
제6강

다중변수 자료의 탐색

Section 01

산점도

Section 02

상관 분석

Section 03

선그래프

3. 선그래프

1. 선그래프의 작성 아래는 월별로 지각생 통계를 나타낸 것이다.

month	1	2	3	4	5	6	7	8	9	10	11	12
late	5	8	7	9	4	6	12	13	8	6	6	4

코드 6-6

```
month = 1:12
late = c(5,8,7,9,4,6,12,13,8,6,6,4)
plot(  month,
      late,
      main="지각생 통계",
      type= "l",
      lty=1,
      lwd=1,
      xlab="Month",
      ylab="Late cnt"
    )
```

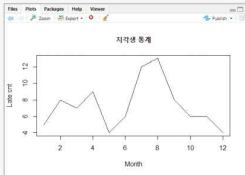
자료 입력
자료 입력
x data
y data
제목
그래프의 종류 선택(알파벳)
선의 종류(line type) 선택
선의 굵기 선택
x축 레이블
y축 레이블

```
> month = 1:12          # 자료 입력
> late = c(5,8,7,9,4,6,12,13,8,6,6,4)  # 자료 입력
```

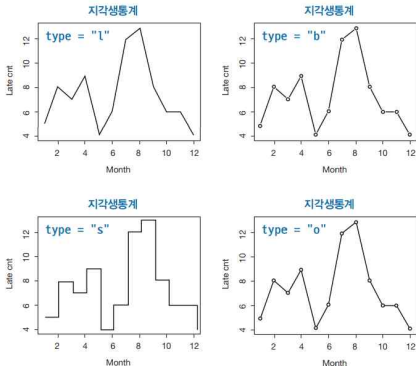
3. 선그래프

```
> plot(month,
+   late,
+   main="지각생 통계",
+   type="l",
+   lty=1,
+   lwd=1,
+   xlab="Month",
+   ylab="Late cnt"
+ )
```

x data
y data
제목
그래프의 종류 선택(알파벳)
선의 종류(line type) 선택
선의 굵기 선택
x축 레이블
y축 레이블



3. 선그래프



좌측 그림은 type에 따라 선그래프가 달리 보이지만 중요한 것은 분석을 해보면 지각생수가 5월 급감했다가 7,8월에 급증을 했다는 사실인 것이다.

이것을 토대로 7,8월 지각생을 줄이기 위해서 방안을 수립하는 것이 데이터 분석하는 이유가 될 것이다.

그림 6-6 매개변수 타입에 따른 다양한 선그래프

3. 선그래프



그림 6-7 선그래프에서의 선의 종류

- 다중변수 자료의 변수 중 하나가 연월일과 같이 시간을 나타내는 값을 갖는 경우 x축을 시간 축으로 하여 선그래프를 그리면 시간의 변화에 따른 자료의 증감 추이를 쉽게 확인할 수 있음
- 시간의 변화에 따라 자료를 수집한 경우, 이를 시계열 자료(times series data)라고 함
- 선그래프는 시계열 자료의 내용을 파악하는 가장 기본적인 방법

3. 선그래프

2. 복수의 선그래프의 작성

- 어느 학급의 월별 지각생 통계

month	1	2	3	4	5	6	7	8	9	10	11	12
late1	5	8	7	9	4	6	12	13	8	6	6	4
late2	4	6	5	8	7	8	10	11	6	5	7	3

코드 6-7

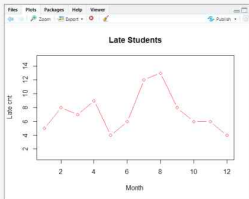
```
month = 1:12
late1 = c(5,8,7,9,4,6,12,13,8,6,6,4)
late2 = c(4,6,5,8,7,8,10,11,6,5,7,3)
plot(  month,                               # x data
       late1,                               # y data
       main="Late Students",
       type="b",                            # 그래프의 종류 선택(알파벳)
       lty=1,                              # 선의 종류(line type) 선택
       col="red",                          # 선의 색 선택
       xlab="Month ",                     # x축 레이블
       ylab="Late cnt",                   # y축 레이블
       ylim=c(1, 15),                    # y축 값의 (하한, 상한)
       )
```

3. 선그래프

```
lines( month,           # x data
       late2,          # y data
       type = "b",     # 선의 종류(line type) 선택
       col = "blue")   # 선의 색 선택
```

```
> month = 1:12
> late1 = c(5,8,7,9,4,6,12,13,8,6,6,4)
> late2 = c(4,6,5,8,7,8,10,11,6,5,7,3)
> plot(month,           # x data
+      late1,          # y data
+      main="Late Students",
+      type= "b",      # 그래프의 종류 선택(알파벳)
+      lty=1,          # 선의 종류(line type) 선택
+      col="red",      # 선의 색 선택
+      xlab="Month ",  # x축 레이블
+      ylab="Late cnt", # y축 레이블
+      ylim=c(1, 15)   # y축 값의 (하한, 상한)
+ )
```

3. 선그래프

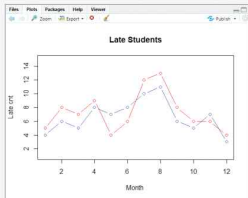


```
> lines(month,
+       late2,
+       type = "b",
+       col = "blue")
```

x data
y data
선의 종류(line type) 선택
선의 색 선택

**lines()함수는 plot()함수로 작성한 그래프 위에 선을 겹쳐서 그리는 역할을 한다.
여기서 사용한 자료는 2반의 자료인 late2인 것이다.**

3. 선그래프



그래프에서 보다시피 역시 2반도 비슷한 추이가 보인다.

감사합니다.