데이터융합SW과
김규석 교수

Data Analysis with Java

데이터 분석 프로그래밍03

Objective of Today's Class

Regression Analysis

- Estimating the Relationships between a dependent variable and independent variables

Regression Analysis(Cont'd)

RA(Regression Analysis)

- A set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

Purpose

- First, for prediction and forecasting, where its use has substantial overlap with the field of machine learning.

- Second, for inferring causal relationships between the independent and dependent variables.

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots\ldots\ldots + \epsilon$$

Source：https://en.wikipedia.org/wiki/Regression_analysis

Regression Analysis(Cont'd)

Simple Linear Regression Analysis

```
import org.apache.commons.math3.stat.regression.SimpleRegression;

public class Main {

    public static void main(String[] args) {
        SimpleRegression simpleRegression = new SimpleRegression(true);
        simpleRegression.addData(new double[][] {
                {12, 20},
                {23, 30},
                {32, 40},
                {43, 50},
                {52, 60}
        });
        double N = 50.0;
        System.out.println("Intercept = " + simpleRegression.getIntercept());
        System.out.println("Prediction for " + N + "= " + simpleRegression.predict(N));
        System.out.println("R square " + simpleRegression.getRSquare());
        System.out.println("Significance Level " + simpleRegression.getSignificance());
    }
}
```

Problems Javadoc Declaration Console Coverage

<terminated> Main (2) [Java Application] C:₩Users₩CTC₩.p2₩pool₩plugins₩or

```
Intercept = 7.6388333999201
Prediction for 50.0= 57.57890531362366
R square 0.9988014382740711
Significance Level 1.761715204118275E-5
```

P1

Multiple Linear Regression Analysis

- Collect any data that you want to analyze

- Calculate Multiple Linear Regression Analysis with VIF

- The result should include the following values
  - Intercept
  - R square
  - Significance level for each independent variable
  - VIF for each variable

P2

Practicing RA

- Find any related data from the internet
- The sample size should be more than 100
- The number of independent variables should be more than 5