

R ML_STAT_02_모델만들기

01 데이터 준비

```
kor <- c(50,70,80,90,100)
eng <- c(77,88,99,100,70)
math <- c(55, 75, 85, 95, 105)
```

```
all_score <- data.frame(kor, eng, math)
all_score
```

```
##   kor eng math
## 1  50  77   55
## 2  70  88   75
## 3  80  99   85
## 4  90 100   95
## 5 100  70  105
```

02. 모델 생성(1)

- lm(패턴, 데이터)

```
model1 <- lm(math~kor, data=all_score)
model1
```

```
##
## Call:
## lm(formula = math ~ kor, data = all_score)
##
## Coefficients:
## (Intercept)          kor
##           5           1
```

- 모델 $y = 1 * x_1(\text{kor}) + 5$

02. 모델 생성(2)

- lm(패턴, 데이터)

```
model2 <- lm(math~kor+eng, data=all_score)
model2
```

```
##
## Call:
## lm(formula = math ~ kor + eng, data = all_score)
##
## Coefficients:
## (Intercept)          kor          eng
##  5.000e+00  1.000e+00 -8.979e-17
```

03. 모델을 이용한 예측

- lm(패턴, 데이터) 모델을 이용하여 생성 후, 우리는 predict()를 이용하여 예측값을 구할 수 있다.

```
# 하나의 변수로 이용한 예측
dat <- data.frame(kor=c(80,85))
pred1 <- predict(model1, newdata=dat)
pred1
```

```
## 1 2
## 85 90
```

```
# 두개의 변수로 이용한 예측
dat <- data.frame(kor=c(80,85), eng=c(50,60))
pred2 <- predict(model2, newdata=dat)
pred2
```

```
## 1 2
## 85 90
```

04. 모델 평가

- p-value 가 0.05보다 적어 가설이 유의하다.(OK)
- 결정계수 1로 100% 설명이 가능하다.

```
summary(model1)
```

```
## Warning in summary.lm(model1): essentially perfect fit: summary may be
## unreliable
```

```
##
## Call:
## lm(formula = math ~ kor, data = all_score)
##
## Residuals:
##      1      2      3      4      5
## 3.753e-15 -6.901e-15  7.413e-16  4.573e-16  1.950e-15
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 5.000e+00  9.758e-15  5.124e+14  <2e-16 ***
## kor         1.000e+00  1.222e-16  8.186e+15  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.7e-15 on 3 degrees of freedom
## Multiple R-squared:      1, Adjusted R-squared:      1
## F-statistic: 6.7e+31 on 1 and 3 DF, p-value: < 2.2e-16
```

- p-value 가 0.05보다 적어 가설이 유의하다.(OK)
- 결정계수 1로 100% 설명이 가능하다.

```
summary(model2)
```

```
## Warning in summary.lm(model2): essentially perfect fit: summary may be
## unreliable
```

```
##
## Call:
## lm(formula = math ~ kor + eng, data = all_score)
##
## Residuals:
##      1      2      3      4      5
## 3.005e-15 -6.755e-15  1.827e-15  1.586e-15  3.373e-16
##
## Coefficients:
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept)  5.000e+00  2.067e-14  2.419e+14  <2e-16 ***
## kor          1.000e+00  1.435e-16  6.967e+15  <2e-16 ***
## eng         -8.979e-17  2.083e-16 -4.310e-01    0.708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.506e-15 on 2 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 2.441e+31 on 2 and 2 DF, p-value: < 2.2e-16
```

(실습 해보기1) 국어, 영어, 역사 데이터를 입력 후, 모델을 만들고 아래 새로운 데이터에 대한 예측을 해 보자.

(실습 해보기2) Bike 데이터를 불러온 후, Bike 데이터를 이용하여 선형 회귀 모델을 만들어 보자.