Lecture14_2025

March 2, 2025

0.1 Lecture 14 EEP118

[1]:
```
#install packaged and load the data

# Load the 'pacman' package
install.packages("pacman")
library(pacman)
#packages to use load them now using the pacman "manager"
p_load(dplyr, haven, readr)
#Another great feature of p_load(): if you try to load a package that is not␣
 ↪installed on your machine, p_load()
#installs the package for you rather than throwing an error. For instance,␣
 ↪let's install and load one final package named ggplot2.
p_load(ggplot2)

pacman::p_load(lfe, lmtest, haven, sandwich, tidyverse)
# lfe for running fixed effects regression
# lmtest for displaying robust SE in output table
# haven for loading in dta files
# sandwich for producing robust Var-Cov matrix
# tidyverse for manipulating data and producing plots



#set scientific display off, thank you Roy
options(scipen=999)

#read in a Stata dataset
my_data <- read_dta("Lecture14hprice1.dta")
head(my_data)
```

```
Installing package into '/srv/r'
(as 'lib' is unspecified)
```

A tibble: 6 × 10

price <dbl>	assess <dbl>	bdrms <dbl>	lotsize <dbl>	sqrft <dbl>	colonial <dbl>	lprice <dbl>	lassess <dbl>	llotsize <dbl>	lsqrft <dbl
300.000	349.1	4	6126	2438	1	5.703783	5.855359	8.720297	7.798
370.000	351.5	3	9903	2076	1	5.913503	5.862210	9.200593	7.638
191.000	217.7	3	5200	1374	0	5.252274	5.383118	8.556414	7.225
195.000	231.8	3	4600	1448	1	5.273000	5.445875	8.433811	7.277
373.000	319.1	4	6095	2514	1	5.921578	5.765504	8.715224	7.829
466.275	414.5	5	8566	2754	1	6.144775	6.027073	9.055556	7.920

[2]:
```
#regression

reg14 <- lm(price~bdrms+lotsize+sqrft, my_data)
summary(reg14)
```

```
Call:
lm(formula = price ~ bdrms + lotsize + sqrft, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-120.026  -38.530   -6.555   32.323  209.376 

Coefficients:
               Estimate  Std. Error t value           Pr(>|t|)    
(Intercept) -21.7703081  29.4750419  -0.739              0.46221    
bdrms        13.8525217   9.0101454   1.537              0.12795    
lotsize       0.0020677   0.0006421   3.220              0.00182 ** 
sqrft         0.1227782   0.0132374   9.275 0.0000000000000166 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 59.83 on 84 degrees of freedom
Multiple R-squared:  0.6724,        Adjusted R-squared:  0.6607 
F-statistic: 57.46 on 3 and 84 DF,  p-value: < 0.00000000000000022
```

[3]:
```
#add an interaction into the model
#gen interaction
#gen bdrms_sqrft=bdrms*sqrft
#reg price bdrms sqrft bdrms_sqrft lotsize
my_data$bdrms_sqrft<-my_data$bdrms*my_data$sqrft

reg14int <- lm(price~bdrms+lotsize+sqrft+bdrms_sqrft, my_data)
summary(reg14int)
```

```
Call:
lm(formula = price ~ bdrms + lotsize + sqrft + bdrms_sqrft, data = my_data)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-127.099  -30.781   -6.305   24.895  200.602

Coefficients:
               Estimate  Std. Error t value Pr(>|t|)
(Intercept) 165.4264973  87.7301516   1.886  0.06284 .
bdrms       -33.7153369  22.8229135  -1.477  0.14339
lotsize       0.0019927   0.0006279   3.174  0.00211 **
sqrft         0.0337926   0.0414616   0.815  0.41739
bdrms_sqrft   0.0218268   0.0096631   2.259  0.02652 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 58.42 on 83 degrees of freedom
Multiple R-squared:  0.6913,        Adjusted R-squared:  0.6765
F-statistic: 46.48 on 4 and 83 DF,  p-value: < 0.00000000000000022
```

0.1.1 How to obtain an estimate of the Prediction for the average value of all houses with bdrms=3, sqrft=2000, and lotsize=9000, that is get a prediction for the value of a house on average with certain characteristics?

The trick is to transform the data and run a regression with the transformed data so that then the estimate of the constant is the average prediction you want and you also gets its standard error

[4]:
```
#trick
#Prediction for the average value of all houses with bdrms=3, sqrft=2000, and␣
 ↪lotsize=9000:
#generate transfored variables such that then the estimated constant gives us␣
 ↪what we need
#gen bdrms0=bdrms-3
#gen sqrft0=sqrft-2000
#gen lotsize0=lotsize-9000
#reg price bdrms0 sqrft0 lotsize0

my_data$bdrms0<-my_data$bdrms-3
my_data$sqrft0<-my_data$sqrft-2000
my_data$lotsize0<-my_data$lotsize-9000
reg14rev <- lm(price~bdrms0+lotsize0+sqrft0, my_data)
summary(reg14rev)
```

```
Call:
lm(formula = price ~ bdrms0 + lotsize0 + sqrft0, data = my_data)

Residuals:
```

```
     Min       1Q   Median       3Q      Max
-120.026  -38.530   -6.555   32.323  209.376

Coefficients:
               Estimate  Std. Error t value             Pr(>|t|)
(Intercept) 283.9529869   8.1210704  34.965 < 0.0000000000000002 ***
bdrms0       13.8525217   9.0101454   1.537              0.12795
lotsize0      0.0020677   0.0006421   3.220              0.00182 **
sqrft0        0.1227782   0.0132374   9.275   0.0000000000000166 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 59.83 on 84 degrees of freedom
Multiple R-squared:  0.6724,        Adjusted R-squared:  0.6607
F-statistic: 57.46 on 3 and 84 DF,  p-value: < 0.00000000000000022
```

From the output we can get a confidence interval

t table got 5% critical from the t 88-4=84 degrees of freedom,, tc14

95% conf interval for predicted average prices of a 3 bd rooom house 2000 sqrt, 9000lot

2.840e+02- tc14 * 8.121e+00 ; 2.840e+02+ tc14 * 8.121e+00

The critical value for the 95% confidence interval is 2 (between 1.987 and 2.000 to be exact). The CI for average price E is thus

283.95 ±2 (8.12) 267.7, 300.2 .

We predict the mean price to be between $267,700 and $300,200. This was the CI for the average house E (of the above characteristics). The CI for the average house E is not the same as the CI for p, the price of a particular house of the above characteristics!!!

0.2 What about a prediction for a house?

For a house

There are two cources randomness

First, the variance of the average prediction Var(Eprediction) = 8.121*8.121, like from above se was 8.12 for an average house

Second the variance of the unobservable disturbance = 59.83*59.83

So, the std error for a predicted house price= square root (varEprice+varU)== se_pp

95% conf interval for predicted prices of a 3 bd rooom house 2000 sqrt, 9000lot

is

2.840e+02- tc14 * sqrt{8.121e+00$^{2+59.83}$2} ; 2.840e+02+ tc14 * sqrt{8.121$^{2+59.83}$2}

since tc=2

and se of predicted price=sqrt{8.121$^{2+59.83}$2}=60.38

which is equal to the interval then of

$283.95 \pm 2\ (60.38) \quad 163.187\ ,\ 404.713\ .$