

Lab _ Multiple Linear Regression

Dr. Farshid Alizadeh-Shabdiz
Spring 2023

- Use Boston Housing data, which is called “Boston” part of MASS library.
 - `library(MASS)`
 - `library(ISLR)`
- Predict “medv” as a linear function of rest of the columns.
- Call your linear model “myModel”.

Question 1: Apply following functions into your model and make sure you understand the returned values and be ready to share your observations

- `anova(myModel)`
- `coefficients(myModel)`
- `coef(myModel)`
- `confint(myModel)`
- `fitted(myModel)`
- `residuals(myModel)`
- `resid(myModel)`
- `deviance(myModel)`
- `summary(myModel)`

What does the following command do?

- `myModel2 = lm(medv ~ ., subset= 1:100 , data=Boston)`

- Question 2:
 - 1. Predict “medv” from “lstat”.
 - 2. Is lstat significant with 95% confidence?
 - 1. Calculate F-test
 - 2. Calculate T-test
 - 3. Is F-test and T-test are the same?
 - 3. Calculate p-value?
 - 4. Calculate critical value for F-Test and T-test.
 - 5. Plot residual as a function of medv
 - 6. Plot residual as a function of lstat
 - 1. Comment on the residual plots
 - 7. Predict medv for “lstat” equal to values 2 to 15 (2:15).
 - 8. Do you have any suggestions to make the model better?

- Question 3:

1. Run multiple linear regression

1. `myModelQ1 = lm(medv ~ age + lstat, data=Boston)`

1. Are the parameters significant?

2. `myModelQ2 = lm(medv ~ ., data=Boston)`

1. Which parameters are significant?

2. Select 80% of samples randomly and build your model of linear model of medv as a function of age and lstat

3. Try following command

```
par(mfrow=c(2,2))  
plot(myModelQ2)
```

• Question 4: Look the result of the following models and be ready to discuss your observations

1. `myModelQ1 = lm(medv ~ age + lstat, subset= 1:100, data=Boston)`

1. Which parameters are significant?

2. `myModelQ2 = lm(medv ~ ., subset= 1:100, data=Boston)`

1. Which parameters are significant?

1. Have a linear model for $(medv^2)$ as a function of $(crim + zn + chas + nox + rm + age + lstat)$.

2. Have a linear model for $(medv)$ as a function of $(crim + zn + chas + nox + rm + age + lstat + I(rm+lstat))$ and comment on your observation.

3. Have a linear model for $(medv)$ as a function of $(crim + zn + chas + nox + rm + age + lstat + rm * lstat)$ and comment on your observation

4. Have a linear model for $(medv)$ as a function of $crim + zn + chas + nox + rm + age + lstat$ and also add squared of $lstat$ to the model

- Question 5:

Run multiple linear regression

```
myModelQ1 = lm(medv ~ age + indus , data=Boston)
```

1. Is the model significant?
2. What is the p-value of the model?

Assume 99% and 95% confidence interval

Bonus Question

- We observe exponential relationship between Y and X.
- Find solution for $medv = \beta_1 2^{\beta_2 lstat}$

$$medv = \beta_1 2^{\beta_2 lstat}$$