5.5

Table 5.8 Output for Exercise 5.5

Dependent Variable: *VALUE*
Included observations: 506

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	28.4067	5.3659	5.2939	0.0000
CRIME	−0.1834	0.0365	−5.0275	0.0000
NITOX	−22.8109	4.1607	−5.4824	0.0000
ROOMS	6.3715	0.3924	16.2378	0.0000
AGE	−0.0478	0.0141	−3.3861	0.0008
DIST	−1.3353	0.2001	−6.6714	0.0000
ACCESS	0.2723	0.0723	3.7673	0.0002
TAX	−0.0126	0.0038	−3.3399	0.0009
PTRATIO	−1.1768	0.1394	−8.4409	0.0000

This question is concerned with the value of houses in towns surrounding Boston. It uses the data of Harrison, D., and D. L. Rubinfeld (1978), ‘‘Hedonic Prices and the Demand for Clean Air,’’ Journal of Environmental Economics and Management, 5, 81–102. The output appears in Table 5.8. The variables are defined as follows:

VALUE = median value of owner-occupied homes in thousands of dollars
CRIME = per capita crime rate
NITOX = nitric oxide concentration (parts per million)
ROOMS = average number of rooms per dwelling
AGE = proportion of owner-occupied units built prior to 1940
DIST = weighted distances to five Boston employment centers
ACCESS = index of accessibility to radial highways
TAX = full-value property-tax rate per $10,000
PTRATIO =pupil–teacher ratio by town

(a) Report briefly on how each of the variables influences the value of a home.

The estimated equation:
$\widehat{VALUE}$ = 28.4067 - 0.1834*CRIME - 22.8109*NITOX + 6.3715*ROOMS – 0.0478*AGE – 1.33533*DIST + 0.2723*ACCESS – 0.0126*TAX – 1.1768*PTRATIO

- Per capita crime rate increases by 1 unit leads to the home value decreases by $183.4.
- A one unit increase in the nitric oxide concentration leads to a decline in value of 22,811.
- An increase in one room leads to an increase of $6,372 in the home value.

- An increase in the proportion of owner-occupied units built prior to 1940 leads to a decline in the home value.
- the weighted distances to the five Boston employment centers leads to the lower the home value by $1,335 for every unit of weighted distance.
- The higher the tax rate per $10,000 leads to the lower the home value.
- the higher the pupil-teacher ratio leads to the lower the home value.

(b) Find 95% interval estimates for the coefficients of CRIME and ACCESS.

```
[1] "b_2+t_(0.975, 497)*se(b2) = -0.1114"
[1] "b_2-t_(0.975, 497)*se(b2) = -0.2554"
[1] "95% confidence interval for the coefficient of CRIME is [-0.2554, -0.1114]"
```

```
[1] "b_2+t_(0.975, 497)*se(b2) = 0.4143"
[1] "b_2-t_(0.975, 497)*se(b2) = 0.1303"
[1] "95% confidence interval for the coefficient of ACCESS is [0.130, -0.414]"
```

(c) Test the hypothesis that increasing the number of rooms by one increases the value of a house by $7,000.

We want to test H_0 : β_{room} = 7, H_1 : $\beta_{room} \neq$ 7. The value of the t statistic is t=(b_{room}-7)/se(b_{room}) = (6.3715-7)/0.3924 = -1.6017.
when α= 0.05, we reject 0 H if the absolute calculated t is greater than 1.965. Since | 1.6017| < 1.965 , we do not reject H_0. The data is consistent with the hypothesis that increasing the number of rooms by one increases the value of a house by $7000.

(d) Test as an alternative hypothesis H1 that reducing the pupil–teacher ratio by 10 will increase the value of a house by more than $10,000.

We want to test H_0 : $\beta_{ptratio}$ >= -1, H_1 : $\beta_{ptratio}$ < -1. The value of the t statistic is t=($b_{ptratio}$+1)/se($b_{ptratio}$) = (-1.1768+1)/0.1394 = -1.2683.
When α= 0.05 , we reject H_0 if the calculated t is less than the critical value $t_{(0.05,497)}$=-1.648. Since -1.2683 > -1.648, we do not reject H_0. We cannot conclude that reducing the pupil-teacher ratio by 10 will increase the value of a house by more than $10,000.

5.9

Table 5.9 **Wage Equation with Quadratic Experience**

Variable	Coefficient	Std. Error	t-Stat	Prob.
C	−13.4303	2.0285	−6.621	0.000
EDUC	2.2774	0.1394	16.334	0.000
EXPER	0.6821	0.1048	6.507	0.000
$EXPER^2$	−0.0101	0.0019	−5.412	0.000

Covariance Matrix for Least Squares Estimates

	C	*EDUC*	*EXPER*	$EXPER^2$
C	4.114757339	−0.215505842	−0.124023160	0.001822688
EDUC	−0.215505842	0.019440281	−0.000217577	0.000015472
EXPER	−0.124023160	−0.000217577	0.010987185	−0.000189259
$EXPER^2$	0.001822688	0.000015472	−0.000189259	0.000003476

When estimating wage equations, we expect that young, inexperienced workers will have relatively low wages and that with additional experience their wages will rise, but then begin to decline after middle age, as the worker nears retirement. This lifecycle pattern of wages can be captured by introducing experience and experience squared to explain the level of wages. If we also include years of education, we have the equation

WAGE = β_1+β_2*EDUC+β_3*EXPER+β_4*EXPER^2 + e

(a) What is the marginal effect of experience on wages?

β_3*EXPER+2*β_4*EXPER

(b) What signs do you expect for each of the coefficients β_2, β_3, and β_4? Why?

β_2 expect positive: workers have a higher level of education should have higher wages

β_3 expect positive: additional experience leads to increase in their wages

β_4 expect negative: expect wages to decline with experience as a worker gets older and their productivity decline.

A negative β_3 and a positive β_4 gives a quadratic function with these properties.

(c) After how many years of experience do wages start to decline? (Express your answer in terms of β's

Find the β_3+2*β_4*EXPER = 0

EXPER = -β_3/2*β_4

(d) The results from estimating the equation using 1000 observations in the file cps4c_small.dat are given in Table 5.9 on page 204. Find 95%interval estimates for

(i) The marginal effect of education on wages

$\sigma\widehat{WAGE}/\sigma\widehat{EDUC}\ =\ b_2$ = 2.2774

95% interval estimates

```
[1] "b_2+t_(0.975, 998)*se(b2) = 2.551"
[1] "b_2-t_(0.975, 998)*se(b2) = 2.0038"
[1] "95% confidence interval for the coefficient of ACCESS is [2.0038, 2.551]"
```

(ii) The marginal effect of experience on wages when EXPER = 4

$\sigma\widehat{WAGE}/\sigma\widehat{EXPER}\ =\ b_3$+2*$b_4$*EXPER = 0.6821-2*0.0101*4 = 0.6013

To compute an interval estimate, the standard error of this quantity which is given by Table 5.9

se(b_3+8*b_4) = ($\widehat{var(b_3})$+8^2$\widehat{var(b_4})$+2*8*$\widehat{cov(b_{3,}}b_4)$)^(1/2)

= (0.010987185+64*0.000003476-16*0.000189259)^(1/2)

= 0.09045

```
[1] "b_3+8*b_4+t_(0.975, 998)*se(b_3+8*b_4) = 0.7788"
[1] "b_3+8*b_4-t_(0.975, 998)*se(b_3+8*b_4) = 0.4238"
[1] "95% confidence interval for the coefficient is [0.4238, 0.7788]"
```

(iii) The marginal effect of experience on wages when EXPER = 25

$\sigma\widehat{WAGE}/\sigma\widehat{EXPER}\ =\ b_3$+2*$b_4$*EXPER = 0.6821-2*0.0101*25 = 0.1771

se(b_3+50*b_4) = ($\widehat{var(b_3})$+50^2$\widehat{var(b_4})$+2*50*$\widehat{cov(b_{3,}}b_4)$)^(1/2) = 0.02741

```
[1] "b_3+50*b_4+t_(0.975, 998)*se(b_3+50*b_4) = 0.2309"
[1] "b_3+50*b_4-t_(0.975, 998)*se(b_3+50*b_4) = 0.1233"
[1] "95% confidence interval for the coefficient is [0.1233, 0.2309]"
```

5.12

The file *cocaine.dat* contains 56 observations on variables related to sales of cocaine powder in northeastern California over the period 1984–1991. The data are a subset of those used in the study Caulkins, J. P. and R. Padman (1993), ‘‘Quantity Discounts and Quality Premia for Illicit Drugs,’’ Journal of the American Statistical Association, 88, 748–757. The variables are

PRICE = price per gram in dollars for a cocaine sale

QUANT = number of grams of cocaine in a given sale

QUAL = quality of the cocaine expressed as percentage purity

TREND = a time variable with 1984 ¼ 1 up to 1991 ¼ 8

Consider the regression model:

PRICE = β_1 + β_2*QUANT + β_3*QUAL + β_4*TREND + e

(a) What signs would you expect on the coefficients β_2, β_3, and β_4?

β_2 expect negative: the number of grams in a given sale increases, the price per gram should decrease, implying a discount for larger sales.

β_3 expect positive: the purer the cocaine, the higher the price.

β_4 expectation will depend on if a fixed demand and an increasing supply will lead to a fall in price. A fixed supply and increased demand will lead to a rise in price.

(b) Use your computer software to estimate the equation. Report the results and interpret the coefficient estimates. Have the signs turned out as you expected?

```r
cocaine <- read.dta("cocaine.dta") # input Stata file
mod1 <- lm(price~quant+qual+trend, data=cocaine)
summary(mod1)
```

```
Call:
lm(formula = price ~ quant + qual + trend, data = cocaine)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.479 -12.014  -3.743  13.969  43.753 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 90.84669    8.58025  10.588 1.39e-14 ***
quant       -0.05997    0.01018  -5.892 2.85e-07 ***
qual         0.11621    0.20326   0.572   0.5700    
trend       -2.35458    1.38612  -1.699   0.0954 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.06 on 52 degrees of freedom
Multiple R-squared:  0.5097,	Adjusted R-squared:  0.4814 
F-statistic: 18.02 on 3 and 52 DF,  p-value: 3.806e-08
```

$\widehat{PRICE}$ = 90.8467- 0.0600*QUANT + 0.1162*QUAL – 2.3546*TREND

(se) =(8.5803) (0.0102) (0.2033) (1.3861)

(t) =(10.588) (-5.892) (0.5717) (-1.6987)

All the signs turn out according to our expectations, with β_4 implying supply has been increasing faster than demand.

(c) What proportion of variation in cocaine price is explained jointly by variation in quantity, quality, and time?

According to the result, R^2 = 0.5097 = 51%

(d) It is claimed that the greater the number of sales, the higher the risk of getting caught. Thus, sellers are willing to accept a lower price if they can make sales in larger quantities. Set up H0 and H1 that would be appropriate to test this hypothesis. Carry out the hypothesis test.

Test H_0 : β_2 >= 0, H_1 : β_2< 0. The value of the *t* statistic is = -5.892.
we reject H_0 if the calculated *t* is less than the critical value $t_{(0.05,52)}$=-1.675. Since -1.675 > -5.892., we reject H_0. So, conclude that sellers are willing to accept a lower price if they can make sales in larger quantities.

(e) Test the hypothesis that the quality of cocaine has no influence on price against the alternative that a premium is paid for better-quality cocaine.

Test H_0 : β_3 <= 0, H_1 : β_3> 0. The calculated t is 0.5717. When α= 0.05, $t_{(0.95,52)}$= 1.675. Since 1.675 > 0.5717. we do not reject H_0. We cannot conclude that a premium is paid for better quality cocaine.

(f) What is the average annual change in the cocaine price? Can you suggest why price might be changing in this direction?

The average annual change in the cocaine price is given by the value of b_4=-2.3546 . It has a negative sign suggesting that the price decreases over time. A possible reason for a decreasing price is the development of improved technology for producing cocaine, such that suppliers can produce more at the same cost.

5.13

The file br2.dat contains data on 1,080 houses sold in Baton Rouge, Louisiana, during mid-2005.We will be concerned with the selling price (PRICE), the size of the house in square feet (SQFT), and the age of the house in years (AGE).

(a) Use all observations to estimate the following regression model and report the results

PRICE = β_1+ β_2*SOFT+ β_3*ACE + e

Run the regression model.

```{r}
br2 <- read.dta("br2.dta") # input Stata file
mod2 <- lm(price~sqft+age, data=br2)
summary(mod2)
```

```
Call:
lm(formula = price ~ sqft + age, data = br2)

Residuals:
    Min      1Q  Median      3Q     Max 
-358116  -33259   -6111   27242  936754 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -41947.696   6989.636  -6.001 2.67e-09 ***
sqft            90.970      2.403  37.855  < 2e-16 ***
age           -755.041    140.894  -5.359 1.02e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 78810 on 1077 degrees of freedom
Multiple R-squared:  0.5896,	Adjusted R-squared:  0.5888 
F-statistic: 773.6 on 2 and 1077 DF,  p-value: < 2.2e-16
```

(i) Interpret the coefficient estimates.

b_1 could be interpreted as the average price of land, since a negative price is unrealistic, we view the equation as a poor model for data values id SQFT = 0 and AGE = 0.

b_2= 90.97 implies that holding age if one square foot increase in the size of the house increases the selling price by 90.97 dollars

b_3= -755.04 implies that holding SQFT if an increase in the age of the house by 1 year decreases the selling price by 755.04 dollars.

(ii) Find a 95% interval estimate for the price increase for an extra square foot of living space—that is, δPRICE/δSQFT.

b_2= 90.97

```
[1] "b_2+t_(0.975, 1077)*se(b_2) = 95.6853"
[1] "b_2-t_(0.975, 1077)*se(b_2) = 86.2547"
[1] "95% confidence interval for the coefficient is [86.2547, 95.6853]"
```

(iii) Test the hypothesis that having a house a year older decreases price by 1000 or less (H_0 : β_3>= -1000) against the alternative that it decreases price by more than 1000 (H_1 : β_3< -1000).

b_3=-755.04, se (b_3) = 140.894

t= (b_3-(-1000))/se (b_3) = 1.7386

5% significance level is $t_{(0.05,1077)}$= -1.646.

The rejection region is t <= -1.646. Since 1.7386 > -1.646, p-value > 0.05, we fail to reject H_0. So, conclude that the estimated equation that an extra year of age decreases the price by $1000 or less.

(b) Add the variables $SQFT^2$and AGE^2 to the model in part (a) and re-estimate the equation. Report the results.

$\widehat{PRICE}$ = 170149.65 – 55.784*SQFT + 0.023*SQFT^2 – 2797.788*AGE+30.16*AGE^2

```
5.13(b)
```{r}
mod3 <- lm(price~sqft+I(sqft^2)+age+I(age^2),data=br2)
smod3 <- summary(mod3)
tabl <- data.frame(xtable(smod3))
names(tabl) <- c("Estimate",
"Std. Error", "t", "p-Value")
kable(tabl, digits=3, align='c',
caption="The quadratic version of the $br2$ model")
```
```

	Estimate	Std. Error	t	p-Value
(Intercept)	170149.649	10432.256	16.310	0
sqft	-55.784	6.389	-8.731	0
I(sqft^2)	0.023	0.001	24.013	0
age	-2797.788	305.116	-9.170	0
I(age^2)	30.160	5.071	5.948	0

(i) Find estimates of the marginal effect δPRICE/δSQFT for the smallest house in the sample, the largest house in the sample, and a house with 2300 SQFT. Comment on these values. Are they realistic?

The marginal effect of *SQFT* on *PRICE* is given by

$\delta PRICE/\delta SQFT = \beta_2 + 2\beta_3 SQFT$

- The estimated marginal for the smallest house:
  SQFT = 662, -55.784+2*0.023*662 = **-25.332**
- The estimated marginal for the 2300 SQFT house:
  SQFT = 2300, -55.784+2*0.023*2300 = **50.016**
- The estimated marginal for the largest house:
  SQFT = 7897, -55.784+2*0.023*2300 = **307.478**

The result for small houses is unrealistic. But it is possible that more square feet lead to a higher price increase in larger houses.

(ii) Find estimates of the marginal effect δPRICE/δAGE for the oldest house in the sample, the newest house in the sample, and a house that is 20 years old. Comment on these values. Are they realistic?

The marginal effect of AGE on PRICE, $\delta PRICE/\delta AGE = \beta_4 + 2\beta_5 AGE$

- the oldest house (AGE = 80): -2797.788+2*30.16*80 = 2027.812
- the house (AGE = 20): -2797.788+2*30.16*20 = -1591.388
- the newest house (AGE = 1): -2797.788+2*30.16*1 = -2737.468

When a house is new, extra years of age have the greatest negative effect on price. Aging has a smaller and smaller negative effect as the house gets older. This result is as expected. it is unrealistic for the oldest houses to increase in price as they continue to age, or the house have other condition need to consider.

(iii) Find a 95% interval estimate for the marginal effect δPRICE/δSQFT for a house with 2300 square feet.

```
5.13(b)(iii)
```{r}
varb2 <- vcov(mod3)[2,2]
varb3 <- vcov(mod3)[3,3]
varb4 <- vcov(mod3)[4,4]
varb5 <- vcov(mod3)[5,5]
covb2b3 <- vcov(mod3)[2,3]
covb4b5 <- vcov(mod3)[4,5]
varb2
varb3
varb4
varb5
covb2b3
covb4b5
```

 [1] 40.82499
 [1] 9.296015e-07
 [1] 93095.48
 [1] 25.71554
 [1] -0.005870334
 [1] -1434.561
```

$t_{(0.975,1075)}$ = 1.962

se(me) = (40.82499+4600^2*9.296015*10^(-7)+9200*(-0.005870334))^(1/2) = 2.5472

me+$t_{(0.975,1075)}$* se(me) = 55.72

me-$t_{(0.975,1075)}$* se(me) = 45.72

Interval estimate = [45.72, 55.72]

(iv) For a house that is 20 years old, test the hypothesis

H_0: δPRICE/δAGE >= -1000, H_1: δPRICE/δAGE < -1000

H_0 : β_4+40*β_5>= -1000, H_1 : β_4+40*β_5< -1000

se(b4+40b5) = (var(b4) +40^2*var(b5)+2*40cov(b4,b5))^(1/2) = (93095.48+40^2*25.71554+2*40*(-1434.561))^(1/2) = 139.55

t =(b4+40*b5-(-1000))/se(b4+40b5) = (-2791.788+40*30.16+1000)/139.554 = -585.388/139.554 = -4.1947

The critical value for a 5% significance level is $t_{(0.05,1075)}$ = -1.646. The rejection region is t < -1.646 . Since -4.1947 < -1.646 , we reject H_0. We conclude that, for a 20-year old house, an extra year of age decreases the price by more than $1000.

(c) Add the interaction variable SQFT * AGE to the model in part (b) and re-estimate the equation. Report the results. Repeat parts (i), (ii), (iii), and (iv) from part (b) for this new model. Use SQFT = 2300 and AGE = 20.

$\widehat{PRICE}$ = 114597 – 30.729*SQFT + 0.022185*SQFT^2 – 442.03*AGE+26.519*AGE^2-0.93062*(SQFT*AGE)

```
Call:
lm(formula = price ~ sqft + I(sqft^2) + age + I(age^2) + (sqft *
    age), data = br2)

Residuals:
    Min      1Q  Median      3Q     Max
-796617  -21537    -439   17825  623609

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.146e+05  1.214e+04   9.437  < 2e-16 ***
sqft        -3.073e+01  6.898e+00  -4.455 9.27e-06 ***
I(sqft^2)    2.218e-02  9.425e-04  23.537  < 2e-16 ***
age         -4.420e+02  4.106e+02  -1.077    0.282
I(age^2)     2.652e+01  4.939e+00   5.370 9.66e-08 ***
sqft:age    -9.306e-01  1.124e-01  -8.277 3.72e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 61470 on 1074 degrees of freedom
Multiple R-squared:  0.751,     Adjusted R-squared:  0.7499
F-statistic:   648 on 5 and 1074 DF,  p-value: < 2.2e-16
```

i. Find estimates of the marginal effect δPRICE/δSQFT for the smallest house in the sample, the largest house in the sample, and a house with 2300 SQFT. Comment on these values. Are they realistic?

The marginal effect of *SQFT* on *PRICE* is given by

δPRICE/δSQFT = $\beta_2 + 2\beta_3 SQFT + \beta_6 AGE$

- The estimated marginal for the smallest house:
  SQFT = 662, -30.7289 +2*0.022185*662-0.93062*20= **-19.97**
- The estimated marginal for the 2300 SQFT house:
  SQFT = 2300, -30.7289 +2*0.022185*2300-0.93062*20 =**52.71**
- The estimated marginal for the largest house:
  SQFT = 7897, -30.7289 +2*0.022185*7897-0.93062*20 = **301.04**

Similar conclusions to those reached in part 5.13(b)(i), the result for small houses is unrealistic. But it is possible that more square feet lead to a higher price increase in larger houses. It would be more realistic if the quadratic reached a minimum before the smallest house in the sample.

ii. Find estimates of the marginal effect δPRICE/δAGE for the oldest house in the sample, the newest house in the sample, and a house that is 20 years old. Comment on these values. Are they realistic?

The marginal effect of AGE on PRICE, δPRICE/δAGE = $\beta_4 + 2\beta_5 AGE + \beta_6 SQFT$

- the oldest house (AGE = 80): -442.0336+2*26.519*80-0.93062*2300 = **1660.6**
- the house (AGE = 20): -442.0336+2*26.519*20-0.93062*2300 = **-1521.7**
- the newest house (AGE = 1): - 442.0336+2*26.519*1-0.93062*2300 = **-2529.4**

Similar conclusions to those reached in part 5.13(b)(ii).

iii. Find a 95% interval estimate for the marginal effect δPRICE/δSQFT for a house with 2300 square feet.

A 95% interval for the marginal effect of SQFT on PRICE

when SQFT = 2300 and AGE = 20, $t_{(0.975,1074)}$ = 1.962

```
. lincom sqft+2*sqft2*2300+sa*20

 ( 1)  sqft + 4600*sqft2 + 20*sa = 0
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	52.70796	2.482473	21.23	0.000	47.83692	57.57901

$me_2 + t_{(0.975,1074)}$*se(me_2) = 52.708+1.962*2.4825 = 57.58

$me_2 - t_{(0.975,1074)}$*se(me_2) = 52.708-1.962*2.4825 = 47.84

Interval estimate = [47.84, 57.58]

iv. **For a house that is 20 years old, test the hypothesis**

H_0: δPRICE/δAGE >= -1000, H_1: δPRICE/δAGE < -1000

H_0 : β_4+40*β_5+2300*β_6>= -1000, H_1 : β_4+40*β_5+2300*β_6< -1000

se(b4+40b5+2300b6) =138.630

t =(b4+40*b5+2300b6-(-1000))/se(b4+40b5+2300b6) =-521.701/135.63 = -3.847

5% significance level is $t_{(0.975,1074)}$=-1.646 . The rejection region is t <= -1.646. Since -3.847 < -1.646, we reject H_0. We conclude that, for a 20-year old house with SQFT = 2300, an extra year of age decreases the price by more than $1000.

(d) From your answers to parts (a), (b), and (c), comment on the sensitivity of the results to the model specification.

The results from the two quadratic specifications in parts (c) and (d) are similar, but they are vastly different from those from the linear model in part (a).

1. the marginal effect of SQFT in part (a) is constant, but in parts (b) and (c), it is different from -20 to +300.
2. The marginal effect of AGE is constant in part (a) but varies from approximately -2600 to +1800 in parts (b) and (c).

These differences on the interval estimates for the marginal effect of SQFT and to the hypothesis tests on the marginal effect of AGE. The marginal effects are clearly not constant and so the linear function is inadequate. Both quadratic functions are improved but have some unrealistic results for old and small houses.

Compare the part (a) to part (b) and (c), the linear regression model cannot provide a perfect predict to the house selling price. The larger house does not increase the selling price, in the meanwhile, the older house might increase the selling price. Therefore, the non-linear regression model is better to provide better information to predict the house selling price.