STAT 500 Homework 10

Yifan Zhu

December 2, 2016

1. (a)

$$SalePrice = \beta_1 LivingArea + \beta_2 Age + \epsilon$$

β_1 means with Living Area increasing by 1 square feet, the change (positive or negative) of conditional expectation of SalePrice.

β_2 mean with Age increasing by 1 year, he change (positive or negative) of conditional expectation of Sale Price.

(b) The R^2 value is 0.6865. This means that 68.65% of the variation in Sale Price can be explained by the multiple linear regression model with Living Area and Age of house.

(c)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.283227E13	6.416137E12	3199.94	< .0001
Error	2922	5.85885E12	2005082271		
Corrected Total	2924	1.869112E13			

The null and alternative hypotheses are:

$$H_0 : \beta_1 = \beta_2 = 0,\ H_a : \text{at least one } \beta_i \neq 0$$

$F = 3199.94$ with p-value < 0.0001. Since the p-value is so small, we will reject the null hypothesis and conclude at least one of the explanatory variables is significant in explaining the response variable Sale Price.

(d)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
LivingArea	1	102.62349	1.74084	58.95	< .0001

The null and alternative hypotheses are:

$$H_0 : \beta_1 = 0,\ H_a : \beta_1 \neq 0$$

The test statistic $t = 58.95$ with p-value< .0001. Since the p-value is small, we will reject the null hypothesis and conclude the variable Living Area is statistically significant in the model that also includes Age.

(e)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Age	1	-1072.88363	28.21248	-38.03	< .0001

The null and alternative hypotheses are:

$$H_0 : \beta_2 = 0,\ H_a : \beta_2 \neq 0$$

The test statistic $t = -38.03$ with p-value< .0001. Since the p-value is small, we will reject the null hypothesis and conclude the variable Age is statistically significant in the model that also includes Living Area.

2. (a) After adding Basement Area and Total Room to the model with Living Area and Age, the sum of squares for error is reduced by 5.85885E12 - 4.360771E12 = 1.498E12.

(b) After adding Basement Area and Total Room to the model with Living Area and Age, the value of R^2 increases by 0.7666 - 0.6865 = 0.0801.

(c) The null and alternative hypotheses are:

$$H_0 : \beta_3 = \beta_4 = 0,\ H_a : \text{at least one of } \beta_3 \text{ and } \beta_4 \text{ is non-zero}$$

The test statistic $F = \frac{1.498E12}{1493926329} = 1002.7$ and p-value $< .0001$. Since the p-value is so small, we will reject the null hypothesis and conclude that at least one of Basement Area and Total Room is statistically significant in the model that includes Living Area and Age.

3. (a) After adding interaction term between Living Area and Total Room to the model that includes Living Area, Age, Basement Area, and Total Room, the sums of squares for error is reduced by 4.360771E12 - 4.201687E12 = 0.159E12.

(b) After adding interaction term between Living Area and Total Room to the model that includes Living Area, Age, Basement Area, and Total Room, the value of R^2 increases by 0.7751 - 0.7666 = 0.0085.

(c)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr $> \|t\|$
Intercept	1	113188	8526.14565	13.28	$< .0001$
LivingArea	1	45.00482	5.91264	7.61	$< .0001$
Age	1	-794.81037	25.87570	-30.72	$< .0001$
BasementArea	1	61.38687	1.95935	31.33	$< .0001$
TotalRoom	1	-17415	1355.06909	-12.85	$< .0001$
interaction	1	7.52734	0.71614	10.51	$< .0001$

The t-test statistic is$t = 10.51$ with p-value $< .0001$. Since the p-value is small, we will reject the null hypothesis and conclude the interaction term is statistically significant in the model that includes Living Area, Age, Basement Area and Total Room.