0398	(0.0433, 0.2589)	3.80	0.003
Lujiazui - Hongqiao	0.1649	0.0345	(0.0714, 0.2583)	4.78	0.000
Lujiazui - Jingan Temple	0.0138	0.0440	(-0.1054, 0.1330)	0.31	0.989

Individual confidence level = 98.94%



If an interval does not contain zero, the corresponding means are significantly different.

Tukey Pairwise Comparisons: Stars

Grouping Information Using the Tukey Method and 95% Confidence

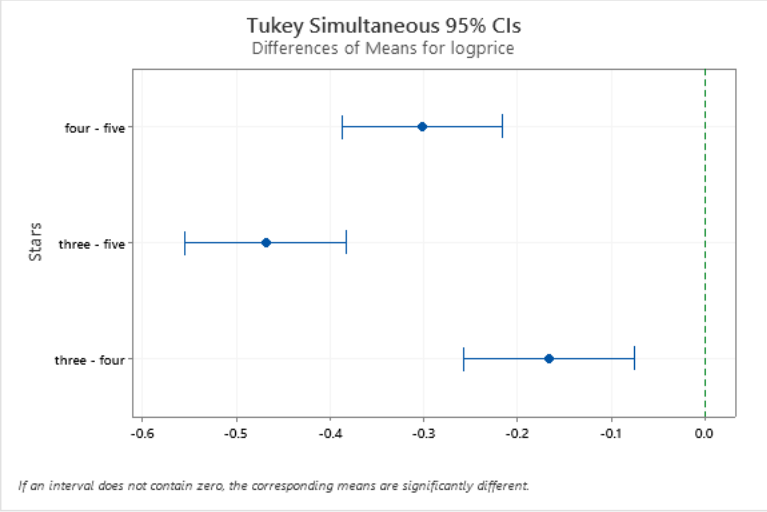
Stars	N	Mean	Grouping
five	15	3.00056	A
four	12	2.69906	B
three	12	2.53209	C

Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

Difference of Stars Levels	Difference of Means	SE of Difference	Simultaneous 95% CI	T-Value	Adjusted P-Value
four - five	-0.3015	0.0348	(-0.3868, -0.2162)	-8.67	0.000
three - five	-0.4685	0.0351	(-0.5545, -0.3825)	-13.36	0.000
three - four	-0.1670	0.0372	(-0.2582, -0.0758)	-4.49	0.000

Individual confidence level = 98.04%



We can use our WLS model to do some predictions about the price for different star rating hotels in different areas. Here are the prediction intervals for some categories:

TESTLEVENE

Prediction for logprice

General Linear Model Information

Terms
Area Stars

Settings

Variable	Setting
Area	Hongqiao
Stars	three

Prediction

Fit	SE Fit	95% CI	95% PI
2.45360	0.0301924	(2.39217, 2.51503)	(2.27204, 2.63516)

Weight = 1.36220

Settings

Variable	Setting
Area	Hongqiao
Stars	five

Prediction

Fit	SE Fit	95% CI	95% PI
2.92207	0.0271420	(2.86685, 2.97729)	(2.74251, 3.10163)

Weight = 1.36220

Settings

Variable	Setting
Area	Lujiazui
Stars	five

Prediction

Fit	SE Fit	95% CI	95% PI
3.08695	0.0326314	(3.02056, 3.15334)	(2.86225, 3.31166)

Weight = 0.862875

Settings

Variable	Setting
Area	Lujiazui
Stars	three

Prediction

Fit	SE Fit	95% CI	95% PI
2.61848	0.0364337	(2.54436, 2.69261)	(2.39138, 2.84559)

Weight = 0.862875

The estimated price for a three-star hotel in Hongqiao is $284.2(10^{2.4536})$. The estimated price for a three-star hotel in Lujiazui is $415.4(10^{2.61848})$. The estimated price for the five-star

hotel in Lujiazui is 1221.7($10^{3.08695}$). The estimated price for a five-star hotel in Hongqiao is 835.7 ($10^{2.92207}$). The hotels in Lujiazui are generally more expensive.

The prediction interval for a three-star hotel in Hongqiao is (187.1, 431.7) (after antilogging). The prediction interval for a three-star hotel in Lujiazui is (246.3, 700.8) (after antilogging). The prediction interval for a five-star hotel in Hongqiao is (552.7, 1263.7) (after antilogging). The prediction interval for five-star hotel in Lujiazui is (728.2, 2049.6) (after antilogging). We can see that the prediction interval for the five-star hotel is wider than the prediction interval for the three-star hotel. The prediction interval for hotels in Lujiazui is wider than the prediction interval for hotels in Hongqiao. This is consistent with our previous finding that there is higher variability for higher star-rating hotels and there is higher variability for hotels in Lujiazui.

To conclude, we will adopt the WLS model only using two main effects. The model does prove that the price of hotels is closely associated with the hotel level (star-rating) and areas. A higher star rating is associated with a higher price. Also, we can see the areas can be divided into two groups. Hotels in Lujiazui and Jing'an Temple are relatively more expensive than the hotels in Hongqiao. This is probably because the former is located in the downtown area of Shanghai and the latter is located in the suburb. Therefore, if you value more on comfort, hotels in suburb area are a good choice. If you value more on traveling convenience, then you might need to sacrifice some living experience. Since our model only has two categories, we still have some limitations in predicting hotel price accurately. Therefore, in terms of future improvement, we could add more predictors to have a more accurate prediction for the real-life cases.