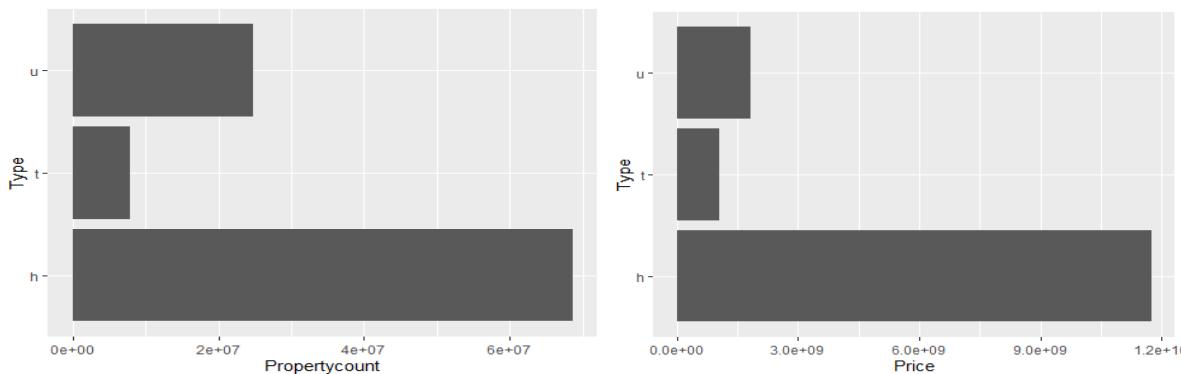


Individual Assignment4- solution

Problem statement:

From basic exploratory analysis of Melbourne housing dataset, I observed that properties of type "house" are more in number as compare to other types and prices for house are more than that of unit and town-houses. "Does type of house in an area plays an important role in determining the price if all other things (like number of rooms, bathroom, locality, area, etc.) are the same?" To answer this question we can propose null hypothesis as type of house has no influence over price and alternative hypothesis as the type is statistically significant while deciding the price. For the sake of analysis we are considering housing data of 658 samples (which is enough sample size for 6 predictors), across 5 popular sub-urban areas (Bentleigh East, Brunswick, Preston, Reservoir, Richmond) and considering the housing types to be House, NonHouse.



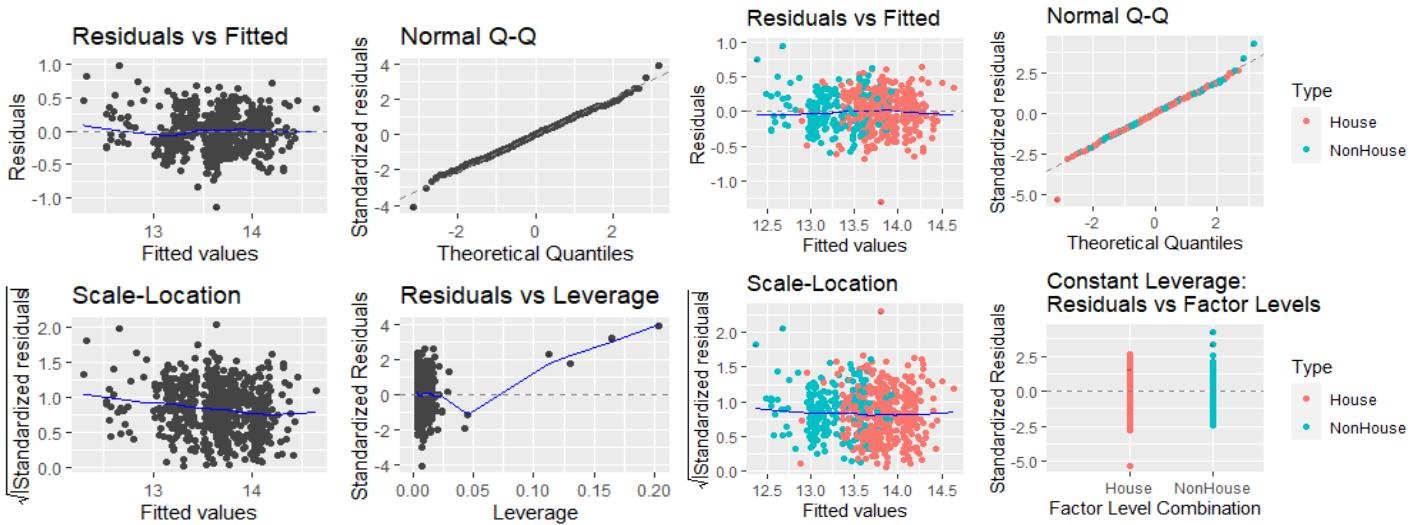
Planning:

Initially 2 models were created, both of them accounting for basic features like building area, location code, no of rooms & bathrooms and distance from CBC and later model considers one extra feature i.e. type of the house. After fitting the models and visual inspection for assumptions [appx- 5], both of the models were showing normal distribution of errors (residuals), however other assumptions like homoscedasticity and linearity were not met. In order to overcome this, log transformation for all numeric predictors and outcome variable is performed and then those models were fitted on the transformed data.

Assumptions:

Assumptions can be accessed from graphs below (first graph represents model without house type and later with house type)

1. Predictor variable are quantitative, categorical and outcome variable is continuous.
2. Linearity: residuals are distributed equally about the zero and the blue line is approximately horizontal at zero. Graph 1 (Residual vs Fitted)
3. Normally distributed errors: From graph 2-QQ plot, residuals are almost following the straight line and hence assumption is met.
4. Non-zero variance: From [Appx- 4b] Predictors are not having zero variance.
5. Homoscedasticity: Graph 3 (Scale-Location), as we can see most of the residuals are spread equally and line is quite horizontal, hence assumption is followed.
6. Outliers /Influential points: few residuals are beyond ± 2.5 standard deviation, however cook's distance for all of them is less than 1 [Appx-7] so we can conclude that none of them is influential.
7. Multicollinearity: [Appx-8] VIF values are all well below 10 and the tolerance statistics all well above 0.2 and average VIF is close to 1. So, there is no collinearity within our data.



Analysis

By comparing both models [Appx-6], there is significant improvement in adjusted R² value which indicates considering house type while determining price would yield more accurate results. Here coefficient corresponding to non-house type unit is negative which means that for same configuration, house will be more expensive as compare to (NH) unit or town-house. Almost all the variables are statistically significant p value much less than .05 (except bathroom for model 1).

MODEL	Price ~ BuildingArea + Postcode + Rooms + Bathroom + Distance				Price ~ BuildingArea + Postcode + Rooms + Bathroom + Distance + Type			
Adj R ²	.64				.71			
P	< 2.2e-16				< 2.2e-16			
	B	SE B	β	p	B	SE B	β	p
Constant	-39.27	7.16		6.09e-08	-	46.49		1.29e-12
BuildingArea	.19	.02	.23	1.54e-14	.15	.02	.18	1.56e-10
Postcode	6.46	.89	.17	1.25e-12	7.42	.80	.20	< 2e-16
Rooms	.78	.04	.60	< 2e-16	.56	.04	.43	< 2e-16
Bathroom	.02	.03	.02	0.63	.10	.03	.08	0.001
Distance	-.33	.02	-.38	< 2e-16	-.31	.02	-.36	< 2e-16
Type(NH)					-.32	.02		< 2e-16

By further comparison using ANOVA [Appx-10]. We got significant result ($F(1, 651) = 167.2, p < 2.2e-16$). So, at the 5% confidence interval, we can reject the null hypothesis that the significance of type of house is equal to 0, and accept the alternative hypothesis. Thus, type of house plays important role in determining the price of the house when other parameters like area, location, no of rooms & bathrooms etc are same. This can be verified by predicting prices for house and non-house type property by keeping all other things same [Appx-11].

Conclusion:

"From given data, we can conclude that type of property has significant impact on the price. For example, for a property of 3 room, 2 bathroom, 100sq meter area which is 12km away from CBD, we would be paying 606579\$ for town house or unit, which will increase to 833424\$ for house type."