

## 8강. 모형개발

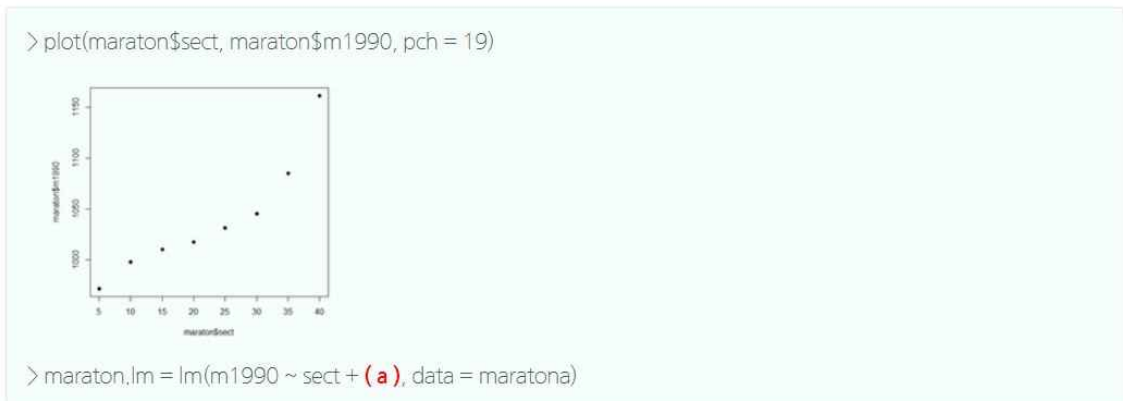
◆ 담당교수 : 김성수 교수

### ■ 주요용어

용어	해설
다항회귀모형	반응변수와 설명변수와의 관계가 다음과 같은 회귀모형을 $d$ 차 다항회귀모형( $d$ -th order polynomial regression model)이라고 정의한다. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_d X^d + \epsilon$
가변수 회귀모형	가변수(dummy variable)는 0과 1 두 가지 값을 갖는 변수를 말한다. 가변수 회귀모형은 독립변수에 이산형 변수가 포함되어 있는 경우에 사용하는 모형으로, 연속인 독립변수를 $X$ , 이산형 변수를 구분하는 가변수를 $D$ 라고 하면 다음과 같은 모형을 적합할 수 있다. 교호작용을 고려한 모형 : $Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 XD + \epsilon$ 교호작용이 없는 모형 : $Y = \beta_0 + \beta_1 X + \beta_2 D + \epsilon$

### ■ 연습문제

1. 두 변수의 산점도가 다음과 같다. maraton 데이터에서 m1990을 반응변수로, sect를 설명변수로 한 2차 다항회귀모형을 적합하고자 한다. ( a ) 명령은 ?



정답 및 해설 : ( a ) = I(sect^2)

2. 다음과 같은 soup 데이터에서 변수 D의 0 = "Line0", 1 = "Line2" 로 범주화하려고 한다.

함수 factor(...)에서 ( b ) 에 들어갈 명령은 ?

```
> soup = read.table("c:/data/reg/soup.txt", header = T)
> soup[c(1,15,16,27),]
  Y  X  D
```

```
1 218 100 1
15 367 265 1
16 140 105 0
27 410 295 0
> soup$D = factor((soup$D, levels = c(0,1), ( b ) = c("Line0", "Line1"))
```

정답 및 해설 : ( b ) = label

3. 다음과 같이 두 생산라인을 고려한 회귀모형을 적합하였다. 생산라인 1과 생산라인 0의 절편의 차이는 ?

```
> soup.lm = lm(Y ~ X + D, data = soup)
> summary(soup.lm)
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 27.28179 15.40701  1.771  0.0893 .
X            1.23074  0.06555 18.775 7.48e-16 ***
DLine1      53.12920  8.21003  6.471 1.08e-06 ***
```

정답 및 해설 : 53.129

#### ■ 참고사이트

- 강명옥,김영일,안철환,이용구, 『회귀분석』, 을곡출판사, 1996.
- 박성현, 『회귀분석』 (제3판), 민영사, 2007.
- Faraway, J.J. (2002), Practical Regression and Anova Using R, (www.google.com에서 검색 후, pdf 파일로 다운받을 수 있음)
- Peter Dalgaard (2005), Introductory Statistics with R, Springer, (www.google.com에서 검색 후, pdf 파일로 다운받을 수 있음)
- R 사이트 <http://www.r-project.org/>
- R Studio 사이트 <https://www.rstudio.com/>