4th homework

Manvydas Sokolovas, p3170190

11/30/2017

1 1st regression summary

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	15.451402	5.547786	2.7851475	0.0058518
NROOM	1.150767	1.242893	0.9258777	0.3555982
NBATH	8.330520	2.200704	3.7853893	0.0002015
PATIO	17.278060	3.468174	4.9818891	0.0000013
FIREPL	17.131196	3.026747	5.6599370	0.0000001
AC	12.767345	2.832506	4.5074377	0.0000110

r.squared
0.5151698

1.1 Bootstrap for the first regression

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = balt, statistic = bootfun, R = 10000)
##
##
## Bootstrap Statistics :
##      original      bias    std. error
## t1* 15.451402 -0.56665855    6.903820
## t2*  1.150767  0.11977407    1.376159
## t3*  8.330520  0.06635571    2.446317
## t4* 17.278061 -0.16282453    4.604999
## t5* 17.131196 -0.10874296    3.662683
## t6* 12.767345 -0.09416763    2.876943
```

1.2 Bootstrap for R^2 confidence intervals

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = balt, statistic = function(data, indices) summary(lm(PRICE ~
##     NROOM + NBATH + PATIO + FIREPL + AC, data[indices, ]))$r.squared,
##     R = 10000)
```

```
##
##
## Bootstrap Statistics :
##      original     bias    std. error
## t1* 0.5151698 0.01155303  0.05187626
```

```
##      2.5%     97.5%
## 0.4209248 0.6238646
```

2 Confidence intervals of the first regression estimates

	2.5%	97.5%
(Intercept)	0.796	27.548
NROOM	-1.410	4.004
NBATH	3.783	13.446
PATIO	8.930	26.831
FIREPL	10.026	24.460
AC	7.033	18.242

3 2nd regression summary

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.681666	5.714221	1.169305	0.2436320
NROOM	2.950607	1.288858	2.289319	0.0230743
NBATH	10.154219	2.335224	4.348285	0.0000216
PATIO	19.249235	3.701457	5.200448	0.0000005
AC	14.396878	3.022653	4.762994	0.0000036

r.squared
0.4394065

3.1 Bootstrap for the second regression

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = balt, statistic = bootfun2, R = 10000)
##
##
## Bootstrap Statistics :
##      original      bias    std. error
## t1*  6.681666 -0.47612452    7.112058
## t2*  2.950607  0.07017626    1.454398
## t3* 10.154219  0.17181843    2.696709
```

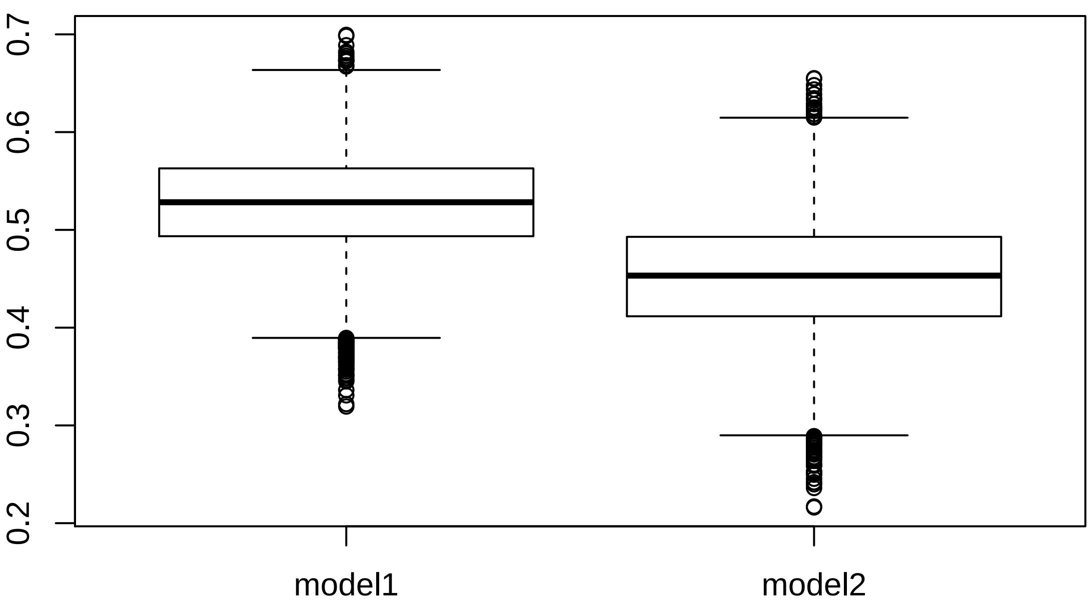
```
## t4* 19.249235 -0.16075103    5.218658
## t5* 14.396878 -0.19708717    3.180646
```

3.2 Bootstrap to get R^2 confidence interval

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = balt, statistic = function(data, indices) summary(lm(PRICE ~
##     NROOM + NBATH + PATIO + AC, data[indices, ]))$r.squared,
##     R = 10000)
##
##
## Bootstrap Statistics :
##      original    bias    std. error
## t1* 0.4394065 0.0118973  0.06018694
```

```
##      2.5%     97.5%
## 0.3295502 0.5634698
```

3.3



- In the second model R^2 is smaller
- R^2 confidence intervals from both models are overlapping, but not a lot.
- Standart errors in the second model got bigger. Thus, assuming all those statements I would prefer the first model.

3.4 Confidence intervals of second model estimates

	2.5%	97.5%
(Intercept)	-8.648	19.281
NROOM	0.197	5.897
NBATH	5.332	15.813
PATIO	9.627	30.258
AC	8.013	20.475

4 Better models by R-square

$mod1 = PRICE = \beta_0 + NROOM\beta_1 + NBATH\beta_2 + PATIO\beta_3 + FIREPL\beta_4 + AC\beta_5$

$mod2 = PRICE = \beta_0 + NROOM\beta_1 + NBATH\beta_2 + PATIO\beta_3 + AC\beta_4$

$mod3 = PRICE = \beta_0 + NBATH\beta_1 + PATIO\beta_2 + FIREPL\beta_3 + SQFT\beta_4$

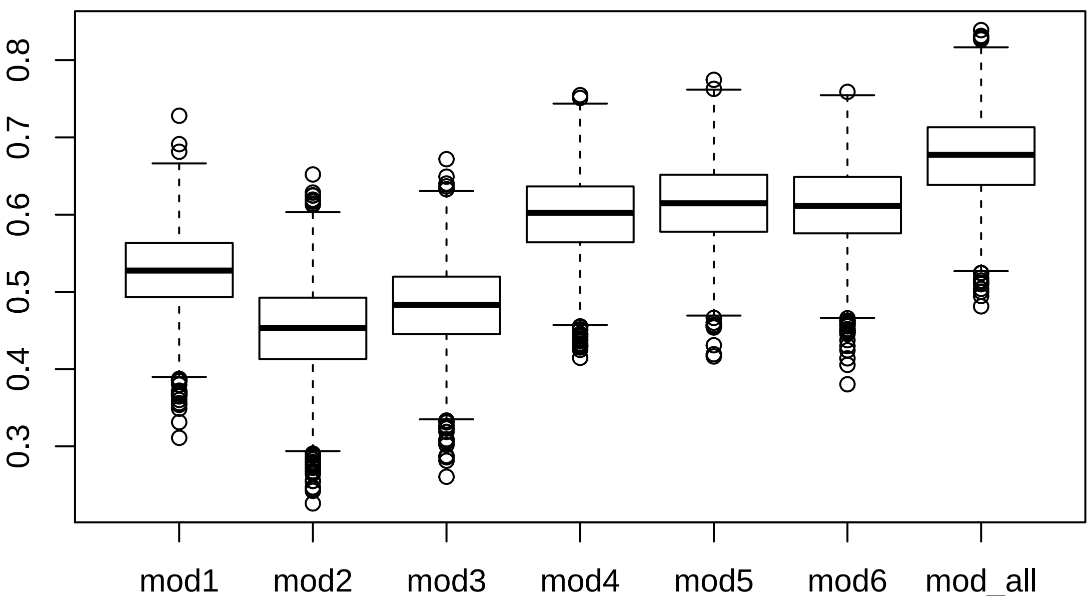
$mod4 = PRICE = \beta_0 + PATIO\beta_1 + FIREPL\beta_2 + AC\beta_3 + BMENT\beta_4 + GAR\beta_5 + AGE\beta_6 + SQFT\beta_7$

$mod5 = PRICE = \beta_0 + NROOM\beta_1 + NBATH\beta_2 + PATIO\beta_3 + FIREPL\beta_4 + AC\beta_5 + GAR\beta_6 + AGE\beta_7 + SQFT\beta_8$

$mod6 = PRICE = \beta_0 + NBATH\beta_1 + PATIO\beta_2 + FIREPL\beta_3 + AC\beta_4 + GAR\beta_5 + AGE\beta_6 + SQFT\beta_7$

$mod\_all = PRICE = \beta_0 + NROOM\beta_1 + NBATH\beta_2 + PATIO\beta_3 + FIREPL\beta_4 + AC\beta_5 + BMENT\beta_6 + NSTOR\beta_7 + GAR\beta_8 + AGE\beta_9 + SQFT\beta_10$

NROOM + NBATH + PATIO + FIREPL + AC + BMENT + NSTOR + GAR + AGE + SQFT



- $mod5$ R^2 confidence interval is just a bit higher than $mod6$. But $mod5$ have one more variable. Thus, that could happen because R^2 gets bigger when you add variables.