

THIS IS THE HOTTEST PPT  
TEMPLATE

# R Machine Learning

2017010715 허지혜

# 목차

부제 / 추가내용은 여기

1. Linear Regression Analysis
2. R code
3. Example

# 1. Linear Regression Model

부제 / 추가내용은 여기

통계 모델이란 ?

복수의 변량간의 관계를 나타내는 일종의 함수

$$Y = f(X_1, X_2, \dots) + \varepsilon$$

변수

변량

$Y, X_1, X_2, \dots$ 는 변량,  $\varepsilon$ 는 오차항

# 1. Linear Regression Model

부제 / 추가내용은 여기

$$Y = \alpha + \beta X_1 + \beta X_2 + \dots + \beta_p X_p + \varepsilon$$

$Y$  = 종속 변수

$X_1, X_2, \dots$  = 독립 변수

$Y$  = 피설명 변량,  
 $X_1, X_2, \dots$ 는 설명 변량,

=>  $Y$ 라는 변량을  $X$ 라는 여러 변량들로 설명이 가능

# 1. Linear Regression Model

부제 / 추가내용은 여기

설명 변량을 하나로 줄이면!

=> 단순 선형 회귀 모델

$$Y = \alpha + \beta X + \varepsilon$$

## 2. R code - lm()

부제 / 추가내용은 여기

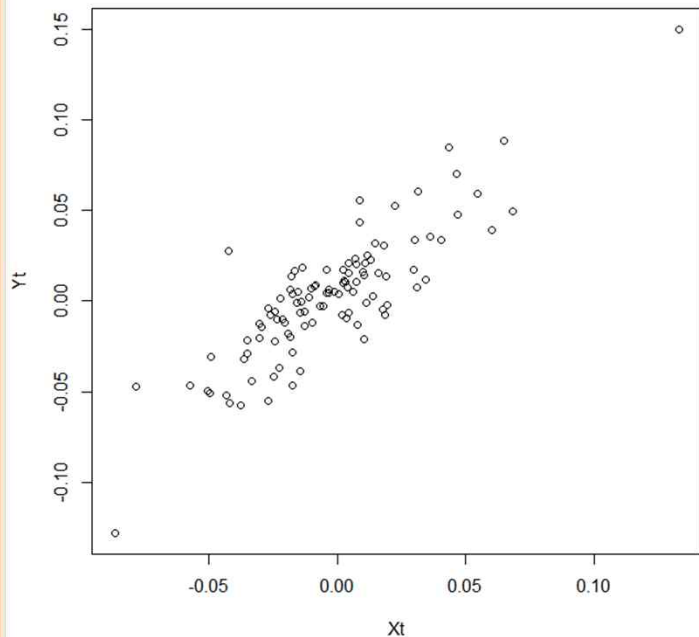
미지수인  $\alpha$ 와  $\beta$ 를 추정  
모델의 계수를 데이터로부터 추정하는 것 = model fitting

$$Y \sim X$$

lm(model식, data = 이용할 데이터 프레임 이름)

## 2. R code - plot()

부제 / 추가내용은 여기



예) 가로축 : 주식 X의 주식 수익률  
=  $X_t$

세로축 : 동일한 시점에서 Y의  
주식 수익률 =  $Y_t$

`Plot( $X_t$ ,  $Y_t$ )`

그래프를 그리는 함수  
`Plot(x좌표의 데이터, y좌표의 데이터)`

## 2. R code - result save

부제 / 추가내용은 여기

```
Result = lm( Yt~ Xt )
```

: Lm함수를 이용하여 선형 회귀 모델을 피팅 한 결과를  
result에 저장

```
Summary(확인하고 싶은 결과 객체 이름) : 요약
```



## 2. R code - result save

부제 / 추가내용은 여기

```
> result= lm(Yt~Xt)
> summary(result)
```

```
Call:
lm(formula = Yt ~ Xt)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.049096	-0.011265	0.001178	0.011698	0.063708

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.005502	0.001807	3.045	0.00267
Xt	0.977899	0.058066	16.842	< 2e-16 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$Y_t = 0.005502 + 0.977899 * X_t + u_i$$

```
F-statistic: 283.7 on 1 and 103 DF,  p-value: < 2.2e-16
```

알파 = 절편항의 계수  
=(intercept)의 estimate에서  
확인

베타 = 설명변량  $X_t$ 의 계수  
= $X_t$ 의 estimate에서 확인

잔차

## 결과값 살펴보기

```
> result= lm(Yt~Xt)
> summary(result)
```

부제 / 추가내용은 여기

```
Call:
lm(formula = Yt ~ Xt)
```

Box plot

```
Residuals:
      Min       1Q   Median       3Q      Max
-0.049096 -0.011265  0.001978  0.011698  0.063708
```

귀무가설 p값

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.005502   0.001807   3.045  0.00295 *
Xt           0.977899   0.058062  16.842 < 2e-16 *
```

0.05보다 작기 때문에  
귀무가설 : 기울기 = 0 을 기각함

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

추정치를 표준오차로 나눈 t값

```
Residual standard error: 0.01844 on 103 degrees of freedom
Multiple R-squared:  0.7336,    Adjusted R-squared:  0.731
F-statistic: 263.7 on 1 and 103 DF,  p-value: < 2.2e-16
```

결정계수가 0.5보다 높으면 설명력이 있다고 판단

## 2. R code - abline()

부제 / 추가내용은 여기

아까 산포도에다가 직선을 그리고 싶을때

`abline(직선의 정의, lty = 직선의 종류, col = 색지정)`

X축에 평행한 수평 직선 = h

y축에 평행한 수평 직선 = v

Lty = 1

실선

Lty=2

파선

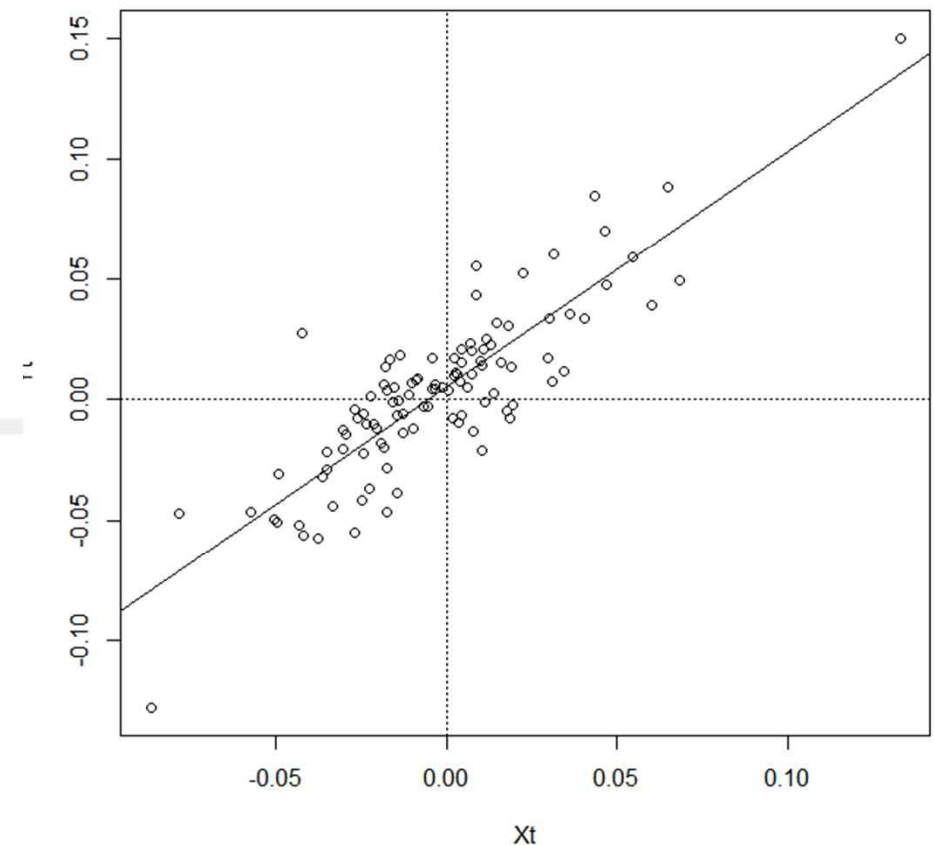
Lty=3

점선

## 2. R code - model fitting result

부제 / 추가내용은 여기

```
> plot(Xt,Yt)
> abline(a=0.005502,b=0.977899,lty=1,col=4)
> plot(Xt,Yt)
> abline(h=0,lty=3)
> abline(v=0,lty=3)
> abline(coef(result))
> |
```



### 3. Example

부제 / 추가내용은 여기

나이에 따른 평균 키	
나이(month)	평균 키 (cm)
18	76.1
19	77
20	78.1
21	78.2
22	78.8
23	79.7
24	79.9
25	81.1
26	81.2
27	81.8
28	82.8
29	83.5

나이에 따른 평균 키에 대하여  
R을 이용하여 코드를 입력하고 선형 회귀선을 알아보자

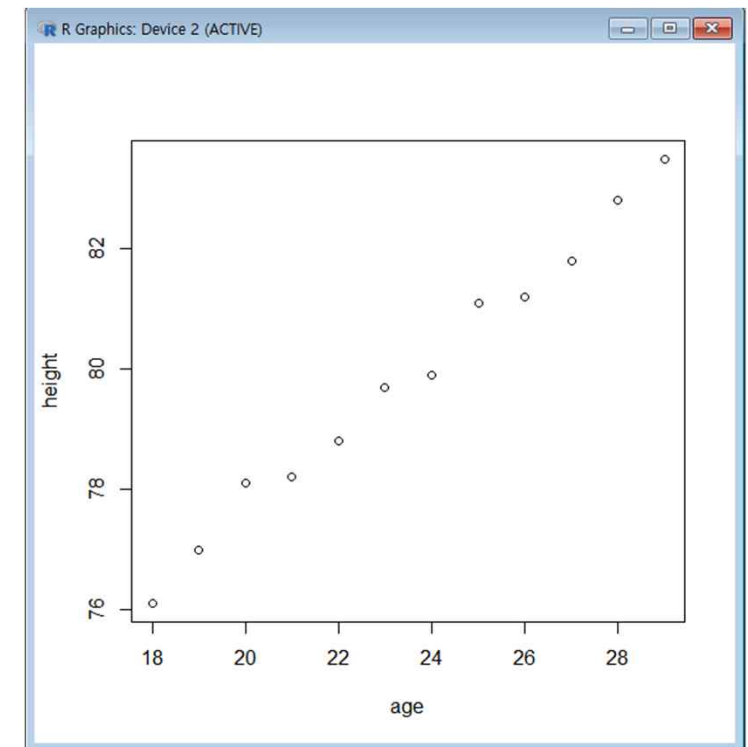
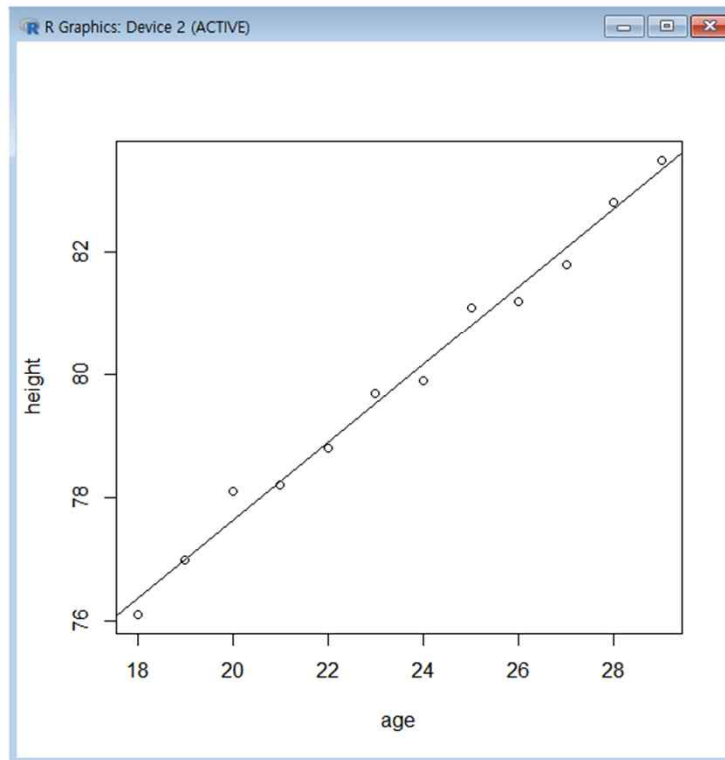
```
> age = 18:29
```

```
> height = c(76.1, 77, 78.1, 78.2, 78.8, 79.7, 79.9, 81.1, 81.2, 81.8, 82.8, 83.5)
```

# 3. Example

부제 / 추가내용은 여기

```
> plot(x=age,y=height)
> |
> a = lm(height ~ age)
> abline(a)
> |
```



### 3. Example

부제 / 추가내용은 여기

```
> summary(a)
```

$$\text{Height} = 64.9283 + 0.6350 * \text{age}$$

```
Call:
```

```
lm(formula = height ~ age)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-0.27238	-0.24248	-0.02762	0.16014	0.47238

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	64.9283	0.5084	127.71	< 2e-16 ***
age	0.6350	0.0214	29.66	4.43e-11 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.256 on 10 degrees of freedom
```

```
Multiple R-squared:  0.9888,    Adjusted R-squared:  0.9876
```

```
F-statistic:   880 on 1 and 10 DF,  p-value: 4.428e-11
```

예로 태어난 지 24개월 된 아이의 키는  
80.168이 된다는 것을 알 수 있습니다!