Name: Hoang Vo

SID: 1671058

Lab 2

Library(MASS)

Library(ISLR)

Task #1: Copy-paste the output you got from the second lm.fit line

> lm.fit = lm(medv~lstat, data=Boston)

> lm.fit

Call:

lm(formula = medv ~ lstat, data = Boston)

Coefficients:

(Intercept) lstat

34.55 -0.95

Task #2: Using the output of Summary(lm.fit), write down the following values (separated by “,”)

> summary(lm.fit)

Call:

lm(formula = medv ~ lstat, data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.168 -3.990 -1.318 2.034 24.500

Coefficients:

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

(Intercept) 34.55384 0.56263 61.41 <2e-16 \*\*\*

lstat -0.95005 0.03873 -24.53 <2e-16 \*\*\*

---

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

Residual standard error: 6.216 on 504 degrees of freedom

Multiple R-squared: 0.5441, Adjusted R-squared: 0.5432

F-statistic: 601.6 on 1 and 504 DF, p-value: < 2.2e-16

1. median of residuals: -1.318
2. standard error of lstat co-efficient estimate: 0.03873
3. multiple R-squared: 0.5441

Task #3: Copy-paste the output of the interval=”prediction” call. Which intervals are wider-con\_dence or prediction (YES/NO)

> predict(lm.fit, data.frame(lstat = c(5,10,15)), interval = "prediction")

fit lwr upr

1 29.80359 17.565675 42.04151

2 25.05335 12.827626 37.27907

3 20.30310 8.077742 32.52846

> predict(lm.fit, data.frame(lstat = c(5,10,15)), interval = "confidence")

fit lwr upr

1 29.80359 29.00741 30.59978

2 25.05335 24.47413 25.63256

3 20.30310 19.73159 20.87461

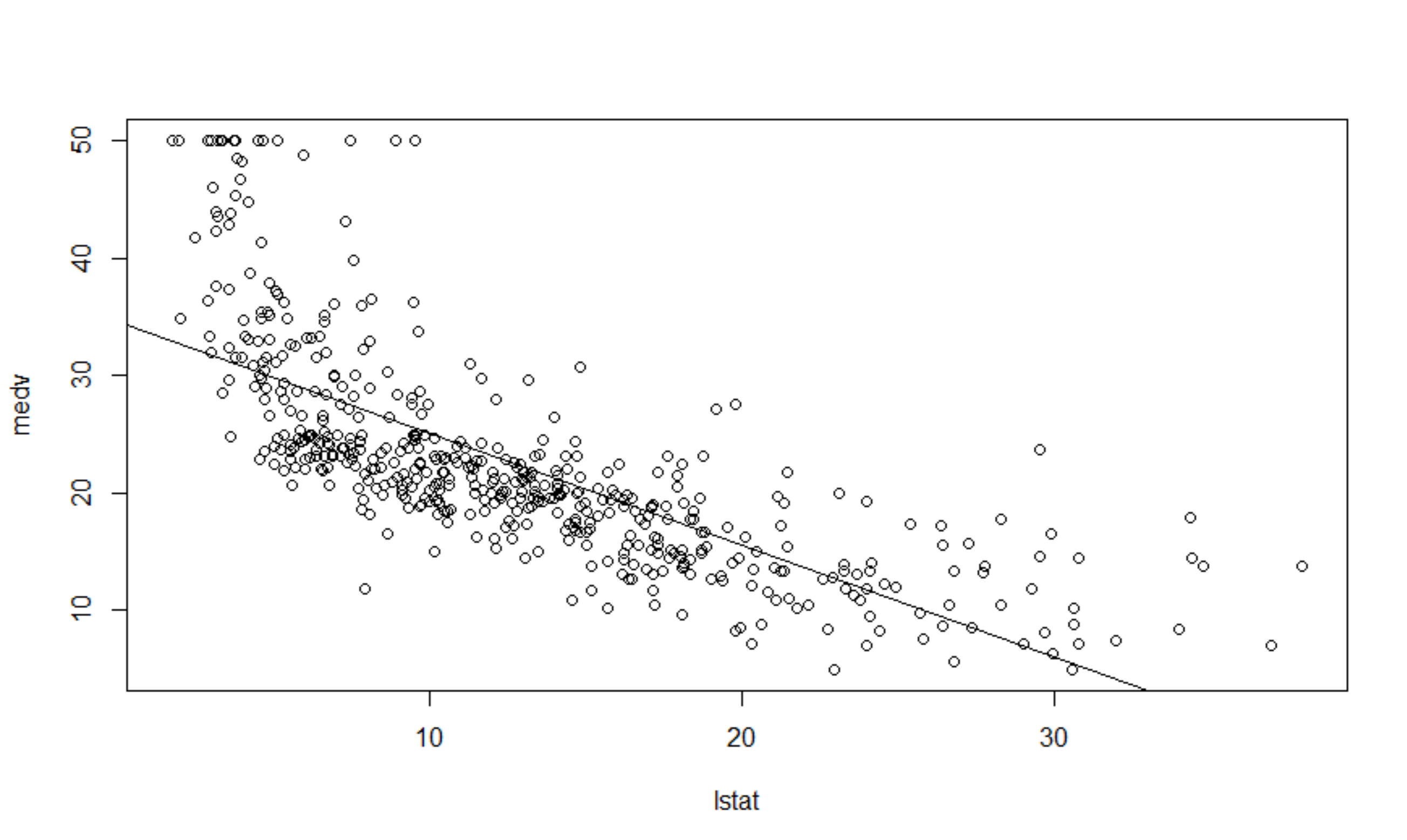
The Intervals of **prediction** are wider than confidence.

Task #4: Copy-paste the plot resulting from the previous operations

> attach(Boston)

> plot(lstat, medv)

> abline(lm.fit)



Task #5: Looking at output of summary(lm.fit), provide the names of variables that not significant, and hence potentially need to be dropped from the model.

> lm.fit = lm(medv~. , data=Boston)

> summary(lm.fit)

Call:

lm(formula = medv ~ ., data = Boston)

Residuals:

Min 1Q Median 3Q

-15.595 -2.730 -0.518 1.777

Max

26.199

Coefficients:

Estimate

(Intercept) 3.646e+01

crim -1.080e-01

zn 4.642e-02

indus 2.056e-02

chas 2.687e+00

nox -1.777e+01

rm 3.810e+00

age 6.922e-04

dis -1.476e+00

rad 3.060e-01

tax -1.233e-02

ptratio -9.527e-01

black 9.312e-03

lstat -5.248e-01

Std. Error t value

(Intercept) 5.103e+00 7.144

crim 3.286e-02 -3.287

zn 1.373e-02 3.382

indus 6.150e-02 0.334

chas 8.616e-01 3.118

nox 3.820e+00 -4.651

rm 4.179e-01 9.116

age 1.321e-02 0.052

dis 1.995e-01 -7.398

rad 6.635e-02 4.613

tax 3.760e-03 -3.280

ptratio 1.308e-01 -7.283

black 2.686e-03 3.467

lstat 5.072e-02 -10.347

Pr(>|t|)

(Intercept) 3.28e-12 \*\*\*

crim 0.001087 \*\*

zn 0.000778 \*\*\*

indus 0.738288

chas 0.001925 \*\*

nox 4.25e-06 \*\*\*

rm < 2e-16 \*\*\*

age 0.958229

dis 6.01e-13 \*\*\*

rad 5.07e-06 \*\*\*

tax 0.001112 \*\*

ptratio 1.31e-12 \*\*\*

black 0.000573 \*\*\*

lstat < 2e-16 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’

0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.745 on 492 degrees of freedom

Multiple R-squared: 0.7406, Adjusted R-squared: 0.7338

F-statistic: 108.1 on 13 and 492 DF, p-value: < 2.2e-16

Variables that are not significant: age, indus

Task #6: Except for the Call: lm(formula::😊 part, are the results of the two above operations identical? (YES/NO)

YES

> summary(lm(medv~lstat\*age, data=Boston))

Call:

lm(formula = medv ~ lstat \* age, data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.806 -4.045 -1.333 2.085 27.552

Coefficients:

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

(Intercept) 36.0885359 1.4698355 24.553 < 2e-16 \*\*\*

lstat -1.3921168 0.1674555 -8.313 8.78e-16 \*\*\*

age -0.0007209 0.0198792 -0.036 0.9711

lstat:age 0.0041560 0.0018518 2.244 0.0252 \*

---

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

Residual standard error: 6.149 on 502 degrees of freedom

Multiple R-squared: 0.5557, Adjusted R-squared: 0.5531

F-statistic: 209.3 on 3 and 502 DF, p-value: < 2.2e-16

Warning messages:

1: In doTryCatch(return(expr), name, parentenv, handler) :

display list redraw incomplete

2: In doTryCatch(return(expr), name, parentenv, handler) :

invalid graphics state

3: In doTryCatch(return(expr), name, parentenv, handler) :

invalid graphics state

> summary(lm(medv~lstat + age + lstat:age, data=Boston))

Call:

lm(formula = medv ~ lstat + age + lstat:age, data = Boston)

Residuals:

Min 1Q Median 3Q Max

-15.806 -4.045 -1.333 2.085 27.552

Coefficients:

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

(Intercept) 36.0885359 1.4698355 24.553 < 2e-16 \*\*\*

lstat -1.3921168 0.1674555 -8.313 8.78e-16 \*\*\*

age -0.0007209 0.0198792 -0.036 0.9711

lstat:age 0.0041560 0.0018518 2.244 0.0252 \*

---

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

Residual standard error: 6.149 on 502 degrees of freedom

Multiple R-squared: 0.5557, Adjusted R-squared: 0.5531

F-statistic: 209.3 on 3 and 502 DF, p-value: < 2.2e-16