1.(a)

>setwd("D:/2016\_MSBA UTD/2016 Fall Classes/ECON 6306\_Applied Econometrics/Problems/Problem 2")

>cps = read.csv("cps4.csv")

1. All males :

> male <- subset(cps,female==0,c("wage", "educ"))

> mean(male$wage)

[1] 22.25795

> var(male$wage)

[1] 181.5091

> summary(male$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.00 12.75 19.10 22.26 28.05 173.00

> library(pastecs)

> stat.desc(male$wage)

nbr.val nbr.null nbr.na min max range

2.395000e+03 0.000000e+00 0.000000e+00 1.000000e+00 1.730000e+02 1.720000e+02

sum median mean SE.mean CI.mean.0.95 var

5.330780e+04 1.910000e+01 2.225795e+01 2.752938e-01 5.398389e-01 1.815091e+02

std.dev coef.var

1.347253e+01 6.052906e-01

1. All females :

> female <- subset(cps,female==1,c("wage", "educ"))

> mean(female$wage)

[1] 18.0538

> var(female$wage)

[1] 124.4743

> summary(female$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.14 10.00 15.00 18.05 22.04 96.17

> stat.desc(female$wage)

nbr.val nbr.null nbr.na min max range

2.443000e+03 0.000000e+00 0.000000e+00 1.140000e+00 9.617000e+01 9.503000e+01

sum median mean SE.mean CI.mean.0.95 var

4.410543e+04 1.500000e+01 1.805380e+01 2.257242e-01 4.426307e-01 1.244743e+02

std.dev coef.var

1.115680e+01 6.179755e-01

1. All whites :

> white = subset(cps, white==1,c("wage","educ"))

> mean(white$wage)

[1] 20.48479

> var(white$wage)

[1] 159.7229

> summary(white$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.14 11.64 17.00 20.48 25.61 173.00

1. All blacks :

> black = subset(cps, black==1,c("wage","educ"))

> mean(black$wage)

[1] 16.44389

> var(black$wage)

[1] 102.7358

> summary(black$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.00 10.00 13.45 16.44 20.00 72.13

1. White males :

> whitemale = subset(cps, white==1&female==0, c("wage","educ"))

> mean(whitemale$wage)

[1] 22.83416

> var(whitemale$wage)

[1] 186.886

> summary(whitemale$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.50 13.23 19.24 22.83 28.83 173.00

6) White females :

> whitefemale = subset(cps, white==1&female==1, c("wage","educ"))

> mean(whitefemale$wage)

[1] 18.11937

> var(whitefemale$wage)

[1] 121.2942

> summary(whitefemale$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.14 10.24 15.00 18.12 22.04 96.17

7) Black males :

> blackmale = subset(cps, black==1&female==0, c("wage","educ"))

> mean(blackmale$wage)

[1] 16.21332

> var(blackmale$wage)

[1] 90.10874

> summary(blackmale$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.00 10.00 13.48 16.21 19.23 72.13

8) Black females :

> blackfemale = subset(cps, black==1&female==1, c("wage","educ"))

> mean(blackfemale$wage)

[1] 16.62075

> var(blackfemale$wage)

[1] 112.7078

> summary(blackfemale$wage)

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.75 9.75 13.45 16.62 20.09 72.13

(b)

1) All males :

> log.male.wage = log(male$wage,)

> log\_male= data.frame(log.male.wage, male$educ)

> reg\_male=lm(log.male.wage~male$educ,data=log\_male)

> summary(reg\_male)

Call:

lm(formula = log.male.wage ~ male$educ, data = log\_male)

Residuals:

Min 1Q Median 3Q Max

-2.79305 -0.31649 -0.00222 0.36355 2.18350

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.732617 0.049857 34.75 <2e-16 \*\*\*

male$educ 0.088370 0.003565 24.79 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.51 on 2393 degrees of freedom

Multiple R-squared: 0.2043, Adjusted R-squared: 0.204

F-statistic: 614.5 on 1 and 2393 DF, p-value: < 2.2e-16

Interpretation : For males, the increase of 1 year in education leads to wage increase by 8.8%

2) All females :

> log.female.wage = log(female$wage)

> log\_female=data.frame(log.female.wage,female$educ)

> reg\_female=lm(log.female.wage~female$educ, data=log\_female)

> summary(reg\_female)

Call:

lm(formula = log.female.wage ~ female$educ, data = log\_female)

Residuals:

Min 1Q Median 3Q Max

-2.81403 -0.31591 -0.00338 0.30661 1.97180

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.242679 0.055943 22.21 <2e-16 \*\*\*

female$educ 0.106399 0.003927 27.09 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4896 on 2441 degrees of freedom

Multiple R-squared: 0.2312, Adjusted R-squared: 0.2309

F-statistic: 734 on 1 and 2441 DF, p-value: < 2.2e-16

Interpretation : For females, the increase of 1 year in education leads to wage increase by 10.6%

1. All whites :

> log.white.wage = log(white$wage,)

> log\_white=data.frame(log.white.wage,white$educ)

> reg\_white=lm(log.white.wage~white$educ, data=log\_white)

> summary(reg\_white)

Call:

lm(formula = log.white.wage ~ white$educ, data = log\_white)

Residuals:

Min 1Q Median 3Q Max

-2.91828 -0.33543 -0.00356 0.34890 2.28609

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.592440 0.041147 38.7 <2e-16 \*\*\*

white$educ 0.091054 0.002909 31.3 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5152 on 4114 degrees of freedom

Multiple R-squared: 0.1923, Adjusted R-squared: 0.1921

F-statistic: 979.8 on 1 and 4114 DF, p-value: < 2.2e-16

Interpretation : For whites, the increase of 1 year in education leads to wage increase by 9.1%

1. All blacks :

> log.black.wage = log(black$wage,)

> log\_black = data.frame(log.black.wage,black$educ)

> reg\_black=lm(log.black.wage~black$educ, data=log\_black)

> summary(reg\_black)

Call:

lm(formula = log.black.wage ~ black$educ, data = log\_black)

Residuals:

Min 1Q Median 3Q Max

-2.50777 -0.31036 -0.02286 0.26157 1.98107

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.245588 0.127762 9.749 <2e-16 \*\*\*

black$educ 0.105182 0.009398 11.192 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4659 on 491 degrees of freedom

Multiple R-squared: 0.2033, Adjusted R-squared: 0.2016

F-statistic: 125.3 on 1 and 491 DF, p-value: < 2.2e-16

Interpretation : For blacks, the increase of 1 year in education leads to wage increase by 10.5%

1. White males :

> log.whitemale.wage = log(whitemale$wage,)

> log\_whitemale=data.frame(log.whitemale.wage,whitemale$educ)

> reg\_whitemale=lm(log.whitemale.wage~whitemale$educ,data=log\_whitemale)

> summary(reg\_whitemale)

Call:

lm(formula = log.whitemale.wage ~ whitemale$educ, data = log\_whitemale)

Residuals:

Min 1Q Median 3Q Max

-2.4192 -0.3045 -0.0012 0.3555 2.1564

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.790900 0.052184 34.32 <2e-16 \*\*\*

whitemale$educ 0.086145 0.003725 23.13 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5041 on 2063 degrees of freedom

Multiple R-squared: 0.2059, Adjusted R-squared: 0.2055

F-statistic: 534.9 on 1 and 2063 DF, p-value: < 2.2e-16

Interpretation : For white males, the increase of 1 year in education leads to wage increase by 8.6%

1. White females :

> log.whitefemale.wage = log(whitefemale$wage,)

> log\_whitefemale = data.frame(log.whitefemale.wage,whitefemale$educ)

> reg\_whitefemale=lm(log.whitefemale.wage~whitefemale$educ, data=log\_whitefemale)

> summary(reg\_whitefemale)

Call:

lm(formula = log.whitefemale.wage ~ whitefemale$educ, data = log\_whitefemale)

Residuals:

Min 1Q Median 3Q Max

-2.81464 -0.31031 0.00296 0.31047 1.75570

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.254094 0.061650 20.34 <2e-16 \*\*\*

whitefemale$educ 0.105723 0.004317 24.49 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4906 on 2049 degrees of freedom

Multiple R-squared: 0.2264, Adjusted R-squared: 0.226

F-statistic: 599.6 on 1 and 2049 DF, p-value: < 2.2e-16

Interpretation : For white females, the increase of 1 year in education leads to wage increase by 10.6%

1. Black males :

> log.blackmale.wage = log(blackmale$wage,)

> log\_blackmale= data.frame(log.blackmale.wage, blackmale$educ)

> reg\_blackmale=lm(log.blackmale.wage~blackmale$educ,data=log\_blackmale)

> summary(reg\_blackmale)

Call:

lm(formula = log.blackmale.wage ~ blackmale$educ, data = log\_blackmale)

Residuals:

Min 1Q Median 3Q Max

-2.56621 -0.31152 -0.01395 0.32666 1.40756

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.65210 0.21048 7.849 2.04e-13 \*\*\*

blackmale$educ 0.07618 0.01584 4.809 2.88e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4908 on 212 degrees of freedom

Multiple R-squared: 0.09835, Adjusted R-squared: 0.09409

F-statistic: 23.12 on 1 and 212 DF, p-value: 2.878e-06

Interpretation : For black males, the increase of 1 year in education leads to wage increase by 7.6%

1. Black females :

> log.blackfemale.wage = log(blackfemale$wage,)

> log\_blackfemale = data.frame(log.blackfemale.wage,blackfemale$educ)

> reg\_blackfemale=lm(log.blackfemale.wage~blackfemale$educ,data=log\_blackfemale)

> summary(reg\_blackfemale)

Call:

lm(formula = log.blackfemale.wage ~ blackfemale$educ, data = log\_blackfemale)

Residuals:

Min 1Q Median 3Q Max

-1.18094 -0.30997 -0.04592 0.23565 2.07730

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.93953 0.15918 5.902 1.04e-08 \*\*\*

blackfemale$educ 0.12616 0.01151 10.958 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4406 on 277 degrees of freedom

Multiple R-squared: 0.3024, Adjusted R-squared: 0.2999

F-statistic: 120.1 on 1 and 277 DF, p-value: < 2.2e-16

Interpretation : For black males, the increase of 1 year in education leads to wage increase by 12.6%

(c) Fit equally?

R-square of each subgroup : 0.2043, 0.2312, 0.1923, 0.2033, 0.2059, 0.2264, 0.09835, 0.3024,

The 7th subgroup (black male) has significantly smaller R-square, the model does not fit equally for each subgroup.

(d)H0 : b2=0.1, H1:b2 != 0.1

1) All males :

T = (0.08837 -0.1)/ 0.003565 = -3.26

Tc(0.975, 2393) = 1.96

T is not located within 95% interval [-1.96, 1.96], so we reject H0, and accept H1 :the return to education is NOT 10%

2) All females :

T = (0.106399- 0.1)/ 0.003927 = 1.6294

Tc(0.975, 2441) = 1.96

T is within interval [-1.96, 1.96], we can not reject H0

3) All whites :

T = (0.091054- 0.1)/ 0.002909 = -3.075

Tc(0.975, 4114) = 1.96

T is not within interval [-1.96, 1.96], we reject H0 and accept H1 :the return to education is NOT 10%

4) All blacks :

T = (0.105182- 0.1)/ 0.009398 = 0.551

Tc(0.975, 491) = 1.96

T is within interval [-1.96, 1.96], we cannot reject H0

5) White males :

T = (0.086145- 0.1)/ 0.003725 = -3.719

Tc(0.975, 2063) = 1.96

T is not within interval [-1.96, 1.96], we reject H0 and accept H1 :the return to education is NOT 10%

6) White females :

T = (0.105723 - 0.1)/ 0.004317 = 1.326

Tc(0.975, 2049) = 1.96

T is within interval [-1.96, 1.96], we cannot reject H0

7) Black males :

T = (0.07618 - 0.1)/ 0.01584 = -1.50

Tc(0.975, 212) = 1.96

T is within interval [-1.96, 1.96], we cannot reject H0

8) Black females :

T = (0.12616 - 0.1)/ 0.01151 = -2.27

Tc(0.975, 277) = 1.96

T is not within interval [-1.96, 1.96], we reject H0 and accept H1 :the return to education is NOT 10%

===========================================================

2.(a)

CRIME : increase in per capita crime rate by 1 will lead to the decrease in value of a home by $183.4

NITOX : increase in nitric oxide concentration by 1 will lead to decrease in value of a home by $22810.9

ROOMS : increase in number of rooms per dwelling by 1 will lead to increase in value of a home by $6371.5

AGE : increase in AGE(proportion of owner-occupied units built prior to 1940) by 1 will lead to decrease in value of a home by $47.7

DIST : increase in weighted distance to 5 Boston employment centers by 1 will lead to decrease in value of a home by $1335.3

ACCESS : increase in index of accessibility to radial highways by 1 will lead to increase in value of a home by $272.3

TAX : increase in full-value property-tax rate per $10,000 by 1 will lead to decrease in value of a home by $12.6

PTRATIO : increase in pupil-teacher ratio by by 1 will lead to decrease in value of a home by $1176.8

(b) 95% interval, N=506, K=9

CRIME :

T(0.975, 497)=1.96

-0.1834 + 1.96\*0.0365 = -0.11186

-0.1834 - 1.96\*0.0365 = -0.25494

95% interval estimate for CRIME = [-0.25494,-0.11186]

ACCESS :

T(0.975, 497)=1.96

0.2723 + 1.96\*0.0723 = 0.414

0.2723 - 1.96\*0.0723 = 0.131

95% interval estimate for ACCESS = [0.131, 0.414]

(c) H0 : 4=7, H1 : 4!=7

 = 0.05 level of significance

T(0.975, 497)=1.96

T = (6.3715-7)/0.3924=-1.60

T is within interval [-1.96, 1.96], so we can not reject H0

(d) H0 : 9 >= -1, H1 : 9 < -1

 = 0.05 level of significance

T(0.05, 497)=-1.645

T = (-1.1768+1)/0.1394 = -1.268

T is larger than -1.645, so we cannot reject H0. Reducing pupil-teacher ratio by 10 may NOT increase the value of a house by more than $10000.

3.(a)

> br2=read.csv("br2.csv")

> reg\_br2 = lm(br2$price~br2$sqft+br2$Age, data=br2)

> summary(reg\_br2)

Call:

lm(formula = br2$price ~ br2$sqft + br2$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 \*\*\*

br2$sqft 90.970 2.403 37.855 < 2e-16 \*\*\*

br2$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

1. The increase of 1 square feet leads to increase of selling price by $90.97.

The increase in the age of the house by 1 year leads to decrease of selling price by $755

1. Degree of freedom = 1077, T(0.975,1077) = 1.96

90.97- 1.96\*2.403 = 86.260, 90.97 + 1.96\*2.403 = 95.680,

95% interval = [86.260, 95.680]

1. H0 : 3>= -1000, H1 : 3 < -1000,

T = (-755.041 – (-1000))/ 140.894 = 1.739

= 0.05, Tc = T(0.05, 1077) = -1.645

T > Tc, so we can not reject H0.

(b) PRICE = 1 + 2 SQFT + 3AGE +4SQFT2 + 5AGE2 + e

> sqft2 = br2$sqft \* br2$sqft

> Age2 = br2$Age \* br2$Age

> model2 = data.frame(br2$price,br2$sqft,br2$Age,sqft2,Age2)

> reg\_model2= lm(br2$price~br2$sqft+br2$Age+sqft2+Age2,data=model2)

> summary(reg\_model2)

Call:

lm(formula = br2$price ~ br2$sqft + br2$Age + sqft2 + Age2, data = model2)

Residuals:

Min 1Q Median 3Q Max

-805011 -23873 -1375 18067 659703

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.701e+05 1.043e+04 16.310 < 2e-16 \*\*\*

br2$sqft -5.578e+01 6.389e+00 -8.731 < 2e-16 \*\*\*

br2$Age -2.798e+03 3.051e+02 -9.170 < 2e-16 \*\*\*

sqft2 2.315e-02 9.642e-04 24.013 < 2e-16 \*\*\*

Age2 3.016e+01 5.071e+00 5.948 3.68e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 63370 on 1075 degrees of freedom

Multiple R-squared: 0.7352, Adjusted R-squared: 0.7342

F-statistic: 746 on 4 and 1075 DF, p-value: < 2.2e-16

PRICE = 170100 – 55.78\*SQFT - AGE +SQFT2 + 30.16\*AGE2

1. PRICESQFT = -55.78 + 0.046\*SQFT

The smallest house = 662 SQFT, The largest house = 7898 SQFT

> min(model2$br2.sqft)

[1] 662

> max(model2$br2.sqft)

[1] 7897

 For the smallest house, PRICESQFT = -55.78 + 0.046 \* 662 = -25.328

For the largest house, PRICESQFT = -55.78 + 0.046 \* 7897 = 307.482

For the house with 2300SQFT, PRICESQFT = -55.78 + 0.046 \* 2300= 50.02

For the smallest house, the increase of 1 SQFT will lead to decrease of $25.328 in selling price. This is unrealistic.

For the largest house, the increase of 1 SQFT will lead to increase of $307.482 in selling price. This is OK.

For the house with 2300 SQFT, the increase of 1 SQFT will lead to increase of $50.02 in selling price. This is OK.

1. PRICEAGE = -2798 + 60.32\*AGE

The newest house = 1 AGE, The newest house = 80 AGE

> min(model2$br2.Age)

[1] 1

> max(model2$br2.Age)

[1] 80

For the newest house, PRICEAGE = -2798 + 60.32 \* 1 = -2737.68

(The increase of 1 AGE will lead to decrease of $2737.68 in selling price. This is OK.)

For the oldest house, PRICEAGE = -2798 + 60.32 \* 80 = 2027.6

(The increase of 1 AGE will lead to increase of $2027.6 in selling price. This is unrealistic.)

For the house of 20 years old, PRICEAGE = -2798 + 60.32 \* 20 = - 1591.6

(The increase of 1 AGE will lead to decrease of $1591.6 in selling price. This is OK.)

1. For the house with 2300SQFT, PRICESQFT = -55.78 + 0.046 \* 2300= 50.02

95% interval : because standard error of SQFT = 6.389, T(0.975, 1075) = 1.96

-55.78 + 1.96\*6.389 = -43.258,

-55.78 - 1.96\*6.389 = -68.302

95% interval for PRICESQFT = [-68.302, -43.258]

1. For a house of 20 years old, H0: PRICE/AGE >= -1000, H1: PRICE/AGE < -1000,

Because the estimate of PRICE/AGE = -2798, standard error = 305.1

For  = 0.05, T = (-2798 – (-1000)) / 305.1 = -5.89

Tc = T(0.05, 1075) = -1.645

T < Tc, so we reject H0 and accept H1

(c) PRICE = 1 + 2 SQFT + 3AGE +4SQFT2 + 5AGE2 + 6AGE\*SQFT + e

> Age.sqft = br2$Age \* br2$sqft

> model3 = data.frame(br2$price,br2$sqft,br2$Age,sqft2, Age2, Age.sqft)

> reg\_model3= lm(br2$price~br2$sqft+br2$Age+sqft2+Age2+Age.sqft, data=model3)

> summary(reg\_model3)

Call:

lm(formula = br2$price ~ br2$sqft + br2$Age + sqft2 + Age2 + Age.sqft, data = model3)

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 \*\*\*

br2$sqft -3.073e+01 6.898e+00 -4.455 9.27e-06 \*\*\*

br2$Age -4.420e+02 4.106e+02 -1.077 0.282

sqft2 2.218e-02 9.425e-04 23.537 < 2e-16 \*\*\*

Age2 2.652e+01 4.939e+00 5.370 9.66e-08 \*\*\*

Age.sqft -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

1. PRICESQFT = b2 + 2b4 \* SQFT + b6\*AGE = -30.73 + 2\*0.0222\*SQFT – 0.9306\*AGE

For the house of SQFT=2300 and AGE=20,

PRICESQFT = -30.73 + 0.0444\*2300 -0.9306\*20 = -30.73 + 102.12-18.612 =52.778

1. PRICEAGE = b3 + 2\*b5\*AGE + b6\*SQFT = -442 +2\*26.52\*AGE – 0.9306\*SQFT

For the house of SQFT=2300 and AGE=20,

PRICEAGE = -442 + 53.04\*20 – 0.9306\*2300 = -442+1060.8-2140.38=-1521.58

1. The estimate of PRICESQFT =52.778, standard error = 6.898,

95% interval : T(0.975,1074) = 1.96,

52.778 + 1.96\*6.898 = 66.298

52.778 - 1.96\*6.898 = 39.258

95% interval = [39.258, 66.298]

1. H0: PRICE/AGE >= -1000, H1: PRICE/AGE < -1000,

The estimate of PRICEAGE = -1521.58 , standard error = 410.6,

Tc = T(0.05,1074) = -1.645

T = (-1521.58-(-1000)) / 410.6 = -1.270

T > Tc, so we can not reject H0

(d)

In part (a) , R-squared=0.5896, 59% of variation in PRICE can be explained by SQFT, AGE. The p-values of SQFT coefficient and AGE coefficient : < 2e-16 and 1.02e-07, saying that these 2 variables are significant.

In part (b) , R-squared=0.7352, 73.5% of variation in PRICE can be explained by the variables SQFT, AGE, SQFT2, AGE2, increasing 14.5% from (a). The p-values of the coefficients of SQFT, AGE, SQFT2, AGE2  are : < 2e-16, < 2e-16, < 2e-16, and 3.68e-09, small enough to say that these 4 variables are significant.

In part (c) , R-squared=0.751, 75.1% of variation in PRICE can be explained by the variables SQFT, AGE, SQFT2, AGE2, AGExSQFT. The increase of only 1.6% from (b) says this model cannot fit much better than model (b) does by including extra variable AGExSQFT. The p-values of SQFT, AGE, SQFT2, AGE2, AGExSQFT : 9.27e-06, 0.282, < 2e-16, < 9.66e-08, 3.72e-16 . We can see the p-value for the coefficient of AGE becomes too large (0.282), making the variable AGE not significant.

4.(a)

> boston = read.csv("boston.csv")

> reg\_boston=lm(boston$VALUE~boston$CRIME+boston$NITOX+boston$ROOMS+boston$AGE+boston$DIST+boston$ACCESS+boston$TAX+boston$PTRATIO)

> library(sandwich)

> sqrt(diag(vcovHC(reg\_boston, type = "HC1")))

(Intercept) boston$CRIME boston$NITOX boston$ROOMS boston$AGE

7.37996160 0.03472230 4.35949994 0.66549321 0.01077200

boston$DIST boston$ACCESS boston$TAX boston$PTRATIO

0.19023409 0.07471355 0.00284265 0.12358891

> sqrt(diag(vcovHC(reg\_boston, type = "const")))

(Intercept) boston$CRIME boston$NITOX boston$ROOMS boston$AGE

5.365948055 0.036488720 4.160741151 0.392386610 0.014101810

boston$DIST boston$ACCESS boston$TAX boston$PTRATIO

0.200146828 0.072276042 0.003770155 0.139415353

So we reestimate the equation with White’s standard error as below table:

Coefficients:

Estimate Std. Error White Std. Error(HC1)

(Intercept) 28.40666 5.36595 7.37996

boston$CRIME -0.18345 0.03649 0.03472 (smaller)

boston$NITOX -22.81088 4.16074 4.35950 (larger)

boston$ROOMS 6.37151 0.39239 0.66550 (larger)

boston$AGE -0.04775 0.01410 0.01077 (smaller)

boston$DIST -1.33527 0.20015 0.19023 (smaller)

boston$ACCESS 0.27228 0.07228 0.07471 (larger)

boston$TAX -0.01259 0.00377 0.00284 (smaller)

boston$PTRATIO -1.17679 0.13942 0.12359 (smaller)

(b)95% interval in problem 2

CRIME :

T(0.975, 497)=1.96

-0.18345 + 1.96\*0.03649 = -0.1119

-0.18345 - 1.96\*0.03649 = -0.2550

95% interval estimate for CRIME = [-0.2550, -0.1119]

ROOMS :

6.37151 + 1.96\*0.39239 = 7.1406

6.37151 - 1.96\*0.39239 = 5.6024

95% interval estimate for ROOMS = [5.6024, 7.1406]

AGE :

-0.04775 + 1.96\*0.0141 = -0.0201

-0.04775 - 1.96\*0.0141 = -0.0754

95% interval estimate for AGE =[-0.0754, -0.0201]

TAX :

-0.01259 + 1.96\*0.00377 = -0.0052

-0.01259 - 1.96\*0.00377 = -0.01998

95% interval estimate for TAX =[-0.01998, -0.0052]

95% interval with White’s standard errors:

CRIME :

T(0.975, 497)=1.96

-0.18345 + 1.96\*0.03472 = -0.1154

-0.18345 - 1.96\*0.03472 = -0.2515

95% interval estimate for CRIME = [-0.2515, -0.1154]

ROOMS :

6.3715 + 1.96\*0.66550 = 7.67588

6.3715 - 1.96\*0.66550 = 5.06712

95% interval estimate for ROOMS = [5.06712, 7.67588]

AGE :

-0.04775 + 1.96\*0.01077 = -0.02664

-0.04775 - 1.96\*0.01077= -0.06886

95% interval estimate for AGE =[-0.06886, -0.02664]

TAX :

-0.0126 + 1.96\*0.00284 = -0.00703

-0.0126 - 1.96\*0.00284 = -0.01817

95% interval estimate for TAX =[-0.01817, -0.00703]

**Comparison :**

95% Interval Problem 2 95% Interval with White Stan. Error

CRIME [-0.2550, -0.1119] [-0.2515, -0.1154], narrower

ROOMS [5.6024, 7.1406] [5.06712, 7.67588], wider

AGE [-0.0754, -0.0201] [-0.06886, -0.02664], narrower

TAX [-0.01998, -0.0052] [-0.01817, -0.00703], narrower

(c)

When we recognize the presence of heteroscedasticity, our calculation could be more precise. In this case, we get narrower interval for CRIME, AGE, and TAX and avoid incorrect values for test statistics.

(d)

If incorrect standard errors are used, it might lead to wrong inference on hypothesis test.