Obaid Masih

Econometrics

Assignment #5

Question 6.4 (p.246-247)

a)

Equation A

	parameter values	Se	TC	T values	Significant
educ	0.0498	0.0397	2	1.25440806	NO
educ^2	0.00319	0.00169	2	1.887573964	NO
Exper	0.0373	0.0081	2	4.604938272	YES
EXPER^2	-0.000485	0.00009	2	-5.388888889	YES
Exper*educ	-0.00051	0.000482	2	-1.058091286	NO
Hwk_HRS	0.01145	0.00137	2	8.357664234	YES

Equation B

EQUATION B

	parameter values	Se	TC	T values	Significant
educ	0.0289	0.0344	2	0.840116279	NO
educ^2	0.00352	0.0016	2	2.2	YES
Exper	0.03	0.0048	2	6.25	YES
EXPER^2	-0.000456	0.000086	2	-5.302325581	YES
Hwk_HRS	0.01156	0.00137	2	8.437956204	YES

EQUATION C

	parameter values	Se	TC	T values	Significant
educ	0.0366	0.035	2	1.045714286	NO
educ^2	0.0029	0.0017	2	1.705882353	NO
Hwk_HRS	0.0134	0.00136	2	9.852941176	YES

EQUATION D

	parameter values	Se	TC	T values	Significant
Exper	0.0279	0.0054	2	5.166666667	YES
EXPER^2	-0.00047	0.000096	2	-4.895833333	YES
Hwk_HRS	0.01524	0.00151	2	10.09271523	YES

EQUATION E

	parameter values	Se	TC	T values	Significant
educ	0.1006	0.0063	2	15.96825397	YES
Exper	0.0295	0.0048	2	6.145833333	YES
EXPER^2	-0.00044	0.000086	2	-5.11627907	YES
Hwk_HRS	0.01188	0.00136	2	8.735294118	YES

b)

F critical value (from chart in book)
Fc(1, infinity) = 3.84

SSE (u) = 222.4166
SSE(r) = 222.6674
N=1000
J=1
K =7

$$F = \frac{\left(SSE_R - SSE_U\right)/J}{SSE_U/\left(N-K\right)}$$

F value is 1.11

F value is very small which means that Ho (B6 = 0) cannot be rejected.

T value for Experience * Education is the smallest so for equation B we should drop that variable BECAUSE it also means that HO (B6=0) cannot be rejected. Hence we make B6=0 and remove the variable Educ* Exper from the equation.

T statistic for the coefficient with null hypotheses = -.000510/.000482= -1.058.

Critical value of t with 993 degrees of freedom is 1.962

This means that parameter of EDUC*EXPER should be 0. Hence it should be removed from the model

c)

F Critical value is F (3, infinity) = 2.60

SSE (u) = 222.4166
SSE(r) = 233.83
N=1000
J=3

K =7
F = 16.98

This is bigger than the F critical meaning HO will be rejected.

In the above case HO is that Beta(Exper) =0 , Beta(Exper^2)=0 , Beta (Exper *Educ) =0
Question we are trying to answer is to analyze the impact of Experience on LN(WAGE)

Since Null hypothesis is rejected which means that Experience does have an impact on the Ln(wage)

d)

F critical value → F (2, infinity) = 3.0

SSE (u) = 222.4166
SSE(r) = 233.83
N=1000
J=2
K =6

F value = 129.09

Here the HO = All parameters of Educ are zero. Meaning that Educ has no impact of LN(WAGE)

129.09 is a lot more than Fc which means that HO can be rejected.

e)

F critical value → F (2, infinity) = 3.0

SSE (u) = 222.4166
SSE(r) = 223.6716
N=1000
J=2
K =7
F value is 2.799

HO is that Beta(Educ^2) =0 and Beta(Educ * Exp) = 0

HO cannot be rejected.

We are trying to see if Educ has a simple linear relationship with LN(WAGE)
f)

From part a to part e few insights are :

EDUC*EXPER is not an important parameter and can be 0. This is because HO to loose EDUC*EXPER cannot be rejected.

Experience is important probably very important because F value is 16, compared to critical value of 2.60. Meaning that HO that experience has zero impact can be rejected.

Part d reveals that Education is important. It has a F value of 129 compared to critical value of 3.0.

Part e reveals while Educ is important but EDUC^2 and EDUC*EXP is not important,

This leads us to believe that equation should be

LN(WAGE) = B1 + B2.EDUC + B3.EXP+B4.EXP^2+B5.HRSWRK+e

This is equation E. Also t values for equation E are all above Tc=2.

g)

AIC for equation D is

N= 1000
K=7 – parameters in Unrestricted model
SSE = 280.50

AIC = ln(SSE/N) + 2k/N
AIC= -1.25

SC for equation A

N=1000
K=7
SSE = 222.416

Ln(SSE/N) + K.LN(N)/N

-1.45

BY AIC equation B is favored as it has the lowest value

BY SC or BIC equation E is favored as it has the lowest value

Question 6.5 (p.247)

a) Null Hypotheses $H_0 = \beta_2 = \beta_4$ and $\beta_3 = \beta_5$ against
Alternate hypotheses= $\beta_2 \neq \beta_4$ or $\beta_3 \neq \beta_5$ or both

b)
That makes the restricted model to be:
$\ln(WAGE) = \beta_1 + \beta_2 *(EDUC+EXPER) + \beta_3 *(EDUC^2+EXPER^2)$

c)

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}$$

F critical value (2, infinity)= 3.00

SSE.r = 254.17
SSE.U = 222.66
J=2
K = 6
N=1000

F = 70.9

HO can be rejected .

Question 6.15 (p.250) no need for part e

a)

SPRICE= 11154.29+10680*livarea-11.33*age-15552.44*beds-7019.296*baths

Summary of model is as follows:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   11154.29    6555.11   1.702   0.0890 .
livarea       10680.00     273.15  39.100  < 2e-16 ***
beds         -15552.44    1970.01  -7.895 5.59e-15 ***
baths         -7019.30    2903.82  -2.417   0.0158 *
age             -11.33      80.50  -0.141   0.8881


Residual standard error: 37580 on 1495 degrees of freedom
F-statistic:   688 on 4 and 1495 DF,
p-value: < 2.2e-16
```

b)

Intercept = 11154.29
B_age = -11.33
price2 = Intercept + B_age*2

Price with age 10
price10 = Intercept + B_age*10
she can expect a difference of
difference = price2 - price10 = $90

95% interval for price2
tc=1.96
se_age = 80.50

intervalL = 8*(B_age -(tc*se_age))
print(intervalL)

intervalH =8* (B_age +(tc*se_age))

print(intervalH)
[-1352.88,1171.6]

This shows the 95% confidence range is too high. Model needs some refinement.

c)

Intercept = 11154
B_livarea = 10680
SE_livarea = 273.15

B2 is the parameter of living area
Ho: B2 <10,000 because 20,000/2 = 10,000
H1: B2 >=10000

T critical is 1.645

T value = (10,680 – 10,000)/273.15 = 2.48

HO can be rejected and meaning that price increase will be atleast $20,000

d)

B_livarea= 10680
B_bath = -7019
IncreaseInprice = (B_livarea*2) + (B_bath)*1
print(IncreaseInprice) = 5807

For intervals

```
covmat
            (Intercept)      livarea         beds       baths         age
(Intercept)  42969508.8   390445.711 -7707291.996 -7951852.97 -287485.490
livarea        390445.7    74610.431  -170680.218  -477425.43   -2782.794
beds         -7707292.0  -170680.218  3880921.738 -1122463.21    9593.284
baths        -7951853.0  -477425.434 -1122463.209  8432146.17   75436.212
age           -287485.5    -2782.794     9593.284    75436.21    6480.578
```

var.b2 = covmat[2,2]
var.b3 = covmat[3,3]
cov.b3.b2 =covmat[3,2]

temp = 4*var.b2 +var.b3 + 4*cov.b3.b2
se_eq1 = sqrt(temp)
print(se_eq1)

Lo = IncreaseInprice-(1.96*se_eq1)

Hi= IncreaseInprice + (1.96*se_eq1)
print(Lo) = 2140.495
print(Hi)= 9472.62

[2140,9472.62]

Question 6.16 (p.250-251)

a)

SPRICE= 79755.74+2994.652*livarea-830.3785*age-11921.92*beds-4971.063*baths + 169* $livarea^2$ + 14.23*age^2

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  79755.736   8744.293   9.121  < 2e-16 ***
livarea       2994.652    772.295   3.878  0.00011 ***
beds        -11921.923   1927.050  -6.187 7.92e-10 ***
baths        -4971.063   2797.366  -1.777  0.07576 .
age           -830.379    197.779  -4.199 2.85e-05 ***
livareasq      169.092     16.125  10.486  < 2e-16 ***
agesq           14.233      3.356   4.241 2.36e-05 ***
---

Residual standard error: 36080 on 1493 degrees of freedom
F-statistic:   519 on 6 and 1493 DF,
 p-value: < 2.2e-16
```

b)

```
Find F value

Unrestricted model is

              Estimate
(Intercept)  79755.736
livarea       2994.652
beds        -11921.923
baths        -4971.063
age           -830.379
livareasq      169.092
agesq           14.233

SSE.U =   1.9435e+12
```

```
Restricted model is

              Estimate
(Intercept)   11154.29
livarea       10680.00
beds         -15552.44
baths         -7019.30
age             -11.33
```

SSE.r = 2.111122e+12

HO is that Beta of LivingSq and Agesq are 0
H1 Beta of livSq=0, Beta of Agesq =0 , Or they both are 0

J=2
N=1500
K=7
fvalue = ((SSE.r-SSE.u)/J) / (SSE.u/(N-K)) = 64.38379

F criticall value is only 3. Since 64.38 is more than that so HO can be rejected. Meaning we will use Living square and age square in our model.

c)

Redoing the parts in question 6.15

(c) (i)   SPRICE2 = 2(b4 + b5*2)
SPRICE10 = 10(b4 + b5*10)
Expected price difference = SPRICE2 – SPRICE10
Expected price difference = -8b4 – 96b5
Expected price difference = 5276.7
se(-8b4 – 96b5) = 1291.95

Price Range = [2741.9, 7811.5]

(ii) SPRICE2000 = 20(b2 + b3*20)
SPRICE2200 = 22(b2 + b3*22)
Expected price difference = SPRICE2000 – SPRICE2200
Expected price difference = 2b2 + 84b3
Expected price difference = 20193
se(2b2 + 84b3) = 534.55
H0 = 2b2 + 84b3 ≤ $20000
H1 = 2b2 + 84b3 > $20000
t = (2b2 + 84b3 – 20000) / 534.55
t = 0.361
tc = 1.646
|t| < tc therefore we cannot reject the null.

(iii)

Expected increase in price of the house if we add a bedroom of size 200 sft

Total price increase = b4 + 2b2+ 76b6 = -11921.92+2*2994.652+76*169.0916 = 6918.34

95% confidence interval for the price increase with adding a bedroom with 200 sft is between 3382.705 and 10453.98

Question 6.22 (p. 252)

```
model = lm(pizza~age+income+agetimeincome, data=context)
#Coefficients:
#               Estimate    Std. Error   t value   Pr(>|t|)
(Intercept)   161.46543    120.66341   1.338   0.1892
 age          -2.97742      3.35210    -0.888   0.3803
income         6.97991      2.82277     2.473   0.0183 *
agetimeincome -0.12324    0.06672     -1.847   0.0730
```

PART A)

Restricted Model pizza~income

ho → b2=0 , B4=0

```
Fcritical value (2,38) = 3.23
SSE.U = 580608.7
SSE.R = 819285.8
```

J=2

N=40

K=4

Fval = ((SSE.r-SSE.u)/J) / (SSE.u/(N-K))

`print(Fval)  =  7.399458`

HO can be rejected, meaning that AGE is very important to be used in predicting Pizza

PART B)

Tc = 2.028

Marginal effect of income is DP /DI =Beta(income) + age*Beta(age.income)
Find se (done in r) for eacg point estimate

Interval Hi = Marginal effect + tc*SE
Interval low = Marginal effect – tc*SE

Age	**Point Estimate**	**Standard Error**	**95% confidence Interval (Lower)**	**95% confidence Interval (Upper)**
20	4.515	1.5	1.43	7.59
30	3.282	.9	1.447	5.117
40	2.05	.46	1.107	2.993
50	.8179	.70	-.6219	2.25
55	.2017	.99	-1.807	2.211

PART C)

Introduce term (AGE^2 *income) Result is as foll
We anticipate that b4>0 and b5 <0 since we expect that the propensity to spend increases with age up to a point and then falls.

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   109.720767 135.572473   0.809    0.424
age            -2.038273   3.541904  -0.575    0.569
income         14.096163   8.839862   1.595    0.120
agetimeincome  -0.470371   0.413908  -1.136    0.264
AGE_SQ_INCOME   0.004205   0.004948   0.850    0.401
```

The p value is .424 which suggests we cannot reject the null hypotheses that b5 is not significantly different from zero. The same applies for b4 in which we cannot reject null hypotheses. The model suggests that age parameter and age2 is insignificant d)

Below are the point estimates and 95% confidence interval of the marginal propensity to spend based on the new model

Age	*Estimate*	*Standard Error*	*95% confidence Interval (Lower)*	*95% confidence Interval (Upper*
20	6.37	2.66	0.962	11.77
30	3.76	1.073	1.58	5.94
40	2.00	0.469	1.056	2.96
50	1.089	0.781	-0.496	2.67
55	0.945	1.32	-1.743	3.63

We see that propensity to spend on pizza decreases with age with age group 20 and 30 have higher propensity to spend compared to age group 40, 50 and 50.

However we see that confidence interval for age group 20 and 30 is wide which makes the estimate unreliable. All the confidence interval overlap which suggests that irrespective of age group the propensity to spend could be the same. Based on the confidence interval for the quadratic function of age it suggests that quadratic function of age has no effect on the propensity.

PART E)

Null Hypotheses = b2 =0 and b4 =0 and b5 =0 Alternate Hypotheses = At least one of the beta coefficients is not equal to zero.

The coefficients are insignificant at α = 5%, due to high p-values.

H0 = b2 = b4 = b5 = 0

H1 = b2 ≠ 0 or b4 ≠ 0 or b5 ≠ 0

for this hypothesis test, let the restricted model be

$pizza = income + e$

unrestricted model

$pizza = age + income + (age * income) + (age * age * income) + e$

N = 40

K = 5

J = 3

SSER = 819285.8 SSEU = 568869.2

F = ($SSER – SSEU$ / J) $SSEU/(N-k)$

F = (819285.84–568869.2 / 3) 568869.2/(35)

F = 5.13
Fc = 2.874

Since the F > Fc, we can reject the null hypothesis of age having no effect on pizza.

PART F)

Since, the p-value of the coefficients is insignificant, but the F test proved that these variables must be considered for regression. But running the auxiliary model, we find that there is a high correlation between the variables used due to a high R2 value. This indicates that collinearity is causing the problem for high p-values.
On introducing the AGE3 * INCOME the collinearity still exists between the variables since the p-values are still insignificant and the R2 value is as high as 0.99.

****
****
*********** CODE FOR ASSIGNMENT

```
###########################################
## QUESTION 6.15
rm(list=ls(all=TRUE))


library(data.table)
library(dplyr)


context = fread('stckton.csv')
head(context)
## part A
model = lm( sprice~livarea+beds+baths+age,data=context)
mat1<-summary(model)
print(mat1)
SSE.r =   sum((context$sprice-predict(model))^2)
## PartB

## price with age 2
Intercept = 11154.29
B_age = -11.33
price2 = Intercept + B_age*2
```

```
## Pric with age 10
price10 = Intercept + B_age*10

## she can expect a difference of
difference = price2 - price10

## 95% interval for price2
tc=1.96
se_age = 80.50

intervalL = 8*(B_age -(tc*se_age))
print(intervalL)

intervalH =8* (B_age +(tc*se_age))
print(intervalH)
##[-1352.88,1171.6]

##PARTC
## house increased by 200 sqft
Intercept = 11154
B_livarea = 10680
SE_livarea = 273.15

IncreaseInprice = (B_livarea*2)
print(IncreaseInprice)
##21360
#  20,000/2 is that 10,000
# HO is that B2(livarea) < 10,000
#H1 is that B2(livarea) >10,000
## critical value of t = 1.645 (positive)
tval= (B_livarea-10000)/SE_livarea
print(tval)
## It is in negative which means that we reject

## Part D
B_livarea= 10680
B_bath = -15552.44
IncreaseInprice = (B_livarea*2) + (B_bath)*1
print(IncreaseInprice)
## Interval for the price difference goes to
covmat = vcov(model)

var.b2 = covmat[2,2]
```

```
var.b3 = covmat[3,3]
cov.b3.b2 =covmat[3,2]

temp = 4*var.b2 +var.b3 + 4*cov.b3.b2
se_eq1 = sqrt(temp)
print(se_eq1)

Lo = IncreaseInprice-(1.96*se_eq1)
Hi= IncreaseInprice + (1.96*se_eq1)
print(Lo)
print(Hi)

## The model may not be aboslutely correct as it assumes that relations between important
parameteres
## are simple linear

#######################################################################
####################6.16
## Part A
rm(list=ls(all=TRUE))
library(data.table)
library(dplyr)
context = fread('stckton.csv')
head(context)
context1  <- context %>%
        mutate(livareasq=livarea^2) %>%
        mutate(agesq=age^2)

head(context1)
model1 = lm( sprice~livarea+beds+baths+age+livareasq+agesq ,data=context1)
print(summary(model1))
covmat = vcov(model1)

## Part B
## SSE unrestricted is the equation with all the parameters
SSE.u = sum((context1$sprice-predict(model1))^2)
print(SSE.u)
## SSE restricted is the equation with less  parameters
SSE.r = 2.111122e+12
J=2
N=1500
K=7
fvalue = (  (SSE.r-SSE.u)/J  ) /  (  SSE.u/(N-K) )
print(fvalue)
```

```
## HO --> b8 =0 and B9 =0
## H1 --> B8 is not 0, B9 is not 0, Or both are not zero
#f_critical value is f(2,infinity) =3
## 18.53 is well above 3 so HO can be rejected It is better to use Un restricted than restricted

## PART C
## Redoing the parts in question 6.15 but with the new model
## get difference in price with age =2 and age =10
##marginal effect of sprice with age
age=8
difference = -830.37 +(14.233*2)*(age)
print(difference)
## -602
## find the SE for the equation    DP/DA = B5+2B7.age
var.b5 = covmat[5,5]
var.b7 = covmat[7,7]
var.b7.b5 = covmat[7,5]

temp= var.b5 +(16*16*var.b7)+(2*2*8*var.b7.b5)
se1 = sqrt(temp)
print(se1)

hi = difference +(1.96*se1)
print(hi)
li = difference -(1.96*se1)
print(li)
 ##
## Find the difference that 200 sqft will make
## marginal effect of sqft
LVG = 2
diff = 2994.652 + (2*169.092*LVG)
print(diff)
## 3671.02
## H0 -->  DP.DSQ <10,000
## H1 --> DP.DSQ >= 10,000
## calculate DP.DSQ for one sqft (HUNDREDS)
var.b2 = covmat[2,2]
var.b6 = covmat[6,6]
var.b2.b6 = covmat[2,6]

temp1= var.b2 +(4*var.b6)+(2*2*var.b2.b6)
se2 = sqrt(temp1)
print(se2)
```

```
## now that SE is calculated now we can calculate the TVAL

LVG = 1
BETA = 2994.652 + (2*169.092*LVG)
Tval = (BETA - 10000)/se2
print(Tval)
## Tc =1.645 and tval is -8.984 meaning that HO cannot be rejected.

############
## Now 200 sqft increase in house LVG and 1 bed added
## DP/DB+DSQ = B2 +2.B6.LVG + B3
LVG = 2
diff1 = 2994.652 + (2*169.092*LVG)+ (-11921.923)
print(diff1)
## -8250

## find intervals with 95%  so t =1.96
var.b2 = covmat[2,2]
var.b6 = covmat[6,6]
var.b3 =covmat[3,3]
var.b2.b6 = covmat[2,6]
var.b3.b2 =covmat[3,2]
var.b6.b3 =covmat[6,3]

temp2= var.b2 + (16*var.b6) + var.b3 + (8*var.b2.b6) + (2*var.b3.b2) + (8*var.b6.b3)
se3 = sqrt(temp2)
print(se3)

hi = diff1 + (1.96*se3)
print(hi)
li = diff1 -(1.96*se3)
print(li)

#####################################################################################
######################################################
############### 6.22 ###################################################
rm(list=ls(all=TRUE))
library(data.table)
library(dplyr)
context = fread('pizza.csv')
head(context)
context <-  context %>%
        mutate(AGE_SQ_INCOME= (age^2)*income)
```

```
model = lm(pizza~age+income+agetimeincome, data=context)
print(summary(model))
SSE.u =   sum( (context$pizza - predict(model))^2 )
finmat = vcov(model)
#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)   161.46543  120.66341   1.338   0.1892
#age           -2.97742    3.35210  -0.888   0.3803
#income         6.97991    2.82277   2.473   0.0183 *
# agetimeincome  -0.12324    0.06672  -1.847   0.0730 .

### Restricted Model  pizza~income
##Part a
## ho b2=0 , B4=0

## Restricted model is as follows
model.r = lm(pizza~income, data=context)
print(summary(model.r))
SSE.r =   sum( (context$pizza - predict(model.r))^2 )
print(SSE.r)

## Fcritical value (2,38) = 3.23
J=2
N=40
K=4
Fval = ((SSE.r-SSE.u)/J ) / (SSE.u/(N-K))
print(Fval)
## HO can be rejeceted , meaning that AGE is very important to be used in predicting Pizza

### PART B
## Find point estimates
## marginal effect of income is
##DP/DI = Beta(income) + B3.age
age = 20
MPE = 6.9799 +(-0.123*age)
print(MPE)
tc = 2.028
### Find standard error
var.b3 = finmat[3,3]
var.b4 = finmat[4,4]
var.b2.b3 = finmat[4,3]
temp = var.b3 + (age^2)*var.b4 + (2*age*var.b2.b3)
se = sqrt(temp)
print(se)
```

```
hi = MPE + (se*tc)
print(hi)
li = MPE - (se*tc)
print(li)


######### PART C
head(context)
new_model <- lm(pizza~age+income+agetimeincome+ AGE_SQ_INCOME, data =context)
summary(new_model)
matddd = vcov(new_model)

######PART D
### Marginal effect of Price withe income
# DP/DI = B2 + B3*AGE+ B4.AGE^2
age =20
MPE = 14.0961 + (age*-0.47037) + (age^2)*0.004205
print(MPE)
tc=2.028

var.b3 = matddd[3,3]
var.b4 =matddd[4,4]
var.b5 = matddd[5,5]
var.b3.b4 = matddd[3,4]
var.b4.b5 = matddd[4,5]
var.b5.b3 = matddd[3,5]
temp = var.b3 + (age^2)*var.b4 + (age^4)*var.b5  + (2*age)*var.b3.b4 + 2*(age^3)*var.b4.b5
+2*(age^2)*var.b5.b3
se= sqrt(temp)
print(se)

hi = MPE + (se*tc)
print(hi)
li = MPE - (se*tc)
print(li)
```