Obaid Masih

HW 6 Econometrics

Question 7.1 (p.287)

a)

Intercept or B1 = 24,200.
B2 is 1643. Meaning for every one unit increase in GPA the salary goes up by $1,643
B3 is 5033. Meaning if you have taken econometric course than your salary will be $5033 more THAN those who have not taken the course.

Ref group is B1 + B2(GPA) which student who never took econometrics.

b) Add an indicator variable FEMALE to the equation.
If estimator for the variable is negative then we will know that women from same university are paid less when starting out.

c)

Add a variable that is the product of FEMALE & METRICS

Now the estimator for (FEMALE*METRICS) can tell us how salary is impacted when FEMALE take METRICS

Question 7.4 (p.289)

SQFT 72.7878 1.0001 72.7773 0.0000

AGE 179.4623 17.0112 10.5496 0.0000

a) Sqft beta = $72.78. Meaning that for every one unit increase in SQFT the price of home increase by $72.78
   It is significant at 5% level and 1% level with a t value of 72.77

   AGE beta = $179.4623 meaning that for every 1 year of age the house value decreases by $179.4623

   It is significant at 5% level and 1% level with a t value of -10.54

b) All indicator variables from 92 -96 have negative estimators. But as we move from 92 to 96 then impact of years on price becomes more significant. And they are negatively impacting the price of houses.
A house in 92 would decrease the price by $4392 but house in 96 would decrease the price by $23,663.

c) Adding D91 will create an exact collinearity and model will be undefined

Currently d91 is the reference model meaning when house is in 91 then model is

B1 + BETA(SQFT) +BETA(AGE)

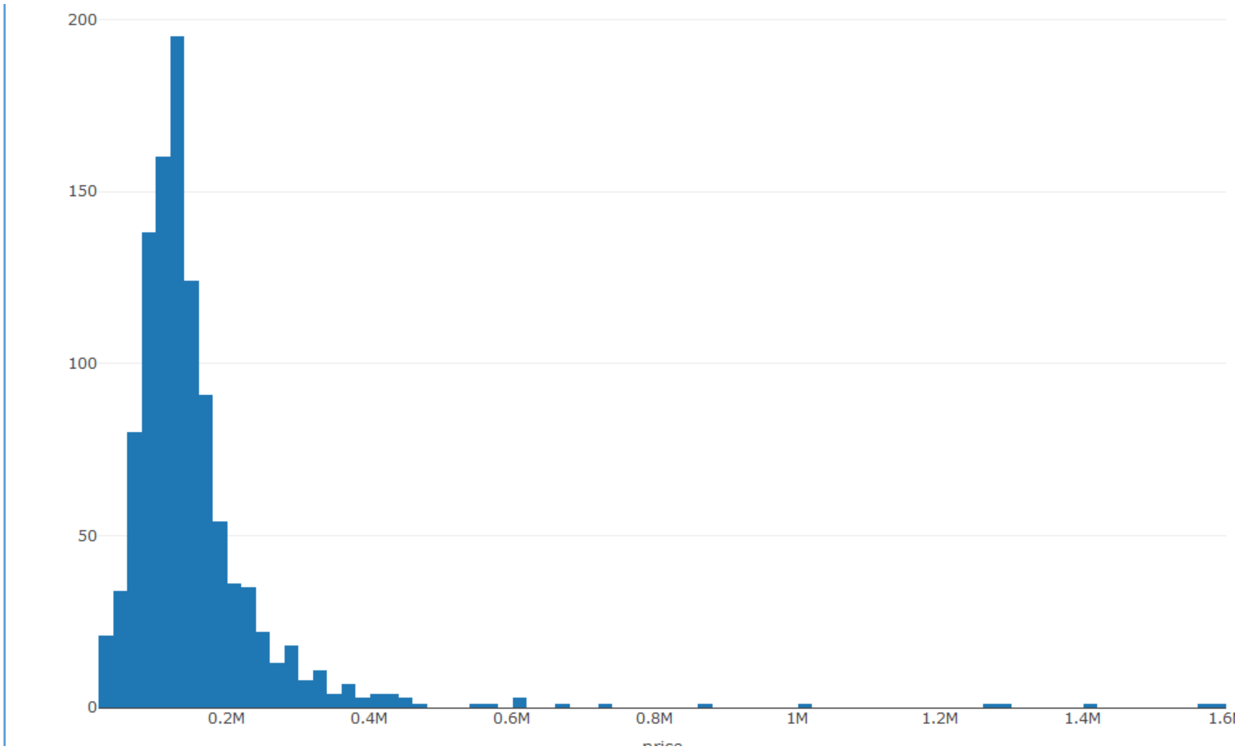
<u>Question 7.15 (p.295). There's a typo in part e of the problem. You need to use the equation estimated in part c, not part d.</u>

A)

```
     price             sqft         bedromms       baths           age            owner            pool
 Min.   :  22000   Min.   : 662   Min.   :1.00   Min.   :1.000   Min.   : 1.00   Min.   :0.0000   Min.   :0.00000
 1st Qu.:  99000   1st Qu.:1604   1st Qu.:3.00   1st Qu.:2.000   1st Qu.: 5.00   1st Qu.:0.0000   1st Qu.:0.00000
 Median : 130000   Median :2186   Median :3.00   Median :2.000   Median :18.00   Median :0.0000   Median :0
 Mean   : 154863   Mean   :2326   Mean   :3.18   Mean   :1.973   Mean   :19.57   Mean   :0.4889   Mean   :0.0796
 3rd Qu.: 170163   3rd Qu.:2800   3rd Qu.:4.00   3rd Qu.:2.000   3rd Qu.:25.00   3rd Qu.:1.0000   3rd Qu.:0.000
 Max.   :1580000   Max.   :7897   Max.   :8.00   Max.   :5.000   Max.   :80.00   Max.   :1.0000   Max.   :1.00000
  traditional       fireplace       waterfront
 Min.   :0.0000   Min.   :0.000   Min.   :0.00000
 1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.00000
 Median :1.0000   Median :1.000   Median :0.00000
 Mean   :0.5389   Mean   :0.563   Mean   :0.07222
 3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:0.00000
 Max.   :1.0000   Max.   :1.000   Max.   :1.00000
```

Price and sqft are continuous variables and their values spread over a big range.
To reduce SSE it is a good idea to take log of price/1000 and divide sqft by 100.

Owner , traditional , fireplace and waterfront are indicator variables



Histogram clearly shows that most values lie in the range of $25,000 - $400,000. Highest frequency is for houses that range from $60,000 - $230,000

Part b)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.9808326  0.0458947  86.738  < 2e-16 ***
sqft         0.0299011  0.0014059  21.269  < 2e-16 ***
bedrooms    -0.0315060  0.0166109  -1.897 0.058135 .
baths        0.1901190  0.0205579   9.248  < 2e-16 ***
age         -0.0062145  0.0005179 -11.999  < 2e-16 ***
owner        0.0674654  0.0177460   3.802 0.000152 ***
pool        -0.0042748  0.0315812  -0.135 0.892353
traditional -0.0560926  0.0170267  -3.294 0.001019 **
fireplace    0.0842748  0.0190150   4.432 1.03e-05 ***
waterfront   0.1099700  0.0333550   3.297 0.001010 **
```

Most coefficents are significant at 5% level and 1% level. But few are not.
Bedrooms shows a negative impact on Price. Meaning that as # of beds increase then price goes down.

Price goes down by 19% per bedroom. This doesn’t make sense. It’s t value is very low and P value is very high meaning that we can reject this estimator .

Having a pool shows a negative impact on price. This is not realistic as pool is considered a luxury feature and buyers will likely pay more to have that feature. According to regression price will decrease 0.4% for having a pool, But t value is very low and P value is really high. We can reject this estimator.

Having a traditional house shows to have a negative impact. This is believable because other architectures and design like French chateau, Mediterranean and villas are more desirable than traditional.
Having a waterfront will increase price by 10.9%.

Part c)

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.9711130  0.0459460  86.430  < 2e-16 ***
sqft             0.0300308  0.0014034  21.399  < 2e-16 ***
bedrooms        -0.0313330  0.0165702  -1.891  0.05890 .
baths            0.1882577  0.0205208   9.174  < 2e-16 ***
age             -0.0061470  0.0005174 -11.881  < 2e-16 ***
owner            0.0683701  0.0177061   3.861  0.00012 ***
pool            -0.0023939  0.0315125  -0.076  0.93946
traditional     -0.0449127  0.0175612  -2.557  0.01068 *
fireplace        0.0873139  0.0190070   4.594 4.87e-06 ***
waterfront       0.1653741  0.0399505   4.139 3.75e-05 ***
trad_waterfront -0.1721747  0.0687162  -2.506  0.01237 *
```

variable trad_waterfront (product of waterfront & traditional) says that having a traditional home with a waterfront will still have a negative impact. In fact homes that have waterfront & traditional style will be 17.2 % less for selling price. T value is high and P value is lower than 0.05. Result is significant at 5% and should not be rejected,

Part d)

model 3 becomes the restricted model

```
model3 <- lm( price~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront
,data=context)
print(summary(model3))
SSE.R=   sum((context$price-predict(model3))^2) = 78.77
```

model 4 becomes the unrestricted model

```
model4 <- lm( price~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront + traditional
+ trad_waterfront +
            trad_sqft + trad_beds + trad_baths + trad_age + trad_owner + trad_pool +
trad_fireplace,
            data=context)
print(summary(model4))
SSE.U =   sum((context$price-predict(model4))^2) = 75.799
```

HO : B10=B11=B12=B13=B14=B15=B16=B17=B18=0

Any estimator of Traditional interaction is 0

H1: Atleast one of the aforementioned is not 0

n =1080 observations
k = 19
J = 9
fVAL = ((SSE.R - SSE.U)/9)/ (SSE.U/1061) = 4.622
fc <- qf(0.95,9,1061) = 1.888

HO can be rejected , so now we know THAT traditional should be included in the model.

fVal is more than F critical

Part E)

use model 2 from part c

sqft= 0.0300308
age =-0.0061470
owner = 0.0683701
fireplace = 0.0873139
trad = -0.0449127
beds = -0.0313330
baths = 0.1882577
Price_partE = 3.971111 + sqft*25 + age*20 + owner + fireplace +trad +beds*3 + baths*2

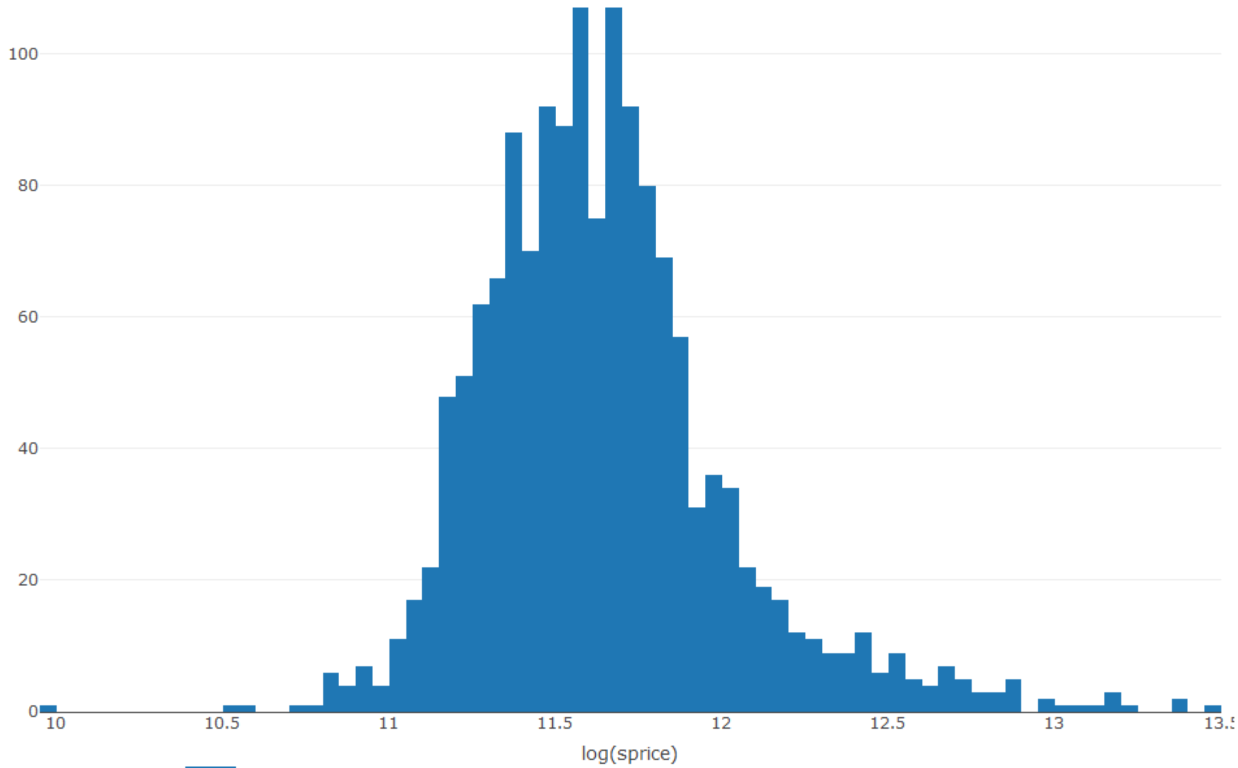
Price_partE = 4.9922

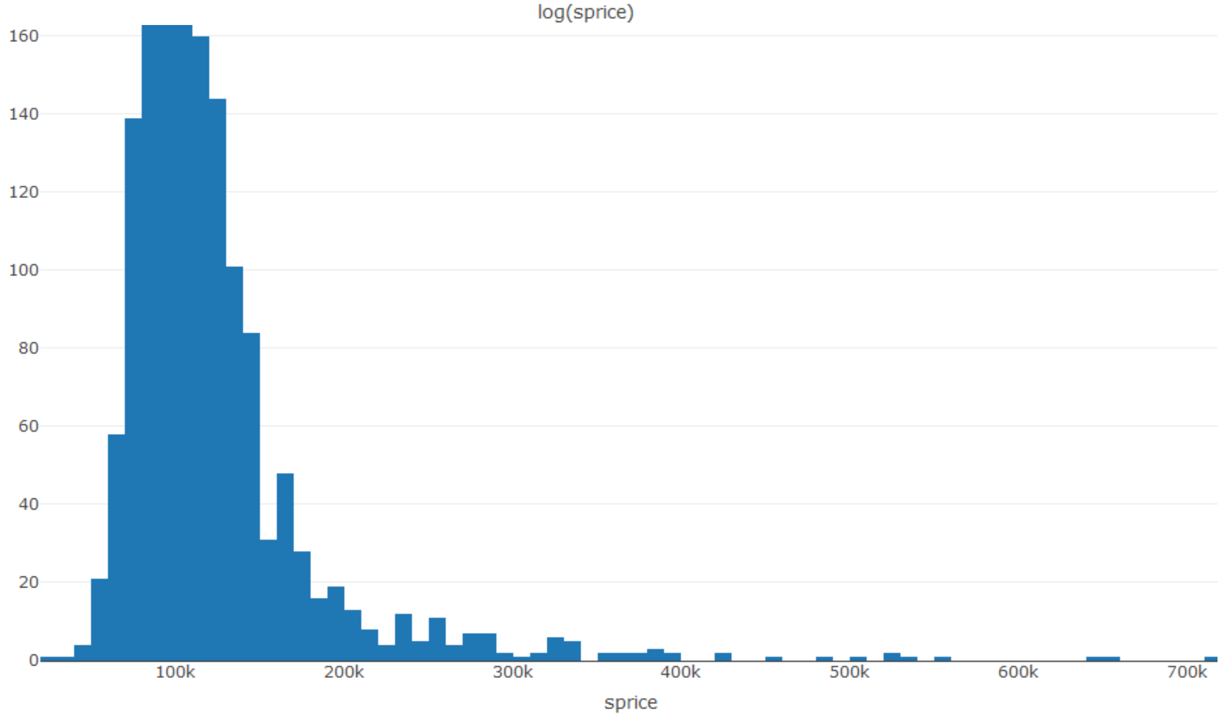
This is the price in natural log . e^(4.9922) = price /1000

Expected price = $147,260.039

<u>Question 7.16 (p.296)</u>

Part a)





Graph of just price histogram shows that most homes are under $200k. Graph is right hand side skewed. But log price shows a bell curved shape. It also shows the outliers more prominent.

Part b)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.9859688  0.0373406 106.746  < 2e-16 ***
livarea      0.0539316  0.0017080  31.576  < 2e-16 ***
beds        -0.0382209  0.0113593  -3.365 0.000786 ***
baths       -0.0102729  0.0165268  -0.622 0.534309
lgelot       0.2530908  0.0255382   9.910  < 2e-16 ***
age         -0.0013113  0.0004601  -2.850 0.004433 **
pool         0.0786611  0.0230548   3.412 0.000662 ***
```

All coeff look as expected with significant at 5%. But baths show very low t value and very high p value.
Ans it is not significant.

Livarea is positive as expected. It shows to add 5.39% in price for every sqft (hundred) added. It is Signiant at 5% level.

bath shows a negative sign showing that every bath causes 1% reduction in price. It is significant at 5% level.

Beds also has a negative estimator. showing a 3.8 % reduction in price for every bedroom added. These two should be added as a polynomial because the relationship may not be linear.

Age has a negative impact. For every year added the price goes down by 1.33%

Pool has a positive impact as expected. It is very significant at 5% and 1%. It shows to increase price by 7.86% when pool is present

Part c)

It has a significant impact. It shows that large lots will add 25% to the selling price. It is significant at 5%,1% and 0.1% level.

Part d)

```
model2 = lm( sprice~livarea + beds+baths+lgelot + age +pool +age+ lge_livarea ,data=context)
print(summary(model2))
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.964941   0.037033 107.064  < 2e-16 ***
livarea      0.058857   0.001864  31.582  < 2e-16 ***
beds        -0.047996   0.011328  -4.237 2.41e-05 ***
baths       -0.020062   0.016398  -1.223 0.221356
lgelot       0.613440   0.063209   9.705  < 2e-16 ***
age         -0.001612   0.000457  -3.527 0.000433 ***
pool         0.085349   0.022795   3.744 0.000188 ***
lge_livarea -0.016125   0.002593  -6.217 6.55e-10 ***
```

when multiplied with livarea then the estimator is negative. It shows that when already on a large lot then added sqft actually have a diminishing return.
Once already a large lot then every sqft will decrease price by 1.6%

Part e)

Carry out the Chow test
Restricted model

```
model3 <- lm( sprice~livarea + beds+baths + age +pool +age ,data=context)
print(summary(model3))
SSE.R=   sum((context$sprice-predict(model3))^2) =  72.06
```

model 4 becomes the unrestricted model

```
model4 <- lm( sprice~livarea + beds+baths+lgelot + age +pool +age +(lgelot*livarea)
             + (lgelot*beds)+(lgelot*baths)+(lgelot*age)+(lgelot*pool),data=context)
print(summary(model4))
SSE.U=   sum((context$sprice-predict(model4))^2) =  65.47
```

$H_0 = \beta_5 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$

All estimators that are an interaction with lgelot are 0

$H_1 = \beta_5 \neq 0$ or $\beta_8 \neq 0$ or $\beta_9 \neq 0$ or $\beta_{10} \neq 0$ or $\beta_{11} \neq 0$ or $\beta_{12} \neq 0$

Atleast one of estimators that are an interaction of lgelot are not 0

n = 1500 observations

k = 13

J =6
fVAL = ((SSE.R - SSE.U)/6)/ (SSE.U/1487) = 24.95

fc <- qf(0.95,6,1487) = 2.10

HO can be rejected because fval is more than fc.

************************CODE FOR the assignment ***********************************

```
############################################
## QUESTION 7.15
rm(list=ls(all=TRUE))


library(foreign)
library(multcomp)
library(data.table)
library(dplyr)
library(plotly)

context = fread('BR2.csv')
View(context)


###PART A
summary(context)
plot_ly(context,x= ~price,type ="histogram")

###PART B

context <- mutate(context,price=log(price/1000))
context <- mutate(context,sqft=sqft/100)

model1 = lm( price~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront,data=context)
print(summary(model1))
SSE =   sum((context$price-predict(model1))^2)
###PART C

context <- mutate(context,trad_waterfront = traditional*waterfront)
model2 = lm( price~sqft+bedrooms+baths+age+owner+pool+traditional+fireplace+waterfront
+trad_waterfront,data=context)
print(summary(model2))

## PART D
## model 3 becomes the restricted model
model3 <- lm( price~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront ,data=context)
print(summary(model3))
SSE.R=   sum((context$price-predict(model3))^2)

### model 4 becomes the unrestricted model
context <- mutate(context,trad_sqft = sqft*traditional)
context <-  mutate(context,trad_beds = bedrooms*traditional)
context <-  mutate(context,trad_baths = baths*traditional)
```

```
context <-  mutate(context,trad_age = age*traditional)
context <-  mutate(context,trad_owner = owner*traditional)
context <-  mutate(context,trad_pool = pool*traditional)
context <-  mutate(context,trad_fireplace = fireplace*traditional)



model4 <- lm( price~sqft+bedrooms+baths+age+owner+pool+fireplace+waterfront + traditional +
trad_waterfront +
          trad_sqft + trad_beds + trad_baths + trad_age + trad_owner + trad_pool + trad_fireplace,
          data=context)
print(summary(model4))
SSE.U =   sum((context$price-predict(model4))^2)

## HO will bE THAT all variables multiplied with traditional are 0
## H1 will be that atleast one of the estimaors ( multilplied with tradition is not 0)


## n =1080 observations
## k = 19
## J = 9
fVAL = ((SSE.R - SSE.U)/9)/ (SSE.U/1061)
fc <-  qf(0.95,9,1061)
## HO can be rejected , so now we know THAT traditional should be included in the model

### PART E

## use model 2 from part c
sqft= 0.0300308
age =-0.0061470
owner = 0.0683701
fireplace = 0.0873139
trad = -0.0449127
beds = -0.0313330
baths = 0.1882577
Price_partE = 3.971111 + sqft*25 + age*20 + owner + fireplace +trad +beds*3 + baths*2
## once convreted it is 147,000

#############################################
## QUESTION 7.16
############################################
rm(list=ls(all=TRUE))


library(foreign)
library(multcomp)
library(data.table)
library(dplyr)
library(plotly)

context = fread('stckton.csv')
View(context)

## part A
```

```
plot_ly(context,x= ~sprice,type ="histogram")

plot_ly(context,x= ~log(sprice),type ="histogram")
## PART B

context <- mutate(context,sprice=log(sprice/1000))

model1 = lm( sprice~livarea + beds+baths+lgelot+age+pool + age,data=context)
print(summary(model1))
SSE =   sum((context$sprice-predict(model1))^2)

## part c
## if they are then increases the price by 4.83%

## part d

context <- mutate(context,lge_livarea=lgelot*livarea)

model2 = lm( sprice~livarea + beds+baths+lgelot + age +pool +age+ lge_livarea ,data=context)
print(summary(model2))
SSE =   sum((context$sprice-predict(model2))^2)

## -0.428% decrease because at that point livaera is giving diminshing returns

## PART E Carry out the Chow test
## Restricted model
model3 <- lm( sprice~livarea + beds+baths + age +pool +age ,data=context)
print(summary(model3))
SSE.R=   sum((context$sprice-predict(model3))^2)

### model 4 becomes the unrestricted model


model4 <- lm( sprice~livarea + beds+baths+lgelot + age +pool +age +(lgelot*livarea)
           + (lgelot*beds)+(lgelot*baths)+(lgelot*age)+(lgelot*pool),data=context)
print(summary(model4))
SSE.U=   sum((context$sprice-predict(model4))^2)


## HO will be that Beta(lgelot and lge_livarea ) are 0
## H1 not 0 , both are not 0


## n = 1500 observations
## k = 13
## J =6
fVAL = ((SSE.R - SSE.U)/6 )/ (SSE.U/1487)
fc <-  qf(0.95,6,1487)
## HO can be rejected , so now we know THAT traditional should be included in the model
```